HỌC VIỆN CHÍNH TRỊ QUỐC GIA HỒ CHÍ MINH

NGUYỄN QUỲNH TRANG

XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM

LUẬN ÁN TIẾN SĨ

NGÀNH: LÝ LUẬN VÀ LỊCH SỬ NHÀ NƯỚC VÀ PHÁP LUẬT

HÀ NỘI - 2026

HỌC VIỆN CHÍNH TRỊ QUỐC GIA HỒ CHÍ MINH

NGUYỄN QUỲNH TRANG

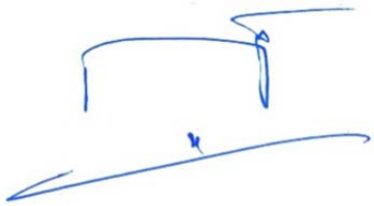
XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM

LUẬN ÁN TIẾN SĨ

NGÀNH: LÝ LUẬN VÀ LỊCH SỬ NHÀ NƯỚC VÀ PHÁP LUẬT

Mã số: 9380106

CHỦ TỊCH HỘI ĐỒNG



PGS.TS. Trần Quang Hiển

NGƯỜI HƯỚNG DẪN KHOA HỌC



1. PGS.TS. Trương Hồ Hải



2. GS.TS. Vũ Công Giao

HÀ NỘI - 2026

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi. Các số liệu, kết quả nêu trong luận án là trung thực, có nguồn gốc rõ ràng và được trích dẫn đầy đủ theo quy định.

Tác giả luận án



Nguyễn Quỳnh Trang

MỤC LỤC

MỞ ĐẦU 1

Chương 1: TỔNG QUAN TÌNH HÌNH NGHIÊN CỨU LIÊN QUAN ĐẾN ĐỀ TÀI LUẬN ÁN 11

1.1. Những công trình nghiên cứu về trí tuệ nhân tạo và phát triển, ứng dụng trí tuệ nhân tạo 11

1.2. Những công trình nghiên cứu về khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo 20

1.3. Những công trình nghiên cứu về xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo 38

1.4. Nhận xét, đánh giá và những vấn đề luận án tiếp tục nghiên cứu 49

1.5. Giả thuyết khoa học và câu hỏi nghiên cứu 53

Chương 2: CƠ SỞ LÝ LUẬN VỀ XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO 55

2.1. Khái niệm, đặc điểm, vai trò và nội dung của xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo 55

2.2. Các nguyên tắc, tiêu chí đánh giá và điều kiện bảo đảm xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo 83

2.3. Kinh nghiệm quốc tế về xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo và giá trị tham khảo cho Việt Nam 91

Chương 3: THỰC TRẠNG XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM 110

3.1. Thực trạng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam 110

3.2. Thực trạng hoạt động xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam 135

Chương 4: QUAN ĐIỂM, GIẢI PHÁP XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM 161

4.1. Quan điểm xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam 161

4.2. Giải pháp xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam 166

KẾT LUẬN 195

DANH MỤC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ ĐÃ CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN 197

DANH MỤC TÀI LIỆU THAM KHẢO 198

PHỤ LỤC 229

DANH MỤC TỪ VIẾT TẮT

AI	: Trí tuệ nhân tạo
BVDLCN	: Bảo vệ dữ liệu cá nhân
CMCN	: Cách mạng công nghiệp
CNCNS	: Công nghiệp công nghệ số
CNTT	: Công nghệ thông tin
CQNN	: Cơ quan nhà nước
ĐMST	: Đổi mới sáng tạo
EU	: Liên minh châu Âu
EVFTA	: Hiệp định thương mại tự do Việt Nam - EU
HCNN	: Hành chính nhà nước
KHCN	: Khoa học công nghệ
LHQ	: Liên hợp quốc
NCS	: Nghiên cứu sinh
NHNN	: Ngân hàng Nhà nước
QLNN	: Quản lý nhà nước
QPPL	: Quy phạm pháp luật
R&D	: Nghiên cứu và phát triển
RCEP	: Hiệp định Đối tác Kinh tế Toàn diện Khu vực
SHTT	: Sở hữu trí tuệ
SME	: Doanh nghiệp vừa và nhỏ
TAND	: Toà án nhân dân
TCĐLCLQG	: Tiêu chuẩn đo lường chất lượng quốc gia
UBND	: Uỷ ban nhân dân
VBQPPL	: Văn bản quy phạm pháp luật
XHCN	: Xã hội chủ nghĩa

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Từ góc độ thực tiễn, trí tuệ nhân tạo (AI) đang định hình lại nền kinh tế, xã hội và hệ thống pháp luật của các quốc gia với tốc độ rất nhanh và phạm vi tác động ngày càng rộng. Thị trường AI toàn cầu được dự báo tăng trưởng 38,1%/năm, từ 93,5 tỷ USD năm 2021 lên khoảng 1.111,8 tỷ USD vào năm 2030 [216]. Từ chẩn đoán y tế, quyết định tín dụng, quản lý giao thông, giáo dục, thương mại điện tử đến kiểm soát hạ tầng quan trọng quốc gia, AI đang từng bước thâm nhập vào nhiều lĩnh vực của đời sống xã hội. Tuy nhiên, cùng với cơ hội, AI cũng đặt ra nhiều vấn đề pháp lý phức tạp, như trách nhiệm khi hệ thống AI gây thiệt hại; bảo vệ dữ liệu cá nhân (BVDLCN); minh bạch và giải trình thuật toán; kiểm soát thiên lệch, phân biệt đối xử; bảo vệ quyền sở hữu trí tuệ (SHTT); bảo đảm an toàn, an ninh mạng; bảo vệ người tiêu dùng; bảo vệ quyền con người, quyền công dân trong môi trường số. Vì vậy, phát triển và ứng dụng AI không chỉ là vấn đề kỹ thuật hoặc kinh tế, mà còn là vấn đề thể chế và pháp luật có ý nghĩa đặc biệt quan trọng. Từ góc độ pháp lý, cộng đồng quốc tế đã bước vào giai đoạn xây dựng và hoàn thiện các khuôn khổ quản trị AI. UNESCO thông qua Khuyến nghị về đạo đức AI năm 2021 [356]; OECD ban hành và cập nhật các nguyên tắc về AI [292]; Liên minh châu Âu thông qua Đạo luật AI [317] theo cách tiếp cận dựa trên rủi ro; Hàn Quốc ban hành đạo luật cơ bản về phát triển AI và xây dựng niềm tin [277]; Trung Quốc ban hành nhiều văn bản điều chỉnh theo từng nhóm công nghệ và ứng dụng AI cụ thể; Nhật Bản, Singapore và các quốc gia ASEAN [154] cũng đang xây dựng mô hình quản trị AI phù hợp với điều kiện của mình. Điểm chung của các mô hình này là đều cố gắng thiết lập sự cân bằng giữa hai yêu cầu: thúc đẩy đổi mới sáng tạo, phát triển thị trường AI, đồng thời phòng ngừa, kiểm soát các rủi ro đối với con người, xã hội và Nhà nước pháp quyền.

Không nằm ngoài xu thế chung của thế giới, từ góc độ chính trị, Đảng Cộng sản và Nhà nước Việt Nam đã sớm nhận thức tầm quan trọng chiến lược của AI và đặt ra yêu cầu hoàn thiện thể chế, pháp luật tương ứng thông qua một loạt văn bản như: Nghị quyết số 52-NQ/TW ngày 27/9/2019 [10], Nghị quyết số 57-NQ/TW ngày 22/12/2024 [11], Quyết định số 1671/QĐ-TTg ngày 28/8/2026. Văn kiện Đại hội XIV của Đảng đã nhấn mạnh yêu cầu: "Xây dựng và hoàn thiện hệ thống pháp luật đáp ứng yêu cầu phát triển nền kinh tế số, nhất là công nghệ tài chính, tài sản số, trí tuệ nhân tạo, thương mại điện tử, các ngành công nghệ mới" [38, tr.88]. Đặc biệt, theo tinh thần bài phát biểu của Tổng Bí thư Tô Lâm, việc làm chủ công nghệ AI là một trong những điều kiện quan trọng để Việt Nam nắm bắt vận hội, thực hiện khát vọng khởi tạo một "kỷ nguyên

vươn mình của dân tộc" [79]. Thực tiễn phát triển AI ở Việt Nam cho thấy nhu cầu hoàn thiện pháp luật đang đặt ra ngày càng rõ. Năm 2025, Việt Nam đứng thứ 45 trên thế giới và thứ 9 trong khu vực Đông Á về phát triển và ứng dụng AI; điểm trung bình của Việt Nam về chỉ số sẵn sàng AI đạt 59,98, cao hơn mức trung bình của khu vực là 50,32 [306]. Có dự báo cho rằng đến năm 2040, AI có thể đóng góp từ 120 đến 130 tỷ USD cho tăng trưởng kinh tế của Việt Nam [70]. Trước thực tiễn đó, sự ra đời của Luật Trí tuệ nhân tạo năm 2025, Nghị định số 142/2026/NĐ-CP hướng dẫn thi hành Luật Trí tuệ nhân tạo năm 2025 [28], Quyết định số 33/2026/QĐ-TTg ban hành danh mục hệ thống AI có rủi ro cao [113] thể hiện sự chủ động của Việt Nam trong việc tạo lập hành lang pháp lý cho một công nghệ mới có tác động sâu rộng.

Từ góc độ lý luận, vấn đề đặt ra hiện nay không chỉ là mô tả các quy định pháp luật đã có về AI, mà là làm rõ cơ sở lý luận, cấu trúc, nội dung và phương thức xây dựng một khung pháp lý phù hợp cho phát triển, ứng dụng AI ở Việt Nam. Trong khi đó, khoa học pháp lý hiện hành về lý luận nhà nước và pháp luật nói chung và pháp luật công nghệ nói riêng chưa xây dựng được một khung lý luận tích hợp có khả năng đồng thời giải quyết bốn vấn đề đặt ra khi vận dụng vào lĩnh vực AI: mở rộng phạm trù nguồn pháp luật để tích hợp có hệ thống các công cụ điều chỉnh phi truyền thống; tái cấu trúc lý luận về cơ chế điều chỉnh pháp luật cho đối tượng có tính bất định và tự tiến hóa; chuyển hóa lý luận xây dựng pháp luật từ mô hình "sự kiện" sang mô hình "quá trình liên tục"; tái diễn giải các nguyên tắc cốt lõi của Nhà nước pháp quyền trong điều kiện mới. Các nghiên cứu ở Việt Nam về AI và pháp luật tuy đã bước đầu đề cập đến một số vấn đề như BVDLCN, trách nhiệm bồi thường thiệt hại, SHTT, đạo đức AI, an ninh mạng hoặc ứng dụng AI trong quản trị nhà nước, nhưng còn thiếu những công trình tiếp cận "khung pháp lý cho phát triển, ứng dụng AI" như một chỉnh thể đa tầng, liên ngành cũng như có phương pháp xây dựng khung pháp lý đó một cách rõ ràng. Trên cơ sở nhận diện khoảng trống đó, luận án xác định bốn nội dung lý luận cụ thể hướng tới bổ sung và phát triển, tạo thành một chỉnh thể có tính hệ thống: mở rộng nguồn pháp luật (thứ nhất) cung cấp nguyên liệu cho tái cấu trúc cơ chế điều chỉnh (thứ hai); cơ chế điều chỉnh mới đòi hỏi quy trình xây dựng pháp luật tương ứng (thứ ba); toàn bộ ba nội dung đó phải được đặt trong khung tham chiếu về Nhà nước pháp quyền để bảo đảm tính chính đáng và bền vững lâu dài của khung pháp lý được xây dựng (thứ tư). Tính hệ thống này là giá trị khoa học cốt lõi mà luận án hướng tới, đồng thời là căn cứ khẳng định tính cấp thiết của đề tài không chỉ dừng lại ở nhu cầu thực tiễn quản lý mà còn ở đòi hỏi phát triển lý luận của chính khoa học pháp lý trước một đối tượng điều chỉnh chưa từng có tiền lệ trong lịch sử lập pháp.

Xuất phát từ những lý do trên, nghiên cứu sinh lựa chọn vấn đề ***"Xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam"*** làm đề tài nghiên cứu luận án tiến sĩ luật học, chuyên ngành Lý luận và Lịch sử nhà nước và pháp luật. Đề tài có ý nghĩa góp phần làm rõ cơ sở khoa học của việc xây dựng pháp luật về AI ở Việt Nam, đồng thời cung cấp luận cứ cho quá trình hoàn thiện thể chế, bảo đảm để AI trở thành động lực phát triển đất nước trong kỷ nguyên số, nhưng vẫn được đặt dưới sự kiểm soát của pháp luật, đạo đức, quyền con người và lợi ích công cộng.

2. Mục đích và nhiệm vụ của luận án

2.1. Mục đích của luận án

Mục đích của luận án là xây dựng cơ sở lý luận khoa học toàn diện về khung pháp lý và xây dựng khung pháp lý cho phát triển, ứng dụng AI; từ đó đánh giá thực trạng và đề xuất hệ thống giải pháp có tính khả thi để xây dựng và hoàn thiện khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam đầy đủ, hiệu quả và bền vững.

2.2. Nhiệm vụ của luận án

Để đạt được mục đích trên, luận án thực hiện các nhiệm vụ sau:

Thứ nhất, tổng quan tình hình nghiên cứu trong và ngoài nước liên quan đến đề tài; đánh giá các công trình đã có và khoảng trống nghiên cứu; xác định các lý thuyết nền tảng, giả thuyết khoa học và câu hỏi nghiên cứu định hướng cho toàn bộ luận án.

Thứ hai, làm rõ cơ sở lý luận của đề tài, bao gồm: hệ thống khái niệm cốt lõi; mô hình kiến trúc của khung pháp lý cho phát triển, ứng dụng AI; các giai đoạn và kỹ thuật xây dựng; các nguyên tắc, tiêu chí và điều kiện bảo đảm xây dựng khung pháp lý cho phát triển, ứng dụng AI; đồng thời phân tích kinh nghiệm quốc tế với giá trị tham khảo cụ thể cho Việt Nam.

Thứ ba, phân tích, đánh giá thực trạng phát triển, ứng dụng AI; thực trạng khung pháp lý; thực trạng xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam; sử dụng tình huống điển hình để cung cấp bằng chứng hỗ trợ kiểm chứng thực nghiệm và xác định nguyên nhân của các hạn chế.

Thứ tư, xác định các quan điểm định hướng và đề xuất hệ thống giải pháp hoàn thiện khung pháp lý và nâng cao chất lượng hoạt động xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam với thứ tự ưu tiên và lộ trình thực hiện cụ thể trong thời gian tới.

3. Đối tượng và phạm vi nghiên cứu

3.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu của luận án là hoạt động xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam và sản phẩm của hoạt động đó là khung pháp lý cho phát triển, ứng dụng AI cũng như mối quan hệ với thực tiễn phát triển, ứng dụng AI cần điều chỉnh.

3.2. Phạm vi nghiên cứu

Thứ nhất, về nội dung nghiên cứu:

Luận án nghiên cứu ba nội dung trung tâm: *(i)* cơ sở lý luận về khung pháp lý cho phát triển, ứng dụng AI và về xây dựng khung pháp lý cho phát triển, ứng dụng AI (khái niệm, đặc điểm, vai trò, nội dung, nguyên tắc, tiêu chí và điều kiện bảo đảm); *(ii)* kinh nghiệm quốc tế về xây dựng khung pháp lý cho phát triển, ứng dụng AI; *(iii)* thực trạng và giải pháp xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam.

Luận án không đi sâu nghiên cứu hai nhóm vấn đề sau: Một là, phân tích nội dung chi tiết của từng ngành luật áp dụng cho AI (pháp luật hình sự, dân sự, sở hữu trí tuệ, lao động, tố tụng trong môi trường AI). Lý do vì những vấn đề đó là đối tượng nghiên cứu của các chuyên ngành luật học chuyên biệt, trọng tâm của luận án là xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam. Hai là, kỹ thuật AI (kiến trúc mô hình học máy, thuật toán cụ thể, hiệu năng kỹ thuật) vì đây là lĩnh vực của khoa học máy tính, không phải pháp luật. Luận án chỉ đề cập đặc trưng kỹ thuật của AI trong phạm vi cần thiết để lý giải yêu cầu đặc thù của khung pháp lý cho phát triển, ứng dụng AI.

Thứ hai, về không gian nghiên cứu:

Trọng tâm nghiên cứu của luận án là khung pháp lý và thực trạng xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam. Luận án có nghiên cứu khung pháp lý của một số tổ chức quốc tế và quốc gia (LHQ/UNESCO, OECD, EU, Hoa Kỳ, Trung Quốc, Nhật Bản, Hàn Quốc, Singapore) nhưng chỉ ở mức khái quát, nhằm mục đích rút ra bài học kinh nghiệm có thể vận dụng vào bối cảnh Việt Nam.

Thứ ba, về thời gian nghiên cứu:

Nghiên cứu của luận án tập trung vào giai đoạn 2017-2026, tức là từ khi Chỉ thị 16/CT-TTg (2017) lần đầu đề cập đến AI như công nghệ chiến lược đến thời điểm Thông tư số 53/2026/TT-NHNN được ban hành (30/9/2026). Điểm mốc 2017 được chọn vì đây là thời điểm Nhà nước Việt Nam bắt đầu có chính sách cụ thể về AI. Giới hạn 9/2026 phản ánh thực tiễn mới nhất luận án có thể cập nhật tại thời điểm hoàn thành. Tương ứng, việc nghiên cứu kinh nghiệm quốc tế trong luận án cũng ưu tiên các văn kiện pháp lý quan trọng nhất trong giai đoạn 2017-2026.

4. Phương pháp luận và phương pháp nghiên cứu

4.1. Phương pháp luận

Luận án được thực hiện trên cơ sở phương pháp luận của chủ nghĩa Mác - Lênin về phép duy vật biện chứng và duy vật lịch sử. Phép duy vật biện chứng được vận dụng để xem xét khung pháp lý về AI như một chỉnh thể đang vận động, trong đó các nguyên tắc, quy phạm, thiết chế, cơ chế thực thi, tiêu chuẩn kỹ thuật và thực tiễn ứng dụng AI có mối liên hệ tác động qua lại với nhau. Cách tiếp cận này giúp luận án

không xem pháp luật về AI như tập hợp rời rạc của các quy định, mà như một hệ thống cần được thiết kế đồng bộ, linh hoạt và có khả năng thích ứng với sự phát triển nhanh của công nghệ. Phép duy vật lịch sử được vận dụng để phân tích sự hình thành và phát triển của chính sách, pháp luật về AI ở Việt Nam trong mối quan hệ với điều kiện kinh tế - xã hội, trình độ khoa học - công nghệ, yêu cầu chuyển đổi số, hội nhập quốc tế và nhu cầu quản trị nhà nước trong từng giai đoạn. Từ đó, luận án có cơ sở để nhận diện tính lịch sử, tính thực tiễn và tính tất yếu của việc xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam hiện nay.

Bên cạnh nền tảng phương pháp luận nêu trên, luận án vận dụng một số lý thuyết chuyên ngành sau:

Thứ nhất, lý thuyết về điều chỉnh pháp luật và quản trị rủi ro được đề xướng bởi Julia Black (trong bài viết The Emergence of Risk-Based Regulation and the New Public Risk Management in the United Kingdom) [248], sau đó được Robert Baldwin và Julia Black tiếp tục phát triển trong bài viết Really Responsive Risk-Based Regulation [320]. Nội hàm cơ bản của lý thuyết này là mức độ can thiệp của pháp luật cần tương ứng với mức độ rủi ro mà một hoạt động, sản phẩm hoặc công nghệ có thể gây ra cho con người, xã hội và lợi ích công cộng. Lý thuyết này được sử dụng trong luận án vì AI không tạo ra mức độ rủi ro giống nhau trong mọi lĩnh vực: AI dùng trong giải trí, quảng cáo, y tế, tín dụng, tuyển dụng, tư pháp hay an ninh quốc gia cần được điều chỉnh bằng các yêu cầu pháp lý khác nhau. Việc vận dụng lý thuyết này giúp luận án xây dựng cơ sở cho việc phân loại rủi ro AI, xác định nghĩa vụ pháp lý tương ứng của các chủ thể và thiết kế cơ chế quản lý vừa chặt chẽ, vừa tránh cản trở đổi mới sáng tạo.

Thứ hai, lý thuyết quản trị thích ứng, được phát triển bởi một số tác giả, trong đó tiêu biểu là Carl Folke và các cộng sự trong bài viết Adaptive Governance of Social-Ecological Systems, trong đó nhấn mạnh năng lực học hỏi, phối hợp đa chủ thể và điều chỉnh thể chế trước các biến đổi phức tạp [167]. Nội hàm của lý thuyết này là pháp luật và thiết chế quản trị phải có khả năng học hỏi, điều chỉnh và cập nhật trước những thay đổi nhanh, phức tạp và khó dự báo của thực tiễn. Lý thuyết này được sử dụng trong luận án vì AI là công nghệ biến đổi nhanh, trong khi quy trình lập pháp truyền thống thường chậm hơn tốc độ phát triển công nghệ. Việc sử dụng lý thuyết quản trị thích ứng giúp luận án luận giải sự cần thiết của các công cụ như cơ chế thử nghiệm có kiểm soát, tiêu chuẩn kỹ thuật linh hoạt, đánh giá tác động, kiểm toán thuật toán, cập nhật nghĩa vụ pháp lý và phối hợp giữa Nhà nước, doanh nghiệp, chuyên gia và xã hội trong quản trị AI.

Thứ ba, lý thuyết tiếp cận dựa trên quyền con người được một số cơ quan trong hệ thống Liên Hợp Quốc đề xuất trong văn kiện The Human Rights Based Approach to Development Cooperation: Towards a Common Understanding Among UN Agencies [360]; trong lĩnh vực AI, cách tiếp cận này được phát triển rõ hơn trong tác phẩm A Human Rights-Based Approach to Responsible AI của Prabhakaran và cộng sự [364]. Nội hàm của lý thuyết này là việc phát triển và ứng dụng công nghệ phải đặt con người ở vị trí trung tâm, bảo đảm các yêu cầu về phẩm giá, tự do, bình đẳng, không phân biệt đối xử, bảo vệ dữ liệu cá nhân, minh bạch, giải trình và khả năng tiếp cận biện pháp khắc phục khi quyền bị xâm phạm. Lý thuyết này được sử dụng trong luận án vì AI có thể tác động trực tiếp đến quyền con người, quyền công dân thông qua các quyết định tự động hoặc bán tự động trong nhiều lĩnh vực đời sống. Việc vận dụng lý thuyết này giúp luận án bảo đảm rằng khung pháp lý về AI không chỉ phục vụ mục tiêu phát triển công nghệ và thị trường, mà còn hướng tới bảo vệ con người, kiểm soát quyền lực và duy trì các giá trị nền tảng của Nhà nước pháp quyền.

Quan hệ giữa ba lý thuyết có thể được mô tả như sau: tiếp cận dựa trên quyền con người tạo ra không gian giới hạn, điều chỉnh dựa trên rủi ro xác định cường độ can thiệp phù hợp cho từng loại đối tượng và quản trị thích ứng bảo đảm cường độ đó tiếp tục phù hợp khi tri thức về rủi ro và bối cảnh công nghệ liên tục thay đổi theo thời gian. Trường hợp ba lý thuyết xung đột về một quyết định cụ thể, cần xác lập thứ tự ưu tiên có điều kiện tùy theo bản chất của xung đột: (1) Ưu tiên tuyệt đối với nhóm quyền không thể giới hạn; (2) Áp dụng bài kiểm tra tỷ lệ khi xung đột liên quan đến quyền có thể giới hạn có điều kiện; (3) Ưu tiên tạm thời cho phòng ngừa rủi ro trong điều kiện bất ổn cao, kèm theo cơ chế rà soát bắt buộc.

4.2. Phương pháp nghiên cứu

Luận án sử dụng kết hợp các phương pháp nghiên cứu sau đây:

Phương pháp phân tích được sử dụng xuyên suốt toàn bộ luận án để làm rõ nội hàm từng khái niệm cốt lõi, từng thành phần của mô hình khung pháp lý, thực trạng từng tầng và từng kỹ thuật xây dựng.

Phương pháp tổng hợp được sử dụng rộng rãi trong việc nghiên cứu các công trình khoa học liên quan đến đề tài, những vấn đề lý luận và thực trạng về khung pháp lý, xây dựng khung pháp lý cho phát triển, ứng dụng AI để từ đó đưa ra các nhận định về những kết quả đạt được, hạn chế trong quá trình xây dựng khung pháp lý về vấn đề này.

Phương pháp luật học so sánh được sử dụng tập trung trong phân tích các tiêu chuẩn và kinh nghiệm quốc tế. Luận án áp dụng khung phân tích ba chiều nhất quán (mô hình khung pháp lý; thực trạng xây dựng về chủ thể, nội dung và quy trình; giá trị

tham khảo cụ thể) cho tám đối tượng so sánh để đảm bảo tính hệ thống và khả năng so sánh. So sánh không dừng ở mô tả quy định nước ngoài mà hướng đến rút ra bài học kiến trúc có thể điều chỉnh và vận dụng vào bối cảnh Việt Nam.

Phương pháp lịch sử được sử dụng khi phân tích tiến trình xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam trong giai đoạn 2017-2026 và khi tổng quan công trình nghiên cứu. Phương pháp này bảo đảm đánh giá thực trạng trong bối cảnh lịch sử cụ thể, không áp đặt tiêu chuẩn của năm 2026 để đánh giá quyết định được đưa ra năm 2019, nhưng cũng không bào chữa cho những hạn chế có thể khắc phục bằng nguồn lực và kiến thức đã có ở từng thời điểm.

Phương pháp hệ thống được vận dụng để phân tích khung pháp lý cho phát triển, ứng dụng AI như một hệ thống với các thành phần có quan hệ nội tại, đặc trưng nổi trội của hệ thống (không thể hiểu từ từng thành phần riêng lẻ) và cơ chế phản hồi (vòng phản hồi từ thực tiễn vận hành đến cập nhật chính sách).

Phương pháp nghiên cứu tình huống của Robert K. Yin (2017) [322] được sử dụng để nghiên cứu tình huống điển hình về làn sóng lừa đảo bằng deepfake và phản ứng pháp lý liên ngành tại Việt Nam giai đoạn 2023-2025. Phương pháp này được áp dụng nhằm khái quát hóa phân tích, cung cấp bằng chứng hỗ trợ kiểm nghiệm tính đúng đắn của việc xây dựng khung pháp lý trong điều kiện thực tiễn Việt Nam.

Phương pháp phỏng vấn 15 chuyên gia chia thành 4 nhóm lĩnh vực chuyên môn chính được thực hiện từ tháng 2 đến tháng 5/2026 để thu thập thông tin thực tiễn không có trong nguồn thứ cấp về quá trình xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam từ góc độ người trong cuộc; đánh giá tính khả thi kỹ thuật của các yêu cầu pháp lý từ góc độ doanh nghiệp AI và nhận định về khoảng trống thực thi từ góc độ cơ quan quản lý. Kết quả phỏng vấn được sử dụng để bổ sung và kiểm chứng phân tích văn bản [Xem Phụ lục 5].

5. Đóng góp mới về khoa học của luận án

Luận án có một số đóng góp mới về khoa học như sau:

Thứ nhất, đóng góp lý luận: Một là, luận án góp phần làm rõ hệ thống khái niệm cơ bản liên quan đến khung pháp lý cho phát triển, ứng dụng AI. Từ đó, luận án đề xuất cách hiểu về "xây dựng khung pháp lý cho phát triển, ứng dụng AI" như một quá trình có mục đích, đa chủ thể và liên tục.

Hai là, luận án đề xuất mô hình khung pháp lý cho phát triển, ứng dụng AI theo cấu trúc đa tầng. Thay vì đồng nhất khung pháp lý với một đạo luật hoặc một nhóm văn bản quy phạm pháp luật chuyên biệt, luận án tiếp cận khung pháp lý cho phát triển, ứng dụng AI như một chỉnh thể gồm nhiều tầng cấu thành. Mô hình này giúp tạo lập một khung phân tích có hệ thống để đánh giá thực trạng khung pháp lý

AI ở Việt Nam, nhận diện các khoảng trống và đề xuất phương hướng hoàn thiện.

Ba là, luận án vận dụng một số lý thuyết chuyên ngành để xây dựng khung phân tích phù hợp với đề tài. Luận án xác lập mối quan hệ bổ trợ giữa ba lý thuyết nền tảng: điều chỉnh dựa trên rủi ro, quản trị thích ứng và tiếp cận dựa trên quyền con người trong quản trị công nghệ. Việc vận dụng các lý thuyết này giúp luận án giải thích đầy đủ hơn mối quan hệ giữa phát triển công nghệ, yêu cầu quản lý nhà nước, bảo vệ quyền con người, bảo đảm lợi ích công cộng và thúc đẩy đổi mới sáng tạo. Đây là cơ sở để luận án tiếp cận khung pháp lý cho phát triển, ứng dụng AI không chỉ như một vấn đề kỹ thuật lập pháp, mà như một vấn đề lý luận về pháp luật, quản trị và phát triển trong bối cảnh chuyển đổi số.

Thứ hai, đóng góp về khung phân tích và phương pháp: luận án xây dựng khung phân tích bốn chiều đối với hoạt động xây dựng khung pháp lý (chủ thể, nội dung, quy trình, kỹ thuật). Bên cạnh đó, luận án góp phần làm rõ các kỹ thuật chủ yếu trong xây dựng khung pháp lý cho phát triển, ứng dụng AI. Trên cơ sở các nguyên tắc, tiêu chí đánh giá, điều kiện bảo đảm, luận án phân tích đối tượng nghiên cứu một cách toàn diện. Đóng góp này có ý nghĩa phương pháp luận, giúp chuyển việc nghiên cứu khung pháp lý AI từ cách tiếp cận thuần túy nội dung sang cách tiếp cận về quy trình, kỹ thuật và điều kiện bảo đảm chất lượng xây dựng.

Thứ ba, đóng góp từ đánh giá thực trạng: Một là, luận án xác định mức độ mất cân đối của kiến trúc khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam. Hai là, luận án xác định bốn nguyên nhân có tính hệ thống lý giải hạn chế trong xây dựng khung pháp lý. Ba là, luận án phát hiện khoảng trống trong quá trình xây dựng, mối quan hệ nhân quả trực tiếp với sự thiếu hụt trong sản phẩm khung pháp lý, đóng góp luận điểm rằng cải thiện khung pháp lý cho phát triển, ứng dụng AI không thể đạt được chỉ bằng bổ sung văn bản mà đòi hỏi cải thiện đồng thời quá trình xây dựng.

Thứ tư, đóng góp kiến nghị có giá trị thực tiễn: luận án đề xuất mô hình kiến trúc phân tầng thích ứng cho Việt Nam sau khi tổng hợp có chọn lọc từ tám mô hình quốc tế đã nghiên cứu. Về các giải pháp đặc thù, luận án đặc biệt nhấn mạnh các giải pháp sau: (1) Mô hình VN-AI Verify theo quy trình đồng thiết kế. (2) Hệ thống TCVN AI rút gọn 12 tháng với tiêu chuẩn tạm thời. (3) Bộ công cụ kỹ thuật lập pháp AI. (4) Chương trình phát triển án lệ AI chủ động của TAND Tối cao. (5) Hệ thống thông tin quản lý AI AMIS như hạ tầng vòng phản hồi chính sách.

Luận án đề xuất giải pháp xây dựng khung pháp lý cho phát triển, ứng dụng AI

ở Việt Nam theo 4 chiều phân tích. *Về chủ thể,* xây dựng năng lực công nghệ-pháp lý; tăng cường vai trò lập pháp chuyên sâu của Quốc hội; thể chế hóa Hội đồng tư vấn AI quốc gia; tăng cường đại diện quốc tế chủ động; thẩm định kỹ thuật bắt buộc. *Về nội dung,* thể chế hóa Chương trình Nghiên cứu Chính sách AI Quốc gia; xây dựng và duy trì Tài liệu kiến trúc pháp lý AI quốc gia; phân tích khoảng trống bảy tầng hàng năm; quy trình rà soát tuân thủ điều ước quốc tế. *Về quy trình,* chuẩn hóa quy trình bốn giai đoạn cho văn bản AI quan trọng; nâng cao chất lượng tham vấn; vận hành cơ chế phản ứng nhanh; thể chế hóa đánh giá hậu kỳ bắt buộc. *Về kỹ thuật,* nâng cấp AI-RIA lên ba chiều đặc thù; hoàn thiện định nghĩa chức năng và thiết lập cơ chế diễn giải; bổ sung đầy đủ ủy quyền linh hoạt phân tầng; hoàn thiện phân định nghĩa vụ theo chuỗi giá trị; thiết kế đầy đủ kỹ thuật miễn trừ tỷ lệ; hoàn thiện luật hóa tiêu chuẩn kỹ thuật.

Nhìn chung, đóng góp khoa học của luận án không nằm ở việc đưa ra một mô hình pháp luật hoàn toàn biệt lập cho AI, mà ở việc xây dựng cách tiếp cận có hệ thống, liên ngành và khả thi đối với quá trình xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam. Cách tiếp cận này vừa góp phần bổ sung cơ sở lý luận cho nghiên cứu pháp luật về công nghệ mới, vừa cung cấp luận cứ khoa học cho việc hoàn thiện thể chế trong bối cảnh chuyển đổi số quốc gia.

6. Ý nghĩa lý luận và thực tiễn của luận án

6.1. Ý nghĩa lý luận

Luận án đóng góp vào khoa học pháp lý theo bốn phương diện.

Thứ nhất, phát triển lý luận về nguồn pháp luật trong môi trường AI: Luận án mở rộng nhận thức về "khung pháp lý" vượt qua ranh giới truyền thống của "pháp luật thành văn" để bao gồm án lệ, tập quán pháp, lẽ công bằng, luật mềm và bộ quy tắc ứng xử như những thành tố có vai trò thiết yếu trong điều chỉnh AI. Đây là đóng góp vào lý luận về nguồn pháp luật trong điều kiện công nghệ thay đổi nhanh hơn chu kỳ lập pháp.

Thứ hai, phát triển lý luận về xây dựng pháp luật công nghệ: kỹ thuật xây dựng và tiêu chí xây dựng khung pháp lý cho phát triển, ứng dụng AI bổ sung vào lý luận xây dựng pháp luật những công cụ đặc thù cho lĩnh vực kỹ thuật - pháp lý cao, không thể áp dụng trực tiếp kỹ thuật lập pháp truyền thống. Điều này có ý nghĩa không chỉ cho AI mà cho toàn bộ lĩnh vực điều chỉnh pháp lý công nghệ.

Thứ ba, đóng góp vào lý luận nhà nước và pháp luật: Mô hình kiến trúc bảy tầng của khung pháp lý cho phát triển, ứng dụng AI là minh chứng cụ thể cho lý luận về sự vận động và phát triển của hệ thống pháp luật trước những thay đổi cách mạng trong điều kiện kinh tế - xã hội - kỹ thuật. Luận án chứng minh rằng phép duy vật biện

chứng đòi hỏi nhìn nhận khung pháp lý không phải như hệ thống tĩnh mà như hệ thống có cơ chế tự thích ứng.

Thứ tư, phát triển phương pháp nghiên cứu luật học so sánh về AI: Khung phân tích ba chiều (mô hình khung pháp lý; thực trạng xây dựng; giá trị tham khảo cụ thể) và tiêu chí lựa chọn đối tượng so sánh dựa trên bối cảnh là đóng góp về phương pháp so sánh pháp luật AI có thể được vận dụng trong các nghiên cứu tiếp theo.

6.2. Ý nghĩa thực tiễn

Đối với hoạt động lập pháp: Kết quả nghiên cứu của luận án, đặc biệt mô hình bảy tầng, sáu kỹ thuật xây dựng và lộ trình ưu tiên cung cấp công cụ thực hành cho Bộ KH&CN, Bộ Tư pháp và các cơ quan liên quan trong tiếp tục xây dựng và hoàn thiện khung pháp lý AI sau Luật AI 2025. Đặc biệt, Bộ công cụ kỹ thuật lập pháp AI với sáu mẫu điều khoản có thể được ứng dụng ngay trong soạn thảo VBQPPL và các hướng dẫn ngành.

Đối với quản lý nhà nước: Phân tích khoảng trống bảy tầng và xác định thứ tự ưu tiên cung cấp cơ sở khoa học cho cơ quan quản lý trong lập kế hoạch và phân bổ nguồn lực. Mô hình vòng phản hồi chính sách qua AMIS là lộ trình xây dựng hạ tầng quản lý AI dựa trên dữ liệu.

Đối với doanh nghiệp AI: Mô hình VN-AI Verify và cấu trúc hướng dẫn ngành ba lớp khi được triển khai sẽ giúp doanh nghiệp, đặc biệt SME và startup biết chính xác mình cần làm gì để tuân thủ Luật AI 2025 mà không cần bộ phận pháp chế lớn.

Đối với tòa án và cơ quan tư pháp: Phân tích về tầng án lệ và lẽ công bằng, cùng đề xuất về chương trình phát triển án lệ AI chủ động của TAND Tối cao, cung cấp định hướng cho hệ thống tư pháp trong giải quyết tranh chấp AI khi chưa có quy phạm thực định đầy đủ.

Đối với đào tạo và nghiên cứu: Luận án là tài liệu tham khảo cho công tác nghiên cứu, giảng dạy, học tập của các cơ sở đào tạo luật và ngành gần trong phạm vi cả nước, đáp ứng nhu cầu đào tạo chuyên gia công nghệ-pháp lý ngày càng cấp bách.

Đối với hội nhập quốc tế: Kết quả nghiên cứu về mô hình kiến trúc phù hợp với điều kiện Việt Nam và đề xuất tham gia chủ động định hình chuẩn mực quốc tế từ góc độ quốc gia đang phát triển có ý nghĩa thiết thực cho hoạt động đối ngoại và hội nhập trong lĩnh vực AI.

7. Kết cấu của luận án

Ngoài phần mở đầu, kết luận, danh mục công trình khoa học của tác giả đã công bố liên quan đến luận án, danh mục tài liệu tham khảo và phụ lục, nội dung luận án gồm 4 chương, 12 tiết.

Chương 1
TỔNG QUAN TÌNH HÌNH NGHIÊN CỨU LIÊN QUAN ĐẾN ĐỀ TÀI LUẬN ÁN

1.1. NHỮNG CÔNG TRÌNH NGHIÊN CỨU VỀ TRÍ TUỆ NHÂN TẠO VÀ PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO

1.1.1. Những công trình nghiên cứu về trí tuệ nhân tạo

AI là một lĩnh vực nghiên cứu liên ngành, được tiếp cận từ nhiều góc độ khác nhau như khoa học máy tính, kinh tế học, xã hội học, triết học, đạo đức học, quản trị công và khoa học pháp lý. Trong luận án này, việc tổng quan các công trình nghiên cứu về AI có ý nghĩa nền tảng, bởi trước khi phân tích khung pháp lý cho phát triển, ứng dụng AI, cần làm rõ AI là gì, có những đặc trưng nào, tác động ra sao đến đời sống xã hội và vì sao AI đặt ra yêu cầu điều chỉnh pháp luật khác với nhiều công nghệ trước đây. Các công trình được tổng quan có thể chia thành bốn nhóm chính sau đây.

Thứ nhất, nhóm công trình nghiên cứu về khái niệm, lịch sử phát triển, nguyên lý và ứng dụng cơ bản của AI. Đây là nhóm tài liệu cung cấp nền tảng nhận thức chung về AI, giúp luận án xác định đúng đối tượng công nghệ cần được pháp luật điều chỉnh.

Trong số các công trình thuộc nhóm này ở nước ngoài, nghiên cứu của Nils J. Nilsson với tiêu đề *Principles of Artificial Intelligence (Nguyên tắc của AI)* [289] được xem là một trong những tài liệu nền tảng về AI thời kỳ đầu. Tác phẩm phân tích các nguyên tắc tính toán, cấu trúc dữ liệu, cách tổ chức hoạt động và cơ chế kiểm soát trong hệ thống AI. Giá trị của công trình này đối với luận án là giúp nhận diện AI như một hệ thống có cấu trúc, có logic vận hành riêng và từ đó gợi mở yêu cầu pháp luật không chỉ điều chỉnh kết quả đầu ra, mà còn phải quan tâm đến quá trình thiết kế, vận hành và kiểm soát hệ thống. Tiếp đó, Stuart J. Russell và Peter Norvig trong nghiên cứu với tiêu đề *Artificial Intelligence: A Modern Approach (AI: Cách tiếp cận hiện đại)* [338] cung cấp một nền tảng toàn diện hơn về AI hiện đại, bao gồm logic, xác suất, học máy, nhận thức, lập luận, học tập và hành động. Đặc biệt, việc hai tác giả hệ thống hóa các định nghĩa về AI theo bốn nhóm: suy nghĩ như con người, hành động như con người, suy nghĩ có lý trí và hành động có lý trí, có ý nghĩa quan trọng đối với luận án khi phân tích sự đa dạng trong cách hiểu về AI. Chính sự đa dạng này là một trong những nguyên nhân khiến việc xây dựng định nghĩa pháp lý về AI gặp nhiều khó khăn. Ngoài ra, còn có một số công trình khác có tính phổ biến tri thức rộng hơn, như bộ sách *Artificial Intelligence for Humans (AI cho con người)* của Jeff Heaton [239] và *The AI Workshop: The Complete Beginner's Guide to AI (Hướng dẫn AI cho người mới bắt đầu)* của Milo Foster [274], giúp người đọc không chuyên tiếp cận các thuật toán cơ bản, học sâu, mạng nơ-ron và các ứng dụng AI thông qua cách trình bày dễ hiểu. Các tài liệu này không trực tiếp bàn về pháp luật AI, nhưng có giá trị tham khảo

trong việc giải thích các khái niệm kỹ thuật phức tạp theo ngôn ngữ dễ tiếp cận, qua đó hỗ trợ việc thiết kế các quy định pháp luật có tính khả thi và dễ hiểu đối với nhiều nhóm chủ thể.

Bên cạnh các công trình kỹ thuật, một số nghiên cứu ở nước ngoài tiếp cận AI từ góc nhìn xã hội rộng hơn. Max Tegmark trong *Life 3.0 - Loài người trong kỷ nguyên trí tuệ nhân tạo* [271] đặt ra các câu hỏi lớn về tác động của AI đối với việc làm, tội phạm, chiến tranh, trật tự xã hội và tương lai của con người. Kai-Fu Lee trong *Các siêu cường AI - Trung Quốc, Thung lũng Silicon và trật tự thế giới mới* [250] phân tích cuộc cạnh tranh công nghệ giữa Hoa Kỳ và Trung Quốc, các làn sóng phát triển AI và khả năng AI tái định hình trật tự kinh tế - chính trị toàn cầu. Các công trình này tuy không phải là nghiên cứu pháp lý chuyên sâu, nhưng góp phần làm rõ bối cảnh chính trị, kinh tế và xã hội của sự phát triển AI, từ đó cho thấy khung pháp lý AI không thể được thiết kế tách rời khỏi cạnh tranh công nghệ, chiến lược phát triển quốc gia và quản trị toàn cầu.

Ở Việt Nam, cuốn *Trí tuệ nhân tạo (Các phương pháp giải quyết vấn đề và kỹ thuật xử lý tri thức)* của Nguyễn Thanh Thuỷ [118] là một trong những tài liệu sớm giới thiệu có hệ thống về AI. Tác phẩm trình bày lịch sử phát triển, các khái niệm cơ bản, các lĩnh vực nghiên cứu và ứng dụng chủ yếu của AI. Dù chủ yếu tiếp cận từ góc độ kỹ thuật, công trình này có ý nghĩa tham khảo quan trọng đối với việc nhận diện những thành tố cơ bản của AI và phân biệt AI với các công nghệ thông tin thông thường. Ngoài ra, các đề tài cấp Bộ do Trần Minh Tuấn chủ nhiệm về định hướng thúc đẩy phát triển AI ở Việt Nam [104] và do Nguyễn Đức Thuỷ chủ nhiệm về định hướng chuẩn hoá AI ở Việt Nam [116] có giá trị tham khảo trực tiếp hơn đối với luận án. Các công trình này phân tích hiện trạng, xu hướng phát triển AI trên thế giới, nhu cầu nghiên cứu và phát triển AI ở Việt Nam, đồng thời đề cập đến tiêu chuẩn hóa AI, kiến trúc hệ thống tiêu chuẩn, lộ trình xây dựng tiêu chuẩn trong nước và yêu cầu hài hòa với tiêu chuẩn quốc tế. Đây là những nội dung có liên hệ chặt chẽ với luận án, nhất là khi luận án tiếp cận khung pháp lý AI không chỉ bao gồm văn bản quy phạm pháp luật, mà còn bao gồm cả tiêu chuẩn kỹ thuật, hướng dẫn, quy tắc ứng xử và các công cụ pháp lý mềm.

Bên cạnh những nghiên cứu kể trên, còn có một số bài viết chuyên khảo về khái niệm AI được công bố trong và ngoài nước cũng cung cấp cơ sở quan trọng cho việc xác định đối tượng nghiên cứu của luận án. Ví dụ, Trần Minh Tâm trong bài "Tìm hiểu về trí tuệ nhân tạo" [99] giới thiệu khái niệm, nội dung cơ bản và triển vọng phát triển AI ở Việt Nam; Rex Martinez trong *"Artificial intelligence: Distinguishing between types and definitions" (Phân biệt định nghĩa và các loại AI)* [318] phân tích quá trình phát triển của AI, các loại AI, tính tự chủ, AI mạnh, AI yếu, đồng thời chỉ ra những hạn

chế trong việc định nghĩa AI dưới góc độ pháp lý; Pei Wang trong *"On defining artificial intelligence"* (*Định nghĩa AI)* [308] tiếp cận vấn đề định nghĩa AI một cách hệ thống, nhấn mạnh rằng một định nghĩa tốt cần phù hợp với cách sử dụng thông dụng, xác định được ranh giới, gợi mở nghiên cứu và có độ đơn giản cần thiết; Kok-Lim Alvin Yau và cộng sự trong nghiên cứu về trí tuệ tăng cường [255] phân tích mối quan hệ giữa trí tuệ con người và AI, qua đó gợi mở cách nhìn AI không chỉ như công cụ thay thế con người, mà còn như công cụ bổ trợ năng lực con người; Sitki Selim Dolanay [334] phân loại AI thành AI hẹp, AI mạnh và siêu AI, từ đó cho thấy mức độ rủi ro pháp lý của AI phụ thuộc đáng kể vào năng lực và phạm vi tự chủ của hệ thống.

Nhìn chung, nhóm công trình thứ nhất giúp luận án xác lập nền tảng khái niệm và kỹ thuật về AI. Tuy nhiên, phần lớn các công trình này chưa đi sâu vào vấn đề khung pháp lý cho phát triển, ứng dụng AI. Do đó, luận án kế thừa các phân tích về bản chất, đặc điểm và phân loại AI, nhưng tiếp tục phát triển theo hướng nghiên cứu những hệ quả pháp lý của AI và yêu cầu xây dựng khung pháp lý phù hợp.

Thứ hai, nhóm công trình nghiên cứu về bản chất, đặc trưng và thách thức của AI từ góc độ xã hội - pháp lý, đạo đức và quản trị. Nhóm công trình này có ý nghĩa quan trọng vì giúp nhận diện AI không chỉ là công nghệ, mà còn là hiện tượng xã hội có tác động trực tiếp đến quyền con người, công bằng xã hội, trách nhiệm pháp lý và niềm tin công chúng.

Trong số các nghiên cứu thuộc nhóm này ở nước ngoài, công trình của Luciano Floridi và cộng sự với tiêu đề *"An Ethical Framework for a Good AI Society" (Khung đạo đức cho một xã hội AI tốt đẹp)* đã đề xuất năm nguyên tắc đạo đức cho một xã hội AI tốt, gồm tính có lợi, không gây hại, tự chủ, công bằng và giải thích được [261]. Đây là một trong những công trình có ảnh hưởng lớn trong thảo luận về AI có trách nhiệm, đồng thời có giá trị tham khảo đối với việc thiết kế các nguyên tắc nền tảng của khung pháp lý AI, đặc biệt là các yêu cầu về công bằng, minh bạch, giải trình và khả năng giải thích. Bên cạnh đó, Stuart Russell trong *Human Compatible: Artificial Intelligence and the Problem of Control (Tương thích với con người: AI và vấn đề kiểm soát)* đặt ra vấn đề kiểm soát AI, nhất là khi các hệ thống AI có thể hành động theo mục tiêu được thiết kế nhưng tạo ra hậu quả ngoài dự kiến của con người [339]. Công trình này có ý nghĩa đối với luận án ở chỗ làm rõ rằng vấn đề pháp lý của AI không chỉ nằm ở việc xử lý thiệt hại sau khi xảy ra, mà còn nằm ở thiết kế cơ chế kiểm soát, phòng ngừa và bảo đảm khả năng can thiệp của con người trong suốt vòng đời của hệ thống AI. Ngoài ra, các nghiên cứu về "hộp đen" và khả năng giải thích của AI, như công trình của Tishby và Schwartz-Ziv về mạng nơ-ron sâu [279], Doshi-Velez và Kim về khoa học học máy có thể diễn giải [209], và Ribeiro, Singh, Guestrin về việc giải thích dự đoán của các bộ phân loại [264], có giá trị quan trọng đối với luận án khi phân tích yêu cầu pháp lý

về minh bạch thuật toán. Các công trình này cho thấy trong nhiều trường hợp, khó có thể áp dụng mô hình trách nhiệm pháp lý truyền thống nếu không hiểu được, hoặc không kiểm tra được, cách thức hệ thống AI đưa ra quyết định. Từ đó, yêu cầu giải thích được, kiểm toán thuật toán, đánh giá tác động và lưu vết quá trình vận hành trở thành những vấn đề trung tâm trong khung pháp lý AI. Xét tổng thể, nhóm công trình này giúp luận án xác định một luận điểm quan trọng: pháp luật về AI không thể chỉ điều chỉnh hành vi của con người sử dụng công nghệ, mà còn phải quan tâm đến thiết kế, dữ liệu, mô hình, mức độ tự chủ, khả năng giải thích và cơ chế giám sát của hệ thống AI. Đây là điểm khác biệt căn bản giữa pháp luật AI và nhiều lĩnh vực pháp luật truyền thống.

Thứ ba, nhóm công trình nghiên cứu về tác động kinh tế - xã hội của AI. Nhóm công trình này cung cấp cơ sở thực chứng và lý luận để đánh giá vì sao phát triển AI cần được khuyến khích, nhưng đồng thời phải được đặt trong khuôn khổ pháp lý nhằm kiểm soát rủi ro.

Tiêu biểu trong số các nghiên cứu thuộc nhóm này ở nước ngoài là các bài viết của Acemoglu và Restrepo về robot, tự động hóa, AI và việc làm đã phân tích tác động của công nghệ đối với thị trường lao động, đặc biệt là nguy cơ thay thế lao động trong một số ngành nghề [181]. Điểm đáng chú ý trong công trình này là sự phân biệt giữa tự động hóa có tính thay thế và tự động hóa có tính bổ trợ. Phân biệt này có ý nghĩa pháp lý quan trọng, bởi chính sách và pháp luật về AI không nên chỉ tập trung hạn chế rủi ro mất việc làm, mà còn cần tạo điều kiện để AI nâng cao năng lực lao động, hình thành việc làm mới và bảo vệ người lao động trong quá trình chuyển đổi. Bên cạnh đó, Bughin và cộng sự, trong báo cáo của McKinsey Global Institute về tác động của AI đối với kinh tế thế giới ước lượng AI có thể tạo thêm giá trị đáng kể cho GDP toàn cầu đến năm 2030 [235]. Công trình này củng cố luận điểm rằng AI là động lực tăng trưởng quan trọng, nhưng đồng thời cũng cho thấy lợi ích từ AI không tự động được phân bổ công bằng nếu thiếu chính sách và pháp luật phù hợp. Cũng theo hướng tiếp cận đó, Köchling và Wehner trong nghiên cứu tổng quan về phân biệt đối xử và công bằng trong ra quyết định thuật toán trong tuyển dụng và phát triển nhân sự đã chỉ ra nguy cơ thiên lệch thuật toán trong lĩnh vực lao động [144]. Công trình này có ý nghĩa trực tiếp đối với yêu cầu thiết kế cơ chế đánh giá, kiểm toán và phòng ngừa phân biệt đối xử trong các hệ thống AI được sử dụng trong tuyển dụng, quản trị nhân sự, tín dụng, giáo dục và dịch vụ công. Taddeo và cộng sự trong nghiên cứu về AI và an ninh mạng phân tích AI như một "con dao hai lưỡi": AI có thể tăng cường năng lực phòng thủ an ninh mạng, nhưng cũng có thể bị lạm dụng để tạo ra các hình thức tấn công mới [265]. Cách tiếp cận này đặc biệt phù hợp với bối cảnh Việt Nam, nơi AI vừa được kỳ vọng hỗ trợ chuyển đổi số, quản trị công và bảo vệ hạ tầng số, vừa làm gia tăng rủi ro

về lừa đảo công nghệ cao, deepfake giả giọng nói, phát tán tin giả và xâm phạm dữ liệu cá nhân.

Nhìn chung, nhóm công trình thứ ba cho thấy yêu cầu xây dựng khung pháp lý AI phải đặt trong sự cân bằng giữa phát triển và kiểm soát rủi ro. Nếu pháp luật quá chặt chẽ, AI có thể bị kìm hãm; nhưng nếu pháp luật quá lỏng lẻo, các rủi ro về lao động, công bằng, an ninh mạng, quyền riêng tư và niềm tin xã hội có thể làm suy giảm chính nền tảng phát triển AI.

Thứ tư, nhóm công trình nghiên cứu về AI agents (tác nhân AI) và agentic AI (hệ thống tác nhân AI). Đây là nhóm nghiên cứu mới, có ý nghĩa đặc biệt đối với luận án vì các hệ thống AI hiện nay đang chuyển từ mô hình công cụ hỗ trợ sang mô hình có khả năng thực hiện chuỗi hành động phức tạp, tương tác với môi trường và theo đuổi mục tiêu nhất định.

Tiêu biểu trong các nghiên cứu ở nước ngoài thuộc nhóm này là bài viết của Yao và cộng sự có tiêu đề *"ReAct: Synergizing Reasoning and Acting in Language Models" (ReAct: Kết hợp lý luận và hành động trong các mô hình ngôn ngữ lớn)* phân tích khả năng kết hợp giữa lập luận và hành động trong các mô hình ngôn ngữ lớn [333]. Cũng theo hướng đó, Park và cộng sự trong bài viết *"Generative Agents: Interactive Simulacra of Human Behavior" (Tác nhân tạo sinh: Mô phỏng tương tác hành vi con người)* đã đi sâu phân tích các tác nhân tạo sinh có khả năng mô phỏng hành vi con người trong môi trường tương tác [245]. Những công trình này cho thấy AI agents có các đặc điểm như tự chủ tương đối, lập kế hoạch, thực hiện chuỗi hành động, ghi nhớ, học từ môi trường và tương tác với con người hoặc hệ thống khác. Đây là những đặc điểm đặt ra nhiều câu hỏi pháp lý mới về năng lực chủ thể, trách nhiệm, kiểm soát, giới hạn ủy quyền và cơ chế giám sát. Đáng chú ý, OECD trong *"The agentic AI landscape and its conceptual foundations" (Bối cảnh hệ thống tác nhân AI và các nền tảng khái niệm)* đã bước đầu phân biệt giữa hệ thống AI nói chung và hệ thống tác nhân AI trong bối cảnh chính sách [301]. R. Sapkota cùng các cộng sự trong *"AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges" (Tác nhân AI với hệ thống tác nhân AI: Phân loại khái niệm, ứng dụng và thách thức)* phân biệt tác nhân AI là các thực thể phần mềm tự động được thiết kế để thực hiện nhiệm vụ hướng mục tiêu trong môi trường kỹ thuật số có giới hạn và hệ thống tác nhân AI đại diện cho một lớp kiến trúc thông minh mới, trong đó nhiều tác nhân chuyên biệt cộng tác để đạt được các mục tiêu phức tạp, cấp cao bằng cách sử dụng suy luận hợp tác và lập kế hoạch nhiều bước, được phối hợp thông qua một bộ điều phối tập trung hoặc một giao thức phi tập trung [314]. Cách phân biệt này có ý nghĩa quan trọng vì khung pháp lý AI trong thời gian tới có thể phải chuyển từ điều chỉnh các hệ thống AI như công cụ sang điều chỉnh các hệ thống tác nhân AI có khả năng hành động với mức độ

tự chủ cao hơn. Đây cũng là cơ sở để luận án phân tích nhu cầu cập nhật khung pháp lý AI theo hướng linh hoạt, thích ứng và dựa trên rủi ro.

Từ tổng quan các công trình nêu trên, có thể thấy nghiên cứu về AI đã phát triển khá phong phú, bao gồm từ nền tảng kỹ thuật, khái niệm, phân loại, nguyên lý vận hành đến các tác động kinh tế, xã hội, đạo đức và quản trị. Các công trình này cung cấp cơ sở quan trọng để luận án nhận diện bản chất của AI, các đặc trưng pháp lý của AI và những rủi ro cần được pháp luật điều chỉnh. Tuy nhiên, phần lớn các công trình mới dừng lại ở việc phân tích AI với tư cách là một công nghệ, một hiện tượng xã hội hoặc một đối tượng quản trị nói chung. Chưa có nhiều nghiên cứu tiếp cận AI từ yêu cầu xây dựng khung pháp lý như một chỉnh thể đa tầng, bao gồm cả văn bản quy phạm pháp luật, tiêu chuẩn kỹ thuật, luật mềm, cơ chế tự quản, thiết chế thực thi và vòng phản hồi chính sách. Đây là khoảng trống mà luận án tiếp tục nghiên cứu trong mối liên hệ trực tiếp với điều kiện chính trị, pháp lý, kinh tế, xã hội và thực tiễn phát triển, ứng dụng AI ở Việt Nam.

1.1.2. Những công trình nghiên cứu về phát triển, ứng dụng trí tuệ nhân tạo

Mục này khảo sát các công trình nghiên cứu về phát triển và ứng dụng AI trong thực tiễn. Những nghiên cứu này có ý nghĩa quan trọng đối với luận án, bởi việc xây dựng khung pháp lý cho AI không thể chỉ xuất phát từ nhận thức lý thuyết về AI, mà còn phải dựa trên thực tiễn AI được phát triển, triển khai, sử dụng và quản trị trong các lĩnh vực kinh tế - xã hội khác nhau. Có thể chia các công trình nghiên cứu dạng này thành các nhóm ở nước ngoài và ở Việt Nam, cụ thể như sau:

Thứ nhất, các công trình nghiên cứu ở nước ngoài về phát triển, ứng dụng và phổ biến AI. Tiêu biểu trong số này là công trình nghiên cứu về phát triển, ứng dụng AI từ lý thuyết lan toả đổi mới của Everett M. Rogers, trong *Diffusion of Innovations (Lan toả đổi mới)* [207]. Nghiên cứu này giải thích vì sao một công nghệ mới được tiếp nhận nhanh hay chậm trong xã hội, thông qua các yếu tố như lợi thế tương đối, mức độ tương thích, độ phức tạp, khả năng thử nghiệm và khả năng quan sát kết quả. Đối với AI, cách tiếp cận này gợi mở rằng việc phát triển và ứng dụng AI không chỉ phụ thuộc vào trình độ công nghệ, mà còn phụ thuộc vào niềm tin xã hội, mức độ dễ sử dụng, khả năng chứng minh hiệu quả và môi trường pháp lý cho phép thử nghiệm an toàn. Bên cạnh đó, nghiên cứu về khung công nghệ - tổ chức - môi trường do Tornatzky, Fleischer và Chakrabarti phát triển trong *The Processes of Technological Innovation (Quá trình đổi mới công nghệ)* [258] cũng có giá trị tham khảo quan trọng. Theo cách tiếp cận trong nghiên cứu này, việc một tổ chức có ứng dụng công nghệ mới hay không phụ thuộc vào ba nhóm yếu tố: đặc điểm của công nghệ, năng lực và cấu trúc của tổ chức, cũng như môi trường bên ngoài như thị trường, cạnh tranh, chính sách và pháp

luật. Khung phân tích này đặc biệt phù hợp với AI, bởi việc ứng dụng AI trong doanh nghiệp, cơ quan nhà nước hay tổ chức xã hội luôn đòi hỏi dữ liệu, nhân lực, hạ tầng, quản trị rủi ro và sự rõ ràng của quy định pháp luật.

Ở cấp độ người sử dụng, các nghiên cứu về chấp nhận và sử dụng công nghệ, tiêu biểu là mô hình UTAUT của Venkatesh và cộng sự trong bài viết *User Acceptance of Information Technology: Toward a Unified View (Lý thuyết thống nhất về chấp nhận và sử dụng công nghệ)* [365], giúp giải thích vì sao cá nhân và tổ chức chấp nhận hoặc từ chối sử dụng một công nghệ mới, qua đó khẳng định các yếu tố như kỳ vọng về hiệu quả, kỳ vọng về mức độ dễ sử dụng, ảnh hưởng xã hội và điều kiện hỗ trợ có ý nghĩa đáng kể đối với việc ứng dụng AI. Đối với luận án, nghiên cứu này gợi mở rằng pháp luật về AI không chỉ cần quy định nghĩa vụ và chế tài, mà còn phải tạo môi trường tin cậy để người dân, doanh nghiệp và cơ quan nhà nước có thể sử dụng AI một cách an toàn, có trách nhiệm.

Ở bình diện thực tiễn doanh nghiệp, Bernard Marr và Matt Ward trong tác phẩm *Ứng dụng trí tuệ nhân tạo để dẫn đầu* [160] phân tích cách các công ty ứng dụng AI và học máy để giải quyết các vấn đề kinh doanh trong nhiều lĩnh vực như ngân hàng, tài chính, truyền thông, marketing và quản trị khách hàng. Công trình này cho thấy AI ngày càng trở thành công cụ chiến lược của doanh nghiệp, không chỉ giúp tự động hóa một số quy trình mà còn hỗ trợ dự báo, cá nhân hóa dịch vụ, tối ưu hóa vận hành và tạo ra mô hình kinh doanh mới. Tuy nhiên, chính sự mở rộng của AI trong doanh nghiệp cũng đặt ra yêu cầu về quản trị dữ liệu, bảo vệ người tiêu dùng, kiểm soát thiên lệch thuật toán và trách nhiệm đối với quyết định được hỗ trợ hoặc tạo ra bởi AI. Bên cạnh đó, báo cáo *State of AI in the Enterprise (Thực trạng AI trong doanh nghiệp),* 2nd Edition của Deloitte Insights [187] cũng có giá trị tham khảo đối với việc nhận diện xu hướng ứng dụng AI trong doanh nghiệp. Báo cáo cho thấy các doanh nghiệp tiên phong tiếp tục tăng đầu tư vào AI, mở rộng các sáng kiến AI và bước đầu thu được lợi ích kinh tế đáng kể. Đồng thời, báo cáo cũng nhấn mạnh yêu cầu quản lý thay đổi, quản trị rủi ro, giảm thiểu rủi ro an ninh mạng, xử lý các vấn đề đạo đức và phát triển nhân lực phù hợp. Điểm đáng chú ý là các rủi ro này không nằm ngoài quá trình ứng dụng AI, mà phát sinh ngay trong quá trình doanh nghiệp thiết kế, triển khai, vận hành và mở rộng hệ thống AI.

Ở cấp độ toàn cầu, *AI Index Report 2025 (Báo cáo chỉ số AI 2025)* của Stanford [336] cung cấp bức tranh dữ liệu tương đối toàn diện về xu hướng phát triển AI, đầu tư tư nhân, nghiên cứu, bằng sáng chế, năng lực tính toán, ứng dụng trong doanh nghiệp và quản trị AI. Các báo cáo dạng này giúp luận án nhìn nhận phát triển, ứng dụng AI như một quá trình có tính toàn cầu, cạnh tranh cao và ngày càng gắn với năng lực thể chế của từng quốc gia. Điều đó củng cố luận điểm rằng khung pháp lý không chỉ có

chức năng kiểm soát rủi ro, mà còn là một bộ phận của năng lực cạnh tranh quốc gia trong kỷ nguyên AI.

Ngoài các công trình có tính khái quát nêu trên, một số nghiên cứu khác cũng cho thấy AI đang được ứng dụng sâu trong các lĩnh vực chuyên môn cụ thể. Chẳng hạn, Marly van Assen và cộng sự trong bài viết *"Artificial intelligence from A to Z: From neural network to legal framework" (AI từ A đến Z: Từ mạng nơ-ron đến khung pháp lý)* [266] phân tích việc ứng dụng AI trong lĩnh vực X-quang, đồng thời đề cập đến những hệ quả pháp lý của việc sử dụng thuật toán AI trong thực hành lâm sàng. Công trình này có ý nghĩa minh họa rõ nét cho đặc điểm của AI trong các lĩnh vực có rủi ro cao: lợi ích chuyên môn rất lớn, nhưng đi kèm các vấn đề về an toàn, trách nhiệm nghề nghiệp, giải thích quyết định và bảo vệ người bệnh.

Nhìn chung, các nghiên cứu nước ngoài về phát triển, ứng dụng AI đã chuyển từ cách tiếp cận mô tả công nghệ sang cách tiếp cận liên ngành, trong đó AI được xem là một quá trình đổi mới công nghệ, tổ chức và thể chế. Các công trình này cung cấp cơ sở lý luận quan trọng để luận án phân tích vì sao khung pháp lý về AI cần đồng thời bảo đảm ba yêu cầu: tạo điều kiện cho phát triển và ứng dụng AI; bảo đảm niềm tin, an toàn và trách nhiệm trong quá trình sử dụng AI; và duy trì khả năng thích ứng của pháp luật trước sự thay đổi nhanh của công nghệ.

Thứ hai, các công trình nghiên cứu ở Việt Nam về phát triển, ứng dụng AI. So với nghiên cứu quốc tế, các công trình ở Việt Nam về phát triển và ứng dụng AI có xu hướng tập trung nhiều vào mô tả hiện trạng, khả năng ứng dụng trong từng lĩnh vực, nhu cầu thúc đẩy chuyển đổi số và một số khuyến nghị chính sách ban đầu. Các công trình này có giá trị thực tiễn, nhưng nhìn chung chưa hình thành một khung lý thuyết đầy đủ về phát triển, ứng dụng AI dưới góc độ pháp lý.

Ví dụ, một số đề tài nghiên cứu cấp quốc gia và cấp bộ đã tiếp cận AI từ yêu cầu ứng dụng trong các ngành cụ thể ở Việt Nam. Chẳng hạn, đề tài trọng điểm quốc gia về ứng dụng AI trong tích hợp dữ liệu địa chất dầu khí do Doãn Ngọc San chủ nhiệm [93], đề tài về ứng dụng AI cho khâu chuẩn bị sản xuất trong nhà máy may do Nguyễn Thị Hường chủ nhiệm [53], cũng như một số luận án, luận văn về ứng dụng AI trong chẩn đoán sự cố máy biến áp [60], dự báo chất lượng không khí [52], kiểm thử phần mềm [47] cho thấy AI đã được nghiên cứu và thử nghiệm trong nhiều lĩnh vực kỹ thuật khác nhau. Tuy nhiên, các công trình này chủ yếu giải quyết bài toán chuyên môn của từng ngành, chưa đặt trọng tâm vào vấn đề pháp lý chung của phát triển và ứng dụng AI.

Bên cạnh đó, các bài viết đăng trên tạp chí chuyên ngành ở Việt Nam cũng phản ánh sự mở rộng của AI trong nhiều lĩnh vực. Có thể kể đến các nghiên cứu về AI trong phát hiện và ngăn chặn gian lận ngân hàng [77], xây dựng thành phố thông minh [51],

quản trị nguồn nhân lực [55], chatbot trong trải nghiệm khách hàng ngành bán lẻ thương mại điện tử [59], trợ lý ảo [33], mô hình trợ giảng số kết hợp AI và IoT [29]. Các công trình này có giá trị trong việc chỉ ra khả năng ứng dụng cụ thể của AI ở Việt Nam, đồng thời làm rõ một số lợi ích thực tiễn như tự động hóa, nâng cao hiệu quả quản trị, cá nhân hóa dịch vụ, hỗ trợ ra quyết định và cải thiện trải nghiệm của người sử dụng. Tuy nhiên, do phạm vi nghiên cứu thường hẹp và gắn với từng lĩnh vực cụ thể, các công trình này chưa trả lời đầy đủ câu hỏi: cần xây dựng khung pháp lý tổng thể nào để vừa thúc đẩy ứng dụng AI trong các lĩnh vực đó, vừa bảo đảm kiểm soát rủi ro phát sinh.

Gần đây, một số báo cáo ở Việt Nam đã cung cấp bức tranh tổng quan hơn về phát triển và ứng dụng AI, có thể kể đến như sau:

- *Báo cáo thường niên về AI tại Việt Nam năm 2025* của Viện Công nghệ Thông tin, Đại học Quốc gia Hà Nội [129] cung cấp cái nhìn tổng thể về tình hình nghiên cứu, phát triển và ứng dụng AI tại Việt Nam dựa trên ba trụ cột: viện/trường, doanh nghiệp và bộ/ngành. Báo cáo này có ý nghĩa quan trọng ở chỗ chuyển cách nhìn về AI từ các ứng dụng đơn lẻ sang hệ sinh thái AI, bao gồm nghiên cứu, đào tạo, doanh nghiệp, cơ quan nhà nước và năng lực triển khai trong thực tiễn.

- *Báo cáo Nền kinh tế AI của Việt Nam 2025* do Trung tâm Đổi mới sáng tạo Quốc gia phối hợp với JICA và BCG thực hiện [70] tiếp cận AI từ góc độ phát triển kinh tế và năng lực cạnh tranh quốc gia. Báo cáo phân tích một số trụ cột của nền kinh tế AI, bao gồm mở rộng ứng dụng AI trong khu vực công và tư, phát triển hệ sinh thái AI, xây dựng nhân lực chuyên sâu, nâng cao nhận thức xã hội về AI, tiếp cận hạ tầng dữ liệu và điện toán chất lượng cao, cũng như yêu cầu có chính sách rõ ràng và thích ứng. Đây là nguồn tham khảo có giá trị đối với luận án, bởi nó cho thấy phát triển AI ở Việt Nam cần được nhìn nhận như một quá trình xây dựng năng lực quốc gia, trong đó pháp luật giữ vai trò định hướng, bảo đảm an toàn và tạo niềm tin cho đổi mới sáng tạo.

- Báo cáo *AI for Business: APAC Trends in AI Platform Adoption (AI cho doanh nghiệp: Xu hướng ứng dụng nền tảng AI tại khu vực châu Á - Thái Bình Dương)* của Deloitte năm 2025 [186] cho thấy ứng dụng AI trong doanh nghiệp tại khu vực châu Á - Thái Bình Dương, trong đó có Việt Nam, đang ngày càng trở thành xu hướng chủ đạo. Báo cáo này gợi mở rằng AI không còn là công nghệ chỉ dành cho một số doanh nghiệp công nghệ lớn, mà đang được các doanh nghiệp ở nhiều quy mô khác nhau quan tâm ứng dụng để tăng doanh thu, giảm chi phí, tối ưu hóa quy trình và nâng cao năng lực cạnh tranh. Đối với luận án, đây là cơ sở để nhấn mạnh rằng khung pháp lý về AI cần quan tâm không chỉ đến các tập đoàn công nghệ lớn, mà cả doanh nghiệp nhỏ và vừa, doanh nghiệp khởi nghiệp và các chủ thể sử dụng AI trong hoạt động thường ngày.

- Các báo cáo của Viện Nghiên cứu Chính sách và Phát triển Truyền thông (IPS), như *Chính sách phát triển và kiểm soát rủi ro đối với AI năm 2024* và *Chính sách về trí tuệ nhân tạo ở Việt Nam* [57], trực tiếp đặt vấn đề cân bằng giữa phát triển AI và kiểm soát rủi ro. Các báo cáo này nhấn mạnh rằng AI có thể thúc đẩy năng suất và tăng trưởng, nhưng cũng có thể xâm hại các giá trị cốt lõi như phẩm giá, tự do, bình đẳng, bảo vệ dữ liệu cá nhân và gắn kết xã hội. Cách tiếp cận cân bằng này phù hợp với định hướng của luận án, nhất là khi luận án coi khung pháp lý AI không chỉ là công cụ thúc đẩy đổi mới, mà còn là cơ chế bảo vệ quyền con người, lợi ích công cộng và an toàn xã hội.

- Một số báo cáo về an ninh mạng, trong đó có báo cáo tình hình an ninh mạng Việt Nam năm 2024 của NCS Group [69], cho thấy môi trường số tại Việt Nam đang đối diện với nhiều rủi ro như tấn công mã độc tống tiền, tấn công có chủ đích, lừa đảo qua email, khai thác lỗ hổng thiết bị, rò rỉ dữ liệu cá nhân và rò rỉ tài liệu nội bộ. Dù các báo cáo này không chỉ nghiên cứu riêng về AI, nhưng chúng có giá trị cảnh báo rằng việc ứng dụng AI trong bối cảnh hạ tầng dữ liệu, an ninh mạng và nhận thức người dùng còn hạn chế có thể làm gia tăng mức độ và quy mô rủi ro. Do đó, phát triển AI phải được gắn với bảo đảm an toàn thông tin, an ninh mạng và BVDLCN.

Xét tổng thể, các công trình trong nước cho thấy nghiên cứu về phát triển và ứng dụng AI ở Việt Nam đã bước đầu phản ánh được sự lan tỏa của AI trong doanh nghiệp, khu vực công và một số lĩnh vực chuyên ngành. Tuy nhiên, phần lớn các nghiên cứu vẫn thiên về mô tả ứng dụng, đánh giá lợi ích kỹ thuật hoặc đề xuất chính sách ở mức khái quát; chưa phân tích đầy đủ mối quan hệ giữa phát triển, ứng dụng AI với yêu cầu xây dựng khung pháp lý đa tầng, bao gồm văn bản quy phạm pháp luật, tiêu chuẩn kỹ thuật, cơ chế thử nghiệm có kiểm soát, luật mềm, bộ quy tắc ứng xử, thiết chế giám sát và trách nhiệm pháp lý. Đây là khoảng trống mà luận án tiếp tục nghiên cứu, nhằm làm rõ cách thức pháp luật có thể vừa tạo điều kiện cho AI phát triển, vừa kiểm soát các rủi ro phát sinh trong quá trình ứng dụng AI ở Việt Nam.

1.2. NHỮNG CÔNG TRÌNH NGHIÊN CỨU VỀ KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO

1.2.1. Nhóm công trình về nội hàm và giới hạn của khung pháp lý

Qua tổng quan tài liệu, luận án khái quát ba cách tiếp cận về phạm vi của khung pháp lý. Cách tiếp cận hẹp chịu ảnh hưởng của H.L.A. Hart [221] đồng nhất khung pháp lý với hệ thống quy phạm xác định trong thứ bậc nguồn pháp luật chính thống, xem các công cụ khác (hướng dẫn, tiêu chuẩn, quy tắc ứng xử) là "ngoài pháp luật". Cách tiếp cận rộng được đại diện bởi các nhà nghiên cứu như Julia Black và Colin Scott [172] mở rộng khung pháp lý sang khái niệm "quản trị theo mạng lưới", trong đó nhà nước chỉ là một trong nhiều mắt xích tạo ra chuẩn mực có

tính ràng buộc thực tế. Cách tiếp cận trung gian có giá trị tham khảo cao nhất được thể hiện qua công trình *The Oxford Handbook of Law, Regulation and Technology (Cẩm nang Oxford về Luật, Quy định và Công nghệ)* của Brownsword, Scotford và Yeung [323], trong đó các tác giả đề xuất khái niệm "hệ sinh thái quy phạm" bao gồm cả luật cứng, luật mềm và cơ chế kỹ thuật. So sánh sự tương đồng và khác biệt, ba cách tiếp cận (hẹp, rộng, trung gian) tương đồng ở điểm xuất phát, đều thừa nhận văn bản quy phạm pháp luật là thành tố cốt lõi không thể thiếu của bất kỳ khung pháp lý nào. Điểm khác biệt căn bản ở tiêu chí xác định ranh giới: tiếp cận hẹp dùng tiêu chí nguồn gốc thẩm quyền (chỉ cơ quan nhà nước có thẩm quyền lập pháp mới tạo ra pháp luật), trong khi tiếp cận rộng dùng tiêu chí hiệu lực hành vi thực tế (bất kỳ công cụ nào định hình hành vi một cách hệ thống đều thuộc phạm trù điều chỉnh), và tiếp cận trung gian của Brownsword, Scotford và Yeung cố gắng dung hòa hai tiêu chí bằng cách phân loại các công cụ theo mức độ chính thức.

Lý luận về quan hệ luật cứng - luật mềm có nguồn gốc từ nghiên cứu luật quốc tế của Abbott và cộng sự [254] với khái niệm "hợp pháp hóa" sau đó được vận dụng rộng rãi vào quy định công nghệ. Trong lĩnh vực AI cụ thể, nghiên cứu của Töller về "quy định tự nguyện" [149] và nghiên cứu về mô hình đồng kiến tạo của EU cung cấp khung phân tích cho câu hỏi tại sao doanh nghiệp tuân thủ tiêu chuẩn không ràng buộc, chủ yếu do áp lực thị trường, áp lực đồng nghiệp và nguy cơ mất uy tín.

Về tiêu chuẩn kỹ thuật, công trình của Majone về "quy định thông qua tiêu chuẩn hóa" [218] là nền tảng lý luận quan trọng nhất, xác định tiêu chuẩn kỹ thuật là cơ chế chuyển giao giữa ngôn ngữ pháp lý trừu tượng và yêu cầu kỹ thuật cụ thể và đề xuất khái niệm "suy đoán tuân thủ" như cơ chế pháp lý kết nối hai thế giới ngôn ngữ khác nhau đó. Nghiên cứu về EU AI Act, đặc biệt các phân tích về cơ chế "tiêu chuẩn hài hòa" theo Điều 40 cung cấp bằng chứng thực nghiệm cho thấy khi tiêu chuẩn kỹ thuật được luật hóa đúng cách, chúng trở thành cơ chế thực thi hiệu quả hơn cả quy phạm cứng vì giải quyết được vấn đề bất đối xứng thông tin giữa nhà làm luật và đối tượng điều chỉnh có kiến thức kỹ thuật.

Hạn chế chung của dòng nghiên cứu này là phần lớn tập trung phân tích một cặp quan hệ (luật cứng - luật mềm, hoặc luật - tiêu chuẩn) mà thiếu một mô hình tích hợp phân tích đồng thời cả ba thành phần trong một cấu trúc thống nhất có thể áp dụng cho thiết kế chính sách cụ thể, đặc biệt thiếu phân tích về thứ tự ưu tiên xây dựng khi cả ba thành phần đều chưa hoàn chỉnh tại một thời điểm.

Về quan hệ giữa luật cứng, luật mềm và tiêu chuẩn kỹ thuật, lý thuyết ba trục hợp pháp hóa của Abbott và cộng sự và lý thuyết tiêu chuẩn hóa của Majone tương đồng ở việc đều xem quan hệ luật cứng - luật mềm là một chuỗi liên tục thay vì hai

phạm trù tách biệt, nhưng khác biệt về trọng tâm phân tích. Abbott và cộng sự tập trung vào động cơ lựa chọn của chủ thể ban hành (tại sao nhà nước chọn mức độ ràng buộc thấp hơn một cách có chủ đích, thường vì lý do chi phí đàm phán hoặc bảo toàn chủ quyền), trong khi Majone tập trung vào cơ chế kỹ thuật kết nối (làm thế nào để một tiêu chuẩn không ràng buộc trở thành có hiệu lực gián tiếp thông qua thiết kế pháp lý). Sự khác biệt trọng tâm này phản ánh hai truyền thống nghiên cứu khác nhau, Abbott và cộng sự xuất phát từ luật quốc tế và quan hệ quốc tế, Majone xuất phát từ nghiên cứu quy định châu Âu và luận án cần lưu ý khi vận dụng đồng thời cả hai không được giả định chúng hoàn toàn tương thích về đơn vị phân tích.

Nghiên cứu các công trình trên, luận án kế thừa mô hình "hệ sinh thái quy định" của Brownsword, Scotford và Yeung như điểm xuất phát khái niệm, đồng thời kế thừa cơ sở pháp lý về mở rộng nguồn pháp luật (án lệ, lẽ công bằng) đã có trong hệ thống pháp luật Việt Nam. Luận án phản biện trường phái hẹp đồng nhất khung pháp lý với văn bản quy phạm pháp luật cho rằng cách tiếp cận này không đủ khả năng lý giải hiệu lực thực tế của các công cụ phi chính thức trong điều chỉnh AI, dẫn đến đánh giá sai lệch về mức độ "đã có luật" đồng nghĩa với "đã có khung pháp lý đầy đủ". Trên cơ sở đó, luận án phát triển khái niệm khung pháp lý theo hướng xây dựng một mô hình kiến trúc bảy tầng có mối quan hệ qua lại với nhau.

Luận án cũng kế thừa khung phân tích ba trục hợp pháp hóa của Abbott và Snidal, đặc biệt kế thừa lý luận của Majone về tiêu chuẩn kỹ thuật như cơ chế chuyển giao và suy đoán tuân thủ. Luận án phản biện xu hướng nghiên cứu tách rời từng cặp quan hệ (chỉ cứng-mềm hoặc chỉ luật-tiêu chuẩn) mà không đặt trong một hệ thống tổng thể do không đủ để giải quyết bài toán thực tiễn của một quốc gia phải xây dựng đồng thời cả ba thành phần trong điều kiện nguồn lực hạn chế.

1.2.2. Nhóm công trình về nền tảng lý luận của khung pháp lý trí tuệ nhân tạo

Nghiên cứu về khung pháp lý cho phát triển, ứng dụng AI không thể tách rời các nghiên cứu lý luận về quan hệ giữa pháp luật, công nghệ và quản trị rủi ro. Nếu các công trình về AI nói chung giúp nhận diện bản chất và tác động của AI, thì nhóm công trình về pháp luật và AI cung cấp cơ sở lý luận để trả lời các câu hỏi: vì sao AI cần một cách tiếp cận pháp lý riêng; pháp luật nên can thiệp ở mức độ nào; khung pháp lý AI nên bao gồm những thành tố nào; và làm thế nào để pháp luật vừa thúc đẩy đổi mới sáng tạo, vừa bảo vệ con người, xã hội và lợi ích công cộng.

Liên quan đến nhóm công trình nghiên cứu này, Roger Brownsword và Morag Goodwin trong *Law and the Technologies of the Twenty-First Century: Text and Materials (Luật và Công nghệ của thế kỷ 21: Văn bản và Tài liệu)* đã đặt nền tảng quan

trọng cho lý luận về vai trò của pháp luật trong bối cảnh công nghệ mới [324]. Công trình này không chỉ xem pháp luật như tập hợp các quy phạm điều chỉnh hành vi, mà đặt pháp luật trong một môi trường điều chỉnh rộng hơn, bao gồm thị trường, chuẩn mực xã hội, đạo đức, kỹ thuật và thiết kế công nghệ. Cách tiếp cận này có ý nghĩa trực tiếp đối với luận án, bởi khung pháp lý AI không thể chỉ được hiểu là một đạo luật hoặc một nhóm văn bản quy phạm pháp luật, mà cần được nhìn nhận như một cấu trúc điều chỉnh đa tầng, kết hợp giữa pháp luật cứng, pháp luật mềm, tiêu chuẩn kỹ thuật, thiết chế thực thi và cơ chế tự quản.

Tiếp nối hướng tiếp cận này, *The Oxford Handbook of Law, Regulation and Technology (Cẩm nang Oxford về Luật, Quy định và Công nghệ)* do Roger Brownsword, Eloise Scotford và Karen Yeung chủ biên là một công trình có tính nền tảng về quan hệ giữa pháp luật, điều chỉnh và công nghệ [323]. Giá trị của công trình này nằm ở chỗ phân tích công nghệ không chỉ như đối tượng bị pháp luật điều chỉnh, mà còn như một yếu tố làm thay đổi phương thức điều chỉnh pháp luật. Đối với nghiên cứu về AI, cách tiếp cận này giúp lý giải vì sao khung pháp lý AI cần vượt ra ngoài tư duy lập pháp truyền thống, đồng thời phải quan tâm đến tiêu chuẩn kỹ thuật, thiết kế hệ thống, quản trị dữ liệu, trách nhiệm giải trình và năng lực của thiết chế thực thi.

Ở cấp độ chính sách AI, Ryan Calo trong bài viết *Artificial Intelligence Policy: A Primer and Roadmap (Chính sách AI: Khái niệm cơ bản và lộ trình)* đã xây dựng một bản đồ tương đối sớm và có ảnh hưởng về các vấn đề chính sách mà AI đặt ra, bao gồm công bằng, quyền riêng tư, an toàn, trách nhiệm, lao động, thiết chế quản lý và năng lực của Nhà nước [326]. Công trình này có giá trị đối với luận án vì cho thấy pháp luật AI không phải là một lĩnh vực biệt lập, mà là điểm giao thoa giữa nhiều lĩnh vực pháp luật và chính sách công. Từ đó, việc xây dựng khung pháp lý AI cần được tiếp cận như một quá trình thiết kế chính sách pháp luật liên ngành, thay vì chỉ bổ sung một số quy định rời rạc về công nghệ.

Đặc biệt, Matthew U. Scherer trong bài viết *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies (Điều chỉnh các hệ thống AI: Rủi ro, Thách thức, Năng lực và Chiến lược)* đã phân tích một cách hệ thống các rủi ro, thách thức và năng lực thể chế cần thiết để điều chỉnh hệ thống AI [270]. Công trình này có ý nghĩa nền tảng vì chỉ ra rằng AI đặt ra những khó khăn đặc thù cho pháp luật, nhất là do tính phức tạp, khả năng tự học, khó dự báo và khó quy trách nhiệm của hệ thống AI. Từ đó, Scherer gợi mở yêu cầu thiết kế cơ chế điều chỉnh phù hợp với rủi ro và năng lực của cơ quan quản lý. Đây là một trong những cơ sở lý luận quan trọng cho cách tiếp cận dựa trên rủi ro trong xây dựng khung pháp lý AI.

Trong các công trình chuyên khảo gần đây, cuốn *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice (Luật và AI: Điều*

chỉnh AI và ứng dụng AI trong thực tiễn pháp lý) do Bart Custers và Eduard Fosch-Villaronga chủ biên là một công trình có phạm vi tương đối toàn diện về pháp luật và AI [157]. Công trình này phân tích cả hai chiều cạnh: pháp luật điều chỉnh AI và AI được ứng dụng trong thực tiễn pháp lý. Về chiều cạnh thứ nhất, cuốn sách đề cập đến nhiều lĩnh vực như luật hiến pháp, quyền con người, luật hình sự, luật thuế, trách nhiệm dân sự, cạnh tranh và bảo vệ người tiêu dùng. Đối với luận án, công trình này có giá trị ở chỗ cho thấy khung pháp lý AI phải có tính liên ngành, bởi AI tác động đồng thời đến nhiều ngành luật và khó có thể được điều chỉnh đầy đủ bằng một đạo luật đơn lẻ.

Marwala và Mpedi trong *Artificial Intelligence and the Law* (AI và Pháp luật) tiếp cận vấn đề từ góc độ tác động của AI đối với pháp luật, đạo đức và hệ thống thực thi pháp luật [343]. Công trình này giúp làm rõ rằng AI không chỉ là đối tượng điều chỉnh của pháp luật, mà còn có thể làm thay đổi chính hoạt động thực thi pháp luật, từ điều tra, truy tố, xét xử, thi hành án đến quản trị hành chính. Vì vậy, khung pháp lý AI cần bao gồm cả các quy định về việc phát triển và sử dụng AI trong khu vực công, nhất là các lĩnh vực có khả năng ảnh hưởng trực tiếp đến quyền con người, quyền công dân và nguyên tắc pháp quyền. Ở bình diện khu vực, cuốn *The Challenges of Artificial Intelligence for Law in Europe (Những thách thức của AI đối với pháp luật ở châu Âu)* do Marton Varju và Peter Mezei chủ biên cung cấp bức tranh về cách tiếp cận của châu Âu đối với điều chỉnh AI [268]. Công trình này có giá trị tham khảo vì châu Âu là khu vực đi đầu trong xây dựng khung pháp lý AI theo mô hình dựa trên rủi ro, gắn kết giữa bảo vệ quyền cơ bản, quản trị dữ liệu, đạo đức AI và thúc đẩy đổi mới sáng tạo. Tuy nhiên, đối với luận án, kinh nghiệm châu Âu cần được tiếp thu có chọn lọc, bởi điều kiện chính trị, pháp lý, năng lực thực thi và trình độ phát triển thị trường AI của Việt Nam có nhiều điểm khác biệt so với EU.

Cùng hướng nghiên cứu liên ngành, *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence (Cẩm nang Cambridge về Luật, Đạo đức và Chính sách AI)* do Nathalie A. Smuha chủ biên là một công trình tổng hợp mới bao quát nhiều vấn đề pháp lý, đạo đức và chính sách của AI [287]. Điểm đáng chú ý của công trình này là không tách rời pháp luật khỏi đạo đức và chính sách công, mà xem đây là ba trụ cột có quan hệ chặt chẽ trong quản trị AI. Cách tiếp cận này phù hợp với luận án, bởi khung pháp lý AI ở Việt Nam cần kết hợp giữa quy phạm pháp luật bắt buộc, tiêu chuẩn kỹ thuật, quy tắc đạo đức, hướng dẫn chuyên môn và cơ chế giám sát thực thi.

Một nguồn có ý nghĩa lịch sử trong tranh luận pháp lý về AI và robot là *Report with Recommendations to the Commission on Civil Law Rules on Robotics (Báo cáo*

kèm theo các khuyến nghị gửi Ủy ban về Quy tắc Luật Dân sự đối với Robot) của Ủy ban Pháp lý Nghị viện châu Âu, gắn với báo cáo của Mady Delvaux và cộng sự năm 2017 [202]. Dù đây không phải là công trình học thuật thuần túy, văn kiện này có giá trị vì đã sớm đặt ra các vấn đề như trách nhiệm dân sự, bảo hiểm, đăng ký robot, trách nhiệm của nhà sản xuất, người vận hành và tranh luận về "nhân cách pháp lý điện tử". Các tranh luận này sau đó không được tiếp nhận nguyên vẹn trong pháp luật châu Âu, nhưng vẫn có ý nghĩa tham khảo khi nghiên cứu trách nhiệm pháp lý đối với các hệ thống AI có mức độ tự chủ cao.

Ở Việt Nam, các công trình nghiên cứu về nền tảng lý luận của khung pháp lý AI còn tương đối hạn chế và thường được đặt trong bối cảnh rộng hơn của chuyển đổi số, phát triển khoa học - công nghệ, đổi mới sáng tạo và hoàn thiện thể chế. Đề tài cấp Bộ trọng điểm năm 2024-2025 *Nghiên cứu hoàn thiện thể chế phát triển khoa học và công nghệ ở Việt Nam trong bối cảnh mới* do PGS.TS Tào Thị Quyên chủ nhiệm [92] là một công trình có giá trị tham khảo quan trọng. Công trình này phân tích thể chế phát triển khoa học và công nghệ như một bộ phận của quá trình thể chế hóa chủ trương, đường lối của Đảng thành pháp luật, gắn với chiến lược phát triển khoa học, công nghệ, đổi mới sáng tạo và phát triển kinh tế - xã hội. Đáng chú ý, công trình nhấn mạnh cơ chế thử nghiệm có kiểm soát cần được thiết kế như một thiết chế gắn với hệ sinh thái đổi mới sáng tạo, làm cầu nối giữa thử nghiệm công nghệ, thử nghiệm chính sách và thương mại hóa sản phẩm. Đây là luận điểm có ý nghĩa trực tiếp đối với việc xây dựng khung pháp lý AI ở Việt Nam, nhất là trong bối cảnh AI phát triển nhanh, khó dự báo và cần không gian thử nghiệm an toàn.

Bài viết "Khung pháp lý để thực hiện các mục tiêu chuyển đổi số trong bối cảnh sự trỗi dậy của quốc gia Việt Nam" của Lê Lan Anh và Lê Thị Kim Oanh [4] tiếp cận khung pháp lý từ yêu cầu thực hiện chuyển đổi số đồng thời chỉ ra những thách thức về sự thiếu đồng bộ, chồng chéo và chậm trễ của quy định pháp luật liên quan đến AI, tài sản số và các mô hình kinh doanh mới. Bài viết cũng đề cập đến nhu cầu thiết kế các cơ chế thử nghiệm pháp lý nhằm thúc đẩy đổi mới sáng tạo trong khi vẫn kiểm soát rủi ro. Tuy nhiên, vì phạm vi của bài viết là chuyển đổi số nói chung, vấn đề khung pháp lý AI mới chỉ được đặt ra như một bộ phận của khung pháp lý chuyển đổi số, chưa được phân tích như một chỉnh thể pháp lý độc lập với cấu trúc, nguyên tắc và kỹ thuật xây dựng riêng.

Tổng quan nhóm công trình nêu trên cho thấy nghiên cứu quốc tế về pháp luật và AI đã phát triển khá mạnh, từ lý thuyết chung về pháp luật và công nghệ đến các nghiên cứu chuyên sâu về điều chỉnh AI, trách nhiệm pháp lý, đạo đức AI, pháp luật mềm, quyền con người, quản trị rủi ro và kinh nghiệm châu Âu. Trong khi đó, các nghiên cứu ở Việt Nam mới chủ yếu tiếp cận AI trong bối cảnh chuyển đổi số, phát

triển khoa học - công nghệ hoặc hoàn thiện thể chế đổi mới sáng tạo; chưa có nhiều công trình nghiên cứu khung pháp lý cho phát triển, ứng dụng AI như một chỉnh thể đa tầng, bao gồm pháp luật cứng, pháp luật mềm, tiêu chuẩn kỹ thuật, cơ chế thử nghiệm, tự quản, thiết chế thực thi và cơ chế phản hồi chính sách. Khoảng trống này là cơ sở để luận án tiếp tục nghiên cứu, nhằm xây dựng nền tảng lý luận phù hợp cho khung pháp lý AI ở Việt Nam.

1.2.3. Nhóm công trình về phân tích so sánh mô hình khung pháp lý trí tuệ nhân tạo

Bên cạnh các công trình đặt nền tảng lý luận chung về pháp luật và AI, một hướng nghiên cứu quan trọng khác tập trung phân tích, so sánh các mô hình quản trị và khung pháp lý AI. Nhóm công trình này có ý nghĩa trực tiếp đối với luận án, bởi việc xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam cần tham khảo kinh nghiệm quốc tế, nhưng không thể sao chép nguyên mẫu một mô hình pháp lý cụ thể. Các nghiên cứu so sánh cho thấy khung pháp lý AI trên thế giới đang vận động theo nhiều hướng khác nhau: mô hình dựa trên rủi ro của châu Âu; mô hình linh hoạt, thúc đẩy đổi mới sáng tạo của Hoa Kỳ và Anh; mô hình quản trị theo chiến lược quốc gia, tiêu chuẩn và kiểm soát nền tảng của Trung Quốc; mô hình luật mềm, hướng dẫn đạo đức và đồng quản trị của một số quốc gia châu Á; và mô hình hợp tác quốc tế nhằm xử lý các rủi ro xuyên biên giới của AI.

Một trong những nghiên cứu sớm và có ảnh hưởng về chính sách AI là bài viết "*Artificial Intelligence Policy: A Primer and Roadmap*" *(Chính sách AI: Khái niệm cơ bản và lộ trình)* của Ryan Calo [326]. Công trình này đề xuất một bản đồ chính sách tương đối toàn diện đối với AI, bao gồm các nhóm vấn đề như công lý và bình đẳng, sử dụng vũ lực, an toàn và chứng nhận, quyền riêng tư, thuế và tác động của AI đối với lao động. Giá trị của bài viết không nằm ở việc xây dựng một mô hình pháp lý hoàn chỉnh, mà ở chỗ chỉ ra AI là vấn đề chính sách công đa ngành, đòi hỏi Nhà nước phải chủ động xây dựng năng lực chuyên môn, sử dụng chính sách đầu tư và mua sắm công, tháo gỡ rào cản pháp lý và định hình nhận thức xã hội về AI. Cách tiếp cận này gợi mở cho luận án rằng khung pháp lý AI cần được đặt trong một chiến lược quản trị rộng hơn, bao gồm cả chính sách phát triển, chính sách dữ liệu, chính sách lao động, an ninh, quyền riêng tư và bảo vệ nhóm yếu thế.

Ở cấp độ quản trị toàn cầu, Allan Dafoe trong "*AI Governance: A Research Agenda*" *(Quản trị AI: Chương trình nghiên cứu)* [142] đề xuất chương trình nghiên cứu về quản trị AI với ba trụ cột lớn: bối cảnh kỹ thuật của AI, bối cảnh chính trị - chiến lược của phát triển AI và mô hình quản trị AI lý tưởng. Công trình này có giá trị vì không chỉ xem AI như vấn đề quản lý trong phạm vi quốc gia, mà còn như vấn đề

quản trị toàn cầu, gắn với cạnh tranh công nghệ, hợp tác quốc tế, phân bổ lợi ích và kiểm soát rủi ro dài hạn. Đối với luận án, hướng tiếp cận này giúp làm rõ rằng khung pháp lý AI của Việt Nam cần được thiết kế trong mối liên hệ với các chuẩn mực, thỏa thuận và xu hướng quản trị quốc tế, nhất là trong bối cảnh AI có tính xuyên biên giới rất cao.

Cũng từ góc độ quản trị toàn cầu, Erdélyi và Goldsmith trong "*Regulating Artificial Intelligence: Proposal for a Global Solution*" *(Điều chỉnh AI: Đề xuất một giải pháp toàn cầu)* [303] lập luận rằng các thách thức do AI đặt ra khó có thể được xử lý hiệu quả nếu thiếu điều phối quốc tế. Hai tác giả đề xuất hình thành một cơ chế quản trị AI ở cấp quốc tế, trước hết có thể dưới dạng một tổ chức liên chính phủ linh hoạt, kết hợp cơ chế luật mềm, tiêu chuẩn chung và mức độ ràng buộc pháp lý phù hợp. Công trình này có ý nghĩa tham khảo vì cho thấy khung pháp lý AI quốc gia không thể tách rời nhu cầu hài hòa tiêu chuẩn, chia sẻ thông tin và phối hợp quản trị giữa các quốc gia. Tuy nhiên, đối với Việt Nam, các đề xuất về cơ chế quản trị toàn cầu cần được tiếp thu một cách thận trọng, trên cơ sở bảo đảm chủ quyền số, an ninh quốc gia và lợi ích phát triển của quốc gia đang phát triển.

Đáng chú ý, Gasser và Almeida trong "*A Layered Model for AI Governance*" *(Mô hình quản trị AI theo tầng)* [363] đề xuất thay vì chỉ tập trung vào một công cụ điều chỉnh duy nhất, mô hình này nhìn quản trị AI như một cấu trúc gồm nhiều lớp, từ chuẩn mực xã hội, đạo đức, thiết kế kỹ thuật, tiêu chuẩn, quy định pháp luật đến thiết chế giám sát. Đây là nguồn tham khảo quan trọng đối với luận án, bởi cách tiếp cận theo tầng giúp tránh đồng nhất khung pháp lý AI với một đạo luật chuyên biệt. Nó cũng củng cố cơ sở lý luận cho việc luận án tiếp cận khung pháp lý AI ở Việt Nam như một chỉnh thể đa tầng, bao gồm pháp luật cứng, pháp luật mềm, tiêu chuẩn kỹ thuật, cơ chế tự quản và thiết chế thực thi.

Về mô hình châu Âu, Veale và Zuiderveen Borgesius trong "*Demystifying the Draft EU Artificial Intelligence Act*" *(Làm rõ Dự thảo Đạo luật AI của EU)* [273] phân tích những điểm mạnh, hạn chế và hệ quả pháp lý của dự thảo Đạo luật AI của Liên minh châu Âu. Công trình này có ý nghĩa vì làm rõ nền tảng của mô hình điều chỉnh dựa trên rủi ro: các hệ thống AI không được điều chỉnh bằng cùng một mức độ nghĩa vụ, mà được phân loại theo mức độ rủi ro để xác định các yêu cầu pháp lý tương ứng. Tuy nhiên, các tác giả cũng chỉ ra những vấn đề đáng lưu ý như nguy cơ quy định quá phức tạp, khó thực thi, khả năng hài hòa tối đa làm hạn chế chính sách quốc gia và khoảng cách giữa mục tiêu bảo vệ quyền với cơ chế thực thi thực tế. Đối với luận án, công trình này giúp tiếp cận kinh nghiệm EU một cách cân bằng hơn: EU AI Act là nguồn tham khảo rất quan trọng, nhưng không nên được xem là khuôn mẫu duy nhất cho Việt Nam.

Ở chiều cạnh so sánh ngoài châu Âu, Roberts và cộng sự trong "*The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation*" *(Cách tiếp cận của Trung Quốc đối với AI: Phân tích về chính sách, đạo đức và quy định)* [224] phân tích cách tiếp cận của Trung Quốc đối với phát triển và quản trị AI thông qua chiến lược quốc gia, chính sách công nghiệp, quản trị dữ liệu, an ninh, tiêu chuẩn đạo đức và điều chỉnh theo từng nhóm công nghệ. Công trình này có giá trị vì cho thấy một mô hình khác với châu Âu. Thay vì bắt đầu bằng một đạo luật AI tổng quát, Trung Quốc phát triển khung quản trị AI thông qua chiến lược nhà nước mạnh, các quy định chuyên biệt đối với thuật toán khuyến nghị, phân tích sâu, AI tạo sinh và quản lý nền tảng số. Đối với Việt Nam, kinh nghiệm này có ý nghĩa tham khảo ở khía cạnh điều chỉnh theo công nghệ và theo lĩnh vực rủi ro, nhưng cũng cần được đánh giá thận trọng từ góc độ bảo vệ quyền con người, minh bạch, trách nhiệm giải trình và kiểm soát quyền lực công.

Một hướng nghiên cứu khác tập trung vào mối quan hệ giữa AI, bảo vệ dữ liệu và minh bạch thuật toán. Wachter, Mittelstadt và Russell trong "*Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR*" *(Giải thích phản sự thật không cần mở "hộp đen": Quyết định tự động và GDPR)* [329] là nghiên cứu có ảnh hưởng lớn về quyền được giải thích trong bối cảnh các quyết định tự động. Các tác giả cho rằng việc mở "hộp đen" của hệ thống AI trong mọi trường hợp có thể gặp nhiều rào cản kỹ thuật và pháp lý, nhưng vẫn có thể cung cấp các giải thích phản thực tế nhằm giúp cá nhân hiểu vì sao họ nhận được một kết quả bất lợi, có căn cứ để khiếu nại và biết cần thay đổi điều gì để đạt kết quả khác trong tương lai. Công trình này có ý nghĩa trực tiếp đối với khung pháp lý AI, bởi minh bạch trong AI không chỉ là minh bạch về mã nguồn hoặc mô hình kỹ thuật, mà còn phải bảo đảm khả năng hiểu, khả năng phản đối và khả năng tiếp cận biện pháp khắc phục của cá nhân chịu tác động.

Tổng quan nhóm công trình nêu trên cho thấy các nghiên cứu so sánh về khung pháp lý AI đã hình thành một số hướng tiếp cận quan trọng. Thứ nhất, AI cần được quản trị bằng một cấu trúc liên ngành và đa tầng, thay vì chỉ bằng một đạo luật chuyên biệt. Thứ hai, mô hình điều chỉnh dựa trên rủi ro đang có ảnh hưởng lớn, nhưng không phải là mô hình duy nhất và cũng không tránh khỏi khó khăn trong phân loại, thực thi và cập nhật. Thứ ba, quản trị AI quốc gia phải gắn với hợp tác quốc tế, do dữ liệu, mô hình, sản phẩm và tác động của AI đều có tính xuyên biên giới. Thứ tư, bảo vệ quyền con người, minh bạch, giải trình và khả năng tiếp cận biện pháp khắc phục là những tiêu chí quan trọng để đánh giá chất lượng của mọi mô hình khung pháp lý AI.

Tuy nhiên, các công trình quốc tế nói trên chủ yếu được xây dựng trên cơ sở kinh nghiệm của các nước phát triển hoặc các trung tâm quyền lực công nghệ lớn như

EU, Hoa Kỳ, Anh và Trung Quốc. Các nghiên cứu này chưa giải quyết đầy đủ câu hỏi về việc một quốc gia đang phát triển như Việt Nam, với điều kiện thể chế, năng lực thực thi, thị trường AI và hạ tầng dữ liệu còn đang hình thành, nên lựa chọn và kết hợp các yếu tố nào để xây dựng khung pháp lý AI phù hợp. Đây là khoảng trống mà luận án tiếp tục nghiên cứu, thông qua việc tiếp thu có chọn lọc các mô hình quốc tế, đồng thời đặt chúng trong điều kiện chính trị, pháp lý, kinh tế, xã hội và yêu cầu phát triển của Việt Nam.

1.2.4. Nhóm công trình về phân tích khung pháp lý trí tuệ nhân tạo ở các quốc gia cụ thể

Nhóm công trình nghiên cứu này tập trung phân tích chính sách, pháp luật và mô hình quản trị AI ở một số quốc gia, khu vực cụ thể. So với nhóm công trình nghiên cứu về ở mục trên, các nghiên cứu trong nhóm này có ý nghĩa ở chỗ đi sâu hơn vào bối cảnh thể chế, chính sách, kỹ thuật lập pháp và cách thức tổ chức thực thi của từng quốc gia hoặc khu vực. Đây là nguồn tham khảo quan trọng đối với luận án, bởi việc xây dựng khung pháp lý AI ở Việt Nam cần dựa trên sự tiếp thu có chọn lọc kinh nghiệm quốc tế, đồng thời phải tính đến điều kiện chính trị, pháp lý, kinh tế, xã hội và năng lực thực thi của Việt Nam.

Các công trình nghiên cứu ở nước ngoài về khung pháp lý AI ở một số quốc gia, khu vực tiêu biểu khá phong phú. Ví dụ, về Trung Quốc, Huw Roberts và cộng sự trong "*The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation*" *(Cách tiếp cận của Trung Quốc đối với AI: Phân tích về chính sách, đạo đức và quy định)* đã phác họa bức tranh tương đối toàn diện về chính sách, đạo đức và điều chỉnh AI ở Trung Quốc [224]. Công trình này cho thấy Trung Quốc tiếp cận AI trước hết như một vấn đề chiến lược quốc gia, gắn với mục tiêu trở thành trung tâm đổi mới AI hàng đầu thế giới, đồng thời sử dụng kết hợp chính sách công nghiệp, quản trị dữ liệu, tiêu chuẩn đạo đức và các quy định chuyên biệt. Giá trị của công trình này đối với luận án là làm rõ một mô hình khác với cách tiếp cận của châu Âu: thay vì bắt đầu bằng một đạo luật AI tổng quát, Trung Quốc phát triển khung quản trị AI thông qua chiến lược nhà nước mạnh, các văn bản chính sách và các quy định theo từng nhóm công nghệ, lĩnh vực hoặc rủi ro cụ thể. Bổ sung cho hướng tiếp cận này, Yang và Huang trong "*Quantitative Mapping of the Evolution of AI Policy Distribution, Targets and Focuses over Three Decades in China*" *(Phân tích định lượng sự tiến hóa của việc phân bổ, mục tiêu và trọng tâm chính sách AI trong ba thập kỷ qua tại Trung Quốc)* sử dụng phương pháp định lượng và trắc lượng thư mục để phân tích sự phát triển chính sách AI ở Trung Quốc trong thời gian dài [176]. Công trình này có giá trị vì không chỉ mô tả nội dung chính sách, mà còn chỉ ra sự thay đổi về tần suất ban hành, chủ thể ban hành, mục tiêu chính sách và trọng tâm điều chỉnh qua các giai đoạn. Qua đó, nghiên

cứu giúp nhận diện chính sách AI như một quá trình tích lũy thể chế, chứ không phải là kết quả của một văn bản pháp lý đơn lẻ. Một nghiên cứu đáng lưu ý nữa về Trung Quốc là bài viết "Kế hoạch trở thành quốc gia dẫn đầu thế giới về trí tuệ nhân tạo của Trung Quốc" của Bạch Tân Sinh [94] là nguồn tham khảo hữu ích về mặt chính sách phát triển AI, đặc biệt trong việc hiểu cách Trung Quốc kết hợp chiến lược công nghiệp, đầu tư nhà nước, phát triển doanh nghiệp công nghệ và quản trị dữ liệu.

Đối với Liên minh châu Âu, Veale và Zuiderveen Borgesius trong "*Demystifying the Draft EU Artificial Intelligence Act*" *(Làm rõ Dự thảo Đạo luật AI của EU)* và Smuha trong "*From a 'Race to AI' to a 'Race to AI Regulation': Regulatory Competition for Artificial Intelligence*" *(Từ "Cuộc đua AI" đến "Cuộc đua quy định về AI": Cạnh tranh về quy định đối với AI)* là hai công trình quan trọng về cách tiếp cận pháp lý của EU đối với AI. Veale và Zuiderveen Borgesius đi sâu phân tích cấu trúc kỹ thuật của dự thảo Đạo luật AI, đặc biệt là cách phân loại rủi ro, các nghĩa vụ đối với hệ thống AI rủi ro cao và vấn đề thực thi [273]. Trong khi đó, Smuha đặt EU AI Act trong bối cảnh cạnh tranh điều chỉnh toàn cầu, cho thấy cuộc đua phát triển AI đang dần đi kèm với cuộc đua thiết lập tiêu chuẩn pháp lý và đạo đức cho AI [286]. Hai công trình này giúp luận án nhận diện EU như một mô hình điều chỉnh dựa trên rủi ro, có tham vọng định hình chuẩn mực toàn cầu, nhưng đồng thời cũng đặt ra thách thức về độ phức tạp của quy định, chi phí tuân thủ và năng lực thực thi.

Đối với Vương quốc Anh, Drake và cộng sự trong "*Legal Contestation of Artificial Intelligence-Related Decision-Making in the United Kingdom: Reflections for Policy*" *(Tranh chấp pháp lý về hệ thống tự động hóa ra quyết định liên quan đến AI tại Vương quốc Anh: Gợi ý chính sách)* phân tích các tranh chấp, phản biện và thách thức pháp lý liên quan đến việc sử dụng AI và ra quyết định tự động trong bối cảnh chính sách của Anh [152]. Công trình này nhấn mạnh rằng sự thiếu minh bạch, hạn chế về nhận thức của người bị tác động và mức độ tin cậy của công chúng là những vấn đề trung tâm khi AI được sử dụng trong các quyết định có ảnh hưởng đến cá nhân. Cùng hướng này, Roberts và cộng sự trong "*Artificial Intelligence Regulation in the United Kingdom: A Path to Good Governance and Global Leadership?*" *(Quy định về AI tại Vương quốc Anh: Con đường hướng tới quản trị tốt và vai trò lãnh đạo toàn cầu?)* phân tích cách tiếp cận "thân thiện với đổi mới" và dựa vào các cơ quan quản lý ngành của Anh [225]. Hai công trình này cho thấy mô hình của Anh có ưu điểm ở tính linh hoạt, tránh tạo thêm một tầng quản lý tập trung cứng nhắc; nhưng cũng có hạn chế là phụ thuộc nhiều vào năng lực, thẩm quyền và sự phối hợp của các cơ quan quản lý hiện có.

Đối với Nhật Bản và Singapore, các nghiên cứu và văn kiện chính sách gần đây cho thấy một cách tiếp cận mềm dẻo hơn, nhấn mạnh hướng dẫn, tiêu chuẩn, trách

nhiệm tự nguyện và đồng quản trị. Nhật Bản thường được mô tả như mô hình ưu tiên thúc đẩy nghiên cứu, phát triển và ứng dụng AI, đồng thời sử dụng hướng dẫn đạo đức, quy tắc thực hành và diễn giải pháp luật hiện hành để xử lý rủi ro [256]. Singapore lại nổi bật với cách tiếp cận thực dụng, xây dựng các khuôn khổ quản trị AI và công cụ kiểm định, đánh giá nhằm hỗ trợ doanh nghiệp triển khai AI có trách nhiệm [231]. Các kinh nghiệm này đặc biệt có giá trị đối với Việt Nam vì đều cho thấy khung pháp lý AI không nhất thiết phải bắt đầu bằng các nghĩa vụ pháp lý quá cứng, mà có thể kết hợp pháp luật cứng với luật mềm, tiêu chuẩn kỹ thuật, cơ chế thử nghiệm và công cụ tự đánh giá, kiểm định.

Ở Việt Nam cũng đã có một số công trình nghiên cứu ở về kinh nghiệm pháp luật AI của một số quốc gia, chủ yếu được thực hiện dưới dạng bài viết chuyên khảo, bình luận chính sách hoặc nghiên cứu kinh nghiệm quốc tế để rút ra bài học cho Việt Nam. Các công trình này có giá trị tham khảo, nhất là trong giai đoạn Việt Nam đang xây dựng và hoàn thiện pháp luật về AI, nhưng phần lớn còn thiên về mô tả chính sách quốc gia, chưa phát triển thành khung phân tích so sánh đầy đủ. Tiêu biểu cho các công trình này là bài viết "Chính sách pháp luật về phát triển trí tuệ nhân tạo của Mỹ và một số bài học kinh nghiệm cho Việt Nam" của Phạm Thanh Tùng, Nguyễn Thủy Nguyên và Nguyễn Tiến Nhật Quang [106] phân tích chính sách pháp luật của Hoa Kỳ về thúc đẩy phát triển AI, nhất là vai trò của khu vực tư nhân, đầu tư công nghệ và nhu cầu xây dựng khuôn khổ quản lý phù hợp. Bài viết có giá trị ở chỗ gợi mở bài học cho Việt Nam từ mô hình Hoa Kỳ, nơi AI được phát triển mạnh nhờ doanh nghiệp công nghệ, đại học, nghiên cứu quốc phòng, cơ chế đầu tư và hệ sinh thái đổi mới sáng tạo. Tuy nhiên, khi vận dụng kinh nghiệm Hoa Kỳ, cần lưu ý rằng mô hình này dựa nhiều vào cơ chế thị trường, năng lực tự điều chỉnh của doanh nghiệp, vai trò của tòa án và hệ thống thực thi pháp luật phân tán, những yếu tố không hoàn toàn tương đồng với điều kiện Việt Nam. Một bài viết khác có tiêu đề "Chiến lược phát triển công nghệ trí tuệ nhân tạo của Nhật Bản hiện nay" của Phạm Thu Thuỷ [120] làm rõ bối cảnh Nhật Bản từng là một quốc gia tiên phong về công nghệ AI nhưng phải điều chỉnh chiến lược để đáp ứng cạnh tranh mới từ Hoa Kỳ, Trung Quốc và châu Âu. Bài viết có giá trị trong việc chỉ ra cách Nhật Bản gắn AI với yêu cầu phục hồi năng lực cạnh tranh, giải quyết các vấn đề xã hội như già hóa dân số, năng suất lao động và đổi mới công nghiệp. Tuy nhiên, trọng tâm của bài viết là chiến lược phát triển công nghệ, còn khía cạnh pháp lý và thiết chế quản trị AI mới chỉ được đề cập ở mức khái quát. Bên cạnh đó, bài viết "Quản lý trí tuệ nhân tạo: kinh nghiệm quốc tế và gợi mở với Việt Nam" của Nguyễn Đức Thuỷ [115] có phạm vi so sánh rộng hơn, đề cập kinh nghiệm thúc đẩy và quản lý AI của một số quốc gia như Hoa Kỳ, Trung Quốc, Pháp, đồng thời rút ra một số bài

học cho Việt Nam. Công trình này có giá trị ở chỗ đặt vấn đề quản lý AI trong tương quan giữa thúc đẩy phát triển và kiểm soát rủi ro, nhấn mạnh các nội dung như quản trị đạo đức, khả năng giải thích, kiểm soát rủi ro và trách nhiệm của các chủ thể. Tuy nhiên, do phạm vi rộng, bài viết chưa phân tích sâu cấu trúc nội tại của từng mô hình khung pháp lý AI cũng như điều kiện để chuyển hóa các kinh nghiệm quốc tế vào bối cảnh Việt Nam.

Một số công trình nghiên cứu ở Việt Nam còn tiếp cận kinh nghiệm nước ngoài từ các lĩnh vực pháp luật cụ thể. Ví dụ, bài viết "Nhận biết những tác động của trí tuệ nhân tạo đến pháp luật sáng chế của Liên minh châu Âu" của Đặng Minh Phương và Nguyễn Như Hà [78] phân tích tác động của AI đối với pháp luật sáng chế EU, đặc biệt là các vấn đề về đối tượng sáng chế, chủ thể sáng chế và điều kiện bảo hộ. Công trình này có giá trị ở chỗ cho thấy AI không chỉ đòi hỏi một đạo luật chung, mà còn tác động sâu đến các ngành luật cụ thể như sở hữu trí tuệ. Tuy nhiên, do phạm vi hẹp, công trình này chủ yếu có giá trị minh họa cho tác động ngành của AI, hơn là phân tích khung pháp lý AI tổng thể. Bài viết "Chiến lược xây dựng đạo đức trí tuệ nhân tạo tại CHLB Đức và một số gợi mở cho Việt Nam" của Hoa Hữu Cường và Chử Thị Nhuần [19] phân tích các định hướng đạo đức AI của Đức, nhất là yêu cầu bảo vệ dữ liệu, quyền riêng tư, phòng ngừa thiên lệch và bảo đảm trách nhiệm của doanh nghiệp cũng như cơ quan quản lý nhà nước. Công trình này có giá trị tham khảo vì Đức là một trong những quốc gia nhấn mạnh mạnh mẽ đến đạo đức, quyền con người, dữ liệu cá nhân và trách nhiệm xã hội trong phát triển AI. Tuy nhiên, trong bối cảnh hiện nay, kinh nghiệm của Đức cần được đặt trong khuôn khổ rộng hơn của EU AI Act và hệ thống pháp luật châu Âu về bảo vệ dữ liệu, cạnh tranh và quyền cơ bản.

Nhìn chung, các công trình về khung pháp lý AI ở các quốc gia cụ thể đã cung cấp nhiều tư liệu có giá trị về EU, Trung Quốc, Hoa Kỳ, Anh, Nhật Bản, Singapore, Đức và một số quốc gia khác. Các công trình này cho thấy không tồn tại một mô hình duy nhất cho điều chỉnh AI. EU thiên về mô hình pháp luật cứng dựa trên rủi ro; Anh thiên về mô hình điều chỉnh theo ngành và thân thiện với đổi mới; Trung Quốc kết hợp chiến lược nhà nước mạnh với các quy định chuyên biệt theo công nghệ; Nhật Bản và Singapore chú trọng luật mềm, hướng dẫn, tiêu chuẩn và đồng quản trị; Hoa Kỳ dựa nhiều vào mô hình ngành, tiêu chuẩn kỹ thuật, thực thi của các cơ quan hiện có và vai trò lớn của khu vực tư nhân. Khoảng trống còn lại là chưa có nhiều nghiên cứu phân tích một cách hệ thống việc Việt Nam nên lựa chọn, kết hợp và điều chỉnh những yếu tố nào từ các mô hình đó để xây dựng một khung pháp lý AI vừa phù hợp với điều kiện trong nước, vừa tương thích với các chuẩn mực và xu hướng quản trị AI quốc tế.

1.2.5. Nhóm công trình về các vấn đề pháp lý chuyên biệt của trí tuệ nhân tạo

Bên cạnh các nghiên cứu về mô hình khung pháp lý AI nói chung, một bộ phận lớn công trình tập trung vào các vấn đề pháp lý chuyên biệt do AI đặt ra. Đây là nhóm tài liệu có ý nghĩa quan trọng đối với luận án, bởi khung pháp lý cho phát triển, ứng dụng AI không chỉ cần xác định nguyên tắc và cấu trúc tổng thể, mà còn phải xử lý các vấn đề cụ thể như tư cách pháp lý, trách nhiệm dân sự, trách nhiệm hình sự, quyền SHTT, BVDLCN, quyền con người, đạo đức AI, tiêu chuẩn kỹ thuật, quản trị rủi ro, sandbox và việc sử dụng AI trong khu vực công. Nhóm công trình này bao gồm cả những nghiên cứu công bố ở nước ngoài và nghiên cứu công bố ở Việt Nam.

Với các công trình nghiên cứu nước ngoài, một trong những hướng nghiên cứu sớm và có ảnh hưởng là vấn đề tư cách pháp lý và trách nhiệm đối với AI, trong đó Lawrence B. Solum trong "*Legal Personhood for Artificial Intelligences*" *(Tư cách pháp nhân cho AI)* là công trình kinh điển khi đặt câu hỏi liệu AI có thể trở thành một chủ thể pháp luật hay không [257]. Dù được viết từ năm 1992, công trình này vẫn có giá trị nền tảng vì đã chỉ ra các điều kiện lý luận để thảo luận về tư cách pháp lý của AI, đồng thời giúp phân biệt giữa việc thừa nhận AI như một công cụ, một tài sản, một hệ thống kỹ thuật, hay một chủ thể có năng lực pháp lý nhất định. Sau đó, Claudio Novelli và cộng sự trong "*A conceptual framework for legal personality and its application to AI*" *(Khái niệm tư cách pháp nhân và ứng dụng cho AI)* [171] tiếp tục phát triển hướng nghiên cứu này bằng cách phân tích khái niệm tư cách pháp nhân và khả năng áp dụng đối với hệ thống AI. Các công trình này cho thấy, mặc dù việc thừa nhận tư cách pháp lý độc lập cho AI còn gây nhiều tranh luận, vấn đề chủ thể pháp luật vẫn có ý nghĩa quan trọng trong việc xác định trách nhiệm, quyền sở hữu, năng lực hành vi và giới hạn của các quan hệ pháp luật có yếu tố AI.

Nhìn từ góc độ khác, Ryan Abbott trong *The Reasonable Robot: Artificial Intelligence and the Law (Người máy biết suy luận: AI và pháp luật)* lập luận ủng hộ nguyên tắc trung lập của AI, theo đó pháp luật không nên phân biệt bất hợp lý giữa hành vi của con người và hành vi của AI khi chúng tạo ra kết quả tương đương [325]. Đây là một cách tiếp cận có ảnh hưởng trong thảo luận về trách nhiệm, quyền sở hữu trí tuệ và khuyến khích đổi mới sáng tạo. Tuy nhiên, nguyên tắc này cũng đặt ra câu hỏi cần cân bằng thế nào giữa trung lập công nghệ với yêu cầu bảo vệ con người trước các rủi ro đặc thù của AI. Ở lĩnh vực trách nhiệm dân sự, báo cáo *Liability for Artificial Intelligence and Other Emerging Digital Technologies (Trách nhiệm pháp lý đối với AI và các công nghệ kỹ thuật số mới nổi khác)* của Nhóm chuyên gia về trách nhiệm và công nghệ mới của Ủy ban châu Âu [199] là một nguồn tham khảo quan trọng. Báo cáo chỉ ra rằng AI làm phức tạp hóa việc xác định lỗi, quan hệ nhân quả, khả năng dự

báo thiệt hại và phân bổ trách nhiệm giữa nhà sản xuất, nhà cung cấp, người triển khai và người sử dụng. Đây là cơ sở quan trọng cho việc nghiên cứu cơ chế trách nhiệm pháp lý trong khung pháp lý AI.

Về sở hữu trí tuệ, các nghiên cứu quốc tế tập trung chủ yếu vào quyền tác giả, sáng chế và quyền đối với các sản phẩm do AI tạo ra. Daniel J. Gervais trong "*The Machine as Author*" *(Cái máy là tác giả)* phân tích một cách có hệ thống khả năng bảo hộ quyền tác giả đối với sản phẩm do AI tạo ra [179]. Tác giả cho rằng cần thận trọng khi mở rộng bảo hộ quyền tác giả cho sản phẩm không xuất phát từ lựa chọn sáng tạo của con người, bởi điều đó có thể làm suy yếu nền tảng nhân văn và mục tiêu khuyến khích sáng tạo của luật quyền tác giả. Công trình này có ý nghĩa trực tiếp đối với các nghiên cứu ở Việt Nam về tác phẩm do AI tạo ra, chủ thể quyền, tính nguyên gốc và bảo hộ đầu tư trong lĩnh vực sở hữu trí tuệ.

Về quyền con người, đạo đức và luật quốc tế, Themistoklis Tzimas trong *Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective (Những thách thức pháp lý và đạo đức của AI từ góc nhìn luật quốc tế)* [352] phân tích các thách thức pháp lý và đạo đức của AI từ góc độ luật quốc tế, bao gồm AI và quyền con người, AI và an ninh quốc tế, vũ khí tự động sát thương, trách nhiệm pháp lý và khả năng thảo luận về tư cách pháp nhân của AI. Amir Rabet Ghannadi trong "*Artificial Intelligence and International Law: Challenges and Opportunities*" *(AI và pháp luật quốc tế: Thách thức và cơ hội)* cũng tiếp cận AI từ công pháp quốc tế, nhấn mạnh các vấn đề như trách nhiệm trong hệ thống tự động, luật nhân đạo quốc tế, quyền riêng tư, phân biệt đối xử thuật toán và sở hữu trí tuệ [145]. Các công trình này cho thấy AI đã vượt khỏi phạm vi điều chỉnh của pháp luật quốc gia và trở thành một vấn đề của luật quốc tế, nhất là khi AI được sử dụng trong an ninh, quân sự, quản trị dữ liệu xuyên biên giới và các nền tảng toàn cầu.

Về công bằng thuật toán, minh bạch và chống phân biệt đối xử, Andrew D. Selbst và cộng sự trong "*Fairness and Abstraction in Sociotechnical Systems*" *(Tính công bằng và Sự trừu tượng hóa trong các Hệ thống Kỹ thuật - Xã hội)* là một nghiên cứu có ảnh hưởng lớn [147]. Công trình này phê phán xu hướng coi công bằng thuật toán như một vấn đề thuần túy kỹ thuật, tách rời bối cảnh xã hội, quyền lực và thể chế. Đối với pháp luật AI, luận điểm này có ý nghĩa quan trọng: kiểm soát thiên lệch AI không thể chỉ dựa vào chỉ số kỹ thuật, mà cần được đặt trong hệ thống bảo vệ quyền, trách nhiệm giải trình, kiểm toán độc lập, khả năng khiếu nại và biện pháp khắc phục. Cùng hướng này, Cath Corinne trong "*Governing artificial intelligence: ethical, legal and technical opportunities and challenges*" *(Quản trị AI: cơ hội và thách thức về mặt đạo đức, pháp lý và kỹ thuật)* [169] nhấn mạnh yêu cầu thiết kế và quản lý AI theo

hướng đáng tin cậy, công bằng, minh bạch, kết hợp giữa giải pháp đạo đức, pháp lý và kỹ thuật.

Về phạm vi điều chỉnh pháp luật AI, Jonas Schuett trong "*Defining the scope of AI regulations*" *(Xác định phạm vi của các quy định về AI)* [243] lập luận rằng pháp luật không nên chỉ dựa vào thuật ngữ "AI" một cách chung chung, mà cần xác định rõ nguồn rủi ro và năng lực của hệ thống, chẳng hạn công nghệ được sử dụng, ứng dụng cụ thể hoặc khả năng tác động vật lý, xã hội, kinh tế của hệ thống. Công trình này có giá trị trực tiếp đối với luận án, bởi việc xác định đúng phạm vi điều chỉnh là điều kiện tiên quyết để xây dựng khung pháp lý AI có tính khả thi, tránh vừa quá rộng gây cản trở đổi mới, vừa quá hẹp khiến rủi ro không được kiểm soát.

Về khu vực công và quyết định tự động, *Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration (Quy tắc mẫu về đánh giá tác động của hệ thống ra quyết định dựa trên thuật toán được sử dụng trong hành chính công)* của European Law Institute [201] là một nguồn tham khảo đáng chú ý. Tài liệu này đề xuất mô hình đánh giá tác động đối với các hệ thống ra quyết định thuật toán được sử dụng trong hành chính công, nhấn mạnh các yêu cầu về quản trị tốt, minh bạch, trách nhiệm giải trình, không phân biệt đối xử, tuân thủ pháp luật và khả năng kiểm soát của con người. Đây là nguồn có ý nghĩa thực tiễn đối với Việt Nam, nhất là trong bối cảnh AI đang được thúc đẩy trong cung ứng dịch vụ công, quản lý hành chính, phân tích dữ liệu lớn và hỗ trợ ra quyết định của cơ quan nhà nước.

Về công cụ điều chỉnh linh hoạt, Yordanova và Bertels trong "*Regulating AI: Challenges and the Way Forward Through Regulatory Sandboxes*" *(Quản lý AI: Những thách thức và hướng đi phía trước thông qua các mô hình thử nghiệm pháp lý)* [252] phân tích sandbox như một công cụ quản lý phù hợp với công nghệ đột phá như AI. Các tác giả cho rằng sandbox có thể tạo không gian thử nghiệm có kiểm soát, giúp cơ quan quản lý học hỏi từ thực tiễn, giảm bất định pháp lý và cân bằng giữa nguyên tắc đổi mới với bảo vệ quyền con người. Tuy nhiên, sandbox cũng có rủi ro nếu thiếu minh bạch, thiếu tiêu chí lựa chọn, thiếu giám sát độc lập hoặc bị doanh nghiệp lớn chi phối. Đây là những gợi ý quan trọng cho việc thiết kế cơ chế thử nghiệm có kiểm soát trong khung pháp lý AI ở Việt Nam.

Nhìn chung, các công trình quốc tế về vấn đề pháp lý chuyên biệt của AI đã hình thành một hệ thống nghiên cứu khá đa dạng, bao gồm tư cách pháp lý, trách nhiệm pháp lý, sở hữu trí tuệ, quyền con người, luật quốc tế, công bằng thuật toán, phạm vi điều chỉnh, khu vực công, tiêu chuẩn kỹ thuật và sandbox. Các công trình này có giá trị ở chỗ cung cấp cơ sở lý luận và phương pháp phân tích cho từng vấn đề cụ thể. Tuy nhiên, phần lớn các nghiên cứu được xây dựng trong bối cảnh các quốc gia có

hệ thống pháp luật, năng lực thực thi và thị trường công nghệ phát triển hơn Việt Nam; do đó, cần được tiếp thu có chọn lọc khi vận dụng vào điều kiện Việt Nam.

Ở Việt Nam cũng đã có một số công trình nghiên cứu về những vấn đề pháp lý chuyên biệt của AI, theo ba xu hướng chính. Một là, các nghiên cứu tổng hợp, có tính liên ngành, bàn về AI trong mối quan hệ với pháp luật, quyền con người, đạo đức và phát triển có trách nhiệm. Hai là, các nghiên cứu tập trung vào từng lĩnh vực pháp luật cụ thể như sở hữu trí tuệ, trách nhiệm dân sự, trách nhiệm hình sự, dữ liệu cá nhân, luật tư, nghề luật và dịch vụ công. Ba là, các nghiên cứu về tiêu chuẩn kỹ thuật, quản trị rủi ro và công cụ thực thi như tiêu chuẩn quốc gia, sandbox, quy tắc đạo đức và cơ chế tự quản.

Trước hết, nhóm công trình có tính tổng hợp đã góp phần đặt nền móng cho nghiên cứu pháp luật AI ở Việt Nam. Sách tham khảo "Trí tuệ nhân tạo với pháp luật và quyền con người" do Nguyễn Thị Quế Anh, Vũ Công Giao và Mai Văn Thắng đồng chủ biên [5] phân tích những vấn đề AI đặt ra đối với dịch vụ pháp lý, nghề luật, tiếp cận công lý, xây dựng pháp luật, chính sách, tư cách pháp lý, trách nhiệm pháp lý, đạo đức của robot và quyền con người. Sách chuyên khảo "Phát triển trí tuệ nhân tạo có trách nhiệm: Tiếp cận từ khía cạnh đạo đức, xã hội và pháp lý" do Nguyễn Thị Quế Anh và Nguyễn Bích Thảo đồng chủ biên [8] tiếp cận AI có trách nhiệm từ các nguyên tắc của OECD, UNESCO và kinh nghiệm của EU, Australia, Hoa Kỳ, Nhật Bản, Hàn Quốc, Trung Quốc; đồng thời khảo sát một số lĩnh vực như giáo dục, y tế, nông nghiệp. Các công trình này có giá trị ở chỗ đặt vấn đề AI trong mối liên hệ giữa pháp luật, đạo đức, xã hội và quyền con người, qua đó tạo nền tảng cho hướng tiếp cận liên ngành trong luận án.

Một số đề tài nghiên cứu trong nước cũng đã tiếp cận AI như một đối tượng cần hoàn thiện pháp luật. Đề tài NAFOSTED "Pháp luật Việt Nam trước sự phát triển của trí tuệ nhân tạo: nghiên cứu, đánh giá và hoàn thiện" do Trần Việt Dũng chủ nhiệm [34] nhấn mạnh các thách thức như tính phức tạp, khó giải thích, sai lệch dữ liệu, bảo mật, quyền riêng tư và trách nhiệm pháp lý khi xảy ra sự cố. Đề tài cấp bộ "Những khía cạnh pháp lý của trí tuệ nhân tạo ở Việt Nam: Lý luận và thực tiễn" do Chu Thị Hoa chủ nhiệm [44] nhận diện các vấn đề về quyền đối với dữ liệu, BVDLCN, quyền tác giả, tin giả và định kiến, tội phạm sử dụng AI, thuật toán và hệ thống AI trong doanh nghiệp. Các công trình này có giá trị bao quát, nhưng chủ yếu dừng ở việc nhận diện các vấn đề pháp lý đặt ra, chưa xây dựng đầy đủ một mô hình khung pháp lý đa tầng cho phát triển, ứng dụng AI.

Nhóm nghiên cứu về tiêu chuẩn kỹ thuật và quản trị rủi ro AI ở Việt Nam cũng rất đáng chú ý. Các đề tài cấp Bộ về hướng dẫn quản lý rủi ro AI [95], quản trị sử dụng AI trong tổ chức [109], thiên vị trong hệ thống AI và AI hỗ trợ ra quyết định [121],

khung cho hệ thống AI sử dụng học máy [117] cho thấy nhu cầu tiêu chuẩn hóa AI đã được đặt ra tương đối sớm. Điểm chung của nhóm công trình này là nhấn mạnh vai trò của tiêu chuẩn, quy chuẩn kỹ thuật trong bảo đảm AI an toàn, minh bạch, công bằng, có trách nhiệm và phù hợp với thông lệ quốc tế. Đối với luận án, nhóm công trình này có ý nghĩa đặc biệt vì củng cố luận điểm rằng khung pháp lý AI không thể chỉ gồm văn bản quy phạm pháp luật, mà phải bao gồm cả tiêu chuẩn kỹ thuật, cơ chế đánh giá rủi ro, kiểm định, hướng dẫn và giám sát thực thi.

Trong lĩnh vực sở hữu trí tuệ, các nghiên cứu trong nước tập trung vào hai vấn đề nổi bật: bảo hộ sản phẩm do AI tạo ra và bảo hộ sáng chế, quyền tác giả trong bối cảnh AI. Luận án của Hồ Thuý Ngọc về bảo hộ quyền sở hữu trí tuệ đối với sản phẩm tạo ra từ AI [72] đề xuất tiếp cận bảo hộ từ góc độ đầu tư, trong đó tính nguyên gốc của tác phẩm nên được đánh giá dựa trên chính tác phẩm thay vì cá nhân tác giả. Các bài viết của Nguyễn Thị Quế Anh [7], Nguyễn Ngọc Hồng Dương [36], Nguyễn Hải An và cộng sự [2], Đặng Thái Bình và Nguyễn Trọng Luận [9], Quách Ánh Nguyệt [74] đều cho thấy pháp luật sở hữu trí tuệ hiện hành gặp khó khăn khi xử lý các sản phẩm do AI tạo ra hoặc có sự tham gia đáng kể của AI. Nhóm công trình này có giá trị trong việc nhận diện khoảng trống của pháp luật sở hữu trí tuệ Việt Nam, nhưng cũng cho thấy cần có cách tiếp cận liên ngành hơn, kết nối giữa sở hữu trí tuệ, dữ liệu, trách nhiệm của nhà phát triển AI và chính sách khuyến khích đổi mới sáng tạo.

Trong lĩnh vực trách nhiệm pháp lý, nhiều nghiên cứu trong nước đã phân tích trách nhiệm bồi thường thiệt hại ngoài hợp đồng, trách nhiệm dân sự, trách nhiệm hình sự và khả năng xác định chủ thể chịu trách nhiệm khi AI gây thiệt hại. Các bài viết của Chu Thị Thanh An và Phạm Thị Hiền [1], Đỗ Giang Nam và Đào Trọng Khôi [67], Nguyễn Thị Hoa [45], Lê Minh Hùng và Nguyễn Thiện Tâm [50], Nguyễn Hoàng Thái Hy và Nguyễn Phượng An [56] đều cho thấy mô hình trách nhiệm truyền thống gặp khó khăn trước tính tự động, khả năng học hỏi, tính phức tạp và sự tham gia của nhiều chủ thể trong vòng đời AI. Ở lĩnh vực hình sự, bài viết của Trịnh Tiến Việt [132] và nghiên cứu của Vũ Ngọc Diệu Linh, Quản Thiện Hưng, Kim Thăng Long [61] về hành vi sử dụng AI thao túng thị trường tài chính cho thấy AI cũng có thể làm phát sinh các dạng hành vi nguy hiểm mới, đòi hỏi pháp luật hình sự và cơ chế giám sát công nghệ phải được cập nhật. Nhóm công trình này có ý nghĩa trực tiếp đối với việc thiết kế cơ chế phân bổ trách nhiệm trong khung pháp lý AI.

Trong lĩnh vực quyền con người, dữ liệu cá nhân và quản trị công, một số nghiên cứu đã bước đầu tiếp cận AI từ góc độ bảo vệ cá nhân trước các hệ thống tự động. Luận văn của Đoàn Văn Nhật [76] phân tích trách nhiệm bảo vệ quyền con

người trong ứng dụng AI và đề xuất các giải pháp thể chế, kỹ thuật, xã hội. Luận văn của Lê Trang Thu Phương [80] tiếp cận quyền được bảo hộ dữ liệu cá nhân trong bối cảnh AI. Bài viết của Trần Kiên và Hồ Minh Thành [58, tr.38-48] phân tích trách nhiệm pháp lý liên quan đến AI trong xử lý dữ liệu cá nhân từ kinh nghiệm EU. Các nghiên cứu về ChatGPT [30], nghề luật [82], cung ứng dịch vụ công [31], điều hành Chính phủ trên nền tảng AI [46] cũng cho thấy việc ứng dụng AI trong khu vực công, giáo dục, pháp luật và dịch vụ công đặt ra yêu cầu mới về minh bạch, giải trình, bảo vệ dữ liệu và duy trì vai trò kiểm soát của con người.

Ngoài ra, một số công trình trong nước tiếp cận AI từ góc độ chiến lược phát triển, hệ sinh thái và quản lý nhà nước. Các bài viết của Nguyễn Mạnh Quân [81], Nguyễn Thanh Thuỷ và cộng sự [119], Trịnh Quang Vũ và cộng sự [136], Hà Quang Thuỵ và cộng sự [114], Hồ Đắc Lộc và Huỳnh Châu Duy [63], Nguyễn Thị Thu Hương [54], Trần Anh Tú và Đinh Hải Đăng [102], Nguyễn Thanh Tùng [105], Nguyễn Thị Hảo và Nguyễn Văn Trọng [41] đều nhấn mạnh nhu cầu xây dựng chính sách, tiêu chuẩn, nguồn nhân lực, hệ sinh thái đổi mới sáng tạo và cơ chế quản lý nhà nước đối với AI. Các nghiên cứu này giúp luận án đặt các vấn đề pháp lý chuyên biệt trong bối cảnh rộng hơn của chiến lược phát triển AI quốc gia.

Tổng quan các công trình trong nước cho thấy nghiên cứu về các vấn đề pháp lý chuyên biệt của AI ở Việt Nam đã khá phong phú, bao phủ nhiều lĩnh vực như sở hữu trí tuệ, trách nhiệm dân sự, trách nhiệm hình sự, quyền con người, dữ liệu cá nhân, luật tư, nghề luật, dịch vụ công, tiêu chuẩn kỹ thuật và quản trị rủi ro. Tuy nhiên, điểm hạn chế chung là các nghiên cứu thường đi sâu vào từng vấn đề riêng lẻ, trong khi thiếu một cách tiếp cận tổng thể để liên kết các vấn đề đó trong một khung pháp lý chung cho phát triển, ứng dụng AI. Chính khoảng trống này đặt ra yêu cầu luận án phải vượt ra ngoài việc tổng hợp các vấn đề pháp lý chuyên biệt, để xây dựng một mô hình khung pháp lý đa tầng, có khả năng kết nối giữa quy phạm pháp luật, tiêu chuẩn kỹ thuật, luật mềm, cơ chế tự quản, thiết chế thực thi, trách nhiệm pháp lý và cơ chế phản hồi chính sách.

1.3. NHỮNG CÔNG TRÌNH NGHIÊN CỨU VỀ XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO

So với các nhóm công trình nghiên cứu về AI, pháp luật AI và các vấn đề pháp lý chuyên biệt của AI, nhóm công trình nghiên cứu trực tiếp về "xây dựng khung pháp lý cho phát triển, ứng dụng AI" còn tương đối hạn chế. Phần lớn tài liệu hiện có chủ yếu tập trung trả lời câu hỏi khung pháp lý AI cần điều chỉnh những nội dung gì, chẳng hạn trách nhiệm pháp lý, bảo vệ dữ liệu, quyền sở hữu trí tuệ, minh bạch thuật toán, kiểm soát rủi ro hoặc tiêu chuẩn kỹ thuật. Trong khi đó, câu hỏi xây dựng khung pháp lý đó như thế nào, với chủ thể nào, theo quy trình nào, sử dụng kỹ thuật lập pháp và

công cụ quản trị nào, lại chưa được nghiên cứu đầy đủ. Đây chính là khoảng trống nghiên cứu trung tâm mà luận án hướng tới.

1.3.1. Các công trình nghiên cứu về xây dựng pháp luật

Các công trình nghiên cứu về xây dựng pháp luật nói chung có ý nghĩa nền tảng đối với luận án, bởi xây dựng khung pháp lý AI trước hết vẫn là một quá trình xây dựng pháp luật, xây dựng chính sách và thiết kế công cụ điều chỉnh. Tuy nhiên, do AI là công nghệ có tốc độ phát triển nhanh, tính liên ngành cao và rủi ro khó dự báo, nên việc vận dụng lý luận chung về xây dựng pháp luật vào lĩnh vực AI cần có sự điều chỉnh nhất định.

Ở bình diện quốc tế, Baldwin, Cave và Lodge trong *Understanding Regulation: Theory, Strategy, and Practice (Hiểu về Quy định: Lý thuyết, Chiến lược và Thực tiễn)* tiếp cận điều chỉnh pháp luật như một lĩnh vực liên ngành, kết hợp giữa luật học, kinh tế học, khoa học chính trị, xã hội học và quản trị công [321]. Công trình này có giá trị ở chỗ không đồng nhất điều chỉnh với mệnh lệnh pháp luật, mà phân tích nhiều chiến lược điều chỉnh khác nhau như điều chỉnh bằng quy tắc, tiêu chuẩn, thông tin, khuyến khích, tự quản, đồng quản trị và thực thi dựa trên rủi ro. Đây là cơ sở quan trọng để luận án tiếp cận khung pháp lý AI theo hướng đa công cụ, bao gồm pháp luật cứng, pháp luật mềm, tiêu chuẩn kỹ thuật, bộ quy tắc ứng xử, sandbox và cơ chế giám sát. Về kỹ thuật lập pháp, Helen Xanthaki trong *Drafting Legislation: Art and Technology of Rules for Regulation (Soạn thảo luật: Nghệ thuật và công nghệ của các quy tắc điều chỉnh)* và các công trình về soạn thảo pháp luật đã phát triển cách tiếp cận coi kỹ thuật lập pháp không chỉ là kỹ năng ngôn ngữ pháp lý, mà còn là một phương pháp thiết kế quy phạm nhằm đạt mục tiêu chính sách [222]. Cách tiếp cận này có ý nghĩa đối với luận án vì xây dựng pháp luật về AI đòi hỏi các quy định phải vừa đủ rõ ràng để thực thi, vừa đủ linh hoạt để thích ứng với thay đổi công nghệ; vừa xác định được nghĩa vụ pháp lý, vừa tránh khóa chặt pháp luật vào những mô tả kỹ thuật có thể nhanh chóng lạc hậu.

Nhìn chung, các công trình nước ngoài về xây dựng pháp luật và quản trị điều chỉnh cung cấp ba gợi mở quan trọng cho luận án. Một là, chất lượng pháp luật không chỉ phụ thuộc vào nội dung quy định, mà còn phụ thuộc vào quy trình xây dựng, đánh giá tác động, tham vấn và hậu kiểm. Hai là, trong các lĩnh vực công nghệ mới, pháp luật cần được thiết kế như một hệ thống điều chỉnh đa công cụ, kết hợp giữa quy phạm bắt buộc, tiêu chuẩn kỹ thuật, hướng dẫn, tự quản và giám sát. Ba là, kỹ thuật lập pháp phải bảo đảm sự cân bằng giữa tính rõ ràng, khả thi, ổn định tương đối và khả năng thích ứng.

Ở Việt Nam, các nghiên cứu về xây dựng pháp luật đã hình thành tương đối phong phú, nhất là về chất lượng văn bản quy phạm pháp luật, tính hợp hiến, hợp pháp, tính thống nhất, tính khả thi, kỹ thuật lập pháp, đánh giá tác động chính sách và kiểm soát chất lượng quy định. Tuy nhiên, các nghiên cứu này chủ yếu được xây dựng cho hoạt động lập pháp nói chung, chưa đi sâu vào yêu cầu đặc thù của pháp luật trong các lĩnh vực công nghệ mới như AI. Ví dụ, tham luận của Viện Khoa học pháp lý, Bộ Tư pháp về "Các yếu tố để tạo ra một văn bản quy phạm pháp luật có chất lượng tốt" tại Hội thảo Chất lượng của quy định và sự thuận lợi của môi trường kinh doanh năm 2022 xác định một số tiêu chí cơ bản để đánh giá chất lượng văn bản quy phạm pháp luật, như khả năng giải quyết đúng vấn đề đặt ra, tính rõ ràng, thống nhất với chính sách chung của Nhà nước, tính hợp hiến, hợp pháp, đồng bộ, tương thích với điều ước quốc tế, tính khả thi, minh bạch, dễ hiểu, tính dự báo, tính ổn định tương đối, phù hợp với điều kiện kinh tế - xã hội và tính hợp lý của chế tài [130]. Đây là nguồn có giá trị tham khảo quan trọng đối với luận án khi xây dựng bộ tiêu chí đánh giá chất lượng khung pháp lý AI. TS. Nguyễn Văn Cương trong sách chuyên khảo "*Tính ổn định của pháp luật - Nhận thức, thực tiễn và giải pháp đổi mới phân tích tính ổn định của pháp luật như một yêu cầu quan trọng của Nhà nước pháp quyền"* [18], đồng thời nhấn mạnh tính ổn định không thể được hiểu một cách tuyệt đối. Nếu quá nhấn mạnh ổn định, pháp luật có thể trở nên cứng nhắc, lạc hậu; nếu thay đổi quá nhanh, pháp luật lại làm suy giảm tính dự báo và niềm tin của xã hội. Luận điểm này đặc biệt có ý nghĩa đối với pháp luật AI, bởi đây là lĩnh vực cần đồng thời bảo đảm tính ổn định của nguyên tắc, chính sách cơ bản và tính linh hoạt của các công cụ điều chỉnh cụ thể.

TS. Thái Thị Tuyết Dung và cộng sự trong bài viết "Một số bất cập trong kĩ thuật lập pháp, lập quy từ các ví dụ thực tiễn ở Việt Nam" chỉ ra nhiều bất cập không chỉ ở kỹ thuật diễn đạt quy phạm, mà còn ở tư duy lập pháp, lập quy [32]. Các vấn đề như kỹ thuật liệt kê danh mục cấm, danh mục được phép; phân biệt giữa sửa đổi, bổ sung một số điều và sửa đổi toàn diện; hiệu lực của văn bản ban hành theo thủ tục rút gọn; cũng như tính rõ ràng của quy phạm đều có ý nghĩa đối với việc xây dựng pháp luật trong các lĩnh vực mới. Đối với AI, những vấn đề này càng trở nên quan trọng vì nếu quy định quá chi tiết có thể nhanh chóng lạc hậu, nhưng nếu quy định quá chung lại khó thực thi và khó kiểm soát rủi ro. Phùng Thị Yến và Phạm Bùi Vi Thảo trong bài viết "Chuyển đổi số trong lập pháp và thi hành pháp luật: Trí tuệ nhân tạo và dữ liệu mở - Thách thức và cơ hội" bước đầu tiếp cận vấn đề chuyển đổi số trong hoạt động lập pháp và thi hành pháp luật [137]. Công trình này có giá trị ở chỗ nhìn nhận AI và dữ liệu mở không chỉ là công cụ hỗ trợ kỹ thuật, mà có thể làm thay đổi cách thức xây dựng, vận hành và giám sát hệ thống pháp luật. Đây là một trong số ít nghiên cứu trong

nước đặt vấn đề kỹ thuật lập pháp trong môi trường số, do đó có ý nghĩa tham khảo đối với luận án khi bàn về kỹ thuật xây dựng khung pháp lý AI.

Xét tổng thể, các công trình trong nước cho thấy lý luận và thực tiễn xây dựng pháp luật ở Việt Nam đã cung cấp nhiều tiêu chí quan trọng về chất lượng văn bản, tính ổn định, tính khả thi, tính minh bạch và kỹ thuật lập pháp. Tuy nhiên, khoảng trống còn lại là các tiêu chí và kỹ thuật này chưa được phát triển thành một phương pháp xây dựng khung pháp lý cho AI, nơi pháp luật phải đồng thời đáp ứng yêu cầu phát triển công nghệ, quản trị rủi ro, bảo vệ quyền con người, hài hòa tiêu chuẩn quốc tế và duy trì khả năng cập nhật thường xuyên. Vì vậy, luận án cần kế thừa các nghiên cứu chung về xây dựng pháp luật, đồng thời phát triển thêm các kỹ thuật đặc thù phù hợp với quá trình xây dựng khung pháp lý AI.

1.3.2. Nhóm công trình về lý luận xây dựng quy định công nghệ

Nhóm công trình này có ý nghĩa trực tiếp hơn đối với luận án so với các nghiên cứu chung về xây dựng pháp luật, bởi AI không chỉ là một đối tượng điều chỉnh mới, mà còn là một công nghệ có khả năng làm thay đổi chính cách thức thiết kế, thực thi và đánh giá pháp luật. Vấn đề đặt ra không chỉ là ban hành quy định gì để quản lý AI, mà còn là phải xây dựng quy định về AI theo tư duy nào, bằng công cụ nào, với sự tham gia của những chủ thể nào và với mức độ linh hoạt ra sao.

Trong các nghiên cứu nước ngoài, Baldwin, Cave và Lodge trong *Understanding Regulation: Theory, Strategy, and Practice (Hiểu về Quy định: Lý thuyết, Chiến lược và Thực tiễn)* cung cấp nền tảng lý luận quan trọng về điều chỉnh pháp luật theo nghĩa rộng [321]. Các tác giả không giới hạn "quy định" trong các mệnh lệnh pháp luật bắt buộc, mà nhìn nhận quy định như tổng thể các cơ chế định hướng, kiểm soát và điều chỉnh hành vi xã hội. Cách tiếp cận này có ý nghĩa quan trọng đối với luận án, bởi khung pháp lý AI không thể chỉ được hiểu là một đạo luật hoặc một nhóm quy phạm bắt buộc, mà cần bao gồm nhiều công cụ khác nhau như luật cứng, luật mềm, tiêu chuẩn kỹ thuật, quy tắc đạo đức, cơ chế tự quản, đồng quản trị và giám sát sau ban hành. Julia Black trong các công trình về "decentring regulation" [246], [247], tự điều chỉnh và hoạt động điều chỉnh phi tập trung phát triển thêm cách tiếp cận đa chủ thể trong xây dựng quy định. Theo hướng này, Nhà nước không còn là chủ thể duy nhất tạo lập và vận hành trật tự điều chỉnh, mà cần huy động sự tham gia của doanh nghiệp, hiệp hội nghề nghiệp, chuyên gia kỹ thuật, tổ chức xã hội và các chủ thể bị tác động. Cách tiếp cận đó rất phù hợp với lĩnh vực AI, nơi tri thức kỹ thuật, dữ liệu và năng lực triển khai phần lớn nằm ngoài khu vực nhà nước. Vì vậy, việc xây dựng khung pháp lý AI đòi hỏi không chỉ quy trình lập pháp chính thức, mà còn cơ chế tham vấn thực chất, đối thoại chính sách, thử nghiệm, đánh giá và cập nhật liên tục.

Một hướng nghiên cứu quan trọng khác là quản trị rủi ro công nghệ trong điều kiện bất ổn. Renn trong *Risk Governance: Coping with Uncertainty in a Complex World (Quản trị rủi ro: Ứng phó với sự bất định trong một thế giới phức tạp)* nhấn mạnh sự chuyển dịch từ quản lý rủi ro theo nghĩa hẹp sang quản trị rủi ro theo nghĩa rộng, bao gồm đánh giá, thẩm định, truyền thông, tham gia của các bên liên quan và xử lý bất định [305]. Lý thuyết của Renn là nền tảng tích cực được nhiều nghiên cứu vận dụng, nhưng cũng chính trong dòng nghiên cứu này đã xuất hiện các phê phán quan trọng đối với việc vận dụng thái quá tính "thích ứng". Nghiên cứu của Sabel và Zeitlin về quản trị theo phương pháp thực nghiệm [177] giải thích rằng kiến trúc thực nghiệm trong quản lý được minh họa rõ nét bởi Chỉ thị khung về nước của Liên minh châu Âu (WFD) và Chiến lược thực hiện chung. Bài viết cho rằng phương pháp thực nghiệm dường như đặc biệt phù hợp với các lĩnh vực xuyên quốc gia, nơi không có một thực thể có chủ quyền tối cao nào có thẩm quyền thiết lập các mục tiêu chung ngay cả trên lý thuyết, và nơi sự đa dạng của các điều kiện và thực tiễn địa phương khiến việc áp dụng và thực thi các quy tắc cố định thống nhất thậm chí còn khó khả thi hơn so với trong nước.

Collingridge trong *The Social Control of Technology (Sự kiểm soát xã hội đối với công nghệ)* chỉ ra nghịch lý nổi tiếng của điều chỉnh công nghệ: khi công nghệ còn mới thì dễ điều chỉnh nhưng khó dự báo tác động; khi công nghệ đã phổ biến thì có thể dự báo tốt hơn nhưng lại rất khó điều chỉnh do sự lệ thuộc về kỹ thuật, thị trường và thể chế [184]. Đây là một luận điểm đặc biệt quan trọng đối với AI, bởi nếu chờ đến khi AI đã thâm nhập sâu vào quản trị nhà nước, thị trường lao động, dịch vụ công, giáo dục, y tế, tài chính và tư pháp thì việc kiểm soát rủi ro sẽ khó khăn và tốn kém hơn rất nhiều. Jasanoff trong các công trình về đồng kiến tạo giữa khoa học, công nghệ và trật tự xã hội cung cấp một nền tảng lý luận sâu hơn cho việc nhìn nhận mối quan hệ giữa AI và pháp luật. Theo cách tiếp cận này, công nghệ không đơn thuần là đối tượng bị pháp luật điều chỉnh; ngược lại, công nghệ cũng góp phần định hình lại các khái niệm pháp lý, phương thức quản trị, tiêu chuẩn trách nhiệm và quan hệ giữa Nhà nước, thị trường và xã hội. Do đó, xây dựng khung pháp lý AI không thể chỉ là việc đưa AI vào khuôn khổ pháp luật hiện có mà còn là quá trình tái thiết kế một số nguyên tắc, công cụ và kỹ thuật pháp lý để phù hợp với môi trường công nghệ mới [331].

Từ nền tảng đó, nghiên cứu về quản trị AI đa bên như công trình của Ngai Lim và Danforn Lim [191] phân tích toàn diện các chính sách, khuôn khổ pháp lý, nghiên cứu tình huống và những thay đổi chính sách gần đây, lập luận rằng sự không đầy đủ của mô hình "trách nhiệm duy nhất" đang ngày càng trở thành đặc điểm nổi bật của các hệ thống AI hiện đại. Phân tích này minh họa một sự dịch chuyển toàn cầu chưa từng có hướng tới các khuôn khổ quản trị với trách nhiệm

phân tán, được truyền cảm hứng từ Đạo luật AI 2024 của EU, Khuyến nghị về đạo đức AI năm 2021 của UNESCO và các khuôn khổ quản trị quốc gia đang nổi lên từ Malaysia, Singapore, Úc và Đan Mạch. Bài đánh giá tổng quan này đề xuất một khuôn khổ chính sách toàn diện thực hiện trách nhiệm phân tán thông qua các chiến lược trách nhiệm theo cấp bậc, đánh giá tác động thuật toán bắt buộc và các cơ chế giám sát phối hợp quốc tế. Những phát hiện này đưa ra các giải pháp cân bằng và khả thi trên thực tế để giải quyết những mong muốn mâu thuẫn giữa trách nhiệm giải trình và đổi mới trong bối cảnh công nghệ.

Nghiên cứu về tính chính đáng của đồng quản trị có thể thấy ba tiêu chí đánh giá được nhiều học giả đồng thuận: tính đại diện cân bằng của quá trình tham gia (không chỉ doanh nghiệp lớn mà cả SME, tổ chức xã hội, nhóm bị ảnh hưởng); tính minh bạch của quá trình ra quyết định (biên bản, căn cứ được công bố); và sự tồn tại của cơ chế giám sát nhà nước tối thiểu để chuẩn mực tư nhân không thay thế hoàn toàn vai trò bảo vệ công của nhà nước. Hạn chế của dòng nghiên cứu này là phần lớn dừng lại ở phân tích trường hợp cụ thể (case study) từng quốc gia riêng lẻ mà chưa xây dựng được một bộ tiêu chí đánh giá tính chính đáng có thể áp dụng phổ quát và có thể lượng hóa mức độ đáp ứng.

Ở Việt Nam, các nghiên cứu trực tiếp về lý luận xây dựng quy định công nghệ chưa nhiều và chưa hình thành thành một dòng nghiên cứu độc lập. Tuy nhiên, có thể nhận diện một số nhóm công trình có giá trị tham khảo đối với luận án.

Trước hết, các công trình nghiên cứu về đổi mới tư duy xây dựng pháp luật trong bối cảnh phát triển mới đã cung cấp cơ sở quan trọng cho cách tiếp cận pháp luật kiến tạo phát triển. Võ Khánh Vinh trong các công trình về chính sách pháp luật, chiến lược phát triển pháp luật và đổi mới tư duy xây dựng pháp luật nhấn mạnh yêu cầu pháp luật phải bám sát thực tiễn, có năng lực dự báo, tiếp thu chọn lọc kinh nghiệm quốc tế, đồng thời tạo không gian cho đổi mới sáng tạo và phát triển các động lực tăng trưởng mới [134]. Cách tiếp cận này có ý nghĩa đối với xây dựng quy định công nghệ, bởi pháp luật không chỉ nhằm kiểm soát rủi ro, mà còn phải mở đường, khơi thông nguồn lực, tạo hành lang pháp lý an toàn và linh hoạt cho công nghệ mới phát triển.

Thứ hai, các công trình về sandbox và thử nghiệm có kiểm soát ở Việt Nam đã trực tiếp đặt vấn đề thay đổi tư duy pháp lý đối với công nghệ mới. Chu Thị Hoa trong bài viết về sandbox phân tích cơ chế thử nghiệm áp dụng trong phạm vi hạn chế từ kinh nghiệm quốc tế và đưa ra gợi ý cho Việt Nam [43, tr.58-64]. Các nghiên cứu gần đây của Trần Quyết Thắng và Dương Văn Quý tiếp tục phát triển hướng này khi xem sandbox là công cụ quản trị thích ứng, cho phép Nhà nước chuyển từ tư duy quản lý tĩnh, cấm đoán hoặc hậu kiểm sang tư duy thử nghiệm, học hỏi và điều chỉnh dựa trên bằng chứng [110, tr.109-114]. Đối với luận án, nhóm công trình này có giá trị đặc biệt

vì sandbox là một trong những kỹ thuật pháp lý quan trọng để xử lý nghịch lý Collingridge trong lĩnh vực AI.

Thứ ba, một số nghiên cứu gần đây đã bắt đầu tiếp cận trực tiếp vấn đề xây dựng khung pháp lý AI ở Việt Nam từ góc độ lựa chọn mô hình điều chỉnh. TS. Thái Thị Tuyết Dung và cộng sự trong bài viết về xây dựng khung pháp lý AI phân tích kinh nghiệm của EU, Vương quốc Anh và Nhật Bản, từ đó gợi mở cho Việt Nam cách thiết kế Luật Trí tuệ nhân tạo theo hướng vừa linh hoạt, vừa kiểm soát rủi ro [32, tr.50-64]. Công trình này có giá trị ở chỗ không chỉ mô tả các mô hình nước ngoài, mà còn đặt ra vấn đề lựa chọn cách tiếp cận lập pháp cho Việt Nam: mô hình quy định chi tiết dựa trên rủi ro, mô hình quản trị dựa trên nguyên tắc hay mô hình luật khung định hướng. Đây là một gợi ý gần với trọng tâm của luận án, dù bài viết chưa phát triển đầy đủ thành lý luận về quá trình xây dựng khung pháp lý AI.

Thứ tư, các nghiên cứu về chuyển đổi số trong lập pháp, dữ liệu mở và ứng dụng AI trong xây dựng, thi hành pháp luật bước đầu cho thấy công nghệ không chỉ là đối tượng điều chỉnh mà còn là công cụ hỗ trợ hoạt động lập pháp và thực thi pháp luật. Phùng Thị Yến và Phạm Bùi Vi Thảo đã đặt vấn đề AI và dữ liệu mở như những yếu tố có thể làm thay đổi cách thức xây dựng, vận hành và giám sát hệ thống pháp luật [137]. Hướng nghiên cứu này gợi mở khả năng sử dụng dữ liệu, phân tích tác động, mô phỏng chính sách, tham vấn điện tử và đánh giá hậu kiểm bằng công nghệ trong quá trình xây dựng khung pháp lý AI.

Xét tổng thể, nhóm công trình này cho thấy, ở bình diện quốc tế đã có nền tảng lý luận tương đối rõ về điều chỉnh công nghệ, bao gồm điều chỉnh đa công cụ, điều chỉnh phi trung tâm, quản trị rủi ro, đồng kiến tạo công nghệ - xã hội và cơ chế thử nghiệm thích ứng. Ở Việt Nam, các nghiên cứu liên quan mới chỉ xuất hiện rải rác trong các hướng nghiên cứu về đổi mới tư duy xây dựng pháp luật, sandbox, chuyển đổi số trong lập pháp và xây dựng khung pháp lý AI. Khoảng trống còn lại là chưa có công trình nào hệ thống hóa các lý thuyết này để xây dựng một phương pháp luận riêng cho quá trình xây dựng khung pháp lý AI ở Việt Nam.

Luận án kế thừa lý thuyết quản trị rủi ro thích ứng của Renn như một trong ba trụ cột lý luận, đồng thời kế thừa các tiêu chí thiết kế sandbox hiệu quả từ kinh nghiệm quốc tế và trong nước (tiêu chí đầu vào rõ ràng, thời hạn xác định, cơ chế chuyển đổi). Luận án phản biện trực tiếp khuynh hướng ca ngợi quản trị thích ứng một chiều mà không phân tích đầy đủ giới hạn của nó, cụ thể phản biện quan điểm cho rằng linh hoạt luôn là giá trị tuyệt đối trong quản trị công nghệ bằng cách chỉ ra rằng tính thích ứng phải bị giới hạn bởi các nguyên tắc của Nhà nước pháp quyền, đặc biệt khi đối tượng chịu tác động là quyền con người không thể thử nghiệm như một biến số kỹ thuật. Trên cơ sở thông qua việc kết hợp lý luận điều chỉnh công nghệ

với thực tiễn xây dựng pháp luật, năng lực thể chế và yêu cầu phát triển AI có trách nhiệm ở Việt Nam, luận án phát triển nguyên tắc xử lý xung đột có thứ bậc ưu tiên (trình bày tại phần khung lý luận) xác định rõ khi nào phòng ngừa rủi ro được ưu tiên tạm thời trước tính thích ứng, kèm điều kiện bắt buộc về rà soát có thời hạn.

1.3.3. Nhóm công trình về phương pháp luận xây dựng quy định trí tuệ nhân tạo cụ thể

Nhóm công trình này tiếp cận gần hơn với trọng tâm của luận án, bởi không chỉ bàn về các vấn đề pháp lý do AI đặt ra, mà còn gợi mở những phương pháp, kỹ thuật và cơ chế cụ thể để xây dựng quy định AI. Có thể nhận thấy các nghiên cứu trong nhóm này tập trung vào một số hướng chính: trách nhiệm giải trình và yêu cầu giải thích; trách nhiệm pháp lý trong tương tác giữa con người và robot; trách nhiệm của doanh nghiệp khi triển khai AI; quản trị theo thiết kế; điều tiết thuật toán; mô hình thị trường quy định; sandbox, self-regulation và thiết kế thiết chế quản lý AI.

Trước hết, về trách nhiệm giải trình, Doshi-Velez và cộng sự trong "*Accountability of AI under the Law: The Role of Explanation*" *(Trách nhiệm giải trình của AI theo pháp luật: Vai trò của sự giải thích)* [210] đề xuất công cụ tăng cường trách nhiệm giải trình trong hệ thống AI là "giải thích". Theo các tác giả, giải thích có thể làm sáng tỏ thông tin về các quyết định cá nhân cụ thể mà không nhất thiết phải làm sáng tỏ cơ chế chính xác của toàn bộ quá trình ra quyết định. Giải thích cũng có thể được sử dụng để ngăn ngừa hoặc khắc phục lỗi, tăng cường niềm tin, đồng thời hỗ trợ việc xác định liệu các tiêu chí nhất định đã được sử dụng một cách thích hợp hay không trong trường hợp có tranh chấp. Công trình này có giá trị ở chỗ chuyển yêu cầu minh bạch và trách nhiệm giải trình từ mức nguyên tắc chung sang một công cụ có thể được thiết kế trong quy định pháp luật về AI.

Về trách nhiệm pháp lý trong tương tác giữa con người và robot, tiêu biểu là quyển sách *Human-Robot Interaction in Law and Its Narratives* (*Tương tác giữa con người và robot trong luật pháp và các câu chuyện liên quan)* [327], trong đó Tatjana Hörnle nghiên cứu "*The Challenges of Human-Robot Interaction for Substantive Criminal Law*" *(Những thách thức của tương tác giữa người và robot đối với bản chất luật hình sự)* [97] cung cấp cái nhìn toàn diện về trách nhiệm hình sự của con người đối với robot, trách nhiệm hình sự của chính robot, tự vệ chống lại robot và robot với tư cách là nạn nhân của tội phạm. Trong cùng hướng tiếp cận đó, Janneke de Snaijer [238] và Marta Bo [267] thảo luận về các khía cạnh cụ thể của trách nhiệm hình sự và các trường hợp miễn trừ, trong khi Thomas Weigend [353] phân tích "khoảng trống trách nhiệm" đang hiện hữu và khả năng mở rộng ý tưởng về trách nhiệm hình sự của doanh nghiệp để bao gồm thiệt hại do các thiết bị AI gây ra. Các công trình này cho thấy việc xây dựng

quy định AI cần tính đến cả yêu cầu phòng ngừa, trấn áp và định hướng dài hạn đối với cách luật hình sự có thể định hình tương tác giữa con người và robot.

Ở cấp độ doanh nghiệp, Iris H.-Y. Chiu và Ernest W.K. Lim trong "*Managing corporations' risk in adopting artificial intelligence: A corporate responsibility paradigm*" *(Quản lý rủi ro của doanh nghiệp khi áp dụng AI: Mô hình trách nhiệm xã hội doanh nghiệp)* [232] hướng trực tiếp đến các doanh nghiệp, tập đoàn áp dụng và triển khai công nghệ học máy trong mô hình kinh doanh. Theo các tác giả, việc doanh nghiệp triển khai AI không thể chỉ được coi là một hoạt động "tuân thủ" pháp luật thông thường. Mọi quyết định áp dụng công nghệ này đang bị đặt dưới sự giám sát chặt chẽ của ngành, các bên liên quan và toàn xã hội thông qua các sáng kiến quản trị từ dưới lên. Do đó, để định hướng cho các quyết định chiến lược khi áp dụng học máy, doanh nghiệp cần đặt hành vi của mình trong một khung quản trị tích hợp, bao gồm cả kiểm soát rủi ro pháp lý của riêng doanh nghiệp và trách nhiệm giải trình rộng lớn hơn đối với lợi ích chung của toàn xã hội.

Một hướng tiếp cận quan trọng khác là quản trị theo thiết kế. Mulligan và Bamberger trong "*Saving Governance-by-Design*" *(Bảo vệ quản trị theo thiết kế)* [185] lập luận rằng kỹ thuật thiết kế và quản trị phải được tích hợp ngay từ giai đoạn phát triển sản phẩm AI, qua đó gợi mở cho các kỹ thuật như thiết kế an toàn và bảo mật ngay từ đầu trong khung pháp lý AI. Bài viết xây dựng một khuôn khổ nhằm giúp các nhà hoạch định chính sách bảo vệ các giá trị và quy trình dân chủ khi xem xét việc điều tiết bằng công nghệ, bao gồm bốn yêu cầu bao quát: thiết kế với sự khiêm tốn và kiềm chế để duy trì tính linh hoạt; ưu tiên quyền con người và quyền công cộng; bảo đảm các nhà quản lý có công cụ phù hợp về thẩm quyền, năng lực và chuyên môn kỹ thuật; và duy trì tính công khai của quá trình hoạch định chính sách.

Yeung và Lodge (chủ biên) trong *Algorithmic Regulation (Điều tiết bằng thuật toán)* [251] là công trình học thuật có tính hệ thống về điều tiết thuật toán. Công trình này tiếp cận thuật toán vừa như phương tiện để phối hợp và điều tiết hành động xã hội, hỗ trợ việc ra quyết định, vừa như đối tượng cần được kiểm soát thông qua các cơ chế thể chế phù hợp. Giá trị của công trình này đối với luận án là chỉ ra tính hai mặt của AI và thuật toán: chúng vừa có thể trở thành công cụ của quản trị, vừa phải được đặt dưới sự quản trị của pháp luật và các thiết chế kiểm soát. Khi AI được nhà nước sử dụng để ra quyết định hành chính, hỗ trợ tư pháp hoặc giám sát công dân, hiện tượng gọi chung là "quyền lực thuật toán" thì các nguyên tắc cốt lõi của Nhà nước pháp quyền (tính hợp pháp của hành vi công quyền, quyền được biết và được giải thích, quyền khiếu nại và được xét xử) phải được tái diễn giải như thế nào để không bị vô hiệu hóa bởi đặc trưng kỹ thuật "hộp đen" của AI.

Quyền lực thuật toán là vấn đề nghiên cứu đang phát triển nhanh, với đóng góp nền tảng từ Citron [180] về quy trình thủ tục tố tụng công bằng dựa trên công nghệ lập luận rằng khi AI thay thế hoặc hỗ trợ quyết định hành chính, các bảo đảm thủ tục tố tụng công bằng truyền thống (quyền được thông báo, quyền được điều trần, quyền được giải thích căn cứ quyết định) phải được thiết kế lại cho phù hợp với đặc thù ra quyết định tự động. Nếu không, AI hành chính có nguy cơ tạo ra một hình thức quyền lực công không chịu sự kiểm soát pháp quyền thông thường. Nghiên cứu của Frank Pasquale trong *The Black Box Society (Xã hội hộp đen)* [211] mở rộng phân tích này sang khái niệm trách nhiệm giải trình thuật toán, đề xuất các cơ chế minh bạch có chọn lọc (không đòi hỏi công khai toàn bộ mã nguồn nhưng đòi hỏi giải trình đủ mức để giám sát được) như giải pháp dung hòa giữa bảo vệ bí mật thương mại và yêu cầu giám sát công quyền. So sánh tương đồng và khác biệt, Citron và Pasquale đều xuất phát từ lo ngại về tình trạng quyền lực công sử dụng công cụ kỹ thuật mà công dân không thể giám sát, nhưng khác biệt về giải pháp đề xuất: Citron tập trung vào cải cách thủ tục tố tụng hành chính (mở rộng các bảo đảm thủ tục truyền thống sang bối cảnh tự động hóa), trong khi Pasquale tập trung vào cơ chế minh bạch có chọn lọc ở cấp độ hệ thống (không đòi hỏi minh bạch tuyệt đối mà đòi hỏi đủ thông tin cho giám sát độc lập). Sự khác biệt này phản ánh hai truyền thống pháp lý khác nhau mà hai tác giả xuất phát, Citron gần với truyền thống luật hành chính, Pasquale gần với truyền thống luật thông tin và cạnh tranh.

Liên quan đến vấn đề này, tranh luận về nội hàm cụ thể của "quyền được giải thích" theo Điều 22 GDPR vẫn chưa ngã ngũ sau hơn nửa thập kỷ giữa quan điểm hẹp (chỉ là quyền tiếp cận thông tin về sự tồn tại của xử lý tự động, theo cách đọc chặt chẽ văn bản) và quan điểm rộng (bao gồm quyền nhận giải thích thực chất về logic quyết định, theo cách đọc mục đích luận). Đây là minh chứng cho thấy ngay cả hệ thống pháp luật có kinh nghiệm điều chỉnh dữ liệu số lâu năm nhất (EU) cũng chưa giải quyết được một câu hỏi nền tảng của kiểm soát quyền lực thuật toán. Luận án ghi nhận điều này như căn cứ cho việc không thể kỳ vọng pháp luật Việt Nam giải quyết vấn đề này một cách hoàn chỉnh ngay từ đầu, mà cần thiết kế cơ chế diễn giải linh hoạt để xử lý dần qua thực tiễn áp dụng. Do đó, luận án kế thừa học thuyết của Citron, khái niệm trách nhiệm giải trình thuật toán của Pasquale. Luận án phản biện khoảng trống chuyển hóa từ common law/EU sang civil law Việt Nam mà nghiên cứu quốc tế chưa giải quyết, đồng thời phản biện quan điểm cho rằng quyền được giải thích chỉ có thể vận hành qua cơ chế tố tụng tư pháp truyền thống. Trong điều kiện Việt Nam, cơ chế giám sát hành chính độc lập và giám sát xã hội có vai trò bổ trợ không thể thiếu bên cạnh giám sát tư pháp.

Hadfield và Clark trong "*Regulatory Markets: The Future of AI Governance*" *(Thị trường điều tiết: Tương lai của quản trị AI)* [219] đề xuất mô hình thị trường quy định cho AI, trong đó Chính phủ yêu cầu các đối tượng bị quản lý mua dịch vụ quản lý từ một cơ quan quản lý tư nhân được Chính phủ cấp phép. Cách tiếp cận này có thể khắc phục những hạn chế của cả mô hình quản lý mệnh lệnh - kiểm soát truyền thống lẫn việc ủy quyền quá mức cho ngành công nghiệp tự điều chỉnh. Thị trường quy định cho phép Chính phủ thiết lập các ưu tiên chính sách cho việc quản lý AI, đồng thời dựa vào thị trường và nỗ lực nghiên cứu - phát triển của ngành để đề ra các phương pháp kỹ thuật quản lý tối ưu. Ý tưởng này có thể gợi mở cho thiết kế cơ chế sandbox và self-regulation trong khung pháp lý AI Việt Nam.

Trong bối cảnh châu Âu, Carlo Casonato và Giulia Olivato trong "*AI Regulation in Europe: Exploring the Artificial Intelligence Act*" *(Quy định về AI ở châu Âu: Khám phá Đạo luật AI)* [168] xem xét các đặc điểm cần thiết để một đạo luật AI có thể bền vững. Theo các tác giả, luật AI cần đáp ứng một số yêu cầu cơ bản: được phát triển phù hợp với chuyên môn khoa học và sự đa dạng của xã hội; có khả năng cung cấp sự chắc chắn pháp lý nhưng vẫn linh hoạt trước thay đổi công nghệ; và tập trung vào công nghệ cụ thể được điều chỉnh. Bên cạnh đó, các tác giả cũng trình bày một khung tổng quát cho môi trường thử nghiệm pháp lý, các biện pháp quản trị và thực thi. Đây là những gợi ý quan trọng cho việc thiết kế quy định AI trong điều kiện công nghệ phát triển nhanh và rủi ro khó dự báo.

Theodorou và Dignum trong "*Towards Ethical and Socio-legal Governance in AI*" *(Hướng tới quản trị đạo đức và pháp lý xã hội AI)* [146] nhấn mạnh yêu cầu quản trị AI theo hướng đạo đức và xã hội - pháp lý. Theo các tác giả, quản trị AI cần sự tham gia của nhiều cá nhân, tổ chức trong nhiều vai trò khác nhau, như nhà phát triển, nhà sản xuất, người dùng, người ngoài cuộc và nhà hoạch định chính sách. Công trình này củng cố quan điểm rằng xây dựng quy định AI không thể chỉ là hoạt động đơn tuyến của cơ quan nhà nước, mà cần có cơ chế tham gia, phối hợp và phân bổ trách nhiệm giữa các chủ thể khác nhau trong toàn bộ vòng đời AI.

Ở cấp độ thiết chế, Cihon, Maas và Kemp trong "*Should Artificial Intelligence Governance be Centralised? Design Lessons from History*" *(Có nên quản trị AI tập trung? Những bài học thiết kế từ lịch sử)* [309] phân tích kinh nghiệm lịch sử để rút ra bài học về thiết kế thiết chế quản lý AI, đặc biệt là lợi thế và bất lợi của mô hình cơ quan quản lý tập trung so với mô hình phân tán. Mô hình tập trung có ưu điểm về thống nhất, chuyên môn hóa và điều phối, nhưng có thể chậm thích ứng và thiếu hiểu biết về đặc thù từng lĩnh vực. Trong khi đó, mô hình phân tán linh hoạt hơn và gần thực tiễn hơn, nhưng có nguy cơ phân mảnh, thiếu thống nhất và khó phối hợp. Những

phân tích này có giá trị tham khảo đối với việc thiết kế thiết chế thực thi khung pháp lý AI ở Việt Nam.

Nhìn chung, các công trình quốc tế nêu trên đã phát triển nhiều gợi ý phương pháp luận quan trọng cho việc xây dựng quy định AI, bao gồm giải thích để bảo đảm trách nhiệm giải trình, xử lý khoảng trống trách nhiệm, quản trị rủi ro doanh nghiệp, quản trị theo thiết kế, quy định theo thuật toán, thị trường điều tiết, sandbox, tự điều chỉnh, quản trị đa chủ thể và lựa chọn mô hình thiết chế quản lý. Tuy nhiên, các nghiên cứu này chủ yếu được xây dựng trong bối cảnh các quốc gia có trình độ phát triển công nghệ, thị trường và thiết chế pháp lý tương đối cao; do đó, việc vận dụng vào Việt Nam cần được điều chỉnh phù hợp với điều kiện thể chế, năng lực thực thi và yêu cầu phát triển AI có trách nhiệm. Một số bài viết ngắn mới bước đầu đặt vấn đề. Trần Việt Dũng trong "Bàn về phát triển quy phạm đạo đức trong tiến trình xây dựng pháp luật về trí tuệ nhân tạo" [35] nhấn mạnh các quy tắc đạo đức đóng vai trò quan trọng, tạo nền tảng cho một khung pháp luật linh hoạt và đáng tin cậy về AI, có thể điều chỉnh kịp thời theo sự phát triển của công nghệ. "Xây dựng khung pháp lý về trí tuệ nhân tạo: kinh nghiệm của một số quốc gia và gợi mở cho Việt Nam" của TS. Thái Thị Tuyết Dung và cộng sự [32, tr.50-64] là một trong số ít bài viết đặt vấn đề về "xây dựng" khung pháp lý AI từ góc độ chủ động, nhưng chủ yếu dừng ở mức xác định vấn đề và gợi mở định hướng, chưa phân tích một cách hệ thống quy trình, chủ thể, kỹ thuật và cơ chế vận hành của khung pháp lý AI.

Từ tổng quan trên có thể thấy, khoảng trống nghiên cứu chủ yếu nằm ở việc chưa có công trình nào hệ thống hóa các phương pháp luận xây dựng quy định AI để vận dụng vào điều kiện Việt Nam. Đây là cơ sở để luận án tiếp tục nghiên cứu việc xây dựng khung pháp lý cho phát triển, ứng dụng AI như một quá trình có tính phương pháp luận, bao gồm xác định vấn đề, đánh giá rủi ro, lựa chọn công cụ điều chỉnh, thiết kế cơ chế tham vấn, xây dựng tiêu chuẩn, thử nghiệm chính sách, thiết kế thiết chế thực thi và cơ chế cập nhật pháp luật.

1.4. NHẬN XÉT, ĐÁNH GIÁ VÀ NHỮNG VẤN ĐỀ LUẬN ÁN TIẾP TỤC NGHIÊN CỨU

1.4.1. Nhận xét, đánh giá

Tổng quan các nhóm công trình nghiên cứu nêu trên cho thấy nền tảng tri thức về AI, phát triển và ứng dụng AI, cũng như các vấn đề pháp lý đặt ra đối với AI đã được các học giả trong và ngoài nước nghiên cứu tương đối phong phú, đặc biệt ở bình diện quốc tế. Các nghiên cứu về vấn đề này cung cấp cơ sở nhận thức quan trọng về đối tượng điều chỉnh, bao gồm bản chất kỹ thuật của AI, xu hướng phát triển, phạm vi ứng dụng và những tác động kinh tế, xã hội, pháp lý của công nghệ này. Đặc biệt, các

nghiên cứu về khung pháp lý AI và các vấn đề pháp lý chuyên biệt như địa vị pháp lý, trách nhiệm pháp lý, dữ liệu cá nhân, quyền sở hữu trí tuệ, đạo đức AI, quyền con người, tiêu chuẩn kỹ thuật và quản trị rủi ro cũng đã phát triển khá mạnh, nhất là sau khi Liên minh châu Âu thúc đẩy và thông qua Đạo luật AI.

Tuy nhiên, khi tiếp cận trực tiếp chủ đề xây dựng khung pháp lý cho phát triển, ứng dụng AI, có thể thấy tình hình nghiên cứu còn nhiều khoảng trống. Phần lớn các công trình hiện có mới tập trung trả lời câu hỏi pháp luật AI cần điều chỉnh những vấn đề gì, trong khi câu hỏi xây dựng khung pháp lý AI như thế nào, theo quy trình nào, với sự tham gia của những chủ thể nào, bằng các kỹ thuật pháp lý và công cụ quản trị nào, lại chưa được phân tích đầy đủ. Nói cách khác, nghiên cứu về nội dung của khung pháp lý AI đã tương đối phát triển, nhưng nghiên cứu về quá trình, phương pháp và cơ chế xây dựng khung pháp lý cho phát triển, ứng dụng AI vẫn còn hạn chế, có thể khái quát như sau:

Thứ nhất, đối với vấn đề cơ sở lý luận về khung pháp lý và xây dựng khung pháp lý, nghiên cứu của các học giả quốc tế đã xây dựng được nền tảng lý luận tương đối vững chắc ở ba cấp độ. Ở cấp độ khái niệm, tiêu biểu là mô hình "hệ sinh thái quy định" của Brownsword và cộng sự. Ở cấp độ cơ chế, lý thuyết về tiêu chuẩn hóa của Majone và lý thuyết ba trục hợp pháp hóa của Abbott và cộng sự đã lý giải được cơ chế vận hành của quan hệ luật cứng - luật mềm. Ở cấp độ triết học pháp lý nền tảng, lý thuyết đồng kiến tạo của Jasanoff và lý thuyết quản trị rủi ro thích ứng của Renn đã cung cấp khung giải thích cho mối quan hệ giữa công nghệ, xã hội và pháp luật trong điều kiện bất ổn. Trong nước, nghiên cứu lý luận nhà nước và pháp luật truyền thống đã có nền tảng vững chắc về nguồn pháp luật, quan hệ pháp luật và xây dựng pháp luật nói chung. Tuy nhiên, có thể nhận diện xuyên suốt ba vấn đề lớn còn bỏ ngỏ: Một là, nghiên cứu quốc tế mạnh về từng lý thuyết riêng lẻ nhưng thiếu một mô hình tích hợp có khả năng lượng hóa. Hai là, các lý thuyết nền tảng (điều chỉnh dựa trên rủi ro, quản trị thích ứng, quyền con người) được nghiên cứu chủ yếu độc lập với nhau trong các dạng tài liệu tách biệt, thiếu công trình xác định rõ ràng mối quan hệ bổ trợ có cấu trúc và nguyên tắc xử lý xung đột giữa chúng. Ba là, trong nước, dù đã có cơ sở mở rộng nguồn pháp luật, chưa có công trình hệ thống hóa các thành tố rời rạc đó (án lệ, lẽ công bằng và tiêu chuẩn kỹ thuật) thành một khái niệm tổng quát về khung pháp lý đa tầng áp dụng chuyên biệt cho lĩnh vực công nghệ cao. Vì vậy, việc xây dựng một nền tảng khái niệm rõ ràng là yêu cầu cần thiết để phát triển nghiên cứu về khung pháp lý AI một cách có hệ thống.

Thứ hai, về kinh nghiệm quốc tế trong xây dựng khung pháp lý AI, kết quả đáng chú ý là khối lượng nghiên cứu mô tả và phân tích các mô hình pháp luật AI quốc

gia và khu vực (EU, Hoa Kỳ, Trung Quốc, Nhật Bản, Hàn Quốc, Singapore, cùng khung khổ của OECD và UNESCO) đã phát triển rất nhanh trong giai đoạn 2021-2026, tương ứng với làn sóng lập pháp AI toàn cầu. Nghiên cứu so sánh đã đạt kết quả quan trọng trong việc rút ra các bài học kỹ thuật lập pháp cụ thể. Tuy nhiên, NCS nhận thấy một số hạn chế: Một là, phần lớn nghiên cứu so sánh dừng lại ở việc mô tả và đối chiếu các mô hình mà chưa phát triển đầy đủ phương pháp luận về điều kiện chuyển giao, ngoài các nhận xét định tính chung chung về khác biệt bối cảnh. Hai là, nghiên cứu so sánh quốc tế gần như hoàn toàn tập trung vào các nền kinh tế phát triển, các trung tâm công nghệ và pháp lý lớn mà thiếu vắng phân tích chuyên sâu về kinh nghiệm của các quốc gia đang phát triển có quy mô kinh tế vừa và nhỏ với đặc điểm hệ sinh thái công nghệ tương tự Việt Nam (phụ thuộc công nghệ nước ngoài cao, năng lực thể chế đang xây dựng, hệ sinh thái doanh nghiệp chủ yếu là SME). Ba là, câu hỏi một quốc gia đang phát triển, có năng lực thể chế và nguồn lực thực thi còn hạn chế nhưng đồng thời có nhu cầu mạnh mẽ về đổi mới sáng tạo và chuyển đổi số, nên xây dựng khung pháp lý AI theo cách nào vẫn chưa được nghiên cứu đầy đủ.

Thứ ba, về thực trạng và giải pháp xây dựng khung pháp lý AI Việt Nam, nghiên cứu trong nước đã phát triển nhanh trong giai đoạn 2021-2026 với đa dạng các chủ đề xoay quanh AI. Tuy nhiên, vẫn còn nhiều vấn đề chưa được giải quyết thấu đáo: Một là, phần lớn các nghiên cứu tiếp cận AI theo từng lĩnh vực pháp luật chuyên biệt, chưa đặt vấn đề về kiến trúc tổng thể của khung pháp lý. Các công trình về AI và pháp luật thường tập trung vào các lĩnh vực cụ thể như dân sự, hình sự, sở hữu trí tuệ, dữ liệu cá nhân, quyền con người, trách nhiệm bồi thường thiệt hại, đạo đức AI hoặc tiêu chuẩn kỹ thuật. Cách tiếp cận này có giá trị trong việc nhận diện các vấn đề pháp lý cụ thể, nhưng chưa trả lời đầy đủ câu hỏi các thành tố đó cần được tổ chức, liên kết và vận hành như thế nào trong một khung pháp lý thống nhất. Trong khi đó, đối với AI, tính liên ngành và xuyên lĩnh vực của công nghệ đòi hỏi phải có cách tiếp cận theo kiến trúc tổng thể, thay vì chỉ xử lý từng vấn đề pháp lý riêng lẻ. Hai là, các nghiên cứu còn ít quan tâm đến quá trình và cơ chế xây dựng khung pháp lý AI. Nhiều công trình đề xuất các nội dung cần có trong pháp luật AI, như phân loại rủi ro, minh bạch thuật toán, trách nhiệm giải trình, bảo vệ dữ liệu, kiểm toán AI, sandbox hoặc quản trị đạo đức. Tuy nhiên, còn ít công trình phân tích một cách hệ thống quá trình xây dựng các quy định đó, bao gồm xác định vấn đề chính sách, đánh giá tác động, lựa chọn mô hình điều chỉnh, thiết kế quy phạm, tham vấn các bên liên quan, thử nghiệm chính sách, tổ chức thực thi và cập nhật sau ban hành. Đây là khoảng trống quan trọng, bởi trong lĩnh vực công nghệ phát triển nhanh như AI, chất lượng của khung pháp lý không chỉ phụ thuộc vào nội dung quy định, mà còn phụ thuộc vào quy trình và năng lực thể chế để

xây dựng, vận hành và điều chỉnh khung pháp lý đó. Ba là, các nghiên cứu trong nước chưa trả lời toàn diện câu hỏi Việt Nam cần tiếp tục làm gì sau khi đã có Luật AI 2025. Sự ra đời của Luật AI 2025 là bước phát triển quan trọng trong quá trình hình thành khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam. Tuy nhiên, một đạo luật chuyên biệt mới chỉ là trung tâm của khung pháp lý, chứ chưa phải toàn bộ khung pháp lý. Vấn đề đặt ra hiện nay không chỉ là đánh giá nội dung của Luật AI, mà còn là kiến trúc tổng thể của khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam sau khi luật đã được ban hành: Luật AI cần được kết nối như thế nào với các luật chuyên ngành, văn bản hướng dẫn, tiêu chuẩn kỹ thuật, quy tắc đạo đức, sandbox, cơ chế tự quản, thiết chế thực thi, cơ chế giám sát, đánh giá và cập nhật chính sách. Đây là câu hỏi có ý nghĩa học thuật và thực tiễn cấp bách, nhưng chưa được công trình nào nghiên cứu một cách hệ thống.

Từ những phân tích trên, có thể thấy, khoảng trống không nằm ở việc thiếu hoàn toàn các nghiên cứu về AI và pháp luật AI, mà nằm ở việc thiếu một công trình tiếp cận vấn đề xây dựng khung pháp lý cho phát triển, ứng dụng AI như một quá trình có tính lý luận, có cấu trúc, có phương pháp và gắn với điều kiện cụ thể của Việt Nam. Đây là cơ sở để nghiên cứu sinh xác định hướng nghiên cứu đặc thù của luận án.

1.4.2. Những vấn đề luận án tiếp tục nghiên cứu

Từ việc nhận diện những kết quả và hạn chế của tình hình nghiên cứu, luận án xác định một số vấn đề tiếp tục nghiên cứu như sau:

Thứ nhất, luận án tiếp tục xây dựng cơ sở lý luận về khung pháp lý cho phát triển, ứng dụng AI theo mô hình đa tầng. Nội dung này bao gồm việc làm rõ khái niệm "khung pháp lý AI", phân biệt "khung pháp lý AI" với "khung pháp luật AI", xác định các thành tố cấu thành của khung pháp lý cho phát triển, ứng dụng AI và đề xuất mô hình kiến trúc phù hợp với điều kiện Việt Nam. Trên cơ sở đó, luận án không tiếp cận khung pháp lý AI chỉ như một đạo luật chuyên biệt hay một tập hợp văn bản quy phạm pháp luật, mà như một chỉnh thể có hệ thống các thành tố rời rạc của lý luận nguồn pháp luật Việt Nam vào một khái niệm chuyên biệt cho công nghệ cao. Luận án xác lập quan hệ giữa ba lý thuyết nền tảng kèm ba nguyên tắc xử lý xung đột.

Thứ hai, luận án tiếp tục làm rõ lý luận về xây dựng khung pháp lý cho phát triển, ứng dụng AI. Đây là nội dung có ý nghĩa phân biệt luận án với nhiều nghiên cứu hiện có. Luận án không chỉ nghiên cứu khung pháp lý cho phát triển, ứng dụng AI cần có những nội dung gì, mà tập trung làm rõ quá trình xây dựng khung pháp lý đó. Cụ thể, luận án sẽ phân tích khái niệm "xây dựng" khung pháp lý cho phát triển, ứng dụng AI, phân biệt với các khái niệm gần nghĩa như thiết lập, hoàn thiện, phát triển hoặc sửa đổi pháp luật; xác định các giai đoạn cơ bản của quá trình xây dựng; làm rõ chủ thể, nội

dung, quy trình và các kỹ thuật xây dựng khung pháp lý cho phát triển, ứng dụng AI trong điều kiện công nghệ phát triển nhanh, rủi ro khó dự báo và nhu cầu điều chỉnh liên ngành cao.

Thứ ba, luận án tiếp tục phân tích thực trạng xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam theo nội hàm mới. Việc đánh giá thực trạng không chỉ dừng lại ở việc liệt kê các văn bản pháp luật hiện hành hoặc phân tích nội dung của Luật AI 2025, mà còn xem xét mức độ đầy đủ, thống nhất, khả thi và thích ứng của toàn bộ khung pháp lý AI. Luận án sẽ đánh giá các tầng cấu thành của khung pháp lý, phân tích quá trình xây dựng dưới các góc độ chủ thể, nội dung, quy trình và kỹ thuật, đồng thời sử dụng một số tình huống điển hình để kiểm chứng những vấn đề đặt ra trong thực tiễn ở Việt Nam.

Thứ tư, luận án tiếp tục đề xuất hệ thống giải pháp hoàn thiện khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam theo cả hai chiều: hoàn thiện sản phẩm pháp lý và hoàn thiện quá trình xây dựng khung pháp lý. Về phương diện sản phẩm, luận án hướng tới việc đề xuất giải pháp hoàn thiện cấu trúc các tầng của khung pháp lý cho phát triển, ứng dụng AI, bảo đảm sự kết nối giữa Luật AI, pháp luật chuyên ngành, văn bản dưới luật, tiêu chuẩn kỹ thuật, quy tắc đạo đức, cơ chế tự quản và thiết chế thực thi. Về phương diện quá trình, luận án đề xuất các giải pháp nâng cao chất lượng xây dựng khung pháp lý cho phát triển, ứng dụng AI.

Như vậy, luận án tiếp tục nghiên cứu xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam không chỉ là một vấn đề nội dung pháp luật, mà còn là một vấn đề lý luận và phương pháp luận về xây dựng pháp luật trong bối cảnh công nghệ mới. Cách tiếp cận này cho phép luận án vừa kế thừa các kết quả nghiên cứu hiện có, vừa góp phần lấp đầy khoảng trống về khái niệm, kiến trúc, quy trình, kỹ thuật và giải pháp xây dựng khung pháp lý AI phù hợp với điều kiện Việt Nam.

1.5. GIẢ THUYẾT KHOA HỌC VÀ CÂU HỎI NGHIÊN CỨU

1.5.1. Giả thuyết khoa học

Giả thuyết khoa học của luận án là: *Trong bối cảnh AI phát triển nhanh, tác động xuyên lĩnh vực và làm phát sinh nhiều rủi ro khó dự báo, việc xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam chỉ có thể đạt hiệu quả nếu được tiếp cận như một quá trình thiết kế thể chế đa tầng, thích ứng và dựa trên rủi ro*.

1.5.2. Câu hỏi nghiên cứu

Xuất phát từ mục tiêu nghiên cứu và giả thuyết khoa học của luận án, câu hỏi nghiên cứu khái quát được đặt ra là: *Việt Nam cần xây dựng khung pháp lý cho phát triển, ứng dụng AI theo mô hình, quy trình và giải pháp nào để vừa thúc đẩy đổi mới sáng tạo, phát triển kinh tế - xã hội, vừa kiểm soát hiệu quả các rủi ro pháp lý, đạo đức, xã hội và quyền con người do AI đặt ra?*

Để trả lời câu hỏi nghiên cứu khái quát đó, luận án tập trung làm rõ ba câu hỏi nghiên cứu cụ thể sau đây:

- Khung pháp lý cho phát triển, ứng dụng AI cần được hiểu như thế nào, bao gồm những thành tố nào, nên được thiết kế theo mô hình nào; đồng thời quá trình xây dựng khung pháp lý cho phát triển, ứng dụng AI bao gồm những chủ thể, nội dung, giai đoạn, quy trình và kỹ thuật xây dựng nào?

- Các chuẩn mực, khuyến nghị của các tổ chức quốc tế và kinh nghiệm xây dựng khung pháp lý cho phát triển, ứng dụng AI của một số quốc gia, khu vực tiêu biểu cung cấp những giá trị tham khảo gì cho Việt Nam trong việc lựa chọn mô hình, nguyên tắc, công cụ và kỹ thuật xây dựng khung pháp lý cho phát triển, ứng dụng AI?

- Khung pháp lý cho phát triển, ứng dụng AI cũng như quá trình xây dựng khung pháp lý cho phát triển, ứng dụng AI hiện nay đang bộc lộ những khoảng trống, chồng chéo, bất cập hoặc hạn chế nào; nguyên nhân của những hạn chế đó là gì; và cần có những quan điểm, giải pháp nào để hoàn thiện khung pháp lý AI theo hướng đầy đủ hơn về cấu trúc, khả thi hơn trong thực thi và thích ứng tốt hơn với sự phát triển của công nghệ?

Tổng kết Chương 1

Chương 1 đã tổng quan tình hình nghiên cứu liên quan đến đề tài luận án theo ba nhóm nội dung: nghiên cứu về AI và phát triển, ứng dụng AI; nghiên cứu về khung pháp lý cho phát triển, ứng dụng AI; và nghiên cứu về xây dựng khung pháp lý cho phát triển, ứng dụng AI. Kết quả cho thấy, các công trình trong và ngoài nước đã cung cấp nền tảng tri thức quan trọng về bản chất, ứng dụng, tác động của AI, cũng như nhiều vấn đề pháp lý chuyên biệt như dữ liệu, trách nhiệm pháp lý, quyền SHTT, đạo đức AI, quyền con người và tiêu chuẩn kỹ thuật.

Tuy nhiên, các nghiên cứu hiện có chủ yếu tập trung vào nội dung cần điều chỉnh hoặc kinh nghiệm pháp luật của một số quốc gia, trong khi chưa nghiên cứu một cách toàn diện quá trình xây dựng khung pháp lý AI như một vấn đề lý luận và phương pháp luận của chuyên ngành Lý luận và lịch sử nhà nước và pháp luật. Đặc biệt, còn thiếu các nghiên cứu làm rõ khái niệm, cấu trúc, mô hình kiến trúc, chủ thể, quy trình, kỹ thuật xây dựng và cơ chế thích ứng của khung pháp lý cho phát triển, ứng dụng AI trong điều kiện cụ thể của Việt Nam.

Trên cơ sở đó, Chương 1 xác định khoảng trống nghiên cứu trung tâm của luận án là việc xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam như một quá trình thiết kế thể chế đa tầng, thích ứng và dựa trên rủi ro. Khoảng trống này là cơ sở để luận án tiếp tục xây dựng khung lý luận, phân tích kinh nghiệm quốc tế, đánh giá thực trạng Việt Nam và đề xuất quan điểm, giải pháp hoàn thiện trong các chương tiếp theo.

Chương 2
CƠ SỞ LÝ LUẬN VỀ XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO

2.1. KHÁI NIỆM, ĐẶC ĐIỂM, VAI TRÒ VÀ NỘI DUNG CỦA XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO

2.1.1. Khái niệm, đặc điểm, vai trò của xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

2.1.1.1. Khái niệm khung pháp lý

Trong khoa học pháp lý trên thế giới và ở Việt Nam, "khung pháp lý" (legal framework) là thuật ngữ được sử dụng khá phổ biến khi bàn về điều chỉnh pháp luật đối với một lĩnh vực nhất định, đặc biệt là các lĩnh vực mới phát sinh từ quá trình phát triển kinh tế, xã hội và công nghệ. Tuy nhiên, thuật ngữ này không phải lúc nào cũng được hiểu thống nhất. Trong nhiều trường hợp, "khung pháp lý" được sử dụng gần với các khái niệm như "hệ thống pháp luật", "chính sách pháp luật", "khung pháp luật" hoặc "cơ chế pháp lý". Vì vậy, trước khi xác định khái niệm khung pháp lý cho phát triển, ứng dụng AI, cần làm rõ nội hàm chung của khái niệm "khung pháp lý" và phân biệt với một số khái niệm có liên quan.

Về cơ bản, có thể tiếp cận khái niệm khung pháp lý theo nghĩa hẹp và nghĩa rộng. Theo nghĩa hẹp, khung pháp lý bao gồm các bộ luật, quy định và quy tắc áp dụng ở một quốc gia cụ thể, điều chỉnh mối quan hệ giữa nhà nước và các chủ thể trong xã hội, đồng thời xác định các giới hạn cho hành vi pháp lý một cách chặt chẽ. Với nội hàm này, pháp luật phải phù hợp với các chuẩn mực quốc tế, chắc chắn về mặt pháp lý, được soạn thảo minh bạch, được ban hành công khai và đảm bảo kiểm soát quyền lực [290]. Khung pháp lý được xác định bởi các chính sách và luật pháp cấp quốc gia và địa phương, tạo thành "luật chơi", tạo điều kiện cho tất cả các bên liên quan thực hiện vai trò của mình [189].

Theo nghĩa rộng, khung pháp lý không chỉ bao gồm thành phần cốt lõi của bản thân pháp luật mà còn bao gồm các điều kiện hoặc sự sắp xếp về thể chế, hành chính, kinh tế, chính trị, xã hội, giúp cho pháp luật dễ tiếp cận, có thể thực thi và có hiệu quả [175]. Cách hiểu này cho phép nhìn nhận khung pháp lý không chỉ như một tập hợp các văn bản quy phạm pháp luật, mà như một cấu trúc điều chỉnh rộng hơn, bao gồm cả các chuẩn mực bổ trợ, thiết chế thực thi và cơ chế vận hành cần thiết để pháp luật đi vào đời sống. Cách tiếp cận theo nghĩa rộng này cũng phù hợp với cách sử dụng khái niệm "legal framework" trong nhiều tài liệu quốc tế. Trong các nguồn này, "khung pháp lý" thường không được hiểu đơn thuần là một đạo luật hoặc một nhóm văn bản quy phạm pháp luật, mà là tổng thể các quy tắc, chính sách, thiết chế và cơ chế vận

hành điều chỉnh một lĩnh vực cụ thể. Chẳng hạn, trong lĩnh vực bầu cử, khung pháp lý và quy định được nhìn nhận như tập hợp các quy tắc hiến định, lập pháp, quy định, án lệ và quản lý cùng nhau xác lập cách thức thực hiện quyền bầu cử [140]. Trong lĩnh vực đối tác công tư, khung pháp lý và thể chế được hiểu là tổng thể các chính sách, quy trình, thiết chế và quy tắc xác định cách thức dự án được nhận diện, đánh giá, lựa chọn, ưu tiên, bố trí ngân sách, mua sắm, giám sát, hạch toán và phân bổ trách nhiệm [366]. Trong lĩnh vực định danh số, khung pháp lý được xem là nền tảng của lòng tin và trách nhiệm giải trình, bao gồm các luật, quy tắc, quy định và thực tiễn điều chỉnh, hỗ trợ hệ thống định danh [367]. Trong khuôn khổ các Mục tiêu phát triển bền vững, chỉ số SDG 5.a.2 cũng sử dụng khái niệm "legal framework" theo nghĩa bao gồm cả luật thành văn và tập quán pháp trong việc bảo đảm quyền bình đẳng của phụ nữ đối với đất đai [361]. Những cách tiếp cận này cho thấy khung pháp lý, nhất là trong các lĩnh vực phức tạp và có tính liên ngành, cần được hiểu như một cấu trúc điều chỉnh bao gồm cả quy phạm, thiết chế và cơ chế vận hành.

Từ những phân tích ở trên, trong luận án này, NCS tiếp cận khái niệm khung pháp lý theo nghĩa rộng. Theo đó, *khung pháp lý là một hệ thống có cấu trúc, gồm các công cụ điều chỉnh và các thiết chế bảo đảm thực hiện được tổ chức theo những nguyên tắc nhất định, có mối liên hệ và tác động qua lại để điều chỉnh một lĩnh vực quan hệ xã hội nhằm đạt được mục tiêu đề ra.*

Luận án áp dụng đồng thời ba tiêu chí lựa chọn một thành tố chỉ được xếp vào khung pháp lý khi đáp ứng đầy đủ Tiêu chí 1 và Tiêu chí 2; Tiêu chí 3 là điều kiện củng cố mức độ đầy đủ của tư cách pháp lý. Tiêu chí 1 về nguồn gốc thẩm quyền là thành tố phải được tạo lập thông qua một quy trình hợp pháp, do cơ quan nhà nước có thẩm quyền ban hành theo đúng trình tự luật định (đối với văn bản quy phạm pháp luật), hoặc được chính hệ thống pháp luật thực định thừa nhận là nguồn bổ trợ có giá trị áp dụng (đối với án lệ được công bố theo cơ chế của cơ quan xét xử tối cao và lẽ công bằng được viện dẫn theo quy định của pháp luật dân sự). Tiêu chí 2 về hiệu lực pháp lý là thành tố phải tạo ra quyền, nghĩa vụ hoặc thẩm quyền có tính ràng buộc đối với chủ thể bị điều chỉnh trực tiếp hoặc gián tiếp thông qua cơ chế luật hóa. Tiêu chí 3 về khả năng bảo đảm thực hiện bằng quyền lực nhà nước là việc vi phạm thành tố đó có thể dẫn đến hậu quả pháp lý được bảo đảm thực hiện bằng quyền lực cưỡng chế nhà nước.

Theo định nghĩa rộng nêu trên, có thể chia khung pháp lý ở Việt Nam gồm hai nhóm thành tố, mỗi nhóm được cụ thể hóa thành các tầng như sau: Nhóm một - công cụ điều chỉnh gồm sáu tầng: tầng 1 là nền tảng hiến định và văn kiện quốc tế; tầng 2 là văn bản quy phạm pháp luật chuyên biệt và liên quan; tầng 3 là tiêu chuẩn kỹ thuật có giá trị pháp lý; tầng 4 là luật mềm [103, tr.3-9] và hướng dẫn ngành; tầng 5 là bộ quy

tắc ứng xử và tự quản ngành là tập hợp các chuẩn mực hành vi do tổ chức nghề nghiệp, hiệp hội ngành hoặc chính doanh nghiệp tự xây dựng và cam kết thực hiện; tầng 6 là án lệ [49], tập quán pháp [100] và lẽ công bằng [101, tr.11-16]. Nhóm hai - thiết chế bảo đảm thực hiện là tầng 7, bao gồm các cơ quan nhà nước có thẩm quyền điều chỉnh, giám sát và xử lý vi phạm, cùng các cơ chế vận hành cụ thể như cấp phép, kiểm định, đánh giá tác động, báo cáo sự cố.

Cũng trên cơ sở cách hiểu rộng như trên, có thể phân biệt "khung pháp lý" với một số khái niệm gần như sau:

Thứ nhất, so sánh với "hệ thống pháp luật". Đây là khái niệm bao quát nhất, được hiểu là "tổng thể các quy phạm pháp luật, các nguyên tắc, quan điểm, định hướng, chính sách, mục đích pháp luật có mối liên hệ mật thiết và thống nhất với nhau, được phân định thành các nhóm ngành, các ngành luật, các chế định pháp luật và được thể hiện trong các văn bản pháp luật do các cơ quan nhà nước có thẩm quyền ban hành theo những hình thức, thủ tục luật định và các nguồn khác theo quy định của pháp luật" [98]. Hệ thống pháp luật khác với khung pháp lý ở chỗ: hệ thống pháp luật mô tả toàn bộ kiến trúc pháp luật của một quốc gia, trong khi khung pháp lý chỉ mô tả kiến trúc pháp luật cho một lĩnh vực cụ thể.

Thứ hai, so sánh với "chính sách pháp luật" thường được hiểu là hệ thống các quan điểm, lập trường, tư tưởng, nguyên tắc hành động, chủ trương, định hướng, chương trình và biện pháp hành động của Nhà nước trong việc xác định vai trò và hướng điều chỉnh của pháp luật trong từng giai đoạn hoặc thời kỳ phát triển của đất nước [65, tr.10-14]. Chính sách pháp luật khác với khung pháp lý ở chỗ chính sách là ý chí chính trị và định hướng chiến lược, còn khung pháp lý là cấu trúc quy phạm và thể chế cụ thể để hiện thực hóa ý chí và định hướng đó.

Thứ ba, so sánh với "khung pháp luật" là khái niệm có xu hướng nhấn mạnh hơn đến các văn bản quy phạm pháp luật được Nhà nước ban hành và bảo đảm thi hành. Đây là khái niệm thiên về quy phạm cứng, trong khi "khung pháp lý" có phạm vi rộng hơn, bao gồm cả tiêu chuẩn kỹ thuật, hướng dẫn đạo đức và các thiết chế thực thi.

Thứ tư, khung quản trị (governance framework) của một lĩnh vực quan hệ xã hội được xác định là tổng thể các phương thức, công cụ, chủ thể, quy trình và cơ chế phối hợp được huy động nhằm định hướng, điều phối, giám sát và điều chỉnh hành vi của các chủ thể liên quan trong lĩnh vực đó, hướng tới đạt được mục tiêu chính sách công đã xác định. Khung quản trị có nội hàm rộng hơn khung pháp lý ở ba điểm cụ thể: một là, khung quản trị bao gồm cả công cụ không có bất kỳ liên hệ nào với hệ thống nguồn pháp luật như chương trình đào tạo nội bộ tổ chức, cơ chế thị trường (đánh giá của nhà đầu tư hoặc người tiêu dùng đối với thực hành có trách nhiệm của doanh nghiệp), hoặc áp lực cộng đồng chuyên môn không qua bất kỳ

kênh thể chế nào. Hai là, khung quản trị bao gồm toàn bộ quy trình vận hành nội bộ của bộ máy nhà nước (cách thức phối hợp công tác giữa các đơn vị, quy trình xử lý hồ sơ hành chính, công cụ quản lý dữ liệu tác nghiệp) vốn không tự thân tạo ra hoặc thay đổi quyền và nghĩa vụ pháp lý của chủ thể bên ngoài. Ba là, khung quản trị bao gồm cả năng lực và nguồn lực thể chế (con người, ngân sách, hạ tầng kỹ thuật) cần thiết để vận hành mọi công cụ, kể cả công cụ pháp lý, năng lực và nguồn lực tự thân không phải là quy phạm nhưng là điều kiện để quy phạm phát huy hiệu lực trên thực tế. Từ đó, quan hệ giữa hai khung là quan hệ bao hàm không đầy đủ, không đồng nhất cũng không tách biệt hoàn toàn. Biểu đạt bằng ngôn ngữ tập hợp, khung pháp lý là tập con của khung quản trị với phần giao thoa chiếm tỷ trọng lớn và có vị trí trung tâm.

Tóm lại, so với hệ thống pháp luật, khung pháp lý có phạm vi hẹp hơn vì gắn với một lĩnh vực quan hệ xã hội cụ thể; so với chính sách pháp luật, khung pháp lý có tính quy phạm và thể chế cụ thể hơn; so với khung pháp luật, khung pháp lý có phạm vi rộng hơn vì bao gồm không chỉ văn bản quy phạm pháp luật mà còn cả các chuẩn mực bổ trợ, công cụ điều chỉnh mềm và thiết chế thực thi. Cách phân biệt này có ý nghĩa quan trọng đối với luận án, bởi khung pháp lý cho phát triển, ứng dụng AI không thể chỉ được hiểu là hệ thống các văn bản pháp luật về AI mà cần được nhìn nhận như một cấu trúc điều chỉnh nhiều thành phần.

2.1.1.2. Khái niệm, đặc điểm, phân loại, bản chất pháp lý của trí tuệ nhân tạo

a. Khái niệm trí tuệ nhân tạo

Việc xác định khái niệm trí tuệ nhân tạo (Artificial Intelligence - AI) có ý nghĩa nền tảng đối với xây dựng khung pháp lý về AI, bởi đây là cơ sở để xác định đối tượng điều chỉnh, phạm vi áp dụng, nghĩa vụ của các chủ thể liên quan và cơ chế kiểm soát rủi ro. Trong các từ điển, bách khoa thư và một số tài liệu kỹ thuật, trí tuệ nhân tạo thường được hiểu là khả năng của máy tính, máy móc hoặc hệ thống thuật toán trong việc thực hiện những nhiệm vụ vốn gắn với trí tuệ con người, như học tập, suy luận, nhận thức, giải quyết vấn đề, hiểu ngôn ngữ hoặc đưa ra quyết định [156], [272]. Trong khoa học máy tính và nghiên cứu AI, các học giả cũng đưa ra nhiều cách định nghĩa khác nhau. John McCarthy, người đặt nền móng cho thuật ngữ AI, định nghĩa trí tuệ nhân tạo là khoa học và kỹ thuật tạo ra các máy móc thông minh, đặc biệt là các chương trình máy tính thông minh [242]. Russell và Norvig tiếp cận AI thông qua mô hình tác nhân thông minh, nhấn mạnh khả năng nhận thức môi trường, xử lý thông tin và hành động để đạt mục tiêu [338].

Các định nghĩa kỹ thuật, học thuật nêu trên có giá trị quan trọng trong việc làm rõ bản chất kỹ thuật và logic vận hành của AI nhưng chưa đủ cho mục đích pháp lý, bởi pháp luật không chỉ cần biết AI là gì về mặt kỹ thuật, mà còn phải xác định hệ

thống nào thuộc phạm vi điều chỉnh, chủ thể nào có nghĩa vụ, hành vi nào bị kiểm soát và rủi ro nào cần được phòng ngừa [356]. Chính vì vậy, trong lĩnh vực pháp luật và quản trị công, AI không chỉ được nhận diện như một năng lực hoặc công nghệ trừu tượng, mà thường được thao tác hóa thông qua khái niệm "hệ thống AI" (AI system). Đây là thuật ngữ được sử dụng phổ biến trong các văn kiện quốc tế và pháp luật thực định gần đây [298], [284], [206], vì nó cho phép xác định cụ thể đối tượng điều chỉnh, phạm vi áp dụng, nghĩa vụ của các chủ thể liên quan và cơ chế kiểm soát rủi ro.

Các văn kiện quốc tế và pháp luật thực định gần đây thể hiện tính đồng thuận cao trong cách định nghĩa hệ thống AI. OECD trong phiên bản cập nhật năm 2024 của Bộ nguyên tắc AI định nghĩa hệ thống AI là một hệ thống dựa trên máy móc, vì các mục tiêu rõ ràng hoặc ngầm định, suy luận từ đầu vào nhận được cách tạo ra đầu ra như dự đoán, nội dung, đề xuất hoặc quyết định có thể ảnh hưởng đến môi trường vật lý hoặc ảo [298]. Theo cách tiếp cận tương tự, NIST trong Artificial Intelligence Risk Management Framework (AI RMF 1.0) tiếp cận hệ thống AI như hệ thống kỹ thuật hoặc dựa trên máy móc, có khả năng tạo ra các đầu ra như dự đoán, khuyến nghị hoặc quyết định, tác động đến môi trường thực hoặc ảo, với các mức độ tự chủ khác nhau [284]. Xét về pháp luật thực định, EU AI Act định nghĩa tại Điều 3(1): "hệ thống AI là hệ thống dựa trên máy móc được thiết kế để hoạt động với các mức độ tự chủ khác nhau và có thể thể hiện khả năng thích ứng sau khi triển khai, và, với các mục tiêu rõ ràng hoặc ngầm định, suy ra từ đầu vào mà nó nhận được cách tạo ra đầu ra như dự đoán, nội dung, khuyến nghị hoặc quyết định có thể ảnh hưởng đến các môi trường vật lý hoặc ảo" [206]. AI Basic Act của Hàn Quốc định nghĩa tại Điều 2: AI liên quan đến việc triển khai bằng điện tử các khả năng trí tuệ của con người, chẳng hạn như học tập, suy luận, nhận thức, phán đoán và hiểu ngôn ngữ; hệ thống AI là hệ thống dựa trên AI suy luận ra các kết quả như dự đoán, khuyến nghị và quyết định ảnh hưởng đến môi trường thực và ảo cho một mục tiêu nhất định với các mức độ tự chủ và khả năng thích ứng khác nhau [277]. Ở Việt Nam, theo khoản 1 Điều 3 Luật AI năm 2025, AI là "việc thực hiện bằng điện tử các năng lực trí tuệ của con người, bao gồm học tập, suy luận, nhận thức, phán đoán và hiểu ngôn ngữ tự nhiên" [91]. Theo khoản 2 Điều 3 Luật AI năm 2025, hệ thống AI là "hệ thống dựa trên máy, được thiết kế để thực hiện các năng lực trí tuệ nhân tạo với các mức độ tự chủ khác nhau, có khả năng tự thích nghi sau khi được triển khai; căn cứ vào các mục tiêu được xác định rõ ràng hoặc được hình thành một cách ngầm định, hệ thống suy luận từ dữ liệu đầu vào để tạo ra các đầu ra như dự đoán, nội dung, khuyến nghị hoặc quyết định có thể gây ảnh hưởng tới môi trường vật lý hoặc môi trường số".

Từ các cách tiếp cận trên, trong luận án này, *trí tuệ nhân tạo được hiểu là phương thức công nghệ được thực hiện bằng hệ thống máy móc, máy tính hoặc thuật toán nhằm mô phỏng, hỗ trợ hoặc thực hiện một số năng lực trí tuệ của con người như học tập, suy luận, nhận thức, phán đoán, hiểu ngôn ngữ, tạo lập nội dung, đưa ra khuyến nghị hoặc hỗ trợ ra quyết định*. Trong khi đó, *hệ thống trí tuệ nhân tạo được hiểu là hệ thống dựa trên máy, được thiết kế để hoạt động với mức độ tự chủ nhất định, có khả năng suy luận từ dữ liệu đầu vào để tạo ra các đầu ra như dự đoán, nội dung, khuyến nghị hoặc quyết định có thể tác động đến cá nhân, tổ chức, môi trường vật lý hoặc môi trường số*. Luận án sẽ sử dụng đồng thời hai khái niệm này tuỳ hoàn cảnh. Với những phân tích như trên, việc sử dụng như vậy không làm thay đổi phạm vi của đề tài, vì AI là khái niệm bao quát, dùng để chỉ lĩnh vực công nghệ, chính sách và pháp luật nói chung; còn hệ thống AI là khái niệm có tính thao tác pháp lý, dùng để xác định cụ thể đối tượng chịu sự điều chỉnh, nghĩa vụ của các chủ thể liên quan, cơ chế phân loại rủi ro và trách nhiệm pháp lý. Vì vậy, khi luận án đề cập đến xây dựng khung pháp lý cho phát triển, ứng dụng AI, nội dung điều chỉnh trực tiếp trong nhiều trường hợp chính là việc nghiên cứu, phát triển, cung cấp, triển khai, sử dụng và giám sát các hệ thống AI trong đời sống xã hội.

b. Đặc điểm của trí tuệ nhân tạo

Khung pháp lý chỉ có thể được xây dựng hợp lý khi xuất phát từ bản chất của đối tượng điều chỉnh. AI, đặc biệt là các tác nhân AI thế hệ mới, có một số đặc trưng kỹ thuật làm phát sinh những vấn đề pháp lý mới, khó xử lý bằng các mô hình điều chỉnh truyền thống. Luận án phân chia hệ thống đặc điểm của AI theo hai nhóm có quan hệ nhân quả rõ ràng: Nhóm A là đặc điểm bản chất (nội tại) của công nghệ AI, là nguồn gốc trực tiếp phát sinh các vấn đề pháp lý cốt lõi và Nhóm B là hệ quả pháp lý phát sinh từ điều kiện và quy mô triển khai, là biểu hiện phái sinh khi các đặc điểm nội tại đó được đưa vào vận hành trong thực tiễn kinh tế - xã hội có tính toàn cầu hóa và quy mô lớn.

Nhóm A: Đặc điểm bản chất (nội tại) của AI

Thứ nhất, tính dựa trên dữ liệu: Đây là đặc tính nền tảng và có vị trí trung tâm nhất trong toàn bộ hệ thống đặc điểm của AI. Khác với phần mềm truyền thống vận hành theo logic quy tắc được lập trình bởi con người, trong đó mỗi đầu ra được suy ra trực tiếp từ một quy tắc do lập trình viên viết ra, hệ thống AI hiện đại, đặc biệt các mô hình học máy và học sâu tạo ra năng lực dự đoán, phân loại hoặc suy luận của mình thông qua quá trình học từ dữ liệu huấn luyện, trong đó các tham số của mô hình được điều chỉnh tự động dựa trên mẫu hình thống kê được phát hiện trong khối lượng lớn dữ liệu, không phải được lập trình viên chỉ định trực tiếp. Vị

trí trung tâm của đặc tính này trong hệ thống lý luận của luận án nằm ở chỗ tính dựa trên dữ liệu là gốc rễ trực tiếp phát sinh bốn nhóm vấn đề pháp lý. Một là, *chất lượng dữ liệu*: vì hành vi của hệ thống AI được quyết định bởi dữ liệu huấn luyện, chất lượng, tính đại diện, tính đầy đủ và nguồn gốc hợp pháp của dữ liệu huấn luyện trở thành yếu tố quyết định chất lượng đầu ra. Hai là, *thiên lệch thuật toán*: nếu dữ liệu huấn luyện phản ánh những thiên lệch lịch sử có sẵn trong xã hội, quá trình học từ dữ liệu sẽ khiến mô hình AI học và tái tạo lại chính những thiên lệch đó. Ba là, *quyền riêng tư và bảo vệ dữ liệu cá nhân*: việc huấn luyện mô hình AI đòi hỏi khối lượng dữ liệu lớn, trong đó phần lớn trường hợp có chứa hoặc suy luận được thông tin cá nhân, làm phát sinh câu hỏi pháp lý về căn cứ hợp pháp cho việc thu thập, sử dụng dữ liệu cá nhân vào mục đích huấn luyện. Bốn là, *cơ sở của vấn đề trách nhiệm pháp lý*: chính vì đầu ra của AI được quyết định bởi tương tác giữa dữ liệu huấn luyện, kiến trúc mô hình và dữ liệu đầu vào tại thời điểm vận hành, việc xác định nguyên nhân của một quyết định hoặc đầu ra cụ thể của AI trở nên phức tạp hơn nhiều so với việc truy vết lỗi trong phần mềm truyền thống.

Thứ hai, khả năng tự học và thích nghi liên tục: một số hệ thống AI tiếp tục điều chỉnh hành vi của mình sau khi đã được triển khai vào môi trường vận hành thực tế thông qua cơ chế học liên tục, học tăng cường trong môi trường tương tác hoặc đơn giản là sự trôi dạt phân phối dữ liệu đầu vào theo thời gian làm thay đổi hiệu năng của mô hình dù bản thân tham số mô hình không được cập nhật chủ động. Đây là đặc tính nội tại phân biệt AI với hầu hết các đối tượng điều chỉnh truyền thống của pháp luật như sản phẩm vật chất và cả phần mềm truyền thống. Chính đặc tính này là căn cứ lý luận trực tiếp cho yêu cầu giám sát liên tục sau triển khai và triết lý cơ chế thích ứng thể chế hóa vì đối tượng điều chỉnh tự thân có khả năng biến đổi sau thời điểm quy phạm được áp dụng lần đầu.

Thứ ba, hạn chế về khả năng giải thích được ("hộp đen"): Đặc tính nội tại thứ ba, đặc biệt rõ nét ở các mô hình học sâu có kiến trúc mạng nơ-ron nhiều lớp, là khoảng cách giữa khả năng dự đoán chính xác của mô hình và khả năng diễn giải bằng ngôn ngữ con người có thể hiểu được về lý do dẫn đến một đầu ra cụ thể, được gọi phổ biến là đặc tính "hộp đen" (black box). Đặc tính này là căn cứ lý luận trực tiếp cho yêu cầu về quyền được giải thích của người bị ảnh hưởng bởi quyết định của AI; cho học thuyết "trách nhiệm giải trình thuật toán" của Pasquale (đã phân tích tại Chương 1) và cho yêu cầu về hồ sơ kỹ thuật bắt buộc.

Thứ tư, tính tự chủ hành vi ở mức độ khác nhau nổi bật ở tác nhân AI là khả năng hệ thống AI tự chủ lập kế hoạch và thực thi chuỗi hành động liên tiếp mà không cần phê duyệt của con người cho từng bước, khác biệt về mức độ so với hệ thống AI thụ động chỉ đưa ra một đầu ra đơn lẻ theo yêu cầu (như phân loại một

hình ảnh hay trả lời một câu hỏi) rồi dừng lại chờ tương tác tiếp theo của con người. Đây là đặc tính nội tại vì mức độ tự chủ được quyết định bởi kiến trúc và mục tiêu thiết kế của hệ thống, không phụ thuộc vào quy mô hay phạm vi triển khai. Đặc tính này là căn cứ lý luận cho việc phân tích khoảng trống định danh pháp lý đối với tác nhân AI và cho yêu cầu về kiểm soát có ý nghĩa của con người.

Nhóm B: Hệ quả phát sinh từ điều kiện và quy mô triển khai AI

Thứ nhất, tính xuyên biên giới trong triển khai và tác động thể hiện ở việc nhà phát triển mô hình đặt tại một quốc gia, hạ tầng tính toán đám mây đặt tại quốc gia khác, dữ liệu huấn luyện thu thập từ nhiều quốc gia, và người dùng cuối chịu tác động tại một quốc gia thứ tư không phải là đặc tính nội tại của thuật toán AI (một mô hình học máy về bản chất kỹ thuật không có quốc tịch hay bị giới hạn bởi biên giới lãnh thổ theo cách các đối tượng vật chất bị giới hạn), mà là hệ quả của điều kiện triển khai trong nền kinh tế số toàn cầu hóa. Vấn đề này đòi hỏi công cụ pháp lý về thẩm quyền lãnh thổ mở rộng, hợp tác quốc tế và nghĩa vụ đại diện pháp lý thuộc lĩnh vực tư pháp quốc tế và luật thương mại quốc tế đã có tiền lệ áp dụng cho thương mại điện tử xuyên biên giới.

Thứ hai, khả năng khuếch đại quy mô rủi ro và tác động xã hội: một hệ thống AI đơn lẻ tác động đồng thời đến hàng triệu người dùng là hệ quả của quy mô và cấu trúc thị trường nơi AI được triển khai. Cùng một thuật toán thiên lệch sẽ gây hậu quả xã hội hoàn toàn khác nhau tùy thuộc vào việc nó phục vụ 100 người dùng hay 10 triệu người dùng, sự khác biệt nằm ở quy mô kinh doanh của chủ thể triển khai, không nằm ở bản thân thuật toán. Đặc tính này có ý nghĩa trực tiếp cho việc thiết kế cơ chế phân loại rủi ro sử dụng ngưỡng về số lượng người bị ảnh hưởng như một trong các tiêu chí định lượng.

Thứ ba, chuỗi giá trị đa chủ thể phức tạp: Sự phân tách vai trò giữa nhà phát triển mô hình nền tảng, nhà cung cấp hệ thống ứng dụng, bên triển khai tích hợp vào sản phẩm cụ thể và người dùng cuối là hệ quả của cấu trúc thị trường công nghệ hiện đại, nơi việc phát triển một mô hình nền tảng đòi hỏi vốn đầu tư và năng lực kỹ thuật lớn đến mức tạo ra động lực kinh tế cho việc phân công chuyên môn hóa theo chuỗi giá trị (một số ít doanh nghiệp phát triển mô hình nền tảng, số đông doanh nghiệp khác xây dựng ứng dụng trên nền tảng đó). Đặc điểm này có ý nghĩa cho giải pháp đòi hỏi công cụ pháp lý về phân định trách nhiệm theo chuỗi giá trị dựa trên nguyên lý kiểm soát và lợi ích.

c. Phân loại trí tuệ nhân tạo

Phân loại AI có thể dựa trên nhiều tiêu chí khác nhau, luận án này tập trung vào các cách thường được sử dụng trong cả lĩnh vực công nghệ và pháp lý.

Một là, phân loại theo khả năng là cách phân loại phổ biến nhất, dựa trên mức độ "thông minh" mà hệ thống AI đạt được so với trí tuệ con người. Theo đó, có ba dạng AI là AI yếu/hẹp (weak/narrow AI), AI mạnh/tổng quát (strong/general AI) và siêu AI (super AI) [227].

Hai là, phân loại theo chức năng công nghệ dựa trên công nghệ cơ bản hoặc cách thức hoạt động của hệ thống AI, gồm hai loại: (1) AI phân biệt (Discriminative AI) sử dụng dữ liệu để phân loại, dự đoán hoặc đưa ra quyết định dựa trên các mẫu dữ liệu đã học. (2) AI tạo sinh (Generative AI) tạo ra nội dung mới (văn bản, hình ảnh, âm thanh) từ dữ liệu đã học, thường thông qua các mô hình lớn.

Ba là, phân loại dựa trên mức độ rủi ro (thấp/cao). Đây là cách phân loại rất quan trọng đối với khung pháp lý, vì rủi ro thường phát sinh từ mục đích sử dụng chứ không phải bản thân công nghệ. Phân loại AI theo hướng này có ảnh hưởng cốt lõi và quyết định đến cách xây dựng khung pháp lý, giúp nhà làm luật cân bằng giữa việc khuyến khích phát triển công nghệ (ít can thiệp vào AI rủi ro thấp) và bảo vệ xã hội (kiểm soát chặt chẽ AI rủi ro cao). Phân loại rõ ràng về các loại hình AI cũng giúp đảm bảo nguyên tắc trung lập công nghệ trong luật. Luật chỉ nên điều chỉnh mục đích sử dụng và rủi ro chứ không nên điều chỉnh công nghệ cụ thể, tránh việc luật bị lỗi thời ngay khi công nghệ AI tiến hóa.

Bốn là, phân loại dựa trên mức độ tự chủ (khả năng của hệ thống AI tự khởi tạo, duy trì và hoàn thành chuỗi hành động mà không cần sự phê duyệt liên tục của con người) tương tự như thang phân loại SAE đối với xe tự lái [328]: (1) AI phản ứng thụ động vận hành theo quy tắc được lập trình sẵn, không có khả năng học hỏi hay thích nghi. (2) AI học máy có giám sát học từ dữ liệu có nhãn và cải thiện hiệu suất theo thời gian, nhưng chỉ hoạt động trong phạm vi nhiệm vụ được huấn luyện và cần con người kích hoạt từng lần sử dụng. (3) AI tác nhân tự chủ có mục tiêu không chỉ phản hồi lệnh mà chủ động hoạch định chuỗi bước để đạt mục tiêu, tự gọi các công cụ và dịch vụ bên ngoài, duy trì trạng thái qua nhiều phiên làm việc và có thể khởi tạo các tác nhân con để thực hiện nhiệm vụ con. (4) Hệ thống đa tác nhân là sự kết hợp của nhiều tác nhân AI hoạt động phối hợp, trong đó một hoặc nhiều tác nhân điều phối chỉ đạo các tác nhân thực thi để hoàn thành mục tiêu phức tạp.

d. Bản chất pháp lý của trí tuệ nhân tạo

AI là gì trong hệ thống phạm trù lý luận nhà nước và pháp luật là câu hỏi có tính quyết định đối với toàn bộ kiến trúc điều chỉnh. Câu trả lời cho câu hỏi này xác định ai phải chịu trách nhiệm khi AI gây thiệt hại, AI có thể tham gia vào giao dịch pháp lý như thế nào và khung pháp lý cần được thiết kế theo logic nào. Hiện có 3 quan điểm chính, mỗi quan điểm có nền tảng lý luận và thực tiễn riêng.

Quan điểm thứ nhất cho rằng AI chỉ là một công cụ kỹ thuật phức tạp hơn máy móc công nghiệp, phần mềm máy tính do con người tạo ra, kiểm soát và chịu trách nhiệm. AI không có ý chí, lợi ích hay nhân cách pháp lý; mọi hậu quả pháp lý từ hành vi của AI đều quy về con người sử dụng, sở hữu hoặc phát triển AI đó. Kết quả phỏng vấn cho thấy đa số chuyên gia đồng tình với quan điểm này [Xem Phụ lục 5].

Quan điểm thứ hai coi AI là chủ thể pháp lý. Theo đó, khi AI đạt đến mức độ tự chủ nhất định nên được trao một dạng tư cách pháp lý, có thể là tư cách "pháp nhân điện tử" hay "năng lực pháp lý một phần" để có cơ sở pháp lý quy trách nhiệm trực tiếp cho AI thay vì cho con người. Nghị viện châu Âu năm 2017 đề xuất khái niệm "electronic persons" cho robot tiên tiến [347]. Học giả Đức đề xuất khái niệm "Teilrechtsfähigkeit" - năng lực pháp lý bộ phận [237]. Ả Rập Xê Út trao quyền công dân danh dự cho robot Sophia năm 2017 [369] (dù được coi là hành động biểu tượng hơn là thực chất pháp lý). Ngay sau đó, quận Shibuya (Tokyo, Nhật Bản) cấp giấy chứng nhận cư trú đặc biệt cho chatbot Shibuya Mirai [307]. Đề xuất này đã bị cộng đồng pháp lý quốc tế bác bỏ vì nhiều lý do: tạo ra cơ hội trốn tránh trách nhiệm; AI không có tài sản thực sự để bồi thường thiệt hại và về mặt triết học, AI không có ý thức, cảm xúc hay lợi ích để bảo vệ.

Quan điểm thứ ba xác định AI là đối tượng điều chỉnh đặc thù, là một loại thực thể mới đòi hỏi cách tiếp cận điều chỉnh riêng biệt. AI không được trao tư cách pháp lý nhưng hành vi của AI và hậu quả của nó cần được điều chỉnh bởi hệ thống quy phạm đặc thù, không thể hoàn toàn áp dụng các quy phạm dành cho công cụ thông thường hay con người. Đây là cách tiếp cận mà EU AI Act, Luật AI Hàn Quốc thực chất đang theo, Việt Nam xây dựng khung pháp lý theo hướng này sẽ bảo đảm tính tương thích quốc tế, là điều kiện cần thiết để doanh nghiệp tiếp cận thị trường toàn cầu và để Việt Nam hội nhập vào hệ thống quản trị AI quốc tế.

2.1.1.3. Khái niệm khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Theo Từ điển Tiếng Việt, động từ "phát triển" là biến đổi hoặc làm cho biến đổi từ ít đến nhiều, hẹp đến rộng, thấp đến cao, đơn giản đến phức tạp [131, tr.769], còn "ứng dụng" là đem lý thuyết dùng vào thực tiễn [131, tr.1090].

Từ nền tảng hai khái niệm chung nêu trên và khái niệm AI đã nêu ở mục trên, khái niệm phát triển AI có thể hiểu là *quá trình nghiên cứu, thiết kế, xây dựng và huấn luyện các mô hình, thuật toán và hệ thống máy tính với mục tiêu là tạo ra các công nghệ có khả năng thực hiện các nhiệm vụ thông minh như học hỏi, suy luận, nhận dạng, hiểu ngôn ngữ và ra quyết định*, còn khái niệm ứng dụng AI có thể hiểu là việc *triển khai và tích hợp các mô hình, hệ thống AI đã phát triển vào các sản phẩm, dịch vụ hoặc quy trình nghiệp vụ trong thực tế*. Trong vòng đời tiêu chuẩn của một hệ

thống AI, hoạt động phát triển và ứng dụng AI thường bao gồm các giai đoạn sau: (1) nghiên cứu cơ bản: nghiên cứu các lý thuyết mới, mô hình toán học và thuật toán AI (như học máy, học sâu, xử lý ngôn ngữ tự nhiên); (2) thu thập và xử lý dữ liệu: thu thập, làm sạch, gắn nhãn và chuẩn bị các tập dữ liệu lớn; (3) xây dựng và huấn luyện mô hình: lựa chọn kiến trúc mô hình, lập trình thuật toán và sử dụng dữ liệu đã xử lý để huấn luyện mô hình, giúp nó học cách đưa ra dự đoán hoặc quyết định; (4) kiểm thử và đánh giá: kiểm tra hiệu suất, độ chính xác, tính ổn định và tính công bằng của mô hình AI trong các môi trường khác nhau trước khi triển khai; (5) tích hợp và triển khai: tích hợp mô hình AI vào các ứng dụng, phần mềm hoặc hệ thống vật lý (robot, thiết bị IoT) có sẵn để đưa vào sử dụng thực tế; (6) quản lý và bảo trì: giám sát hiệu suất của mô hình AI sau khi triển khai (theo dõi lỗi, độ trôi của dữ liệu), cập nhật và huấn luyện lại mô hình để duy trì độ chính xác theo thời gian; (7) phát triển ứng dụng cuối: phát triển các giao diện người dùng (chatbot, ứng dụng di động) để người dùng cuối có thể tương tác với công nghệ AI; (8) chấm dứt vận hành: một hệ thống AI có thể bị chấm dứt vận hành khi lỗi thời, hết vòng đời, không còn hiệu quả kinh tế hoặc tiềm ẩn rủi ro về an toàn, đạo đức.

Trên cơ sở đó, *khung pháp lý cho phát triển, ứng dụng AI* có thể được hiểu là *hệ thống có cấu trúc, gồm các công cụ điều chỉnh và các thiết chế bảo đảm thực hiện được tổ chức theo những nguyên tắc nhất định, có mối liên hệ và tác động qua lại nhằm điều chỉnh hoạt động của các chủ thể trong toàn bộ vòng đời của hệ thống AI, qua đó vừa thúc đẩy đổi mới sáng tạo và phát triển kinh tế - xã hội, vừa bảo vệ quyền con người, kiểm soát rủi ro và bảo đảm phát triển, ứng dụng AI có trách nhiệm.*

Ba mục tiêu nêu trong định nghĩa trên (bảo vệ quyền con người, kiểm soát rủi ro và thúc đẩy đổi mới sáng tạo) là ba giá trị cốt lõi cần được xử lý đồng thời trong thiết kế khung pháp lý về AI. Trên thực tế, khó có khung pháp lý nào có thể đạt được tuyệt đối cả ba mục tiêu này trong mọi hoàn cảnh. Do đó, nhiệm vụ của nhà thiết kế chính sách và pháp luật không phải là tối đa hóa một mục tiêu đơn lẻ, mà là xác lập điểm cân bằng hợp lý giữa các giá trị này, phù hợp với điều kiện chính trị, pháp lý, kinh tế, xã hội và năng lực thực thi của từng quốc gia.

Cũng cần nhấn mạnh rằng "khung pháp lý" không đồng nghĩa với một đạo luật hay bộ luật riêng về AI. Một quốc gia có thể xây dựng được khung pháp lý tương đối hiệu quả dù chưa ban hành đạo luật AI chuyên biệt, nếu các quy định pháp luật hiện hành, tiêu chuẩn kỹ thuật, cơ chế quản trị, thiết chế thực thi và công cụ hướng dẫn được tổ chức hợp lý. Ngược lại, việc ban hành một đạo luật AI không tự thân bảo đảm sự hình thành của một khung pháp lý hoàn chỉnh, nếu thiếu các thành tố hỗ trợ như thiết chế thực thi, tiêu chuẩn kỹ thuật, cơ chế giám sát, công cụ đánh giá rủi ro và cơ chế cập nhật chính sách.

Khung pháp lý cho phát triển, ứng dụng AI là một bộ phận cụ thể của khái niệm rộng hơn là khung pháp lý về AI. Sự khác biệt giữa hai khái niệm này chủ yếu nằm ở phạm vi và trọng tâm điều chỉnh. Khung pháp lý về AI có nội hàm rộng, bao quát toàn bộ các quy định, chuẩn mực, thiết chế và cơ chế liên quan đến AI với tư cách là một công nghệ mới cần được quản lý hiệu quả. Trong khi đó, khung pháp lý cho phát triển, ứng dụng AI có phạm vi cụ thể hơn, nhấn mạnh vào các hoạt động của chủ thể trong toàn bộ vòng đời của công nghệ AI, từ nghiên cứu, thiết kế, huấn luyện, kiểm thử, cung cấp, triển khai đến sử dụng, giám sát, cập nhật và chấm dứt vận hành. Cách diễn đạt này phù hợp với định hướng của luận án, bởi trọng tâm nghiên cứu không chỉ là quản lý AI nói chung, mà là xây dựng một khung pháp lý vừa thúc đẩy quá trình phát triển, ứng dụng AI, vừa kiểm soát các rủi ro phát sinh trong quá trình đó.

2.1.1.4. Khái niệm, đặc điểm của việc xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

a. Khái niệm xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Để làm rõ khái niệm xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo, trước hết cần xác định nội hàm của thuật ngữ "xây dựng". Trong tiếng Việt, "xây dựng" được hiểu theo ba nghĩa cơ bản: (1) làm nên công trình kiến trúc theo một kế hoạch nhất định; (2) làm cho hình thành một tổ chức hay một chỉnh thể về xã hội, chính trị, kinh tế, văn hoá theo một phương hướng nhất định; và (3) tạo lập, sáng tạo ra cái có giá trị tinh thần, có ý nghĩa trừu tượng [131]. Với nghĩa này, "xây dựng" không chỉ là hành động tạo ra một yếu tố đơn lẻ, mà là quá trình có mục tiêu, có kế hoạch, có cấu trúc và có phương hướng nhằm hình thành một chỉnh thể nhất định.

Khi vận dụng vào lĩnh vực pháp luật, "xây dựng" gợi mở hình ảnh của một công trình pháp lý - thể chế, trong đó cần có thiết kế trước khi thi công, lựa chọn vật liệu và kỹ thuật phù hợp, tổ chức thực hiện theo quy trình, kiểm tra, nghiệm thu và tiếp tục bảo trì sau khi hoàn thành. Do đó, xây dựng khung pháp lý không thể được hiểu giản đơn là ban hành một văn bản pháp luật, mà bao gồm toàn bộ quá trình từ nhận diện nhu cầu điều chỉnh, thiết kế kiến trúc, xây dựng quy phạm, thiết lập thiết chế, tổ chức thực thi, đánh giá tác động đến cập nhật, điều chỉnh khi điều kiện kỹ thuật và xã hội thay đổi. Đây là một quá trình liên tục, không kết thúc tại thời điểm văn bản pháp luật được ban hành.

Từ cách hiểu trên, có thể phân biệt "xây dựng" với một số thuật ngữ gần nghĩa thường được sử dụng trong nghiên cứu pháp luật [131]. "Thiết kế" là "lập tài liệu kĩ thuật toàn bộ, gồm có bản tính toán, bản vẽ, v.v., để có thể theo đó mà xây dựng công trình, sản xuất thiết bị, sản phẩm, v.v.", nhấn mạnh đến hoạt động tư duy kiến trúc, xác định mô hình tổng thể, logic nội tại và quan hệ giữa các thành phần trước khi triển khai. Vì vậy, thiết kế là một bộ phận quan trọng của xây dựng, nhưng chưa bao

quát toàn bộ quá trình xây dựng. Một khung pháp lý thiếu giai đoạn thiết kế có chủ đích dễ trở thành tập hợp rời rạc của các văn bản, thiếu logic nội tại và khó vận hành hiệu quả. "Tạo lập" là "tạo ra, gây dựng nên", nhấn mạnh đến việc hình thành một yếu tố hoặc một hệ thống từ chỗ chưa có. Thuật ngữ này phù hợp với giai đoạn khởi đầu của quá trình xây dựng, nhất là khi chưa tồn tại hệ thống điều chỉnh tương ứng. "Thiết lập" gần nghĩa với "tạo lập", nhưng thường nhấn mạnh hơn đến việc đặt nền móng cho các nguyên tắc cơ bản, cơ quan quản lý hoặc cơ chế nền tảng để từ đó các yếu tố khác có thể tiếp tục phát triển. Trong khi đó, "hoàn thiện" hàm chứa ý nghĩa đã có một hệ thống hiện hữu và cần được bổ sung, sửa đổi, nâng cao chất lượng để vận hành tốt hơn. "Hoàn thiện" khác với "xây dựng" ở chỗ nó xuất phát từ một nền tảng đã có, còn "xây dựng" có thể bao gồm cả tạo lập mới, thiết kế lại và hoàn thiện hệ thống hiện hữu.

Như vậy, so với các thuật ngữ "thiết kế", "tạo lập", "thiết lập" và "hoàn thiện", "xây dựng" là khái niệm có nội hàm rộng hơn. Nó bao hàm cả hoạt động tư duy kiến trúc, hoạt động tạo lập các thành phần, hoạt động tổ chức quan hệ giữa các thành phần, hoạt động thiết lập năng lực thực thi và hoạt động điều chỉnh, cập nhật sau khi hệ thống đã vận hành. Cách hiểu này đặc biệt phù hợp với lĩnh vực AI, bởi khung pháp lý cho phát triển, ứng dụng AI không thể là sản phẩm tĩnh, mà phải là một cấu trúc điều chỉnh có khả năng thích ứng với sự thay đổi nhanh của công nghệ, thị trường và xã hội.

Trên cơ sở những phân tích trên, có thể hiểu: *xây dựng khung pháp lý cho phát triển, ứng dụng AI là quá trình có mục đích, đa chủ thể và liên tục, bao gồm việc thiết kế kiến trúc tổng thể, tạo lập các thành phần cấu thành, tổ chức quan hệ giữa các thành phần, thiết lập năng lực thực thi và hình thành cơ chế thích ứng, nhằm tạo ra một hệ thống điều chỉnh đầy đủ, nhất quán, hiệu quả và có khả năng cập nhật trước sự thay đổi của điều kiện kỹ thuật, kinh tế, xã hội và pháp lý.*

b. Đặc điểm của việc xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Xây dựng khung pháp lý cho phát triển, ứng dụng AI là hoạt động có tính đặc thù so với xây dựng khung pháp lý trong nhiều lĩnh vực khác của đời sống xã hội. Những đặc điểm này chi phối trực tiếp phương pháp tiếp cận, quy trình thực hiện và chất lượng của sản phẩm pháp lý được tạo lập. Từ cơ sở lý luận về khái niệm của khung pháp lý cho phát triển, ứng dụng AI đã phân tích ở trên, có thể xác định một số đặc điểm của việc xây dựng khung pháp lý cho phát triển, ứng dụng AI như sau.

Thứ nhất, tính đồng kiến tạo với quá trình phát triển công nghệ. Hoạt động xây dựng khung pháp lý AI không diễn ra sau khi đối tượng công nghệ đã định hình ổn định như trường hợp lập pháp điều chỉnh đối tượng truyền thống (một loại phương

tiện giao thông, một loại hóa chất công nghiệp) mà diễn ra đồng thời và tương tác qua lại với chính quá trình công nghệ AI đang được phát triển và định hình. Đây là hệ quả tất yếu, có tính khách quan từ lý thuyết đồng kiến tạo của Jasanoff đã phân tích tại Chương 1 khi công nghệ và trật tự xã hội bao gồm trật tự pháp lý được tạo lập cùng nhau. Biểu hiện cụ thể của đặc điểm này là các quyết định thiết kế kiến trúc pháp lý (như lựa chọn định nghĩa chức năng thay vì liệt kê công nghệ hay lựa chọn ngưỡng phân loại rủi ro) tác động ngược trở lại định hướng đầu tư nghiên cứu và phát triển của doanh nghiệp công nghệ, trong khi hướng phát triển công nghệ mới (như sự trỗi dậy của tác nhân AI) lại tạo áp lực buộc pháp luật phải tái thiết kế các quyết định kiến trúc đã có. Đây là mối quan hệ hai chiều liên tục, không phải quan hệ một chiều "công nghệ có trước, pháp luật theo sau" như mô hình lập pháp truyền thống. Đối với AI, công nghệ thay đổi trong chu kỳ rất ngắn; năng lực và rủi ro của AI ở thời điểm hiện tại có thể khác căn bản so với giai đoạn trước và tiếp tục thay đổi trong tương lai gần. Do đó, xây dựng khung pháp lý cho phát triển, ứng dụng AI không thể được xem là một dự án có thời hạn, mà là một quá trình vận động liên tục. Quá trình này đòi hỏi cơ chế cập nhật, điều chỉnh và học hỏi phải được tích hợp ngay từ đầu vào thiết kế của khung pháp lý.

Thứ hai, tính đa chủ thể và phân tán quyền lực điều chỉnh. Khác với nhiều lĩnh vực pháp luật truyền thống, nơi hoạt động xây dựng pháp luật chủ yếu do cơ quan nhà nước thực hiện, xây dựng khung pháp lý cho phát triển, ứng dụng AI là quá trình có sự tham gia của nhiều chủ thể. Nhà nước, thông qua hoạt động lập pháp, hành pháp và tư pháp, tạo lập tầng quy phạm cứng; khu vực tư nhân, gồm doanh nghiệp và hiệp hội ngành, tham gia hình thành tiêu chuẩn kỹ thuật và cơ chế tự điều chỉnh; cộng đồng quốc tế góp phần kiến tạo chuẩn mực chung; còn người sử dụng và xã hội dân sự tham gia định hình giá trị, phản biện và giám sát độc lập. Trong Khuyến nghị về đạo đức của AI năm 2021, UNESCO xác định ba cấp độ của các bên liên quan khi xem xét các hệ thống AI có trách nhiệm: cá nhân, bao gồm người dùng, nhà phát triển và nhà nghiên cứu; tổ chức thuộc khu vực tư nhân, bao gồm doanh nghiệp, truyền thông; và các bên thứ ba trong và ngoài nước tham gia xây dựng luật, quy tắc và quy định, như Chính phủ, tổ chức liên chính phủ, tổ chức nhân quyền, cộng đồng kỹ thuật, nhà hoạch định chính sách, chuyên gia pháp lý… [356]. Bên cạnh đó, các cơ sở nghiên cứu, đào tạo có vai trò phát triển lý thuyết, thực hiện nghiên cứu cơ bản và đào tạo chuyên gia phục vụ cho quá trình xây dựng khung pháp lý. Ngay trong nội bộ Nhà nước, cũng khó có một cơ quan đơn lẻ nào có đủ thẩm quyền và năng lực để một mình xây dựng toàn bộ khung pháp lý cho phát triển, ứng dụng AI. Với tính chất liên ngành, AI đòi hỏi sự phối hợp giữa cơ quan quản lý công nghệ, cơ quan quản lý ngành, cơ quan lập pháp, cơ quan tư pháp và các thiết chế giám sát. Vì vậy, tính đa chủ thể có thể trở thành

nguồn lực nếu được tổ chức hợp lý, nhưng cũng có thể dẫn đến phân mảnh, chồng chéo và thiếu hiệu quả nếu thiếu cơ chế điều phối đủ mạnh.

Thứ ba, tính phức hợp kỹ thuật - pháp lý cao. Vì đối tượng điều chỉnh (hệ thống AI) có bản chất kỹ thuật phức tạp đồng thời với việc điều chỉnh đối tượng đó tạo ra hệ quả pháp lý đối với quyền và nghĩa vụ của chủ thể có liên quan, hoạt động xây dựng khung pháp lý AI về mặt khách quan đòi hỏi đồng thời hai loại tri thức khác biệt về bản chất là tri thức pháp lý (về nguyên tắc lập pháp, hệ thống nguồn pháp luật, giải thích và áp dụng quy phạm) và tri thức kỹ thuật (về cấu trúc, nguyên lý vận hành và giới hạn năng lực của công nghệ AI) mà một trong hai loại tri thức đó không thể tự thân thay thế cho loại tri thức còn lại. Cần kết hợp với nhận thức kinh tế về thị trường AI cùng tri thức về đạo đức và triết học công nghệ nói chung. Hiếm có chuyên gia đơn lẻ nào có thể bao quát đầy đủ tất cả các lĩnh vực này. Vì vậy, quá trình xây dựng khung pháp lý AI tất yếu phải dựa trên sự cộng tác thực chất giữa các chuyên gia pháp lý, kỹ thuật, kinh tế, đạo đức, quản trị và chính sách công.

Tính phức hợp này cũng dẫn đến một hệ quả quan trọng, tiêu chuẩn kỹ thuật không chỉ là tài liệu hỗ trợ, mà là thành phần không thể thiếu của khung pháp lý. Một quy định yêu cầu "hệ thống AI rủi ro cao phải được kiểm định trước khi triển khai" chỉ có thể vận hành nếu đi kèm các tiêu chuẩn kỹ thuật xác định phương pháp kiểm định, chỉ số đo lường và ngưỡng đánh giá. Nếu thiếu tiêu chuẩn kỹ thuật, quy phạm pháp luật khó có thể được thực thi hiệu quả. Đây là điểm khác biệt quan trọng so với nhiều lĩnh vực pháp luật truyền thống. Nhà nước cũng khó xây dựng được khung pháp lý AI có chất lượng nếu không có sự chia sẻ thông tin thực chất từ phía doanh nghiệp và cộng đồng kỹ thuật. Vì vậy, tham vấn kỹ thuật chuyên sâu không thể chỉ là thủ tục hình thức, mà phải trở thành điều kiện nội dung để tạo ra quy phạm khả thi. Mô hình Nhà nước soạn thảo, doanh nghiệp góp ý hình thức cần được thay thế bằng mô hình đồng thiết kế thực chất hơn.

Thứ tư, tính đa tầng của cấu trúc hoạt động xây dựng. Do chính khung pháp lý AI có cấu trúc đa tầng với nhiều loại nguồn khác nhau, hoạt động xây dựng khung pháp lý đó cũng phải là một hoạt động đa tầng tương ứng, không chỉ là một hành vi lập pháp đơn lẻ tạo ra một văn bản duy nhất. Đây là đặc điểm khách quan mang tính hệ quả logic, nếu đối tượng cần điều chỉnh có nhiều loại nguồn khác nhau với chức năng phân biệt, thì hoạt động tạo lập các loại nguồn đó tất yếu bao gồm nhiều loại hoạt động con tương ứng có tính chất khác nhau. Soạn thảo văn bản quy phạm pháp luật khác về bản chất công việc với xây dựng tiêu chuẩn kỹ thuật, khác với phát triển án lệ qua hoạt động xét xử, khác với xây dựng bộ quy tắc ứng xử qua cơ chế đồng thiết kế với ngành.

Thứ năm, tính không hoàn hảo và không chắc chắn. Trong xây dựng pháp luật truyền thống, mục tiêu lý tưởng thường là dự liệu đầy đủ các tình huống, xây dựng quy

phạm rõ ràng, ổn định trước khi ban hành. Đối với AI, yêu cầu này không còn hoàn toàn phù hợp, bởi công nghệ thay đổi nhanh, rủi ro mới liên tục phát sinh và nhiều hệ quả xã hội chỉ bộc lộ trong quá trình vận hành. Tại thời điểm tiến hành xây dựng hoặc sửa đổi khung pháp lý AI, chủ thể xây dựng thường phải đưa ra quyết định trong điều kiện tri thức về rủi ro, tác động và hệ quả dài hạn của công nghệ AI là không đầy đủ và không thể đầy đủ. Lý do không phải vì thiếu nỗ lực nghiên cứu mà vì tri thức về thế hệ công nghệ hiện tại chưa kịp tích lũy đầy đủ thì thế hệ công nghệ tiếp theo đã xuất hiện. Vì vậy, xây dựng khung pháp lý cho phát triển, ứng dụng AI đòi hỏi sự thay đổi trong tư duy lập pháp: thay vì chờ đến khi có thể giải quyết trọn vẹn mọi vấn đề, cần thiết kế một khung pháp lý đủ rõ để vận hành, nhưng đủ mở để học hỏi, thử nghiệm, điều chỉnh và cập nhật. Một khung pháp lý chưa hoàn hảo nhưng được thực thi nghiêm túc, giám sát thường xuyên và cập nhật kịp thời có thể có giá trị thực tiễn cao hơn một khung pháp lý hoàn chỉnh trên giấy nhưng không theo kịp sự biến đổi của công nghệ.

2.1.1.5. Vai trò của việc xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Xây dựng khung pháp lý cho phát triển, ứng dụng AI không phải là một yêu cầu hành chính hay thủ tục pháp lý đơn thuần, mà là hoạt động có ý nghĩa chiến lược đối với phát triển quốc gia, quản trị rủi ro công nghệ và bảo vệ các giá trị xã hội nền tảng. Từ cơ sở lý luận về khái niệm, đặc điểm của khung pháp lý và việc xây dựng khung pháp lý cho phát triển, ứng dụng AI đã phân tích ở trên, có thể xác định vai trò của việc xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo như sau:

Thứ nhất, vai trò định hình và cụ thể hóa chính sách công về AI: Hoạt động xây dựng khung pháp lý AI đóng vai trò là khâu chuyển hóa giữa chủ trương chính sách ở tầng định hướng chính trị (các nghị quyết, chiến lược, chương trình hành động của Đảng và Nhà nước về AI) với hệ thống quy phạm có khả năng vận hành và tạo ra hiệu lực ràng buộc trong đời sống kinh tế - xã hội.

Vai trò định hình chính sách của hoạt động xây dựng khung pháp lý thể hiện ở ba phương diện cụ thể: Một là, xác định phạm vi và giới hạn của chính sách: hoạt động xây dựng khung pháp lý chuyển hóa các tuyên bố chính sách có tính khái quát thành phạm vi điều chỉnh cụ thể, thực sự tác động đến hành vi của chủ thể trong xã hội. Hai là, lựa chọn và cân bằng giữa các định hướng chính sách cạnh tranh: mọi quyết định thiết kế cụ thể trong xây dựng khung pháp lý AI đều đồng thời tác động đến cả định hướng thúc đẩy đổi mới lẫn kiểm soát rủi ro. Hoạt động xây dựng khung pháp lý, thông qua từng lựa chọn kỹ thuật cụ thể chính là nơi diễn ra hành vi lựa chọn chính sách thực chất giữa các định hướng cạnh tranh. Ba là, tạo phản hồi định hình lại chính sách: quá trình xây dựng khung pháp lý, đặc biệt ở giai đoạn chẩn đoán bằng chứng và giai đoạn

vận hành, thích ứng tạo ra tri thức và bằng chứng thực tiễn có khả năng phản hồi ngược, điều chỉnh lại chính các định hướng chính sách ở tầng cao hơn.

Thứ hai, vai trò hoàn thiện thể chế quản lý nhà nước về AI: Hoạt động xây dựng khung pháp lý AI đóng vai trò là động lực trực tiếp thúc đẩy quá trình hoàn thiện thể chế quản lý nhà nước đối với lĩnh vực AI theo nghĩa mỗi bước tiến trong xây dựng khung pháp lý đều đặt ra yêu cầu tương ứng về sự tồn tại, thẩm quyền và năng lực của bộ máy nhà nước để vận hành khung pháp lý đó, và ngược lại, việc thể chế được hoàn thiện tạo điều kiện để khung pháp lý tiếp tục được xây dựng ở mức độ phức tạp và tinh vi hơn.

Vai trò hoàn thiện thể chế thể hiện ở ba cấp độ cụ thể, tương ứng với ba cấp độ của bộ máy nhà nước chịu tác động: Một là, thúc đẩy phân định lại thẩm quyền quản lý nhà nước: việc xây dựng khung pháp lý AI, do đối tượng điều chỉnh có tính xuyên ngành, buộc thể chế nhà nước phải xem xét lại và phân định rõ ràng hơn ranh giới thẩm quyền giữa các Bộ, ngành vốn trước đây có thể chưa từng phải va chạm nhau trong cùng một lĩnh vực quản lý cụ thể. Hai là, thúc đẩy hình thành thiết chế chuyên trách mới: khi mức độ phức tạp của khung pháp lý AI vượt quá năng lực xử lý của cơ cấu hành chính hiện có theo phương thức thông thường, hoạt động xây dựng khung pháp lý tạo ra áp lực và luận cứ để hình thành thiết chế chuyên trách mới. Vai trò này có tính biện chứng: bản thân yêu cầu về thiết chế mới xuất phát từ quá trình xây dựng khung pháp lý, nhưng một khi thiết chế được hình thành, nó lại tiếp tục thúc đẩy hoàn thiện khung pháp lý ở giai đoạn tiếp theo tạo thành một vòng lặp củng cố lẫn nhau giữa hoàn thiện quy phạm và hoàn thiện thiết chế. Ba là, thúc đẩy nâng cao năng lực và chất lượng đội ngũ cán bộ: quá trình xây dựng khung pháp lý AI với những khó khăn cụ thể được bộc lộ qua thực tiễn trở thành động lực trực tiếp thúc đẩy cải cách công tác đào tạo, tuyển dụng và bố trí cán bộ trong hệ thống hành chính nhà nước, vượt ra ngoài phạm vi riêng của lĩnh vực AI để tác động đến toàn bộ tư duy về xây dựng đội ngũ cán bộ quản lý công nghệ nói chung.

Thứ ba, vai trò huy động và tổ chức sự tham gia của các chủ thể liên quan: Hoạt động xây dựng khung pháp lý AI đóng vai trò là cơ chế tổ chức cho phép các chủ thể có lợi ích và tri thức khác nhau trong xã hội được huy động và tích hợp tiếng nói của mình vào quá trình hình thành chuẩn mực chung. Điểm cốt lõi của vai trò huy động tham gia là hoạt động xây dựng khung pháp lý chuyển hóa sự tham gia từ hình thức phân tán, tự phát và không có tính đại diện sang hình thức có cấu trúc, có quy trình và có trách nhiệm giải trình.

Vai trò huy động tham gia thể hiện ở ba phương diện: Một là, vai trò tổ chức đối thoại giữa nhà nước và khu vực kỹ thuật: hoạt động xây dựng khung pháp lý AI, thông qua các cơ chế tham vấn kỹ thuật chuyên sâu đóng vai trò là kênh chính thức và có cấu

trúc để tri thức kỹ thuật từ cộng đồng chuyên môn được đưa vào quá trình ra quyết định pháp lý. Hai là, vai trò tổ chức phản ánh lợi ích kinh tế đa dạng: hoạt động xây dựng khung pháp lý AI đóng vai trò tổ chức để tiếng nói của các nhóm doanh nghiệp có quy mô và vị thế thị trường khác nhau từ tập đoàn công nghệ lớn đến startup được phản ánh vào thiết kế chính sách theo tỷ trọng phù hợp, tránh tình trạng lợi ích của một nhóm chi phối hoàn toàn quá trình hình thành chuẩn mực chung. Ba là, vai trò tổ chức tiếng nói của nhóm chịu tác động xã hội: hoạt động xây dựng khung pháp lý AI đóng vai trò là kênh để lợi ích và mối quan ngại của người dùng cuối, đặc biệt là nhóm dễ bị tổn thương, được đưa vào quá trình thiết kế chính sách vì nhóm này thường thiếu năng lực tổ chức và nguồn lực để tự vận động chính sách như các nhóm doanh nghiệp có tổ chức.

Thứ tư, vai trò tạo lập nền tảng cho hoạt động quản trị AI: Hoạt động xây dựng khung pháp lý AI đóng vai trò là điều kiện tiên quyết đối với toàn bộ hoạt động quản trị AI nói chung. Vai trò này khác về bản chất với ba vai trò đã trình bày ở trên vì nó mô tả một quan hệ điều kiện - kết quả tương đối một chiều: hoạt động quản trị AI không thể vận hành có căn cứ chính đáng nếu thiếu nền tảng khung pháp lý được xây dựng trước hoặc đồng thời.

Vai trò tạo nền tảng thể hiện ở ba phương diện cụ thể: Một là, tạo nền tảng thẩm quyền cho công cụ quản trị vận hành: mọi công cụ tác nghiệp thuộc khung quản trị (như công cụ tự đánh giá tuân thủ, hệ thống thông tin quản lý AI, quy trình phối hợp nội bộ) chỉ có căn cứ chính đáng để vận hành khi hoạt động xây dựng khung pháp lý đã xác lập thẩm quyền pháp định cho cơ quan triển khai công cụ đó. Hai là, tạo nền tảng chuẩn mực cho hoạt động giám sát: hoạt động giám sát thực thi chỉ có thể vận hành có căn cứ khi có chuẩn mực rõ ràng để đối chiếu (chuẩn mực đó do khung pháp lý xác lập). Không có khung pháp lý xác định rõ nghĩa vụ cụ thể, hoạt động giám sát sẽ thiếu tiêu chí khách quan và có nguy cơ vận hành tùy tiện. Ba là, tạo nền tảng chính đáng cho các cơ chế quản trị mềm: nhiều công cụ quản trị mềm (tiêu chuẩn kỹ thuật, hướng dẫn ngành, bộ quy tắc ứng xử) chỉ đạt được mức độ hiệu lực và tính chính đáng cao nhất khi được "neo" vào nền tảng pháp lý thông qua cơ chế pháp luật hóa. Hoạt động xây dựng khung pháp lý đóng vai trò tạo nền tảng để các công cụ quản trị mềm chuyển hóa từ chỗ chỉ có hiệu lực dựa vào áp lực thị trường hoặc danh tiếng sang chỗ có hiệu lực được củng cố bởi thẩm quyền nhà nước.

Cần nhấn mạnh một giới hạn quan trọng của vai trò này để tránh đề cao quá mức vị trí của hoạt động xây dựng khung pháp lý: vai trò "tạo nền tảng" không có nghĩa là khung pháp lý có thể hoặc nên thay thế toàn bộ hoạt động quản trị AI. Nhiều phương diện quan trọng của quản trị AI hiệu quả như năng lực vận hành cụ thể, chất lượng công cụ kỹ thuật, văn hóa tuân thủ của doanh nghiệp, năng lực giám sát xã hội nằm ngoài phạm vi mà bản thân hoạt động xây dựng khung pháp lý có thể trực tiếp tạo

ra, dù nó tạo điều kiện nền tảng cho những phương diện đó phát triển. Nhận thức đúng giới hạn này là cơ sở lý luận quan trọng lý giải tại sao luận án không chỉ dừng lại ở đề xuất hoàn thiện khung pháp lý mà còn đề xuất song song hệ thống giải pháp về hoạt động xây dựng mang tính quản trị để bổ trợ cho nhau chứ không thể thay thế lẫn nhau.

2.1.2. Nội dung xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Hoạt động xây dựng khung pháp lý cho phát triển, ứng dụng AI được phân tích theo bốn chiều: chủ thể, nội dung, quy trình và kỹ thuật, tương ứng với bốn câu hỏi về ai thực hiện, tạo ra cái gì, theo trình tự nào và bằng phương tiện nào. Cách phân chia này kế thừa lý luận về xây dựng pháp luật, phù hợp với cấu trúc của Luật Ban hành văn bản quy phạm pháp luật.

2.1.2.1. Chủ thể xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Xây dựng khung pháp lý cho phát triển, ứng dụng AI là quá trình đa chủ thể. Do AI là công nghệ phức hợp, liên ngành và thay đổi nhanh, không một chủ thể đơn lẻ nào có thể tự mình xây dựng đầy đủ, hiệu quả toàn bộ khung pháp lý. Các chủ thể tham gia quá trình này có vai trò, thẩm quyền và mức độ ảnh hưởng khác nhau, nhưng cần được tổ chức trong một cơ chế phối hợp thống nhất. Luận án lựa chọn tiêu chí phân loại theo tính chất của hệ quả pháp lý mà sự tham gia của họ tạo ra. Tư cách và thẩm quyền do pháp luật quy định là căn cứ để xác định hệ quả đó. Theo tiêu chí này có ba nhóm. Chủ thể có thẩm quyền quyết định là chủ thể mà hành vi, dựa trên thẩm quyền được pháp luật trao, trực tiếp làm phát sinh, thay đổi hoặc chấm dứt hiệu lực của một thành tố. Chủ thể tham gia chính thức là chủ thể được pháp luật quy định một vai trò trong thủ tục, mà sự tham gia buộc chủ thể quyết định phải tiếp nhận, xem xét hoặc giải trình, nhưng tự nó không làm thay đổi hiệu lực của thành tố. Chủ thể có ảnh hưởng là chủ thể tác động đến nội dung hoặc định hướng của khung pháp lý bằng các phương tiện không có tính ràng buộc pháp lý. Cách phân loại này là đề xuất của nghiên cứu sinh, trên cơ sở kế thừa lý luận về chủ thể của hoạt động xây dựng pháp luật, lý thuyết về quyền năng pháp lý của Hohfeld [368] và Hart [221], và lý thuyết về sự tham gia [332], về chủ thể có liên quan trong hoạch định chính sách [313].

Nhóm thứ nhất, chủ thể quyết định xây dựng khung pháp lý là những chủ thể có thẩm quyền do Hiến pháp hoặc luật trao cho trong quy trình tạo lập, thẩm định, thông qua hoặc thực thi các thành tố của khung pháp lý. Nhóm này bao gồm cơ quan lập pháp (có thẩm quyền thông qua văn bản luật), cơ quan hành pháp có thẩm quyền ban hành văn bản dưới luật (nghị định, quyết định, thông tư) và cơ quan tư pháp có thẩm quyền tạo lập án lệ hoặc áp dụng lẽ công bằng trong xét xử. Hành vi

của các chủ thể này trực tiếp làm phát sinh, thay đổi hoặc chấm dứt hiệu lực của một thành tố cụ thể thuộc khung pháp lý. Đặc trưng chung của nhóm này là tính chịu trách nhiệm trực tiếp về pháp lý hoặc chính trị - hành chính về chất lượng và hậu quả của thành tố mà mình trực tiếp tạo lập.

Nhóm thứ hai, chủ thể tham gia chính thức vào quy trình bao gồm cơ quan thẩm định pháp lý, cơ quan thẩm tra của cơ quan lập pháp, cơ quan hoặc tổ chức được pháp luật trao thẩm quyền ban hành tiêu chuẩn kỹ thuật hoặc công nhận bộ quy tắc ứng xử theo cơ chế pháp luật hóa và các cơ quan, tổ chức, cá nhân được lấy ý kiến theo thủ tục luật định. Đặc trưng phân biệt phân nhóm này với nhóm trên là chủ thể tham gia vào quy trình tạo lập với tư cách chính thức và có ảnh hưởng thực chất đến nội dung cuối cùng (thông qua quyền yêu cầu sửa đổi, quyền từ chối thông qua, quyền công nhận hoặc không công nhận), nhưng không phải là chủ thể có quyền quyết định cuối cùng độc lập, quyết định của họ vẫn cần được chuyển tiếp đến hoặc phối hợp với chủ thể thuộc nhóm trên để hoàn tất hiệu lực pháp lý. Chuyên gia được mời tham vấn chính thức không trực tiếp tạo ra hệ quả đối với thành tố mà tạo ra hệ quả đối với quy trình: cơ quan soạn thảo phải tổng hợp và giải trình việc tiếp thu hay không tiếp thu ý kiến đó.

Nhóm thứ ba, chủ thể có ảnh hưởng đến quá trình xây dựng khung pháp lý là những chủ thể tác động đến nội dung, định hướng hoặc chất lượng của khung pháp lý mà thông qua các kênh gián tiếp như cung cấp tri thức chuyên môn, thiết lập chuẩn mực tham chiếu, chia sẻ kinh nghiệm so sánh, hoặc tạo áp lực chính trị - kinh tế - xã hội buộc chủ thể trực tiếp phải cân nhắc khi ra quyết định. Đặc trưng chung của nhóm này là tính chất khuyến nghị hoặc tham chiếu, sản phẩm mà nhóm này tạo ra không tự thân có hiệu lực ràng buộc đối với khung pháp lý quốc gia cho đến khi được chủ thể thuộc Nhóm thứ nhất tiếp nhận và chuyển hóa thông qua hành vi pháp lý cụ thể. Cần nhấn mạnh việc xếp một chủ thể vào Nhóm thứ ba hoàn toàn không hàm ý chủ thể đó có vai trò kém quan trọng hơn Nhóm thứ nhất và thứ hai. Trên thực tế, nhiều chủ thể thuộc Nhóm thứ ba (đặc biệt tổ chức tiêu chuẩn hóa quốc tế) có ảnh hưởng thực chất đến nội dung cuối cùng của khung pháp lý không kém, thậm chí trong một số trường hợp còn lớn hơn, so với một số chủ thể của Nhóm thứ nhất.

Một là, tổ chức quốc tế liên chính phủ và cơ chế hợp tác đa phương. Các tổ chức liên chính phủ có chức năng xây dựng khuyến nghị, nguyên tắc hoặc công ước quốc tế liên quan đến AI. Phương thức ảnh hưởng của phân nhóm này đến khung pháp lý quốc gia diễn ra qua ba kênh: kênh điều ước khi quốc gia ký kết và phê chuẩn văn kiện quốc tế, nội dung đó được chuyển hóa thành nghĩa vụ nội luật hóa (một quá trình mà bản thân hành vi phê chuẩn và nội luật hóa lại thuộc về chủ thể

trong Nhóm thứ nhất, còn tổ chức quốc tế chỉ đóng vai trò soạn thảo và thúc đẩy văn kiện gốc); kênh khuyến nghị mềm khi các nguyên tắc, khuyến nghị không có tính ràng buộc được quốc gia tham chiếu tự nguyện trong quá trình soạn thảo nội luật; và kênh đánh giá so sánh khi tổ chức quốc tế công bố báo cáo đánh giá xếp hạng hoặc so sánh giữa các quốc gia, tạo áp lực uy tín thúc đẩy quốc gia điều chỉnh khung pháp lý của mình.

Hai là, tổ chức tiêu chuẩn hóa kỹ thuật quốc tế và khu vực có vị trí đặc biệt quan trọng trong hệ thống chủ thể ảnh hưởng vì sản phẩm tiêu chuẩn kỹ thuật có khả năng chuyển hóa trực tiếp và với chi phí thấp nhất thành một bộ phận của khung pháp lý quốc gia thông qua cơ chế pháp luật hóa. Khác với khuyến nghị chính sách chung chung, tiêu chuẩn kỹ thuật do các tổ chức này ban hành có thể được viện dẫn gần như nguyên bản (chỉ cần một hành vi pháp lý là quyết định chỉ định) mà không cần trải qua toàn bộ quy trình soạn thảo nội dung kỹ thuật từ đầu.

Ba là, mạng lưới chuyên gia độc lập và cộng đồng học thuật xuyên quốc gia bao gồm các nhóm nghiên cứu học thuật quốc tế, mạng lưới chuyên gia độc lập, và các sáng kiến tập hợp chuyên gia đa quốc gia (như các nhóm cố vấn kỹ thuật cho tổ chức quốc tế). Phương thức ảnh hưởng của phân nhóm này mang tính chất cung cấp khung phân tích, phương pháp luận đánh giá rủi ro, và bằng chứng thực nghiệm về hiệu quả của các mô hình quản lý khác nhau, có ảnh hưởng gián tiếp nhưng sâu rộng đến tư duy thiết kế của chủ thể trực tiếp xây dựng khung pháp lý, thường thông qua kênh công bố học thuật, hội thảo khoa học, hoặc tham gia với tư cách chuyên gia tư vấn không chính thức.

Bốn là, chủ thể thị trường tạo ra chuẩn mực thực tiễn dẫn dắt bao gồm các doanh nghiệp công nghệ có vị thế dẫn dắt thị trường, mà thông qua thực tiễn kinh doanh và các công cụ tự nguyện của mình (như bộ nguyên tắc đạo đức AI nội bộ, cam kết an toàn tự nguyện), tạo ra những chuẩn mực thực tiễn có ảnh hưởng định hình kỳ vọng của thị trường và xã hội về "thực hành tốt" trong lĩnh vực AI, trước khi và độc lập với việc các chuẩn mực đó được bất kỳ chủ thể nào chính thức hóa thành quy phạm. Đây là phương thức ảnh hưởng đặc thù của lĩnh vực công nghệ cao, nơi tốc độ đổi mới thực tiễn kinh doanh thường vượt trước tốc độ phản ứng của quy trình xây dựng khung pháp lý chính thức. Chuẩn mực thị trường trong nhiều trường hợp trở thành nguyên liệu tham chiếu cho chủ thể trực tiếp xây dựng khung pháp lý khi thiết kế nội dung của tầng bộ quy tắc ứng xử.

2.1.2.2. Nội dung của hoạt động xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Nội dung của hoạt động xây dựng khung pháp lý cho phát triển, ứng dụng AI không chỉ có văn bản quy phạm pháp luật, mà còn bao gồm kiến trúc phân tầng của

khung pháp lý, nguyên tắc và phương pháp tiếp cận điều chỉnh, cơ chế thử nghiệm có kiểm soát, hệ thống quyền và nghĩa vụ của các chủ thể, cơ chế bảo đảm tuân thủ và khắc phục hậu quả, cơ chế rà soát và cập nhật. Dựa trên nền tảng các phân tích về bản chất của AI, nội hàm khung pháp lý, yêu cầu điều chỉnh và chủ thể tham gia, có thể xác định các nội dung cơ bản của hoạt động xây dựng khung pháp lý cho phát triển, ứng dụng AI như sau:

Thứ nhất, kiến trúc phân tầng của khung pháp lý là sự phân bổ vấn đề điều chỉnh giữa các tầng và quan hệ giữa các tầng. Đây là nội dung có ý nghĩa định hướng đối với toàn bộ quá trình xây dựng khung pháp lý cho phát triển, ứng dụng AI. Để xác định các lựa chọn chính sách nền tảng, mô hình điều chỉnh và cách tổ chức các tầng trong một chỉnh thể thống nhất, cần xác định rõ mục tiêu, giá trị và thứ tự ưu tiên của khung pháp lý. Trước khi xây dựng các quy phạm cụ thể, cần làm rõ khung pháp lý AI nhằm đạt mục tiêu gì, bảo vệ giá trị nào và xử lý thế nào khi các mục tiêu xung đột nhau. Tiếp theo, lựa chọn mô hình kiến trúc điều chỉnh có thể bao gồm: tập trung hay phân tán thẩm quyền điều chỉnh; kiểm soát trước thị trường hay xử lý sau thiệt hại; ưu tiên quy phạm cứng hay kết hợp với quy phạm mềm. Không có mô hình tối ưu tuyệt đối; việc lựa chọn cần căn cứ vào truyền thống pháp lý, mô hình nhà nước, trình độ phát triển công nghệ, năng lực thể chế, bối cảnh địa - chính trị và mục tiêu chính sách ưu tiên.

Thứ hai, nguyên tắc và phương pháp tiếp cận điều chỉnh định hình triết lý lập pháp, giải bài toán cân bằng giữa hai mục tiêu tối thượng: thúc đẩy đổi mới sáng tạo, bảo vệ chủ quyền công nghệ và kiểm soát rủi ro, bảo vệ quyền con người cùng trật tự công cộng. Trong đó, cơ quan hoạch định phải lựa chọn phương pháp luận quản trị chủ đạo, tiêu biểu là: (i) Tiếp cận dựa trên rủi ro: Phân loại hệ thống AI theo các cấp độ rủi ro (không thể chấp nhận, rủi ro cao, rủi ro trung bình/thấp) để áp dụng nghĩa vụ tuân thủ tương ứng, tránh đánh đồng mọi ứng dụng AI gây lãng phí nguồn lực xã hội. Để trả lời câu hỏi đối với mỗi loại/mức độ rủi ro của AI, cường độ can thiệp pháp lý nên ở mức nào: cấm tuyệt đối, ràng buộc chặt, hướng dẫn mềm, hay để tự điều chỉnh, cần xác lập tiêu chí phân loại rủi ro, nguyên tắc tỷ lệ giữa mức rủi ro và mức nghĩa vụ và xác lập ngưỡng miễn trừ theo quy mô chủ thể. (ii) Xác định các nguyên tắc điều hướng: Quản trị vị nhân sinh, nguyên tắc thận trọng, bảo đảm tính minh bạch và tính giải thích được. Việc chọn lựa đúng phương pháp luận giúp khung pháp lý vừa có tính định hướng dài hạn, vừa tránh được hai thái cực, một là thắt chặt quá mức làm triệt tiêu động lực đổi mới hoặc hai là buông lỏng dẫn đến rủi ro mất kiểm soát.

Thứ ba, cơ chế thử nghiệm có kiểm soát (Regulatory Sandbox) là một chế định gồm điều kiện, cấp độ, thẩm quyền chấp thuận, giới hạn và miễn trừ, cung cấp dữ liệu cho giai đoạn đánh giá và thích ứng.

Nhằm khắc phục vấn đề khi tốc độ phát triển của công nghệ vượt xa năng lực phản ứng của hệ thống pháp luật, việc thiết kế và vận hành các không gian thử nghiệm pháp lý đóng vai trò là cầu nối thực chứng giữa lý thuyết và thực tiễn. Regulatory Sandbox cho phép doanh nghiệp triển khai thử nghiệm các sản phẩm, dịch vụ AI rủi ro cao hoặc chưa có quy định điều chỉnh (như xe tự lái, AI trong chẩn đoán y tế, tự động hóa tài chính) trong một phạm vi không gian, thời gian và tập khách hàng giới hạn, dưới sự giám sát trực tiếp của cơ quan quản lý và cơ chế miễn trừ trách nhiệm pháp lý có điều kiện. Quá trình này giúp nhà lập quy thu thập dữ liệu thực tế về các rủi ro phát sinh, đo lường phản ứng của thị trường và thử nghiệm tính hiệu quả của các công cụ quản lý dự kiến trước khi chính thức hóa thành luật cứng trên diện rộng.

Thứ tư, hệ thống quyền và nghĩa vụ của các chủ thể, cơ chế bảo đảm tuân thủ và khắc phục hậu quả:

Nội dung hoạt động này bao gồm phân định nghĩa vụ theo vị trí trong chuỗi giá trị; quyền của người bị tác động; và thẩm quyền giữa các cơ quan nhà nước khi đối tượng có tính xuyên ngành. Để bảo đảm nghĩa vụ trên được tuân thủ trên thực tế và người bị thiệt hại có thể được khắc phục, cần lựa chọn công cụ bảo đảm tuân thủ, báo cáo sự cố, đánh giá sự phù hợp, thanh tra và thiết kế cơ chế khắc phục (bồi thường và xử phạt, biện pháp khẩn cấp, quyền khiếu nại).

Thứ năm, cơ chế rà soát và cập nhật:

Do AI có bản chất tiến hóa liên tục (như sự trỗi dậy của AI tạo sinh hay các tác nhân AI tự chủ), hoạt động xây dựng khung pháp lý không dừng lại khi văn bản được ban hành mà vận hành theo chu trình khép kín và liên tục. Cơ quan quản lý phải duy trì hoạt động rà soát định kỳ thông qua: một là, khi các điều khoản tự động hết hiệu lực buộc quy định phải được đánh giá lại sau một khoảng thời gian nhất định để tiếp tục duy trì hoặc sửa đổi; hai là, giám sát sau thị trường theo dõi diễn biến các sự cố do AI gây ra trong thực tiễn để phát hiện kịp thời các lỗ hổng mới; ba là, cập nhật danh mục rủi ro điều chỉnh ranh giới phân loại các hệ thống AI rủi ro cao dựa trên sự xuất hiện của các mô hình nền tảng mới. Nội dung này bảo đảm cho hệ thống pháp luật duy trì tính thích ứng linh hoạt, giúp nhà nước luôn giữ thế chủ động trước những bước ngoặt công nghệ trong tương lai.

2.1.2.3. Quy trình xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Xây dựng khung pháp lý cho phát triển, ứng dụng AI cần được nhìn nhận như một quy trình chính sách - pháp lý liên tục, không chỉ là quá trình soạn thảo và ban hành văn bản. Cũng dựa trên nền tảng các phân tích về bản chất của AI, khung pháp lý, yêu cầu điều chỉnh, chủ thể tham gia và nội dung, có thể xác định quy trình xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo bao gồm bốn giai đoạn cơ bản đó là: chẩn đoán nhu cầu điều chỉnh; thiết kế kiến trúc tổng thể; tạo lập các thành tố; vận hành, đánh giá và thích ứng.

Giai đoạn thứ nhất là chẩn đoán nhu cầu điều chỉnh. Đây là bước nhằm xác định đúng vấn đề trước khi lựa chọn công cụ pháp lý. Đối với AI, chẩn đoán cần tập trung vào năm nội dung chính. Một là, nhận diện rủi ro hiện thực, tức là ưu tiên những rủi ro đang hoặc sắp phát sinh trong thực tế, thay vì chỉ dựa vào các rủi ro lý thuyết. Ở Việt Nam, các rủi ro như lừa đảo bằng deepfake, xâm phạm dữ liệu cá nhân, thiên lệch trong quyết định tự động hoặc sử dụng AI tạo sinh để phát tán thông tin sai lệch cần được nhận diện sớm. Hai là, rà soát văn bản hiện hành xem đang thiếu hay chồng chéo ở đâu. Ba là, đánh giá năng lực thể chế hiện có, bao gồm năng lực của cơ quan quản lý, năng lực kỹ thuật, nhân sự, dữ liệu và công cụ giám sát. Bốn là, tham vấn đa bên với doanh nghiệp, chuyên gia kỹ thuật, chuyên gia pháp lý, người tiêu dùng, tổ chức xã hội và các nhóm có thể bị ảnh hưởng. Năm là, phân tích kinh nghiệm quốc tế có chọn lọc, dựa trên mức độ tương đồng về bối cảnh, năng lực thể chế và mục tiêu chính sách. Về vị trí trình tự, giai đoạn này bắt buộc phải hoàn thành và được công bố công khai trước khi bước sang giai đoạn thiết kế kiến trúc.

Giai đoạn thứ hai là thiết kế kiến trúc tổng thể của khung pháp lý. Đây là giai đoạn lựa chọn và sắp đặt có ý nghĩa quyết định vì xác lập hình dạng, logic và phương thức vận hành của toàn bộ khung pháp lý. Các hoạt động của giai đoạn này bao gồm: chọn phương pháp tiếp cận giữa các phương án đã có, chẳng hạn quản lý dựa trên rủi ro; phân bổ từng vấn đề điều chỉnh vào tầng phù hợp; xác định thứ tự ưu tiên; lập lộ trình ban hành đồng bộ; xác định thiết chế nào cần có và cần năng lực gì và công bố tài liệu thiết kế sơ bộ để nhận phản biện trước khi bước sang giai đoạn soạn thảo chi tiết. Về vị trí trình tự, đây là giai đoạn duy nhất có chức năng cầu nối, chuyển hóa bằng chứng thu thập ở giai đoạn một thành khung định hướng cho việc soạn thảo chi tiết ở giai đoạn ba. Về đóng góp cho nội dung, giai đoạn này không trực tiếp tạo ra nội dung cuối cùng mà tạo ra bản thiết kế định hướng cho việc soạn thảo nội dung của cả ba nhóm thành tố ở giai đoạn tiếp theo.

Giai đoạn thứ ba là tạo lập các thành tố. Các hoạt động của giai đoạn này bao gồm: soạn thảo, thẩm định, thông qua và ban hành luật, nghị định, quyết định, thông

tư; xây dựng và công bố tiêu chuẩn quốc gia, kể cả theo trình tự rút gọn; ban hành hướng dẫn kỹ thuật tham chiếu; khuyến khích hoặc thừa nhận bộ quy tắc ứng xử của hiệp hội; giao nhiệm vụ, công nhận và cấp năng lực cho thiết chế: chỉ định cơ quan đầu mối, công nhận tổ chức đánh giá sự phù hợp, vận hành Cổng thông tin một cửa; phía tòa án, lựa chọn và công bố án lệ. Về vị trí trình tự, đây là giai đoạn có khối lượng hoạt động lớn nhất và là nơi phần lớn nội dung chính thức được soạn thảo thành văn bản cụ thể. Trong thực tế, sai lầm thường gặp là tập trung nhiều vào việc ban hành văn bản mà xem nhẹ thiết chế thực thi.

Giai đoạn thứ tư là vận hành, đánh giá và thích ứng. Đây không phải là giai đoạn kết thúc, mà là giai đoạn học hỏi của khung pháp lý. Các hoạt động chính bao gồm triển khai quy định và hướng dẫn tuân thủ; giám sát việc thực hiện; thu thập dữ liệu từ thông báo phân loại, báo cáo sự cố, kết quả thanh tra, khiếu nại, kết quả thử nghiệm có kiểm soát và bản án; đánh giá tác động thực tế sau ban hành; phát hiện quy định không còn phù hợp và cập nhật kịp thời các thành phần của khung pháp lý; chuyển kết luận đánh giá trở lại giai đoạn 1 để khởi động một vòng mới. Việc đánh giá cần được thực hiện theo chu kỳ phù hợp, luật có thể được rà soát trong chu kỳ dài hơn, trong khi tiêu chuẩn kỹ thuật, hướng dẫn thực hành và công cụ kiểm định cần được cập nhật thường xuyên hơn. Về vị trí trình tự, đây là giai đoạn duy nhất không có điểm kết thúc mà quay vòng trở lại cung cấp bằng chứng đầu vào cho giai đoạn chẩn đoán của chu kỳ tiếp theo. Đối với AI, yêu cầu này đặc biệt quan trọng vì năng lực công nghệ và mức độ ứng dụng AI sẽ thay đổi nhanh; do đó, khung pháp lý cần được thiết kế như một hệ thống có khả năng tự học hỏi và điều chỉnh, thay vì một cấu trúc tĩnh chỉ vận hành theo các văn bản đã ban hành.

Bốn giai đoạn được phân theo chức năng logic. Trên thực tế chúng có thể chồng lên nhau và lặp lại thành các vòng nhỏ, không đòi hỏi các giai đoạn phải nối tiếp nhau một cách tuyến tính.

2.1.2.4. Kỹ thuật xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Kỹ thuật xây dựng khung pháp lý cho phát triển, ứng dụng AI là tập hợp các phương pháp, công cụ và quy trình được áp dụng trong từng giai đoạn xây dựng để bảo đảm sản phẩm đạt các mục tiêu đã đề ra. Nền tảng lý luận của kỹ thuật xây dựng khung pháp lý cho phát triển, ứng dụng AI bao gồm:

Thứ nhất, lý thuyết soạn thảo lập pháp hiện đại được phát triển bởi Garth Thornton [212] và Helen Xanthaki [222], xác lập rằng soạn thảo văn bản pháp lý là một kỹ năng kỹ thuật có thể học, có thể chuẩn hóa và có thể đánh giá chất lượng độc lập với nội dung chính sách. Xanthaki phân biệt ba chiều của soạn thảo lập pháp: *mục tiêu chính sách* thuộc về chủ thể và quy trình, *cấu trúc văn bản* thuộc về kỹ thuật và

hiệu lực đạt được liên quan đến tiêu chí. Kỹ thuật soạn thảo quyết định liệu mục tiêu chính sách đúng đắn có được chuyển hóa thành quy phạm có hiệu lực thực tiễn hay không. Đối với lĩnh vực công nghệ, đặc biệt AI, Xanthaki nhấn mạnh kỹ thuật *định nghĩa* là quan trọng nhất: định nghĩa theo danh sách kỹ thuật cụ thể có nguy cơ lỗi thời ngay khi ban hành; định nghĩa theo chức năng bền vững hơn vì bao quát công nghệ tương lai.

Thứ hai, phương pháp luận *Better Regulation* được tổng hợp trong *OECD Regulatory Policy Outlook* (xuất bản định kỳ) và Khuyến nghị về Chính sách Quy định [291]. Cốt lõi của phương pháp luận này là đánh giá tác động pháp lý (*Regulatory Impact Assessment - RIA*) như một kỹ thuật phân tích bằng chứng bắt buộc trước khi ban hành văn bản quy phạm. OECD xác định RIA là kỹ thuật, không phải nguyên tắc hay quy trình, có thể được thực hiện tốt hoặc kém trong cùng một quy trình soạn thảo; có thể tuân thủ nguyên tắc dựa trên bằng chứng hoặc vi phạm nguyên tắc đó tùy theo phương pháp thực hiện. Baldwin, Cave và Lodge [321] mở rộng phân tích này: RIA chỉ có giá trị thực khi kèm theo phân tích chi phí và lợi ích thực chất, so sánh lựa chọn chính sách thay thế và phân tích tác động phân phối. Đối với AI, lưu ý rằng công nghệ phát triển nhanh đặt ra thách thức đặc thù cho RIA: không chỉ đánh giá tác động của văn bản hiện tại mà còn tác động của việc có hay không điều khoản thích ứng và đánh giá hậu kỳ.

Thứ ba, lý thuyết quản trị kỹ thuật số và "Collingridge Dilemma" [184] cho rằng khi công nghệ còn mới, có thể điều chỉnh dễ dàng nhưng khó đánh giá tác động; khi công nghệ đã phổ biến, có thể đánh giá tác động nhưng rất khó điều chỉnh do đã có sự phụ thuộc công nghệ. Tình trạng này không thể giải quyết bằng nguyên tắc hay tiêu chí mà chỉ giải quyết được bằng *kỹ thuật thiết kế đặc thù*: xây dựng khung pháp lý với cơ chế thích ứng tích hợp từ đầu, điều khoản rà soát bắt buộc và ủy quyền cập nhật linh hoạt.

Brownsword, Goodwin [324] phát triển từ Collingridge Dilemma: đặc trưng "hộp đen" của AI đặt ra yêu cầu kỹ thuật đặc biệt cho soạn thảo điều khoản trách nhiệm, không thể dùng kỹ thuật lập pháp truyền thống chứng minh lỗi khi cơ chế quyết định của AI không thể quan sát được. Thay vào đó cần kỹ thuật phân bổ trách nhiệm theo *vị trí kiểm soát* trong chuỗi giá trị.

Thứ tư, lý thuyết luật hoá tiêu chuẩn: Majone [218] xác lập luận điểm có ảnh hưởng lâu dài: trong điều chỉnh công nghệ, "quy tắc dựa trên nguyên tắc" (*principles-based rules*) và "quy tắc dựa trên tiêu chuẩn" (*standards-based rules*) có đặc tính pháp lý khác nhau và đòi hỏi kỹ thuật soạn thảo khác nhau. Kỹ thuật luật hóa tiêu chuẩn (viện dẫn tiêu chuẩn kỹ thuật trong văn bản pháp luật theo cách biến tuân thủ tiêu chuẩn thành suy đoán tuân thủ pháp luật) là sáng tạo kỹ thuật giải quyết bài toán: quy

phạm pháp luật giữ ở mức nguyên tắc bền vững, còn yêu cầu kỹ thuật chi tiết có thể cập nhật nhanh hơn.

Chalmers và cộng sự [178] ghi nhận rằng EU sử dụng kỹ thuật này rộng rãi trong điều chỉnh công nghệ, Điều 40 EU AI Act [206] ủy quyền cho Ủy ban châu Âu chỉ định tiêu chuẩn hài hòa mà các nhà cung cấp AI rủi ro cao có thể áp dụng để tạo ra suy đoán tuân thủ, đây là bằng chứng thực định rõ ràng nhất về giá trị của kỹ thuật luật hóa tiêu chuẩn.

Trên cơ sở nền tảng lý luận trên, có thể xác định sáu kỹ thuật xây dựng trọng yếu cho khung pháp lý cho phát triển, ứng dụng AI như sau:

Kỹ thuật thứ nhất, đánh giá tác động pháp lý (RIA)

RIA là quy trình phân tích có hệ thống nhằm đánh giá hậu quả của các lựa chọn chính sách trước khi ban hành văn bản pháp lý. Trong xây dựng khung pháp lý cho phát triển, ứng dụng AI, RIA đầy đủ bao gồm bốn bước. *Bước 1 - Xác định vấn đề*: Mô tả vấn đề chính sách cần giải quyết với bằng chứng định lượng về quy mô và bản chất (không phải chỉ nói "deepfake gây hại" mà phải có số liệu về tần suất, quy mô thiệt hại và phân bố đối tượng bị hại). *Bước 2 - Xác định lựa chọn*: Đề xuất ít nhất ba lựa chọn chính sách khác biệt, bao gồm cả lựa chọn "không làm gì" như điểm so sánh. *Bước 3 - Phân tích chi phí - lợi ích*: Đánh giá tác động kinh tế, xã hội và hành chính của từng lựa chọn, bao gồm phân tích tác động phân phối (ai được lợi, ai chịu chi phí và theo tỷ lệ như thế nào). *Bước 4 - Lý giải lựa chọn ưu tiên*: Giải thích công khai tại sao lựa chọn ưu tiên được chọn, có tài liệu phân tích kèm theo.

Đối với khung pháp lý cho phát triển, ứng dụng AI, RIA có yêu cầu đặc thù: đánh giá tác động đến hệ sinh thái đổi mới sáng tạo (yêu cầu quá cao có thể cản trở startup AI nội địa); đánh giá năng lực thể chế thực thi (yêu cầu không khả thi về năng lực thực thi có thể biến quy định thành hình thức). OECD [293] lưu ý rằng RIA cho công nghệ số cần xử lý *bất định kép*: không chỉ bất định về tác động của quy định đề xuất mà còn bất định về tốc độ thay đổi công nghệ. RIA cho khung pháp lý AI vì vậy cần bổ sung hai yếu tố so với RIA thông thường: *(i) phân tích tác động theo kịch bản công nghệ,* đánh giá văn bản dự thảo sẽ hoạt động như thế nào trong ba kịch bản phát triển AI (chậm, vừa, nhanh) trong 5 năm tới; *(ii) phân tích gánh nặng tuân thủ theo quy mô doanh nghiệp,* đặc biệt xác định ngưỡng quy mô nào thì yêu cầu tuân thủ đầy đủ trở thành rào cản thực tiễn đối với SME và startup AI.

Kỹ thuật thứ hai, định nghĩa chức năng

Xanthaki trong *Drafting Legislation* [222] phân biệt ba kiểu định nghĩa trong soạn thảo lập pháp: *định nghĩa liệt kê*, *định nghĩa kiểu loại* và *định nghĩa chức năng*. Đối với công nghệ thay đổi nhanh, định nghĩa chức năng ưu việt hơn vì tránh được

điểm yếu cơ cấu của định nghĩa liệt kê và cho phép văn bản pháp lý tự mở rộng để bao phủ công nghệ tương lai mà không cần sửa đổi.

Kỹ thuật thứ ba, ủy quyền linh hoạt

Trong lý thuyết quản trị thích ứng, Stankey và cộng sự [214] đề xuất nguyên tắc "quản lý thích ứng" - hệ thống quản trị cần thiết kế tích hợp cơ chế học hỏi và cập nhật từ đầu. Dù phát triển trong bối cảnh quản lý tài nguyên thiên nhiên, nguyên tắc này được vận dụng rộng rãi trong quản trị công nghệ: luật khung ổn định (7-10 năm) ủy quyền cho nghị định cập nhật yêu cầu kỹ thuật (2-3 năm), nghị định lại ủy quyền cho tiêu chuẩn cập nhật chi tiết kỹ thuật (12-18 tháng).

Kỹ thuật cụ thể trong soạn thảo: *(i) Ủy quyền có ngày và ủy quyền không có ngày* trong viện dẫn tiêu chuẩn kỹ thuật: viện dẫn có ngày chỉ áp dụng phiên bản tiêu chuẩn cụ thể; viện dẫn không có ngày tự động áp dụng phiên bản cập nhật nhất. Đối với AI, viện dẫn không có ngày phù hợp hơn nhưng cần cơ chế thông báo khi tiêu chuẩn cập nhật. *(ii) Điều khoản danh mục cập nhật theo quy trình rút gọn*: thiết kế điều khoản cho phép Quyết định Thủ tướng cập nhật danh mục AI rủi ro cao theo quy trình 60 ngày thay vì quy trình ban hành văn bản thông thường, Luật AI 2025 đã vận dụng kỹ thuật này đúng hướng, cần mở rộng cho cả cập nhật tiêu chuẩn kỹ thuật. *(iii)* Thiết kế điều khoản quy định thời hạn tự động hết hiệu lực hoặc nghĩa vụ rà soát và báo cáo Quốc hội sau thời hạn xác định.

Kỹ thuật thứ tư, phân định nghĩa vụ theo vị trí chuỗi giá trị

Thuật ngữ "Many hands problem" lần đầu tiên được phân tích bởi Thompson [188] mô tả tình huống khi trách nhiệm bị phân tán qua quá nhiều chủ thể đến mức không ai phải chịu trách nhiệm. Brownsword [324] phân tích vấn đề tương tự trong ngữ cảnh chuỗi tự động hóa kỹ thuật số, trách nhiệm theo mô hình lỗi truyền thống không hoạt động khi không thể xác định chủ thể nào có "lỗi" khi AI quyết định sai.

Giải pháp kỹ thuật là chuyển từ phân bổ theo lỗi sang phân bổ theo kiểm soát và lợi ích. Chủ thể có khả năng kiểm soát hành vi của AI nhiều hơn và hưởng lợi kinh tế nhiều hơn từ AI chịu nghĩa vụ pháp lý cao hơn, bất kể có chứng minh được "lỗi" theo nghĩa truyền thống hay không. EU AI Act tại Điều 9-17 vận hành kỹ thuật này, phân biệt rõ nghĩa vụ của nhà cung cấp mô hình và nhà triển khai, mỗi nhóm có nghĩa vụ khác nhau tương ứng với mức độ kiểm soát.

Kỹ thuật thứ năm, miễn trừ tỷ lệ

Kaplow [259] lập luận rằng chi phí tuân thủ là một biến số độc lập quan trọng trong thiết kế pháp luật, quy định có chi phí tuân thủ vượt quá khả năng của đối tượng sẽ tạo ra loại trừ pháp lý trong thực tiễn. Đối với hệ sinh thái AI, loại trừ pháp lý của startup và SME có hậu quả không chỉ gây bất bình đẳng mà còn làm giảm đổi mới sáng tạo vì đây là lực lượng đổi mới quan trọng nhất.

Kỹ thuật soạn thảo cụ thể là thiết kế ngưỡng hai chiều với tiêu chí cộng dồn, ví dụ doanh nghiệp được áp dụng chế độ tuân thủ đơn giản hóa khi đồng thời đáp ứng: *(a)* doanh thu hàng năm dưới mức X; *(b)* số lượng người dùng thường xuyên của hệ thống AI dưới mức Y. Hai tiêu chí cộng dồn tránh được tình huống doanh nghiệp lớn có doanh thu cao nhưng ít người dùng lách luật. AI Basic Act Hàn Quốc đã vận hành kỹ thuật này, quy định miễn trừ cho doanh nghiệp "nhỏ" theo ngưỡng kết hợp quy mô và doanh thu [277]. Kỹ thuật này cần phân biệt rõ với miễn trừ hoàn toàn, miễn trừ tỷ lệ không loại bỏ nghĩa vụ pháp lý mà điều chỉnh cách thức chứng minh tuân thủ.

Kỹ thuật thứ sáu, luật hóa tiêu chuẩn kỹ thuật

Trong lý luận về nguồn pháp luật, Majone [218] phân biệt giữa "quy tắc" (quy định cứng nhắc) và "tiêu chuẩn" (yêu cầu có thể được đáp ứng bằng nhiều cách khác nhau). Trong điều chỉnh công nghệ cao, tiêu chuẩn ưu việt hơn quy tắc vì tính linh hoạt với sự đa dạng kỹ thuật; có thể cập nhật nhanh hơn và có thể được xây dựng bởi những người có kiến thức kỹ thuật sâu hơn cơ quan lập pháp. Kỹ thuật luật hóa tiêu chuẩn là cách chuyển hóa ưu điểm của tiêu chuẩn vào khung quy phạm có tính bắt buộc.

Hai hình thức viện dẫn và hệ quả pháp lý khác nhau:

Viện dẫn tham chiếu đơn thuần: "nhà cung cấp nên tuân thủ TCVN ISO/IEC 42001" không tạo ra suy đoán tuân thủ, tiêu chuẩn chỉ mang tính chất khuyến nghị.

Viện dẫn tạo suy đoán tuân thủ: "hệ thống AI rủi ro cao được suy đoán là tuân thủ yêu cầu về hệ thống quản lý AI tại Điều X nếu tuân thủ TCVN ISO/IEC 42001 phiên bản hiện hành" tạo ra "safe harbor" (vùng an toàn) kỹ thuật: doanh nghiệp tuân thủ tiêu chuẩn mặc nhiên được coi là tuân thủ yêu cầu pháp lý tương ứng, không cần chứng minh thêm.

Kỹ thuật này có giới hạn chỉ áp dụng được khi tiêu chuẩn kỹ thuật tương ứng đã tồn tại và đủ chi tiết. Khi TCVN chưa có (như trường hợp TCVN về AI agents và watermarking hiện tại), viện dẫn sẽ tạo ra nghĩa vụ pháp lý không có công cụ thực thi. Đây là lý do tầng tiêu chuẩn kỹ thuật phải được xây dựng song song hoặc trước tầng VBQPPL khi áp dụng kỹ thuật này.

2.2. CÁC NGUYÊN TẮC, TIÊU CHÍ ĐÁNH GIÁ VÀ ĐIỀU KIỆN BẢO ĐẢM XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO

2.2.1. Các nguyên tắc xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Nguyên tắc xây dựng khung pháp lý là những điều đặt ra mang tính nền tảng, định hướng toàn bộ quá trình xây dựng từ thiết kế kiến trúc đến tạo lập thành phần, từ thiết lập thiết chế đến vận hành cơ chế thích ứng. Nguyên tắc là những đòi hỏi bắt buộc

phải tuân theo trong suốt quá trình xây dựng, vi phạm nguyên tắc sẽ dẫn đến sai lầm có tính hệ thống, không thể khắc phục chỉ bằng bổ sung nội dung sau đó.

Việc xây dựng khung pháp lý cho phát triển, ứng dụng AI một cách bền vững và có trách nhiệm không chỉ đòi hỏi sự đầy đủ về mặt quy phạm mà trước hết phải được đặt trên những nguyên tắc cơ bản. Trên cơ sở những vấn đề lý luận đã xác lập ở các mục trên, đặc biệt là cách tiếp cận coi khung pháp lý AI là một hệ thống đa tầng và thích ứng, cũng như các yêu cầu đặt ra từ bản chất kỹ thuật - pháp lý của AI và từ mục tiêu bảo vệ quyền con người, kiểm soát rủi ro, thúc đẩy đổi mới sáng tạo có trách nhiệm, có thể xác định khung pháp lý điều chỉnh lĩnh vực này cần được thiết kế trên cơ sở tích hợp các nguyên tắc sau:

Thứ nhất, nguyên tắc đặt quyền con người làm trọng tâm

Nguyên tắc này đòi hỏi trong suốt quá trình xây dựng khung pháp lý AI từ thiết kế kiến trúc đến soạn thảo từng điều khoản cụ thể, bảo vệ và thúc đẩy quyền con người phải là mục tiêu chi phối tối thượng, không phải là một trong nhiều mục tiêu bình đẳng. Ở Việt Nam, cơ sở pháp lý của nguyên tắc này bắt nguồn trực tiếp từ Điều 3 Hiến pháp 2013: "Nhà nước bảo đảm và phát huy quyền làm chủ của Nhân dân; công nhận, tôn trọng, bảo vệ và bảo đảm quyền con người, quyền công dân". Nghĩa vụ hiến định này không bị loại trừ khi xây dựng khung pháp lý cho phát triển, ứng dụng AI, trái lại, đặt ra yêu cầu đặc biệt cao vì AI có tiềm năng vi phạm quyền con người theo quy mô và cách thức chưa từng có. Trên bình diện quốc tế, Khuyến nghị UNESCO về Đạo đức AI [356] khẳng định: "AI nên phục vụ cá nhân và nhân loại nói chung và do đó quyền con người và tự do cơ bản phải được tôn trọng, bảo vệ và thúc đẩy trong suốt vòng đời của hệ thống AI". Bộ nguyên tắc AI của OECD [297] đặt quyền con người và các giá trị dân chủ ở vị trí thứ hai trong danh sách năm nguyên tắc.

Từ lý thuyết về tiếp cận dựa trên quyền con người, trong thực tiễn xây dựng khung pháp lý, nguyên tắc này đòi hỏi: một là, mọi lựa chọn kiến trúc phải được kiểm tra cân nhắc bảo vệ hay tạo rủi ro cho quyền con người như thế nào trước khi được thông qua; *hai là*, khi có xung đột giữa lợi ích kinh tế (thúc đẩy đổi mới, giảm gánh nặng tuân thủ cho doanh nghiệp) và bảo vệ quyền con người, quyền con người phải được ưu tiên; *ba là*, các nhóm dễ bị tổn thương phải được đặc biệt quan tâm trong thiết kế bảo vệ.

Thứ hai, nguyên tắc tương xứng giữa biện pháp điều chỉnh và mức độ rủi ro

Nguồn gốc lý luận của nguyên tắc này từ học thuyết về nguyên tắc tỷ lệ trong luật công, theo cách hiểu của Alexy [319], lý thuyết điều chỉnh dựa trên rủi ro của Julia Black, đặc biệt mô hình điều chỉnh dựa trên rủi ro có tính phản hồi của Black và Baldwin.

Nguyên tắc này đòi hỏi mức độ kiểm soát pháp lý phải tương xứng với mức độ rủi ro thực tế của từng loại AI, không thống nhất áp dụng cùng một mức kiểm soát cho tất cả các hệ thống AI bất kể sự khác biệt về rủi ro giữa chúng. Nhưng nguyên tắc tỷ lệ không chỉ đòi hỏi kiểm soát đúng mức độ mà còn phải thúc đẩy đổi mới sáng tạo, không hy sinh hoàn toàn một mục tiêu cho mục tiêu kia. Đây là thách thức thiết kế trung tâm của mọi khung pháp lý AI: quá chặt sẽ cản trở đổi mới; quá lỏng sẽ khiến người dùng dễ bị phơi bày trước rủi ro. Điểm cân bằng tối ưu không cố định mà phụ thuộc vào bối cảnh lĩnh vực ứng dụng, trình độ phát triển công nghệ và mức độ chín muồi của thị trường.

Thứ ba, nguyên tắc tương thích quốc tế có chọn lọc

Nguyên tắc này đòi hỏi khung pháp lý cho phát triển, ứng dụng AI được xây dựng trên cơ sở có ý thức về bối cảnh quốc tế, tương thích đủ với chuẩn mực quốc tế để tạo điều kiện hội nhập, đồng thời bảo vệ lợi ích và không gian chính sách quốc gia. Nguồn gốc lý luận của nguyên tắc xuất phát từ lý thuyết về tiếp nhận pháp luật của Watson [143]; cảnh báo của Teubner [220] rằng quy tắc vay mượn có thể trở thành "chất kích ứng" trong một hệ thống pháp luật khác và phân tích của Bradford về hiệu ứng Brussels [151] khi quy định của Liên minh châu Âu lan ra ngoài lãnh thổ qua thị trường. Cơ sở pháp lý tại Điều 12 Hiến pháp về tuân thủ điều ước quốc tế và hội nhập, cùng quy định về bảo đảm thực hiện điều ước trong Luật Ban hành VBQPPL năm 2025.

"Tương thích có chọn lọc" phân biệt với hai thái cực: *sao chép toàn bộ* mô hình nước ngoài mà không điều chỉnh cho bối cảnh trong nước dẫn đến khung pháp lý không khả thi với điều kiện thể chế và kinh tế Việt Nam; *tự cô lập* bằng cách xây dựng khung pháp lý hoàn toàn khác biệt với chuẩn mực quốc tế dẫn đến doanh nghiệp Việt Nam khó tiếp cận thị trường quốc tế và khó thu hút đầu tư nước ngoài.

Nguyên tắc này đặt ra yêu cầu kép: cần *phân tích có hệ thống* các chuẩn mực quốc tế để xác định những gì có thể và nên học tập; cần *lý giải có căn cứ* khi quyết định không áp dụng một chuẩn mực quốc tế cụ thể. Thực tiễn chỉ ra rằng hai loại chuẩn mực quốc tế cần được tiếp cận khác nhau: *chuẩn mực kỹ thuật* (tiêu chuẩn ISO/IEC, phương pháp kiểm định) thường có thể áp dụng trực tiếp hoặc với điều chỉnh tối thiểu vì chúng dựa trên kiến thức kỹ thuật phổ quát; *chuẩn mực kiến trúc và chính sách* (mô hình thiết chế, cơ chế thực thi, mức phạt) cần điều chỉnh nhiều hơn vì chúng phụ thuộc sâu vào bối cảnh thể chế và kinh tế quốc gia.

Thứ tư, nguyên tắc nhất quán và phối hợp hệ thống

Nguyên tắc nhất quán đòi hỏi các thành phần của khung pháp lý cho phát triển, ứng dụng AI phải liên kết với nhau theo logic nội tại: không mâu thuẫn, không chồng chéo, không tạo ra khoảng trống điều chỉnh giữa các tầng và giữa các lĩnh vực.

Tính nhất quán cần được bảo đảm theo ba chiều: *chiều dọc* giữa luật, nghị định, thông tư và tiêu chuẩn kỹ thuật; *chiều ngang* giữa khung pháp lý AI và các lĩnh vực pháp luật liên quan (BVDLCN, an toàn thông tin mạng, bảo vệ người tiêu dùng, SHTT) và *chiều thẩm quyền* giữa các cơ quan quản lý nhà nước có thẩm quyền điều chỉnh AI.

Nhất quán không tự nhiên xảy ra mà đòi hỏi thiết kế có chủ đích. Trong thực tiễn xây dựng khung pháp lý, tình trạng thiếu nhất quán xuất hiện khi: các cơ quan soạn thảo văn bản độc lập mà không trao đổi thông tin đủ; khi có áp lực ban hành nhanh dẫn đến bỏ qua kiểm tra tính nhất quán; khi không có "kiến trúc sư hệ thống" - người chịu trách nhiệm tổng thể về tính nhất quán nội tại của toàn bộ khung pháp lý.

Nguyên tắc phối hợp hệ thống bổ sung cho tính nhất quán: ngay cả khi nội dung văn bản nhất quán, khung pháp lý vẫn có thể vận hành kém hiệu quả nếu các cơ quan thực thi không phối hợp với nhau. Phối hợp liên ngành phải được thiết kế vào kiến trúc từ đầu, không phải giải quyết tình huống khi xung đột đã xảy ra.

Thứ năm, nguyên tắc thích ứng và học hỏi từ thực tiễn

Nguyên tắc thích ứng đòi hỏi rằng khung pháp lý AI phải được xây dựng với cơ chế học hỏi từ thực tiễn được tích hợp ngay từ đầu trong thiết kế, không phải là điều chỉnh tình huống khi vấn đề phát sinh, mà là một quy trình thể chế hóa thường xuyên. Liên quan đến nguyên tắc này, lý thuyết "pacing problem" của Marchant [213] chỉ ra rằng công nghệ AI thay đổi nhanh hơn chu kỳ lập pháp truyền thống. Giải pháp là thiết kế khung pháp lý với khả năng tự cập nhật có kiểm soát, không phải là lập pháp nhanh hơn dẫn đến văn bản kém chất lượng.

Trong thực tiễn, nguyên tắc này đòi hỏi ít nhất bốn yếu tố phải được xây dựng đồng thời với văn bản quy phạm: hệ thống thu thập thông tin từ thực tiễn vận hành (báo cáo sự cố, dữ liệu tuân thủ, phản hồi từ đối tượng điều chỉnh); quy trình đánh giá định kỳ dựa trên thông tin thu thập được; cơ chế chuyển hóa kết quả đánh giá thành cập nhật chính sách theo thời hạn xác định; hàng rào bảo vệ để cơ chế học hỏi không bị lạm dụng bởi các nhóm lợi ích.

Sự tích hợp các nguyên tắc trên tạo nên nền tảng cho một khung pháp lý phát triển, ứng dụng AI vừa bảo đảm tính hợp pháp và minh bạch, vừa thích ứng với đặc thù rủi ro của công nghệ, đồng thời đặt lợi ích và quyền con người vào vị trí trung tâm. Đây là điều kiện cần để phát triển, ứng dụng AI theo hướng bền vững và có trách nhiệm ở mọi quốc gia, trong đó có Việt Nam.

2.2.2. Các tiêu chí đánh giá khung pháp lý và xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Xét về cơ sở lý luận, việc đánh giá việc xây dựng khung pháp lý thường dựa trên ba lý thuyết sau đây:

Lý thuyết về chất lượng quy trình lập pháp: Khoa học lập pháp từ lâu đã phân biệt giữa chất lượng của sản phẩm lập pháp (văn bản pháp luật) và chất lượng của quá trình lập pháp (cách thức tạo ra văn bản đó). Wintgens [260] lập luận rằng tính hợp lý và tính chính đáng của pháp luật không chỉ phụ thuộc vào nội dung mà còn vào quy trình tạo ra có tuân thủ các nguyên tắc lý tính thực tiễn không, có dựa trên bằng chứng không, có bảo đảm sự tham gia của những người bị ảnh hưởng không. Trong khi đó, Waldron [240] nhấn mạnh rằng trong một nhà nước pháp quyền, không chỉ nội dung của pháp luật mà cả quy trình tạo ra pháp luật phải tôn trọng phẩm giá của con người và bảo đảm sự tham gia bình đẳng. Áp dụng vào AI, quy trình xây dựng khung pháp lý AI phải bảo đảm tiếng nói của những người sẽ bị ảnh hưởng bởi AI, không chỉ tiếng nói của doanh nghiệp AI và cơ quan nhà nước.

Lý thuyết về tính chính đáng của quy phạm pháp luật: Julia Black [249] phân tích ba nguồn chính đáng của quy phạm điều chỉnh: nội dung tốt, quy trình đúng đắn và được xã hội chấp nhận. Đối với AI, một lĩnh vực kỹ thuật phức tạp mà phần lớn người dân không hiểu sâu. chính đáng thủ tục càng quan trọng: khi người dân không thể đánh giá chất lượng kỹ thuật của quy phạm, họ đánh giá chủ yếu dựa trên việc quy trình tạo ra quy phạm đó có đáng tin cậy không.

Ở Việt Nam, từ góc độ lý luận nhà nước và pháp luật, hoạt động đánh giá việc xây dựng pháp luật thường được tiếp cận qua các tiêu chí về: tính hợp hiến và hợp pháp của quy trình; tính khoa học của căn cứ soạn thảo; tính dân chủ của quá trình tham vấn và tính khả thi của phương án lập pháp được lựa chọn [107, tr.23-29]. Những tiêu chí này được thể chế hóa một phần trong Luật Ban hành VBQPPL 2025 nhưng cần được phát triển và cụ thể hóa thêm cho lĩnh vực AI với những đặc thù kỹ thuật của nó.

Trên cơ sở những phân tích ở trên, có thể xác định các tiêu chí cơ bản để đánh giá việc xây dựng khung pháp lý cho phát triển, ứng dụng AI như sau:

Tiêu chí 1: Tính có căn cứ thực tiễn và dựa trên bằng chứng

Nội dung đánh giá bao gồm: *(i)* đánh giá tác động pháp lý (*RIA*) được thực hiện trước khi soạn thảo văn bản, bao gồm đánh giá chi phí - lợi ích và phân tích các lựa chọn thay thế; *(ii)* nghiên cứu thực tiễn về rủi ro AI tại Việt Nam được thực hiện và sử dụng làm căn cứ; *(iii)* thu thập và phân tích dữ liệu về hành vi của hệ thống AI trong điều kiện thực tế Việt Nam; *(iv)* việc lựa chọn mô hình kiến trúc điều chỉnh được lý giải bằng bằng chứng so sánh từ kinh nghiệm quốc tế.

Tiêu chí 2: Tính khoa học trong thiết kế kiến trúc

Nội dung đánh giá bao gồm: *(i)* tài liệu thuyết minh kiến trúc tổng thể của khung pháp lý, giải thích lý do lựa chọn mô hình kiến trúc, lý do phân tầng văn bản và lý do phân công thẩm quyền; *(ii)* phân tích tính nhất quán nội tại, kiểm tra xem các thành phần của khung pháp lý có mâu thuẫn nhau hoặc tạo ra khoảng trống điều chỉnh

không; *(iii)* so sánh có hệ thống với kinh nghiệm quốc tế theo tiêu chí rõ ràng; *(iv)* xác định rõ lý do lựa chọn mô hình này thay vì các mô hình khác.

Tiêu chí 3: Tính tham vấn thực chất và bao trùm

Nội dung đánh giá bao gồm: *(i)* tham vấn kỹ thuật chuyên sâu với doanh nghiệp AI trong và ngoài nước, thu thập ý kiến kỹ thuật về tính khả thi của yêu cầu pháp lý; *(ii)* cơ chế bảo đảm tiếng nói của người tiêu dùng và công chúng - những người bị ảnh hưởng bởi AI nhưng không phải doanh nghiệp công nghệ; *(iii)* tham vấn các nhóm dễ bị tổn thương (người già, người khuyết tật, người lao động dễ bị thay thế bởi AI); *(iv)* ý kiến tham vấn được xem xét nghiêm túc và phản hồi công khai.

Tiêu chí 4: Tính minh bạch và trách nhiệm giải trình của quy trình

Nội dung đánh giá bao gồm: *(i)* dự thảo văn bản được công bố công khai để lấy ý kiến rộng rãi, thời hạn tham vấn đủ dài để các bên chuẩn bị ý kiến thực chất; *(ii)* công bố tài liệu thuyết minh giải thích lý do lựa chọn chính sách quan trọng, bao gồm lý do chấp nhận một số ý kiến và không chấp nhận ý kiến khác; *(iii)* cơ quan chủ trì có cơ chế trả lời ý kiến từng bên tham vấn không hay chỉ tổng hợp chung chung; *(iv)* sau khi ban hành, có cam kết công bố đánh giá hậu kỳ không.

Tiêu chí 5: Tính bền vững và khả năng cải tiến liên tục của quá trình xây dựng

Dựa trên câu hỏi quá trình xây dựng khung pháp lý cho phát triển, ứng dụng AI có được thiết kế để tự duy trì và cải tiến liên tục không, hay sẽ dừng lại sau khi ban hành văn bản, nội dung đánh giá: *(i)* Có cơ chế thể chế hóa để bảo đảm quá trình xây dựng không kết thúc khi luật được ban hành, kế hoạch cho các giai đoạn tiếp theo, phân công trách nhiệm rõ ràng, nguồn lực được phân bổ; *(ii)* Có hệ thống theo dõi và đánh giá việc thực thi được xây dựng đồng thời với khung pháp lý không, bao gồm chỉ số đo lường hiệu quả cụ thể; *(iii)* Có cam kết thể chế về đánh giá sau khi thực thi theo thời hạn xác định không; *(iv)* Có cơ chế học hỏi từ kinh nghiệm quốc tế được cập nhật không.

Bên cạnh đó, các tiêu chí đánh giá sản phẩm của hoạt động xây dựng là khung pháp lý cho phát triển, ứng dụng AI bao gồm:

Tiêu chí 1, tính đầy đủ và cân đối, đòi hỏi khung pháp lý có đủ bộ phận và phát triển tương xứng với nhau. Hai ngưỡng đánh giá tối thiểu: tầng 1, tầng 2, tầng 7 đã có mặt, nội dung thực chất được quy định ở tầng 2 và đầy đủ: không tầng nào ở mức chưa có và không giai đoạn nào thiếu chủ thể mang nghĩa vụ.

Tiêu chí 2, tính thống nhất và đồng bộ, đòi hỏi các thành tố ăn khớp và tạo ra một trật tự pháp lý ổn định, dễ dự đoán. Hai ngưỡng đánh giá tối thiểu: không có mâu thuẫn trực tiếp, các quy định giao chi tiết đã được thực hiện và đầy đủ: đồng bộ về thời gian và không còn chế độ song song.

Tiêu chí 3, tính phù hợp, đòi hỏi khung pháp lý phục vụ đúng mục tiêu kép trong tên đề tài và phù hợp với bối cảnh Việt Nam. Nội hàm gồm bốn thành phần: (i) Bảo đảm quyền con người và khả năng khắc phục; (ii) Tạo điều kiện cho đổi mới sáng tạo; (iii) Tương xứng trong phân bổ nghĩa vụ; (iv) Phù hợp điều kiện Việt Nam và tương thích quốc tế. Thành phần thứ nhất là mức sàn không thể bù trừ được. Nếu nó không đạt yêu cầu tối thiểu, toàn bộ tiêu chí phù hợp sẽ không đạt.

Tiêu chí 4, tính khả thi, đòi hỏi các quy định thi hành được trên thực tế. Hai ngưỡng đánh giá tối thiểu: mỗi nghĩa vụ chính có thiết chế được giao theo dõi và đầy đủ: thiết chế có năng lực thực tế và không quy định nào bị treo vì thiếu điều kiện.

Tiêu chí 5, tính thích ứng, đòi hỏi bản thân khung pháp lý có sẵn công cụ để tự điều chỉnh khi bằng chứng thay đổi. Ngưỡng đánh giá tối thiểu: có ít nhất một cơ chế cập nhật được quy định và đầy đủ: có điều khoản đánh giá định kỳ toàn bộ khung và dữ liệu sự cố là căn cứ bắt buộc của việc sửa đổi.

2.2.3. Các điều kiện bảo đảm xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Điều kiện bảo đảm xây dựng khung pháp lý cho phát triển, ứng dụng AI là tổng hợp các nhân tố khách quan và chủ quan, vật chất và tinh thần, trong nước và quốc tế, cần phải có hoặc cần phải tạo ra để quá trình xây dựng khung pháp lý diễn ra thành công và khung pháp lý đó vận hành hiệu quả trên thực tiễn. Có thể xác định năm nhóm điều kiện bảo đảm như sau:

Điều kiện về ý chí chính trị và định hướng chiến lược

Xây dựng khung pháp lý không thể thành công nếu thiếu ý chí chính trị rõ ràng và nhất quán từ lãnh đạo cao nhất của hệ thống chính trị. Kinh nghiệm quốc tế cho thấy các quốc gia xây dựng được khung pháp lý AI hiệu quả đều có: *(i)* cam kết chính trị cấp cao được ghi nhận trong văn kiện chính thức, không phải tuyên ngôn chính sách chung chung mà là chỉ tiêu cụ thể, lộ trình rõ ràng và phân công trách nhiệm theo thẩm quyền; *(ii)* tính nhất quán giữa các nhiệm kỳ bởi AI là lĩnh vực đầu tư dài hạn, đòi hỏi sự nhất quán chính sách qua nhiều nhiệm kỳ, không thể bị đảo ngược bởi sự thay đổi lãnh đạo.

Để đánh giá điều kiện này, ngưỡng tối thiểu là có văn kiện định hướng và chiến lược. Ngưỡng đầy đủ là mục tiêu về thể chế được lượng hóa, có lộ trình, có báo cáo định kỳ công khai và gắn với trách nhiệm người đứng đầu.

Điều kiện về năng lực thể chế và nhân lực là điều kiện bảo đảm quan trọng nhất và cũng là thách thức lớn nhất đối với Việt Nam hiện nay, bao gồm ba thành phần chính.

Năng lực kỹ thuật-pháp lý là loại năng lực đặc thù chưa được đào tạo chính thức tại Việt Nam, phải có khả năng kết hợp hiểu biết sâu về kỹ thuật AI với kiến thức pháp

lý đủ để thiết kế quy phạm có thể thực thi và giải thích các vấn đề pháp lý phát sinh từ hành vi kỹ thuật của hệ thống AI.

Năng lực quản lý pháp lý chỉ khả năng của cơ quan nhà nước để thực hiện đầy đủ chức năng quản lý trong lĩnh vực AI từ soạn thảo quy định đến đánh giá tác động, từ cấp phép đến kiểm định, từ giám sát đến xử lý vi phạm. Năng lực này cần được đánh giá thực tiễn trước khi thiết kế khung pháp lý, không thể ban hành yêu cầu tuân thủ mà cơ quan nhà nước không có khả năng kiểm tra.

Năng lực tư pháp: Tòa án phải có đủ kiến thức về AI để giải quyết tranh chấp liên quan một cách công bằng và nhất quán. Kinh nghiệm ở nhiều quốc gia cho thấy sự thiếu năng lực tư pháp dẫn đến phán quyết tùy tiện làm suy yếu tính tiên liệu được của pháp luật và cản trở đầu tư AI.

Để đánh giá điều kiện này, ngưỡng tối thiểu là cơ quan đầu mối đã được xác định và có đơn vị phụ trách. Ngưỡng đầy đủ là mỗi khâu soạn thảo, thẩm định, đánh giá sự phù hợp, thanh tra, xét xử đều có đội ngũ được đào tạo và có nguồn cung nhân lực ổn định.

Điều kiện về hạ tầng pháp lý và kỹ thuật

Hạ tầng pháp lý là môi trường pháp luật và công cụ lập pháp mà hoạt động xây dựng khung dựa vào, bao gồm hệ thống pháp luật BVDLCN, an toàn thông tin mạng, SHTT trong môi trường số và trách nhiệm dân sự có thể áp dụng cho thiệt hại từ AI; tiêu chuẩn kỹ thuật; cơ sở dữ liệu văn bản; bản dịch chính thức; công cụ đánh giá tác động. Hạ tầng kỹ thuật gồm phòng thử nghiệm, bộ dữ liệu chuẩn để kiểm định, năng lực tính toán phục vụ kiểm thử và hệ thống CNTT của các nền tảng quản lý.

Để đánh giá điều kiện bảo đảm này, ngưỡng tối thiểu là pháp luật nền đã có và các nền tảng quản lý đã được thiết lập. Ngưỡng đầy đủ là có năng lực kiểm thử trong nước cho các lĩnh vực thuộc Danh mục rủi ro cao, có bộ dữ liệu chuẩn tiếng Việt để đánh giá thiên lệch và có văn bản dịch chính thức phục vụ doanh nghiệp nước ngoài.

Điều kiện về sự tham gia của các bên liên quan

Kinh nghiệm quốc tế chứng minh rằng khung pháp lý AI được xây dựng với sự tham gia thực chất của nhiều bên liên quan thường có chất lượng cao hơn và tỷ lệ tuân thủ cao hơn so với khung pháp lý được xây dựng theo quy trình đóng, chủ yếu từ phía nhà nước.

Tham vấn thực chất của doanh nghiệp AI: doanh nghiệp phải có cơ hội đưa ra ý kiến về tính khả thi kỹ thuật của yêu cầu pháp lý, và ý kiến đó phải được cân nhắc nghiêm túc trong quá trình soạn thảo. Cơ chế dự thảo công khai và trả lời từng ý kiến của Hoa Kỳ [141], mặc dù không phải là lựa chọn mô hình điều chỉnh tổng thể, lại là thực tiễn tốt trong tham vấn doanh nghiệp có thể học tập.

Tham gia của tổ chức xã hội và học thuật: Các tổ chức bảo vệ quyền người tiêu dùng, tổ chức bảo vệ quyền lao động, tổ chức bảo vệ quyền trẻ em và các cơ sở nghiên cứu độc lập phải có vai trò chính thức trong quy trình tham vấn.

Hợp tác quốc tế trong xây dựng khung pháp lý: Tham gia tích cực vào các diễn đàn quốc tế (OECD, UNESCO, ISO/IEC SC 42, ASEAN) không phải chỉ để học hỏi mà còn để bảo vệ lợi ích quốc gia trong quá trình định hình chuẩn mực quốc tế về AI.

Để đánh giá điều kiện bảo đảm này, ngưỡng tối thiểu là kênh chính thức theo luật định đang hoạt động. Ngưỡng đầy đủ là có kênh đối thoại kỹ thuật thường xuyên, có tổ chức đại diện đủ năng lực và có cơ chế để lắng nghe tiếng nói của nhóm dễ bị tổn thương.

Điều kiện về nguồn lực tài chính

Xây dựng và duy trì khung pháp lý AI hiệu quả đòi hỏi đầu tư tài chính đáng kể và bền vững, không thể thực hiện với ngân sách tùy tiện hoặc không đủ.

Đầu tư cho xây dựng khung pháp lý bao gồm chi phí nghiên cứu chính sách, soạn thảo văn bản, tham vấn đa bên, đào tạo cán bộ và xây dựng hệ thống thông tin. Kinh nghiệm cho thấy các quốc gia xây dựng khung pháp lý AI chất lượng thường đầu tư không kém cho quá trình tham vấn và đánh giá tác động so với soạn thảo văn bản.

Đầu tư cho thực thi bao gồm nhân lực kiểm tra, công nghệ giám sát (bao gồm công cụ phân tích AI để kiểm tra tuân thủ), hạ tầng xử lý khiếu nại và tranh chấp. Đây là khoản đầu tư thường bị cắt giảm nhất khi ngân sách bị thắt chặt nhưng lại là điều kiện quyết định nhất đối với hiệu quả thực thi.

Đầu tư cho quản trị AI phải được bảo đảm bền vững lâu dài, không bị cắt giảm theo chu kỳ ngân sách ngắn hạn. Mô hình doanh nghiệp nộp phí để trang trải chi phí kiểm định AI rủi ro cao của cơ quan nhà nước là một cơ chế tài chính bền vững nên áp dụng.

Để đánh giá điều kiện bảo đảm này, ngưỡng tối thiểu là có cam kết ngân sách chung và nguồn cho các biện pháp hỗ trợ đã được pháp luật quy định. Ngưỡng đầy đủ là có dòng kinh phí xác định cho từng khâu của việc xây dựng và thi hành khung, được công khai và đánh giá hiệu quả.

2.3. KINH NGHIỆM QUỐC TẾ VỀ XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO VÀ GIÁ TRỊ THAM KHẢO CHO VIỆT NAM

2.3.1. Các tiêu chuẩn quốc tế về xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Từ cơ sở lý luận về khung pháp lý AI đã phân tích ở chương trước, có thể thấy các chuẩn mực quốc tế hiện nay không tạo thành một "luật AI toàn cầu" thống nhất, nhưng có giá trị tham chiếu quan trọng đối với việc thiết kế mô hình, quy trình và kỹ

thuật xây dựng khung pháp lý quốc gia. Trong đó, các tiêu chuẩn của Liên hợp quốc, OECD và Liên minh châu Âu thể hiện ba cách tiếp cận tiêu biểu: quản trị định hướng bằng luật mềm, quản trị bằng nguyên tắc và công cụ đo lường, và quản trị bằng pháp luật ràng buộc dựa trên rủi ro.

2.3.1.1. Liên hợp quốc

Liên hợp quốc (LHQ) là tổ chức có tính đại diện toàn cầu với 193 quốc gia thành viên, do đó các chuẩn mực do LHQ và các cơ quan chuyên môn của LHQ (đặc biệt là UNESCO) xây dựng có giá trị tham chiếu rộng rãi, bất kể sự khác biệt về bối cảnh chính trị, pháp lý và trình độ phát triển giữa các quốc gia. Tuy nhiên, cần nhận thức rõ rằng kiến trúc quản trị AI của LHQ không phải là khung pháp lý theo nghĩa truyền thống, bởi không có đạo luật AI ràng buộc chung, không có thiết chế thực thi tập trung và không có chế tài vi phạm trực tiếp. LHQ chủ yếu vận hành thông qua cơ chế luật mềm [193], bao gồm khuyến nghị, nghị quyết, nguyên tắc chỉ đạo và công cụ đánh giá năng lực thực hiện. Đây không phải là sự thiếu hụt, mà phản ánh cấu trúc phân quyền của trật tự quốc tế, trong đó thẩm quyền lập pháp vẫn thuộc về từng quốc gia.

Thứ nhất, về mô hình khung pháp lý, LHQ lựa chọn mô hình quản trị định hướng, phi ràng buộc. Đây là lựa chọn phù hợp trong bối cảnh không tồn tại một chủ thể siêu quốc gia có thẩm quyền ban hành luật AI bắt buộc đối với mọi quốc gia; nguyên tắc chủ quyền quốc gia hạn chế khả năng áp đặt chuẩn mực pháp lý thống nhất; đồng thời sự khác biệt lớn về hệ thống giá trị, năng lực công nghệ và ưu tiên phát triển giữa các quốc gia khiến việc đạt được đồng thuận về một công ước AI ràng buộc trở nên rất khó khăn.

Văn kiện nền tảng và có phạm vi bao quát nhất của LHQ hiện nay là Khuyến nghị về Đạo đức AI của UNESCO [356], được thông qua tháng 11/2021 với sự đồng thuận của toàn bộ 193 quốc gia thành viên. Đây là văn kiện chuẩn mực đa phương đầu tiên trên thế giới về đạo đức AI. Tiếp đó, năm 2024, LHQ ban hành Nghị quyết A/RES/78/311 về quản trị AI toàn cầu và Báo cáo "Quản trị AI vì Nhân loại" [359], trong đó đề xuất thành lập "Sàn giao dịch tiêu chuẩn AI" nhằm thúc đẩy khả năng tương thích kỹ thuật của các hệ thống AI xuyên biên giới. Global Digital Compact [358], được thông qua tại Hội nghị Thượng đỉnh vì Tương lai ngày 22/9/2024, tiếp tục đặt mục tiêu hình thành các tiêu chuẩn AI toàn cầu đến năm 2030, qua đó tạo nền tảng cho môi trường pháp lý AI quốc tế ổn định và có thể dự đoán hơn.

Thứ hai, về xây dựng khung pháp lý, các tiêu chuẩn của LHQ thể hiện rõ ở bốn phương diện. Một là, về chủ thể xây dựng, UNESCO tổ chức quy trình đa chủ thể với sự tham gia của Ban Thư ký, đại diện quốc gia thành viên, nhóm chuyên gia liên ngành về kỹ thuật AI, đạo đức, luật quốc tế, khoa học xã hội và xã hội dân sự toàn cầu. Đây là

quy trình có tính đại diện rộng, nhưng cũng đặt ra thách thức về chất lượng kỹ thuật do yêu cầu đồng thuận giữa 193 quốc gia.

Hai là, về nội dung xây dựng, Khuyến nghị UNESCO 2021 thiết lập cấu trúc gồm bốn giá trị, 10 nguyên tắc và 11 lĩnh vực chính sách. Cấu trúc này cho thấy cách tiếp cận ba tầng: giá trị định hướng, nguyên tắc điều chỉnh và lĩnh vực chính sách cụ thể. Trước khi soạn thảo, UNESCO cũng ủy thác các nghiên cứu nền về AI và đạo đức cho các trường đại học, viện nghiên cứu, qua đó bảo đảm cơ sở tri thức cho quá trình xây dựng chuẩn mực.

Ba là, về quy trình xây dựng, Khuyến nghị UNESCO 2021 được hình thành trong hơn hai năm, từ nghiên cứu cơ sở, soạn thảo bởi nhóm chuyên gia quốc tế, tham vấn khu vực và quốc gia, đến thông qua tại Hội nghị toàn thể UNESCO tháng 11/2021. Đây là quy trình tham vấn địa lý rộng, có phiên tham vấn riêng cho các khu vực như châu Phi, châu Á - Thái Bình Dương, Mỹ Latinh và Caribê, các nước Ả Rập và châu Âu.

Bốn là, về kỹ thuật xây dựng, LHQ và UNESCO sử dụng ba kỹ thuật đáng chú ý. Trước hết là kỹ thuật đánh giá mức độ sẵn sàng. UNESCO phát triển Phương pháp luận Đánh giá Mức độ Sẵn sàng (Readiness Assessment Methodology - RAM) [357], có thể coi là hình thức RIA đa phương dành cho các quốc gia thành viên. RAM không áp đặt một mô hình đánh giá duy nhất, mà cung cấp công cụ tự đánh giá có hướng dẫn, phù hợp với sự khác biệt về năng lực thể chế giữa các quốc gia. Hạn chế chính là thiếu cơ chế kiểm tra chéo hoặc xác minh độc lập kết quả tự đánh giá, làm giảm độ tin cậy so sánh giữa các quốc gia.

Tiếp theo là kỹ thuật đánh giá rủi ro và tác động. Khuyến nghị UNESCO 2021 coi đánh giá tác động đạo đức là công cụ cần thiết đối với hệ thống AI có rủi ro cao trước khi triển khai. Nghị quyết A/RES/78/265 [355] cũng kêu gọi các quốc gia xây dựng cơ chế đánh giá rủi ro AI phù hợp với bối cảnh quốc gia, thay vì áp dụng máy móc mô hình từ bên ngoài. Nghị quyết A/RES/78/311 [354] đồng thời nhấn mạnh rằng bất bình đẳng năng lực AI không thể được giải quyết chỉ bằng quy định pháp lý, mà cần đầu tư vào con người, hạ tầng và tri thức.

Sau cùng là kỹ thuật ủy quyền linh hoạt và miễn trừ theo năng lực. Trong điều kiện không có cơ quan lập pháp tập trung, UNESCO sử dụng cơ chế báo cáo định kỳ, hướng dẫn thực hiện cập nhật và các nhóm công tác chuyên đề để bổ sung nội dung kỹ thuật cho Khuyến nghị. Đồng thời, Khuyến nghị UNESCO 2021 cho phép quốc gia đang phát triển thực hiện từng phần phù hợp với năng lực, với hỗ trợ kỹ thuật từ UNESCO. Đây là kỹ thuật miễn trừ theo năng lực quốc gia, có giá trị tham chiếu đối với Việt Nam khi thiết kế cơ chế thực hiện nghĩa vụ AI trong mối quan hệ với các chuẩn mực quốc tế.

2.3.1.2. OECD

Hoạt động của OECD trong xây dựng khung pháp lý AI tập trung vào việc hình thành các nguyên tắc dựa trên giá trị, kết hợp với công cụ đo lường, phân loại và hỗ trợ chính sách. Khác với LHQ, OECD không có tính đại diện toàn cầu tuyệt đối, nhưng lại có ưu thế về năng lực phân tích chính sách, dữ liệu so sánh và khả năng tạo ra chuẩn mực có ảnh hưởng lớn đối với các quốc gia thành viên cũng như các đối tác ngoài OECD.

Thứ nhất, về mô hình khung pháp lý, OECD lựa chọn mô hình nguyên tắc và công cụ đo lường. Mô hình này không ràng buộc pháp lý trực tiếp, nhưng có ảnh hưởng thực tiễn lớn vì đã được 47 quốc gia thành viên và đối tác chấp thuận, được trích dẫn trong các văn bản như EU AI Act, AI Basic Act của Hàn Quốc, đồng thời được hỗ trợ bởi OECD.AI Policy Observatory - nền tảng dữ liệu so sánh quốc tế về chính sách AI của hơn 70 chính phủ.

Nền tảng của mô hình này là Nguyên tắc OECD về AI, văn kiện liên chính phủ đầu tiên về AI, được thông qua năm 2019 và cập nhật vào tháng 11/2023, tháng 5/2024 [292]. Các tài liệu tiếp theo như hướng dẫn triển khai AI đáng tin cậy [294], Khung phân loại hệ thống AI [295], Báo cáo về cơ chế thử nghiệm AI [296], sáng kiến GPAI hợp tác với OECD [299], chính sách về khung báo cáo chung cho sự cố AI [300] và khuyến nghị về "la bàn đạo đức" cho các tác nhân AI xuyên biên giới [301] cho thấy OECD không dừng lại ở việc nêu nguyên tắc, mà phát triển một hệ sinh thái công cụ phục vụ hoạch định và thực thi chính sách.

Thứ hai, về xây dựng khung pháp lý, các tiêu chuẩn của OECD thể hiện ở bốn phương diện. Một là, về chủ thể xây dựng, OECD dựa vào mạng lưới chuyên gia AI đa ngành, đại diện doanh nghiệp thông qua Business at OECD, đại diện người lao động thông qua Trade Union Advisory Committee và các cơ chế tham vấn chính sách khác. Tuy nhiên, hạn chế của mô hình này là đại diện từ các quốc gia đang phát triển ngoài OECD còn tương đối hạn chế, dẫn đến nguy cơ chuẩn mực phản ánh nhiều hơn quan điểm và năng lực của các quốc gia phát triển.

Hai là, về nội dung xây dựng, OECD xác lập năm nguyên tắc cốt lõi: tăng trưởng toàn diện và phát triển bền vững; quyền con người và giá trị dân chủ; minh bạch và giải thích được; an toàn và bảo mật; trách nhiệm giải trình. Bên cạnh đó là năm khuyến nghị chính sách: đầu tư vào nghiên cứu, nuôi dưỡng hệ sinh thái số, định hình môi trường chính sách thuận lợi, xây dựng năng lực con người và thúc đẩy hợp tác quốc tế.

Ba là, về quy trình xây dựng, OECD dựa trên nghiên cứu nền, tham vấn đa bên trong khoảng 18 tháng, thông qua ở cấp Hội đồng Bộ trưởng năm 2019 và cập nhật định kỳ dựa trên phản hồi thực tiễn, phát triển công nghệ. Điểm mạnh của quy trình

này là tính minh bạch, khả năng cập nhật và mức độ kết nối giữa nghiên cứu học thuật, dữ liệu chính sách và quyết định liên chính phủ.

Bốn là, về kỹ thuật xây dựng, OECD đặc biệt đáng chú ý ở kỹ thuật RIA và kỹ thuật định nghĩa chức năng. Về RIA, OECD nhấn mạnh yêu cầu đánh giá tác động trong lĩnh vực AI phải tính đến các kịch bản phát triển công nghệ khác nhau, bao gồm kịch bản đổi mới nhanh và kịch bản phát triển chậm, nhằm nhận diện những điều khoản có nguy cơ lỗi thời sớm và cần cơ chế ủy quyền linh hoạt. Về định nghĩa chức năng, OECD nhất quán tiếp cận AI theo chức năng của hệ thống thay vì theo kiến trúc kỹ thuật cụ thể. Việc phân biệt "AI system" và "agentic AI" theo chức năng tự chủ thực hiện chuỗi hành động cho thấy kỹ thuật định nghĩa này có khả năng thích ứng tốt hơn với sự biến đổi công nghệ.

2.3.1.3. Liên minh châu Âu

Liên minh châu Âu (EU) là chủ thể có khung pháp lý AI toàn diện và ràng buộc pháp lý rõ nhất hiện nay. EU AI Act được Nghị viện châu Âu thông qua ngày 13/3/2024 và Hội đồng phê chuẩn ngày 21/5/2024 [205], có hiệu lực từ ngày 01/8/2024 và được áp dụng theo lộ trình nhiều giai đoạn: các hành vi AI bị cấm và nghĩa vụ về năng lực AI có hiệu lực từ ngày 02/2/2025; quy tắc quản trị và nghĩa vụ đối với mô hình GPAI có hiệu lực từ ngày 02/8/2025; còn một số quy tắc đối với hệ thống AI rủi ro cao nhúng vào sản phẩm được quản lý có thời gian chuyển tiếp đến ngày 02/8/2027 [195].

Thứ nhất, về mô hình khung pháp lý, EU lựa chọn mô hình điều chỉnh tập trung, toàn diện và dựa trên rủi ro. Lựa chọn này phản ánh ba yếu tố bối cảnh: truyền thống civil law coi trọng quy tắc rõ ràng; nhu cầu thống nhất thị trường nội khối giữa 27 quốc gia thành viên; và chiến lược định vị EU như chủ thể thiết lập tiêu chuẩn toàn cầu về AI. EU AI Act phân loại AI thành bốn tầng rủi ro: bị cấm, rủi ro cao, rủi ro hạn chế và rủi ro tối thiểu, tương ứng với các chế độ pháp lý khác nhau. AI Act không tồn tại đơn lẻ mà gắn với các công cụ pháp lý khác của EU, trong đó có Chỉ thị Bản quyền thị trường đơn kỹ thuật số [204] và cơ chế khai thác văn bản, dữ liệu để huấn luyện AI [203]. Về thiết chế quản lý, AI Act thiết lập Văn phòng AI châu Âu [196], AI Board ở cấp EU và cơ quan AI tại từng quốc gia thành viên. Lộ trình áp dụng theo giai đoạn cho phép EU xây dựng năng lực thực thi trước khi các nghĩa vụ đầy đủ có hiệu lực. Điểm mạnh của mô hình EU là tính toàn diện, chuỗi trách nhiệm rõ ràng, khả năng tạo "Hiệu ứng Brussels" [151] và ưu tiên bảo vệ quyền con người. Điểm hạn chế là chi phí tuân thủ cao, quy trình đánh giá sự phù hợp phức tạp, danh mục rủi ro cao có nguy cơ lỗi thời và mức độ phù hợp chưa cao đối với hệ sinh thái doanh nghiệp nhỏ và vừa.

Thứ hai, về xây dựng khung pháp lý, kinh nghiệm của EU thể hiện ở bốn phương diện: Một là, về chủ thể xây dựng, Ủy ban châu Âu giữ vai trò đề xuất và soạn

thảo, AI HLEG gồm 52 chuyên gia đa ngành xây dựng nền tảng giá trị, Nghị viện châu Âu bổ sung nhiều nội dung quan trọng, Hội đồng châu Âu đại diện cho lợi ích quốc gia thành viên, trong khi các tổ chức xã hội dân sự và doanh nghiệp công nghệ tham gia vận động, phản biện và tham vấn. Bộ Quy tắc Thực hành GPAI [197] cũng được xây dựng thông qua quy trình mở với sự tham gia của gần một nghìn bên liên quan.

Hai là, về nội dung xây dựng, EU triển khai theo các giai đoạn tương đối rõ: giai đoạn chuẩn bị 2018-2020 hình thành nền tảng giá trị thông qua AI HLEG và Sách trắng; giai đoạn soạn thảo năm 2021 với dự thảo gồm 84 điều, 13 phụ lục và Báo cáo đánh giá tác động; giai đoạn thương lượng 2022-2023 bổ sung nội dung về GPAI; và giai đoạn hoàn thiện, ban hành năm 2024 với lộ trình áp dụng nhiều giai đoạn.

Ba là, về quy trình xây dựng, EU mất khoảng sáu năm từ ý tưởng đến ban hành, thể hiện quy trình dài nhưng có chiều sâu. Hai vòng tham vấn công khai, các báo cáo tổng hợp phản hồi và các phiên đàm phán giữa Ủy ban, Nghị viện và Hội đồng cho thấy quá trình xây dựng AI Act không chỉ là soạn thảo kỹ thuật, mà là quá trình thương lượng chính trị - pháp lý phức tạp giữa nhiều nhóm lợi ích và giá trị.

Bốn là, về kỹ thuật xây dựng, EU AI Act cung cấp nhiều kinh nghiệm quan trọng. Trước hết là kỹ thuật RIA. Trước khi soạn thảo AI Act, Ủy ban châu Âu thực hiện báo cáo đánh giá tác động ngày 21/4/2021, phân tích ba lựa chọn chính sách: tiếp cận tự nguyện, điều chỉnh theo ngành và khung quy định ngang dựa trên rủi ro. Báo cáo đánh giá chi phí tuân thủ, tác động đến đổi mới và mức độ bảo vệ người dùng, qua đó tạo căn cứ cho lựa chọn mô hình dựa trên rủi ro.

Tiếp theo là kỹ thuật định nghĩa chức năng. Đề xuất ban đầu năm 2021 liệt kê các nhóm kỹ thuật AI cụ thể, nhưng cách tiếp cận này bị phê phán vì có nguy cơ bỏ sót công nghệ mới và tạo cơ hội né tránh điều chỉnh. Văn bản cuối cùng năm 2024 chuyển sang định nghĩa chức năng tại Điều 3(1). Theo phân tích của Veale và Borgesius, đây là một trong những sửa đổi kiến trúc quan trọng nhất trong quá trình đàm phán AI Act.

Thứ ba là kỹ thuật ủy quyền linh hoạt phân tầng. AI Act thiết kế ba cấp: luật quy định nguyên tắc và danh mục cấm ổn định; Ủy ban châu Âu có thẩm quyền cập nhật phụ lục danh mục AI rủi ro cao theo quy trình rút gọn; và EU AI Office ban hành hướng dẫn kỹ thuật có thể cập nhật linh hoạt. Điều 112 yêu cầu Ủy ban định kỳ đánh giá và báo cáo Nghị viện, Hội đồng châu Âu ba năm một lần, qua đó biến báo cáo thực thi thành cơ chế kích hoạt sửa đổi có căn cứ.

Thứ tư là kỹ thuật phân định nghĩa vụ theo chuỗi giá trị. EU AI Act phân biệt nghĩa vụ của nhà cung cấp, nhà triển khai và các chủ thể khác trong chuỗi giá trị AI. Nhà cung cấp phải quản lý rủi ro, kiểm định trước khi đưa ra thị trường, duy trì tài liệu kỹ thuật và đăng ký hệ thống; nhà triển khai phải sử dụng đúng mục đích, giám sát hoạt

động, báo cáo sự cố và thông báo cho người dùng. Điều 25 quy định trường hợp nhà triển khai có thể phải gánh thêm nghĩa vụ của nhà cung cấp nếu sửa đổi hệ thống vượt quá phạm vi sử dụng đã được kiểm định. Điều 28 xử lý vấn đề AI xuyên biên giới bằng cách bổ sung nghĩa vụ cho nhà triển khai tại EU khi nhà cung cấp ngoài EU đưa sản phẩm vào thị trường EU.

Thứ năm là kỹ thuật miễn trừ tỷ lệ. EU AI Act cho thấy nếu kỹ thuật miễn trừ không đủ mạnh, chi phí tuân thủ có thể trở thành rào cản đối với doanh nghiệp nhỏ và vừa. Mặc dù Recital 49 và Điều 11(3) có một số điều chỉnh cho SME, yêu cầu kiểm định bên thứ ba đối với nhiều hệ thống AI rủi ro cao vẫn có thể tạo gánh nặng đáng kể. Nghiên cứu của Mueller [159] ước tính chi phí tuân thủ đối với SME AI châu Âu có thể lên đến 400.000 EUR cho một hệ thống, vượt quá khả năng của nhiều startup.

Cuối cùng là kỹ thuật luật hóa tiêu chuẩn kỹ thuật. Điều 40 AI Act thiết kế cơ chế ủy quyền cho CEN-CENELEC xây dựng tiêu chuẩn hài hòa; khi được Ủy ban châu Âu chỉ định, việc tuân thủ tiêu chuẩn tạo ra suy đoán phù hợp với yêu cầu pháp lý tương ứng. EU không viện dẫn trực tiếp từng tiêu chuẩn trong văn bản luật, mà ủy quyền cập nhật danh sách tiêu chuẩn trên Official Journal of the EU. Cách làm này kết hợp kỹ thuật ủy quyền linh hoạt với kỹ thuật pháp luật hóa tiêu chuẩn, giúp hệ thống có thể cập nhật mà không cần sửa Quy định gốc. Gói Kỹ thuật số về Đơn giản hóa tiếp tục hướng đến việc làm cho AI Act rõ ràng, đơn giản và thân thiện hơn với đổi mới sáng tạo [200].

2.3.2. Kinh nghiệm của một số quốc gia về xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo

Trong luận án này, các quốc gia được lựa chọn nghiên cứu gồm Hoa Kỳ, Trung Quốc, Nhật Bản, Hàn Quốc và Singapore vì đại diện cho những mô hình xây dựng khung pháp lý AI khác nhau, đồng thời có giá trị tham chiếu trực tiếp cho Việt Nam. Hoa Kỳ đại diện cho mô hình phân tán, dựa nhiều vào tiêu chuẩn kỹ thuật và tự điều chỉnh; Trung Quốc tiêu biểu cho mô hình nhà nước định hướng mạnh, điều chỉnh theo từng công nghệ cụ thể; Nhật Bản thể hiện mô hình linh hoạt dựa trên luật mềm và cơ chế đồng thuận; Hàn Quốc là ví dụ về mô hình luật khung kết hợp thúc đẩy công nghiệp với kiểm soát rủi ro; còn Singapore đại diện cho cách tiếp cận thực dụng, dựa trên công cụ tự đánh giá, sandbox và hợp tác công - tư. Việc khảo cứu các mô hình này giúp rút ra bài học về lựa chọn kiến trúc, quy trình và kỹ thuật xây dựng khung pháp lý AI phù hợp với điều kiện của Việt Nam.

2.3.2.1. Hoa Kỳ

Hoa Kỳ được lựa chọn vì là trung tâm AI hàng đầu thế giới, đứng đầu chỉ số sẵn sàng AI năm 2025 [306], có thị trường AI lớn [217] và tập trung nhiều công ty AI hàng

đầu toàn cầu [158]. Về mặt so sánh, Hoa Kỳ còn có giá trị đặc biệt vì đại diện cho mô hình đối lập với EU: không có luật AI liên bang toàn diện, mà dựa vào tiêu chuẩn kỹ thuật tự nguyện, sắc lệnh hành pháp, quy định chuyên ngành và tự điều chỉnh của doanh nghiệp. Kinh nghiệm của Hoa Kỳ có thể khái quát như sau:

Thứ nhất, về mô hình khung pháp lý, Hoa Kỳ theo đuổi mô hình phân tán. Ở cấp liên bang, không có đạo luật AI thống nhất; các cơ quan như FTC, FDA, SEC, CFPB và các cơ quan chuyên ngành áp dụng thẩm quyền hiện có để xử lý các vấn đề liên quan đến AI. Ở tầng tiêu chuẩn kỹ thuật, NIST AI Risk Management Framework 1.0 là công cụ trung tâm, có tính tự nguyện nhưng ảnh hưởng lớn trong thực tiễn quản trị AI. Ở tầng tiểu bang, làn sóng lập pháp AI phát triển mạnh: năm 2025, tất cả 50 tiểu bang, Puerto Rico, Quần đảo Virgin và Washington, D.C. đều đã đệ trình dự luật liên quan đến AI [280]. Một số bang như California, Texas và Colorado đã ban hành quy định riêng về minh bạch AI, giám sát AI rủi ro cao hoặc chống phân biệt đối xử thuật toán [166], [341], [173].

Sự thay đổi chính sách AI ở Hoa Kỳ cũng cho thấy tính bất ổn của mô hình dựa nhiều vào hành pháp. Năm 2025, Tổng thống Trump ban hành Sắc lệnh số 14179 nhằm gỡ bỏ rào cản đối với vị thế lãnh đạo AI của Hoa Kỳ [351], thu hồi định hướng của Sắc lệnh số 14110 thời Tổng thống Biden. Tiếp đó, Sắc lệnh ngày 11/12/2025 về khung chính sách quốc gia cho AI tiếp tục nhấn mạnh mục tiêu duy trì ưu thế AI toàn cầu và giảm gánh nặng điều chỉnh [350]. Điều này cho thấy ưu điểm của sắc lệnh hành pháp là linh hoạt, nhưng nhược điểm là thiếu ổn định và dễ thay đổi theo nhiệm kỳ chính trị.

Thứ hai, về xây dựng khung pháp lý, chủ thể quan trọng nhất không phải là Quốc hội mà là NIST, với vai trò thiết lập tiêu chuẩn kỹ thuật liên bang. Quốc hội có vai trò tương đối hạn chế, chủ yếu thông qua các ủy ban nghiên cứu và khuyến nghị chính sách. Khu vực tư nhân có ảnh hưởng rất lớn, đặc biệt là các công ty như OpenAI, Google, Microsoft, Meta, Anthropic, vốn tham gia vào các cam kết tự nguyện và quá trình định hình tiêu chuẩn thực hành.

Thứ ba, về nội dung xây dựng, NIST AI RMF 1.0 [284] là công cụ hướng dẫn quan trọng nhất, được tổ chức theo bốn chức năng: Govern, Map, Measure và Manage. Khung này tập trung vào quản lý rủi ro trong toàn bộ vòng đời hệ thống AI, thiên về hỗ trợ doanh nghiệp nhận diện, đo lường và xử lý rủi ro hơn là xây dựng một chế độ quyền công dân theo kiểu EU. Năm 2024, NIST tiếp tục phát triển hồ sơ riêng cho AI tạo sinh; năm 2025, Kế hoạch hành động AI quốc gia đặt NIST vào trung tâm của nhiều nhiệm vụ chính sách [349]. Đạo luật liên bang chuyên biệt đáng chú ý là TAKE IT DOWN Act năm 2025, hình sự hóa hành vi phát tán hình ảnh thân mật không có sự

đồng thuận và đặt nghĩa vụ gỡ bỏ nội dung đối với nền tảng [346]. Hoa Kỳ cũng xử lý mạnh tay các hành vi AI xâm phạm an ninh quốc gia [285].

Thứ tư, về kỹ thuật xây dựng, Hoa Kỳ có một số kinh nghiệm đáng chú ý. Một là, kỹ thuật xây dựng tiêu chuẩn mở: AI RMF được phát triển thông qua RFI, dự thảo công khai, hội thảo và phản hồi đa bên; NIST dự kiến rà soát toàn diện để xây dựng phiên bản 2.0. Hai là, kỹ thuật định nghĩa chức năng, thể hiện ở cách NIST tiếp cận AI theo đặc điểm vận hành hơn là liệt kê công nghệ cụ thể. Ba là, kỹ thuật ủy quyền linh hoạt thông qua sắc lệnh hành pháp, tuy nhanh nhưng thiếu ổn định. Bốn là, kỹ thuật miễn trừ tỷ lệ, tiêu biểu như California SB 942 chỉ áp dụng nghĩa vụ tăng cường đối với các chủ thể rất lớn. Năm là, kỹ thuật luật hóa tiêu chuẩn ở cấp tiểu bang, như TRAIGA cho phép doanh nghiệp viện dẫn việc tuân thủ NIST AI RMF như một dạng "safe harbor".

Kinh nghiệm Hoa Kỳ cho thấy một khung pháp lý AI có thể vận hành mạnh nhờ tiêu chuẩn kỹ thuật, năng lực tư nhân và cơ quan chuyên môn, nhưng nếu thiếu luật liên bang ổn định và cơ chế điều phối chung, hệ thống dễ phân mảnh, thay đổi nhanh theo chính trị và gây khó khăn cho bảo đảm thống nhất pháp lý.

2.3.2.2. Trung Quốc

Trung Quốc được lựa chọn vì có nhiều điểm tương đồng với Việt Nam về truyền thống pháp luật thành văn, mô hình nhà nước định hướng mạnh và chiến lược phát triển công nghệ quốc gia. Đồng thời, Trung Quốc là một trong những quốc gia ban hành quy định thực định về AI sớm nhất, đặc biệt trong các lĩnh vực thuật toán khuyến nghị, deep synthesis và AI tạo sinh.

Thứ nhất, về mô hình khung pháp lý, Trung Quốc xây dựng mô hình tích lũy theo công nghệ cụ thể. Thay vì ban hành một đạo luật AI chung ngay từ đầu, Trung Quốc ban hành các quy định riêng cho từng công nghệ hoặc nhóm rủi ro mới phát sinh, đặt trong định hướng chiến lược của Kế hoạch Phát triển AI Thế hệ mới năm 2017 [337]. Cách tiếp cận này phản ánh đặc trưng quản trị của Trung Quốc: nhà nước định hướng mạnh phát triển công nghệ, đồng thời kiểm soát chặt các ứng dụng có khả năng tác động đến dư luận xã hội, an ninh dữ liệu và an ninh quốc gia.

Khung pháp lý AI Trung Quốc hiện gồm nhiều tầng. Tầng pháp luật nền bao gồm Luật An ninh mạng, Luật An toàn dữ liệu và Luật Bảo vệ thông tin cá nhân. Tầng quy định theo công nghệ gồm Quy định về thuật toán khuyến nghị [162], Quy định về deep synthesis [163], Biện pháp tạm thời về dịch vụ AI tạo sinh [312] và Biện pháp gắn nhãn nội dung do AI tạo sinh năm 2025 [164], bao gồm cả nhãn hiển thị và nhãn ẩn trong metadata [150]. Tầng tiêu chuẩn kỹ thuật được phát triển thông qua hướng dẫn xây dựng hệ thống tiêu chuẩn AI thế hệ mới [276], các tiêu chuẩn về học máy [281], đạo đức thuật toán [282], ghi nhãn dữ liệu [283] và các tiêu chuẩn an toàn AI tạo sinh.

Ngoài ra, Trung Quốc còn thiết lập cơ chế đăng ký thuật toán, nhất là đối với các thuật toán có "tính chất dư luận hoặc khả năng vận động xã hội"; đến tháng 10/2025, hàng nghìn hồ sơ đăng ký đã được phê duyệt [226].

Thứ hai, về xây dựng khung pháp lý, chủ thể trung tâm là Cục Quản lý Không gian mạng Trung Quốc (CAC), phối hợp với nhiều cơ quan khác như Ủy ban Cải cách và Phát triển Quốc gia, Bộ Giáo dục, Bộ Khoa học và Công nghệ, Bộ Công nghiệp và Công nghệ thông tin, Bộ Công an và Cục Phát thanh Truyền hình Quốc gia. Đây là mô hình điều phối đa cơ quan theo lĩnh vực quản lý, không dựa vào một cơ quan AI độc lập. Đáng chú ý, TC260 đóng vai trò cầu nối giữa quy định pháp lý và tiêu chuẩn kỹ thuật, thể hiện ở việc công bố danh sách tiêu chuẩn dự kiến để hỗ trợ thực thi các quy định về AI tạo sinh.

Thứ ba, về kỹ thuật xây dựng, Trung Quốc thể hiện một số đặc điểm rõ nét. Một là, quy trình tham vấn ngắn, ưu tiên tốc độ hơn phân tích tác động đầy đủ; ví dụ Biện pháp AI tạo sinh được đưa ra lấy ý kiến tháng 4/2023 và ban hành tháng 7/2023. Hai là, kỹ thuật định nghĩa theo tên công nghệ, như "deep synthesis", "thuật toán khuyến nghị", "AI tạo sinh"; cách này giúp phản ứng nhanh nhưng dễ lỗi thời khi công nghệ mới xuất hiện. Ba là, kỹ thuật ủy quyền theo cơ quan chủ trì ngành, tạo linh hoạt nhưng có nguy cơ thiếu điều phối tổng thể. Bốn là, kỹ thuật phân định nghĩa vụ chủ yếu theo "nhà cung cấp dịch vụ", đặt nghĩa vụ lớn lên chủ thể cung cấp dịch vụ AI cho người dùng cuối. Năm là, miễn trừ theo phạm vi cung cấp cho công chúng, theo đó dịch vụ nội bộ doanh nghiệp thường không thuộc phạm vi điều chỉnh. Sáu là, luật hóa tiêu chuẩn kỹ thuật nội địa, trong đó Trung Quốc phát triển hệ thống tiêu chuẩn GB/T riêng thay vì chỉ nội luật hóa tiêu chuẩn ISO/IEC.

Kinh nghiệm Trung Quốc có giá trị tham khảo đối với Việt Nam ở ba điểm: khả năng phản ứng nhanh trước công nghệ mới; sự kết nối giữa quy định pháp lý, tiêu chuẩn kỹ thuật và đăng ký thuật toán; và vai trò điều phối mạnh của nhà nước. Tuy nhiên, cách tiếp cận theo từng công nghệ cụ thể cũng cho thấy rủi ro phân mảnh, chồng chéo và khó thích ứng với các hệ thống AI đa năng.

2.3.2.3. Nhật Bản

Nhật Bản được lựa chọn vì đại diện cho mô hình quản trị AI linh hoạt, thân thiện, thúc đẩy bằng khuôn khổ, định hướng bằng hướng dẫn và uy tín. Nhật Bản cũng có truyền thống civil law tương đồng nhất định với Việt Nam và có nhiều kinh nghiệm về sandbox pháp lý, đặc biệt trong các lĩnh vực y tế, giao thông tự hành và nông nghiệp.

Thứ nhất, về mô hình khung pháp lý, Đạo luật về Thúc đẩy Nghiên cứu, Phát triển và Ứng dụng Công nghệ liên quan đến AI năm 2025 [311] đặt trong hệ sinh thái liên kết với Đạo luật Cơ bản về Khoa học, Công nghệ và Đổi mới sáng tạo và Đạo luật

Cơ bản về Hình thành Xã hội Số. Sự ra đời của Đạo luật AI đánh dấu bước chuyển đáng kể từ cách tiếp cận quản trị AI dựa trên hướng dẫn mềm sang khuôn khổ lập pháp chính thức. Ngoài ra, Kế hoạch Cơ bản AI của Nhật Bản là tài liệu chiến lược mô tả cụ thể cách Chính phủ triển khai các nguyên tắc trong Đạo luật thành hành động. Hướng dẫn AI cho doanh nghiệp của Bộ Kinh tế, Thương mại và Công nghiệp (METI) năm 2024, hướng dẫn chuyên ngành của các bộ, Hội đồng AI quốc gia thành lập năm 2023 và sandbox pháp lý trong nhiều lĩnh vực. Đặc trưng kiến trúc phân tầng từ rộng đến hẹp: Đạo luật (nguyên tắc) → Kế hoạch Cơ bản (chiến lược) → Hướng dẫn (thực hành) → Pháp luật ngành (ràng buộc có chọn lọc). Định vị quốc tế là trọng tâm, Đạo luật AI Nhật Bản được thiết kế để tương thích với OECD AI Principles, G7 Hiroshima Process và chuẩn mực quốc tế.

Thứ hai, về xây dựng khung pháp lý, các chủ thể chính gồm Ban Chỉ đạo Chiến lược AI, METI trong lĩnh vực công nghiệp và dịch vụ, Bộ Nội vụ và Truyền thông (MIC) trong lĩnh vực truyền thông, các cơ quan quản lý chuyên ngành, Hội đồng AI quốc gia, doanh nghiệp và các viện nghiên cứu như RIKEN, NII. Quy trình xây dựng chịu ảnh hưởng của phương thức đồng thuận nemawashi: tham vấn kỹ trước khi ban hành hướng dẫn, duy trì hội đồng nghiên cứu trong từng bộ và cập nhật hướng dẫn theo chu kỳ ngắn. Ưu điểm là linh hoạt và có tính chấp nhận cao; hạn chế là mức độ minh bạch không phải lúc nào cũng tương đương quy trình lập pháp chính thức.

Thứ ba, về kỹ thuật xây dựng, Nhật Bản có một số điểm đáng chú ý. Một là, thay vì RIA một lần trước ban hành, Nhật Bản dựa vào khảo sát thực tiễn thường xuyên về triển khai AI trong doanh nghiệp, từ đó cập nhật hướng dẫn theo chu kỳ 12-18 tháng. Hai là, phạm vi điều chỉnh được xác định theo tác động của hệ thống đối với con người, thay vì theo đặc trưng kỹ thuật của AI. Ba là, cơ chế cập nhật hướng dẫn không qua thủ tục lập pháp cho phép METI, MIC và các bộ chuyên ngành điều chỉnh nhanh khi công nghệ thay đổi. Bốn là, nghĩa vụ được phân định theo ngành, phù hợp với đặc thù từng lĩnh vực nhưng có thể gặp khó khăn khi AI hoạt động liên ngành. Năm là, cơ chế comply-or-explain cho phép doanh nghiệp giải thích lý do không áp dụng một yêu cầu cụ thể và đề xuất biện pháp thay thế, kết hợp với "nêu tên và xấu hổ" (*name and shame*) khi doanh nghiệp không tuân thủ hướng dẫn và không có giải thích tương đương, Chính phủ có thể yêu cầu cung cấp thông tin và nếu không được đáp ứng, công khai tên doanh nghiệp vi phạm.

Kinh nghiệm Nhật Bản cho thấy giai đoạn chẩn đoán công khai trước soạn thảo có tác dụng tạo đồng thuận, mô hình Ban Chỉ đạo đa Bộ trong Văn phòng Thủ tướng là hình thức điều phối phù hợp với bối cảnh không muốn thêm thiết chế mới. Vòng phản hồi PDCA (Plan-Do-Check-Act) được thiết kế vào thiết chế là điều kiện để khung pháp

lý AI không lỗi thời. Tận dụng pháp luật ngành hiện có mang lại hiệu quả và tiết kiệm nguồn lực lập pháp.

2.3.2.4. Hàn Quốc

Hàn Quốc được lựa chọn vì AI Basic Act có cấu trúc và triết lý gần với định hướng xây dựng luật AI của Việt Nam vừa thúc đẩy phát triển công nghiệp AI, vừa thiết lập cơ chế quản lý rủi ro và xây dựng niềm tin. Hàn Quốc cũng có truyền thống civil law, mô hình nhà nước phát triển và hệ sinh thái công nghệ đang mở rộng, nên có giá trị tham chiếu trực tiếp cho Việt Nam.

Thứ nhất, về mô hình khung pháp lý, Hàn Quốc xây dựng mô hình luật khung phát triển AI. Khác với EU, Hàn Quốc không chỉ tập trung kiểm soát rủi ro mà còn chủ động kiến tạo hệ sinh thái AI quốc gia. AI Basic Act phân loại AI rủi ro cao, thiết lập cơ chế tự đánh giá có kiểm toán định kỳ, thành lập Ủy ban AI quốc gia, luật hóa quỹ đầu tư AI, quy định yêu cầu minh bạch trong các quyết định có ảnh hưởng lớn và thiết kế cơ chế cập nhật danh mục thông qua nghị định.

Thứ hai, về xây dựng khung pháp lý, Bộ Khoa học và ICT giữ vai trò chủ trì, phối hợp với Bộ Tư pháp, Bộ Kinh tế, Bộ Y tế và Ủy ban Bảo vệ Thông tin cá nhân. Quốc hội Hàn Quốc tham gia sâu thông qua Tiểu ban đặc biệt về AI. Cộng đồng startup được tham vấn riêng về điều khoản miễn trừ tỷ lệ, còn các viện nghiên cứu công lập như ETRI, KISTEP cung cấp nghiên cứu nền. Quy trình xây dựng kéo dài từ các dự luật đầu tiên năm 2020, qua nhiều dự luật chưa được thông qua, đến dự luật hợp nhất năm 2024 và việc chuẩn bị văn bản dưới luật sau khi AI Basic Act được ban hành.

Thứ ba, về kỹ thuật xây dựng, Hàn Quốc có bốn kinh nghiệm nổi bật. Một là, định nghĩa AI và hệ thống AI theo hành vi, chức năng, đồng thời giao Ủy ban AI quốc gia quyền diễn giải khi xuất hiện công nghệ mới. Hai là, phân định nghĩa vụ theo chuỗi giá trị, nhưng giảm nghĩa vụ cho nhà triển khai khi sử dụng nền tảng AI đã được chứng nhận. Ba là, thiết kế miễn trừ tỷ lệ tương đối toàn diện: doanh nghiệp siêu nhỏ và startup mới thành lập được miễn kiểm định bên thứ ba, doanh nghiệp nhỏ được kiểm định đơn giản hóa, doanh nghiệp lớn áp dụng kiểm định đầy đủ. Bốn là, luật hóa tiêu chuẩn kỹ thuật thông qua tiêu chuẩn AI quốc gia và tiêu chuẩn ngành, tạo cơ sở suy đoán tuân thủ trong từng lĩnh vực.

Kinh nghiệm Hàn Quốc đặc biệt hữu ích cho Việt Nam ở ba khía cạnh: cách kết hợp giữa thúc đẩy phát triển và kiểm soát rủi ro; thiết kế cơ quan điều phối liên ngành; và cơ chế miễn trừ tỷ lệ cho doanh nghiệp nhỏ, startup, nhằm tránh biến tuân thủ pháp lý thành rào cản đổi mới sáng tạo.

2.3.2.5. Singapore

Singapore được lựa chọn vì đại diện cho mô hình quản trị AI thực dụng, cởi mở và thân thiện với đổi mới. Đây cũng là quốc gia ASEAN có bối cảnh gần gũi với Việt

Nam, đồng thời có kinh nghiệm nổi bật về công cụ tự đánh giá, sandbox và hợp tác công - tư trong xây dựng tiêu chuẩn AI.

Thứ nhất, về mô hình khung pháp lý, Singapore không ban hành luật AI cứng mà xây dựng hệ thống quản trị tinh vi dựa trên luật mềm, công cụ thực hành và cơ chế thử nghiệm. Các thành tố chính gồm Model AI Governance Framework, AI Verify, hướng dẫn ngành của Monetary Authority of Singapore (MAS) trong tài chính, hướng dẫn của Ministry of Health (MOH) trong y tế và sandbox pháp lý. Logic trung tâm của Singapore là giảm gánh nặng cho nhà nước bằng cách trao cho doanh nghiệp công cụ tự đánh giá, trong khi cơ quan quản lý thực hiện kiểm tra mẫu và can thiệp theo rủi ro.

Thứ hai, về xây dựng khung pháp lý, IMDA giữ vai trò dẫn dắt AI trong nền kinh tế số, MAS dẫn dắt AI trong tài chính, MOH phụ trách AI y tế. Điểm nổi bật nhất là quá trình phát triển AI Verify thông qua hợp tác công - tư thực chất giữa IMDA và hơn 60 doanh nghiệp, trong đó có các công ty công nghệ, ngân hàng và doanh nghiệp dịch vụ lớn. Đây là mô hình đồng thiết kế đáng chú ý, trong đó công cụ quản trị được thử nghiệm với doanh nghiệp trước khi chính thức áp dụng rộng rãi.

Thứ ba, về nội dung và quy trình xây dựng, AI Verify chuyển các nguyên tắc quản trị AI thành bài kiểm tra kỹ thuật có thể tự động hóa. Công cụ này có tính mã nguồn mở, có thể chia sẻ với các quốc gia ASEAN và hỗ trợ doanh nghiệp tự kiểm tra mức độ tuân thủ. Hướng dẫn của MAS trong lĩnh vực tài chính được thiết kế theo ba lớp: nguyên tắc, kỳ vọng quản lý và ví dụ thực hành. Sandbox có tiêu chí tham gia, thời hạn xét duyệt, nghĩa vụ báo cáo và cơ chế học hỏi chính sách tương đối rõ ràng.

Thứ tư, về kỹ thuật xây dựng, Singapore có sáu điểm đáng chú ý. Một là, kiểm thử thực địa trước khi ban hành thay cho RIA thuần túy trên giấy; AI Verify được thử nghiệm với hơn 60 doanh nghiệp trong khoảng 18 tháng trước khi ra mắt. Hai là, định nghĩa phạm vi áp dụng được tích hợp vào bộ câu hỏi kiểm thử, giúp doanh nghiệp xác định hệ thống của mình có thuộc phạm vi điều chỉnh hay không. Ba là, cơ chế cập nhật dựa trên nhóm đồng thiết kế thường trực, gồm cơ quan quản lý, doanh nghiệp, học thuật và xã hội dân sự. Bốn là, phân định nghĩa vụ thông qua công cụ tự xác định vai trò của doanh nghiệp trong chuỗi giá trị AI. Năm là, phổ cập công cụ tự đánh giá miễn phí cho mọi quy mô doanh nghiệp, kết hợp kiểm tra mẫu dựa trên rủi ro. Sáu là, AI Verify được thiết kế để phản ánh các nguyên tắc OECD và AI Governance Framework, hướng tới khả năng công nhận lẫn nhau với các thị trường khác.

Kinh nghiệm Singapore có giá trị đặc biệt đối với Việt Nam vì cho thấy một quốc gia có quy mô nhỏ hơn và nguồn lực quản lý hữu hạn vẫn có thể xây dựng khung quản trị AI hiệu quả nếu biết kết hợp luật mềm, công cụ kỹ thuật, sandbox và đồng thiết kế công - tư. Đây là hướng tiếp cận phù hợp để Việt Nam tham khảo trong giai đoạn đầu xây dựng năng lực quản trị AI.

2.3.3. Giá trị tham khảo cho Việt Nam

Từ việc phân tích các chuẩn mực quốc tế và kinh nghiệm của một số quốc gia về xây dựng khung pháp lý cho phát triển, ứng dụng AI, có thể rút ra một số giá trị tham khảo quan trọng cho Việt Nam. Các giá trị này không nên được hiểu như những mô hình có thể sao chép nguyên trạng, mà là cơ sở để Việt Nam lựa chọn, điều chỉnh và thiết kế khung pháp lý phù hợp với điều kiện chính trị, pháp lý, kinh tế, công nghệ và năng lực thực thi trong nước.

2.3.3.1. Tác động của các tiêu chuẩn quốc tế đối với Việt Nam

Thứ nhất, các tiêu chuẩn quốc tế về AI, đặc biệt từ Liên hợp quốc, UNESCO, OECD và các tổ chức tiêu chuẩn kỹ thuật quốc tế, tạo ra nền tảng chuẩn mực quan trọng cho việc xác định giá trị, nguyên tắc và mục tiêu của khung pháp lý AI ở Việt Nam. Dù phần lớn các chuẩn mực này không có giá trị ràng buộc trực tiếp, chúng góp phần định hình nhận thức chung về AI có trách nhiệm, trong đó các yêu cầu về bảo vệ quyền con người, minh bạch, an toàn, công bằng, trách nhiệm giải trình và phát triển bao trùm ngày càng trở thành chuẩn mực phổ biến.

Thứ hai, các tiêu chuẩn quốc tế giúp Việt Nam bảo đảm tính tương thích của khung pháp lý quốc gia với môi trường pháp lý và công nghệ toàn cầu. AI là công nghệ xuyên biên giới; dữ liệu, mô hình, hạ tầng tính toán và dịch vụ AI thường không giới hạn trong phạm vi lãnh thổ quốc gia. Vì vậy, nếu khung pháp lý của Việt Nam quá tách biệt với các chuẩn mực quốc tế, doanh nghiệp trong nước có thể gặp khó khăn khi tham gia chuỗi giá trị AI toàn cầu, còn cơ quan quản lý có thể gặp trở ngại trong hợp tác quốc tế, chuyển giao công nghệ, công nhận tiêu chuẩn và xử lý rủi ro xuyên biên giới.

Thứ ba, các chuẩn mực quốc tế cung cấp công cụ tham chiếu để Việt Nam xây dựng năng lực quản trị AI. Những công cụ như đánh giá mức độ sẵn sàng, đánh giá rủi ro, đánh giá tác động, tiêu chuẩn kỹ thuật, báo cáo sự cố và hướng dẫn thực hành có thể giúp Việt Nam rút ngắn quá trình thử nghiệm chính sách. Tuy nhiên, việc tiếp thu các chuẩn mực này cần có chọn lọc, tránh sao chép máy móc những mô hình được thiết kế cho các quốc gia có năng lực thể chế, hạ tầng kỹ thuật và thị trường AI phát triển hơn.

Thứ tư, tác động quan trọng nhất của tiêu chuẩn quốc tế đối với Việt Nam không phải là tạo ra nghĩa vụ tuân thủ hình thức, mà là đặt ra yêu cầu nội địa hóa chuẩn mực quốc tế thành các quy phạm, tiêu chuẩn, thiết chế và công cụ thực thi phù hợp. Nói cách khác, Việt Nam cần sử dụng các tiêu chuẩn quốc tế như nguồn tham chiếu để thiết kế khung pháp lý quốc gia, đồng thời bảo đảm khung pháp lý đó phản ánh đúng nhu cầu phát triển, năng lực thực thi và yêu cầu bảo vệ lợi ích quốc gia trong bối cảnh chuyển đổi số.

2.3.3.2. Bài học từ kinh nghiệm của một số quốc gia

Từ Bảng tổng hợp so sánh kinh nghiệm các quốc gia về xây dựng khung pháp lý AI [Xem Phụ lục 1], có thể thấy không tồn tại một mô hình hoàn hảo hoặc có thể áp dụng trực tiếp cho Việt Nam. Mỗi mô hình đều phản ánh điều kiện đặc thù của quốc gia hoặc khu vực tạo ra nó: EU phù hợp với bối cảnh thị trường chung có năng lực thể chế cao; Hoa Kỳ dựa vào hệ sinh thái doanh nghiệp và tiêu chuẩn kỹ thuật mạnh; Trung Quốc thể hiện mô hình nhà nước định hướng và kiểm soát công nghệ; Nhật Bản tập trung vào giai đoạn chẩn đoán công khai trước soạn thảo, mô hình Ban Chỉ đạo đa Bộ và PDCA; Hàn Quốc kết hợp luật khung với chính sách phát triển công nghiệp; Singapore nhấn mạnh công cụ tự đánh giá, sandbox và hợp tác công - tư. Vì vậy, bài học đầu tiên cho Việt Nam là phải tổng hợp có chọn lọc, không sao chép một mô hình đơn lẻ.

Thứ nhất, bài học về chủ thể xây dựng khung pháp lý: Một là, mô hình chủ thể điều phối đặt trong cơ quan hành pháp cao nhất (Văn phòng Thủ tướng của Nhật Bản, Ủy ban AI Quốc gia trực thuộc cấp cao của Hàn Quốc) có hiệu lực điều phối liên ngành cao hơn đáng kể so với mô hình đặt trong một Bộ chuyên ngành ngang cấp với các Bộ khác vì thẩm quyền điều phối chỉ có tính ràng buộc thực chất khi đặt trên cấp các chủ thể cần điều phối. Đây là bài học rút ra trực tiếp từ đối chiếu giữa Nhật Bản, Hàn Quốc (điều phối hiệu quả) với Hoa Kỳ (điều phối phân tán, kém hiệu quả).

Hai là, sự tham gia của khu vực tư nhân có giá trị cao nhất khi được tổ chức theo quy trình có cấu trúc chính thức thay vì tham vấn không chính thức. Kinh nghiệm Singapore về đồng thiết kế công cụ tự đánh giá với quy trình kiểm thử thực địa 18 tháng có tổ chức chặt chẽ tạo ra chất lượng sản phẩm cao hơn nhiều so với tham vấn một chiều qua cổng thông tin điện tử.

Ba là, đại diện chuyên gia học thuật độc lập cần có tư cách tham gia chính thức trong cơ chế điều phối (như Ban điều tra chuyên gia của Nhật Bản do Thủ tướng trực tiếp bổ nhiệm), không chỉ đóng vai trò tư vấn không chính thức từ bên ngoài. Đây là điều kiện để tri thức kỹ thuật độc lập cân bằng được ảnh hưởng của lợi ích hành chính và lợi ích doanh nghiệp trong quá trình ra quyết định.

Thứ hai, bài học về nội dung xây dựng khung pháp lý:

Một là, không mô hình quốc tế nào thành công khi chỉ đầu tư vào một loại nội dung duy nhất (chỉ quy phạm cứng, hoặc chỉ công cụ mềm). EU kết hợp quy phạm cứng với tiêu chuẩn hài hòa và bộ quy tắc ứng xử; Hàn Quốc kết hợp luật khung với cơ chế miễn trừ tỷ lệ và chứng nhận nền tảng; Nhật Bản kết hợp luật cơ bản với Kế hoạch Cơ bản và hướng dẫn ngành.

Hai là, nội dung miễn trừ theo quy mô là thành tố không thể thiếu đối với quốc gia có hệ sinh thái doanh nghiệp AI chủ yếu là SME và startup. Cơ chế ba tầng miễn trừ của Hàn Quốc trực tiếp xuất phát từ đặc điểm cơ cấu doanh nghiệp tương đồng với Việt Nam, trong khi Hoa Kỳ nơi có nhiều doanh nghiệp AI quy mô lớn không phát triển cơ chế tương đương ở mức độ tinh vi như vậy xác nhận mối quan hệ giữa bối cảnh cơ cấu doanh nghiệp và nội dung cụ thể.

Ba là, nội dung cơ chế thích ứng là thành tố xuất hiện nhất quán ở các mô hình được đánh giá là thành công về lâu dài. Điều 112 EU AI Act, cơ chế PDCA thể chế hóa của Nhật Bản xác nhận nội dung xuyên suốt về cơ chế thích ứng là điều kiện bắt buộc để khung pháp lý duy trì hiệu lực.

Thứ ba, bài học về quy trình xây dựng khung pháp lý:

Một là, giai đoạn chẩn đoán bằng chứng công khai trước soạn thảo tạo ra hiệu quả kép, vừa cải thiện chất lượng nội dung, vừa rút ngắn thời gian giai đoạn tạo lập nhờ đồng thuận đã được xây dựng trước. Nhật Bản là minh chứng rõ nét nhất thông qua nhóm Nghiên cứu Thể chế AI công bố công khai kết quả tạm thời trước khi trình dự luật.

Hai là, soạn thảo song song văn bản quy phạm pháp luật với văn bản hướng dẫn chi tiết thay vì mô hình tuần tự là điều kiện để tránh khoảng trống thực thi tạm thời; kinh nghiệm EU về việc Ủy ban châu Âu triển khai song song quy trình xây dựng Danh sách tiêu chuẩn hài hòa ngay khi EU AI Act còn ở giai đoạn hoàn thiện văn bản chính là minh chứng.

Ba là, tham vấn kỹ thuật chuyên sâu có cấu trúc riêng biệt, không lồng ghép chung với tham vấn chính sách rộng tạo ra chất lượng, phát hiện vấn đề kỹ thuật bất khả thi cao hơn hẳn. Kinh nghiệm Hàn Quốc với khảo sát chi phí tuân thủ 200 doanh nghiệp trước khi soạn AI Basic Act là minh chứng cho giá trị của việc tách riêng kênh tham vấn kỹ thuật khỏi kênh tham vấn chính sách chung.

Thứ tư, bài học về kỹ thuật xây dựng khung pháp lý:

Một là, định nghĩa chức năng có tính ổn định thời gian cao hơn hẳn định nghĩa liệt kê công nghệ. Bài học được xác nhận qua thất bại của chính EU khi phải từ bỏ Phụ lục liệt kê kỹ thuật trong dự thảo 2021, đối chiếu với việc OECD, Nhật Bản đều lựa chọn định nghĩa chức năng ngay từ đầu.

Hai là, cơ chế suy đoán tuân thủ qua pháp luật hóa tiêu chuẩn kỹ thuật là kỹ thuật có hiệu lực mạnh nhất để giải quyết đồng thời hai vấn đề chắc chắn pháp lý cho doanh nghiệp và tiêu chí nhất quán cho cơ quan thực thi nhưng đòi hỏi hạ tầng tiêu chuẩn hóa kỹ thuật đã phát triển.

Ba là, kỹ thuật miễn trừ tỷ lệ kết hợp công cụ tự đánh giá tạo ra hiệu ứng khuyến khích tuân thủ tự nguyện tăng lên đáng kể. Bằng chứng phản trực quan từ Hàn Quốc (miễn trừ làm tăng tỷ lệ tham gia kiểm định tự nguyện) là phát hiện quan trọng cần được vận dụng thận trọng, có kiểm chứng, khi thiết kế cơ chế tương tự cho Việt Nam.

Bốn là, kỹ thuật viện dẫn không ngày là công cụ kỹ thuật lập pháp có chi phí thấp nhưng hiệu quả cao để duy trì tính cập nhật của văn bản mà không cần sửa đổi hình thức được vận dụng nhất quán ở cả Nhật Bản và EU, có thể áp dụng ngay cho Việt Nam mà không đòi hỏi điều kiện tiên quyết đặc biệt.

Từ các giá trị tham khảo trên, có thể gợi mở cho Việt Nam mô hình "kiến trúc phân tầng thích ứng". Mô hình này không sao chép nguyên trạng bất kỳ kinh nghiệm quốc tế nào mà kết hợp có chọn lọc giữa tính ổn định của luật, tính linh hoạt của văn bản dưới luật, tính cập nhật của tiêu chuẩn kỹ thuật, tính mềm dẻo của hướng dẫn thực hành và tính thực tiễn của cơ chế thử nghiệm. Về cấu trúc, mô hình này có thể bao gồm bảy tầng cơ bản: nền tảng hiến định và văn kiện quốc tế; luật AI và các luật liên quan; tiêu chuẩn kỹ thuật; luật mềm và hướng dẫn ngành; bộ quy tắc ứng xử và tự quản ngành; án lệ, lẽ công bằng và tập quán pháp; thiết chế thực thi và cơ chế giám sát. Cách tiếp cận này phù hợp hơn với điều kiện của Việt Nam vì hệ thống pháp luật có truyền thống thành văn và tính tập trung cao, hệ sinh thái AI còn đang phát triển, phần lớn doanh nghiệp là vừa và nhỏ, năng lực thực thi cần được tăng cường từng bước, trong khi yêu cầu đặt ra là phải đồng thời thúc đẩy đổi mới sáng tạo, kiểm soát rủi ro, bảo vệ quyền con người và bảo đảm chủ quyền số trong kỷ nguyên AI.

Cần nhận thức rõ bốn chiều phân tích (chủ thể, nội dung, quy trình, kỹ thuật) và mô hình bảy tầng là hai công cụ có chức năng khác nhau được áp dụng nối tiếp nhau. Bốn chiều phân tích được sử dụng với chức năng công cụ so sánh quá trình cho phép đối chiếu cách thức các quốc gia khác nhau đã xây dựng khung pháp lý AI của họ. Mô hình bảy tầng được đề xuất với chức năng đặc tả kiến trúc sản phẩm, xác định khung pháp lý AI Việt Nam nên có những thành phần nào. Bước chuyển từ công cụ thứ nhất sang công cụ thứ hai chính là bước chuyển từ phân tích quá trình của người khác sang thiết kế sản phẩm cho mình và đòi hỏi một quy trình chuyển hóa cụ thể được trình bày dưới đây.

Bước 1, từ "nội dung" sang các tầng cụ thể của mô hình bảy tầng: Khi phân tích nội dung của các đối tượng quốc tế theo bốn nhóm (quy phạm ràng buộc, công cụ chuẩn tắc/kỹ thuật, thiết chế thực thi, cơ chế thích ứng xuyên suốt), luận án nhận thấy các quốc gia thành công nhất (đặc biệt EU, Hàn Quốc, Nhật Bản, Singapore) đều trên thực tế đã phân bổ nội dung vào các lớp có chức năng phân biệt tinh vi hơn bốn nhóm

khái quát: EU phân biệt rõ giữa quy phạm luật khung (AI Act), tiêu chuẩn hài hòa (một lớp riêng), hướng dẫn thực thi của Ủy ban (một lớp khác nữa), và cơ chế xét xử của CJEU (một lớp thứ tư). Tức là trong nội bộ "Nhóm nội dung công cụ chuẩn tắc/kỹ thuật", thực tiễn quốc tế đã tự phân hóa thành ba lớp con phân biệt (tiêu chuẩn kỹ thuật, hướng dẫn ngành, bộ quy tắc ứng xử) mà luận án xác định tương ứng chính xác với Tầng 3, Tầng 4, Tầng 5 của mô hình bảy tầng. Việc đối chiếu thực tiễn phân hóa tự nhiên này giữa nhiều quốc gia độc lập với nhau chính là bằng chứng thực chứng bổ sung, củng cố cho ba tiêu chí phương pháp luận xác định số tầng đã trình bày.

Bước 2, từ "chủ thể, quy trình, kỹ thuật" sang các giải pháp vận hành gắn với từng tầng: Các bài học về chủ thể, quy trình và kỹ thuật không trực tiếp tạo ra thêm tầng mới trong kiến trúc bảy tầng là cấu trúc của sản phẩm, mà được chuyển hóa thành nội dung vận hành cụ thể được gắn vào từng tầng đã có. Đây chính là điểm phân biệt quan trọng cần làm rõ để tránh gây hiểu lầm rằng bốn chiều phân tích tự động sinh ra bảy tầng theo một phép biến đổi cơ học. Cụ thể, bài học về chủ thể (như mô hình Ban Chỉ đạo đa Bộ của Nhật Bản) được chuyển hóa thành nội dung cụ thể của Tầng 7 (thiết chế thực thi) trong đề xuất về cơ chế điều phối AI xuyên ngành; bài học về kỹ thuật suy đoán tuân thủ của EU được chuyển hóa thành nội dung vận hành của Tầng 3 (tiêu chuẩn kỹ thuật); bài học về quy trình soạn thảo song song của EU được chuyển hóa không phải thành một tầng mà thành một nguyên tắc quy trình áp dụng cho việc tạo lập đồng thời nhiều tầng.

Bước 3, từ "mô hình tổng thể riêng lẻ của từng quốc gia" sang "mô hình tổng hợp có chọn lọc": Bước chuyển hóa cuối cùng và có tính quyết định là thay vì sao chép nguyên trạng kiến trúc tổng thể của một quốc gia đơn lẻ, luận án tổng hợp có chọn lọc lấy kiến trúc bảy tầng làm khung xương sống, sau đó với mỗi tầng, lựa chọn nội dung vận hành từ quốc gia có kinh nghiệm phù hợp nhất với đặc thù của tầng đó và với điều kiện chuyển giao đã đánh giá: Tầng 3 (tiêu chuẩn kỹ thuật) học kỹ thuật suy đoán tuân thủ có điều chỉnh từ EU; Tầng 4 (luật mềm) học cấu trúc ba lớp và cơ chế comply-or-explain từ Nhật Bản; Tầng 5 (bộ quy tắc ứng xử) học quy trình đồng thiết kế có bổ sung yếu tố ràng buộc từ Singapore; và toàn bộ mô hình miễn trừ tỷ lệ xuyên suốt các tầng học từ Hàn Quốc. Đây chính là quá trình hình thành mô hình kiến trúc phân tầng thích ứng cho Việt Nam.

Tổng kết Chương 2

Chương 2 đã xây dựng cơ sở lý luận cho việc nghiên cứu đề tài “Xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam”. Trên cơ sở làm rõ các khái niệm nền tảng như khung pháp lý, AI, hệ thống AI, phát triển và ứng dụng AI, chương

này đã xác định khung pháp lý cho phát triển, ứng dụng AI không chỉ là tập hợp các VBQPPL mà là một cấu trúc điều chỉnh đa tầng.

Từ đặc thù của AI, Chương 2 đã chỉ ra rằng việc xây dựng khung pháp lý cho phát triển, ứng dụng AI phải được tiếp cận như một quá trình liên tục, đa chủ thể, có tính thích ứng và dựa trên rủi ro.

Chương 2 cũng đã làm rõ chủ thể, nội dung, quy trình và yêu cầu kỹ thuật đặt ra đối với hoạt động xây dựng khung pháp lý cho phát triển, ứng dụng AI. Việc xây dựng khung pháp lý vì vậy không thể dừng ở khâu ban hành luật, mà cần bao gồm cả thiết kế kiến trúc tổng thể, tạo lập các tầng quy phạm và tiêu chuẩn, thiết lập thiết chế thực thi, vận hành, đánh giá và cập nhật chính sách.

Trên cơ sở phân tích các tiêu chuẩn quốc tế và kinh nghiệm của một số quốc gia, Chương 2 rút ra rằng Việt Nam cần tiếp thu có chọn lọc các giá trị phù hợp với điều kiện chính trị, pháp lý, kinh tế, công nghệ và năng lực thực thi trong nước.

Như vậy, Chương 2 đã xác lập khung lý luận làm nền tảng cho các phân tích tiếp theo trong Chương 3 để đánh giá thực trạng khung pháp lý và thực tiễn xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam, đồng thời làm cơ sở cho Chương 4 trong việc đề xuất quan điểm và giải pháp xây dựng khung pháp lý trong thời gian tới.

Chương 3
THỰC TRẠNG XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM

3.1. THỰC TRẠNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM

3.1.1. Thực trạng phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Thứ nhất, nghiên cứu và phát triển (R&D)

Từ đầu những năm 70 của thế kỷ 20, AI đã được giảng dạy và bước đầu nghiên cứu với thuật ngữ "trí khôn nhân tạo". Những năm vừa qua, các trường đại học hàng đầu của Việt Nam đã bắt đầu triển khai các chương trình đào tạo Khoa học máy tính, định hướng AI và Khoa học dữ liệu. Bên cạnh đó, còn có các hoạt động như bài giảng đại chúng về AI, khoa học dữ liệu và học máy của Viện Toán cao cấp, Cộng đồng VietAI và Công ty AI Academy [66, tr.12-15]. Tổng cộng có 50 trường đại học đào tạo các chuyên ngành liên quan tới AI, trong đó 10 trường có ngành đào tạo AI. Ngoài ra, ở khu vực tư, các tập đoàn, doanh nghiệp lớn cũng thành lập trung tâm nghiên cứu AI phục vụ nội bộ như FPT.AI Quy Nhơn, VinAI, AI Centers (hợp tác giữa Tập đoàn NAVER với Trường Đại học Bách khoa Hà Nội)… Giai đoạn 2010-2025, trong 11 quốc gia Đông Nam Á, Việt Nam xếp thứ 4 về số lượng công bố khoa học trên cơ sở dữ liệu Scopus về AI [302]. Trong khi GDP danh nghĩa của Việt Nam thấp hơn hẳn trong số quốc gia hàng đầu về kinh tế, điều này cho thấy nỗ lực của cộng đồng AI Việt Nam. Việt Nam ưu tiên đầu tư công cho AI thông qua các chương trình nghiên cứu phát triển như: KC01, NAFOSTED, KC-4.0... Bên cạnh đó, phải kể đến các hội nghị quốc tế có chất lượng và chuyên sâu về AI đã được tổ chức ở Việt Nam như: Khai phá dữ liệu PAKDD 2005 và 2015, AI-PRICAI 2008, Học máy-ACML 2014, Xử lý ngôn ngữ tự nhiên - CICLING 2018, Tác tử thông minh - PRIMA 2008. Hội thảo quốc gia về AI (AI4VN) tổ chức mỗi năm một lần là Ngày hội AI đã góp phần tập hợp, kết nối, tụ hội phát triển hệ sinh thái nghiên cứu, ứng dụng AI tại Việt Nam và cho Việt Nam [66, tr.12-15]. Tuy nhiên, thực trạng nghiên cứu cơ bản còn hạn chế. Nghiên cứu chủ yếu tập trung vào các ứng dụng hoặc cải tiến các mô hình đã có của thế giới, thiếu các nghiên cứu cơ bản tạo ra các thuật toán lõi hoặc kiến trúc mô hình mới có tính đột phá toàn cầu. Nguồn kinh phí đầu tư cho nghiên cứu cơ bản còn khiêm tốn. Một vấn đề pháp lý nữa là việc nghiên cứu thường dùng dữ liệu cá nhân và tác phẩm được bảo hộ để thử nghiệm, trong khi ranh giới giữa nghiên cứu và đưa hệ thống ra thị trường chưa rõ.

Thứ hai, thu thập và xử lý dữ liệu

Dữ liệu là "nguyên liệu" sống còn của AI. Việt Nam có nguồn dữ liệu lớn về hành vi người dùng, viễn thông và dữ liệu công (dân cư, đất đai) tiềm năng. Tập đoàn VNPT đã xây dựng, triển khai 8 trung tâm IDC trên toàn quốc và các IDC vệ tinh tại các tỉnh, thành phố, phục vụ chuyển đổi số [135]. Viettel với trung tâm dữ liệu thứ 14 đã nâng tổng số lên 230.000 máy chủ, 81.000 m2 mặt sàn, 11.500 rack; 87MW điện, tương đương một siêu DC của thế giới [128]. Thách thức chủ yếu nằm ở dữ liệu nhà nước và doanh nghiệp còn bị phân mảnh và thiếu các tiêu chuẩn chung về chất lượng, định dạng. Về chất lượng, các bộ dữ liệu lớn thường thiếu tính sạch, chính xác và tính đại diện, dẫn đến nguy cơ thiên lệch thuật toán. Dữ liệu tiếng Việt, bao gồm văn bản và giọng nói, thường không có dấu hoặc sử dụng ngôn ngữ giới trẻ (teencode), điều này gây khó khăn trong quá trình đào tạo AI. Sự thiếu sót về dữ liệu đặc trưng dẫn đến hiệu suất của AI không cao, vì AI chưa hiểu được ngôn ngữ địa phương, từ ngữ lóng, tiếng của các dân tộc và văn hóa vùng miền. Hệ quả là phải phụ thuộc vào các nền tảng từ nước ngoài như Google hay OpenAI [75].

Dữ liệu tiếng Việt chất lượng cao còn khan hiếm, nhất là dữ liệu bằng ngôn ngữ của các dân tộc thiểu số và dữ liệu về các vùng ít được số hóa. Ngược lại, dữ liệu cá nhân tập trung với khối lượng lớn ở ngân hàng, viễn thông và hệ thống định danh điện tử quốc gia. Từ ngày 01/7/2024, Quyết định 2345/QĐ-NHNN yêu cầu xác thực sinh trắc học đối với một số giao dịch trực tuyến, làm khối lượng dữ liệu sinh trắc học được xử lý tăng mạnh. Chưa có quy định riêng về căn cứ pháp lý cho việc tái sử dụng dữ liệu vốn được thu thập cho mục đích khác để huấn luyện trong khi phải áp dụng nguyên tắc đúng mục đích của Luật BVDLCN, nên trên thực tế thường phải xin lại sự đồng ý. Chưa có nghĩa vụ chung về minh bạch nguồn dữ liệu huấn luyện. Hai hệ phân loại rủi ro theo Điều 9 Luật AI và khoản 4 Điều 30 Luật BVDLCN mới được phối hợp ở mức hồ sơ, chưa thống nhất về tiêu chí.

Thứ ba, xây dựng và huấn luyện mô hình

Các tập đoàn công nghệ lớn trong nước như Viettel, FPT và VinGroup đóng vai trò chủ đạo, sử dụng rộng rãi các nền tảng mã nguồn mở toàn cầu để phát triển các mô hình ứng dụng cụ thể (nhận dạng tiếng nói, thị giác máy tính và xử lý ngôn ngữ tự nhiên tiếng Việt). Việt Nam đã gia nhập nhóm những quốc gia hiếm hoi trong khu vực Đông Nam Á sở hữu mô hình ngôn ngữ lớn (LLM) do nội địa phát triển. Đặc biệt, Zalo đã hoàn thành việc huấn luyện và ra mắt một mô hình LLM tập trung vào tiếng Việt vào năm 2023 với sự nghiên cứu và phát triển hoàn toàn từ đội ngũ kỹ sư Việt Nam. Có hai phương pháp huấn luyện LLM, bao gồm phương pháp tinh chỉnh [183] sử dụng các LLM đã được huấn luyện trước đó để tạo ra mô hình mới cho các mục đích cụ thể và phương pháp huấn luyện từ đầu [330] thực hiện tất cả các quy trình từ khởi tạo tham

số, lựa chọn kiến trúc mô hình đến thuật toán huấn luyện trên một tập dữ liệu cụ thể. Zalo đã quyết định sử dụng phương pháp huấn luyện từ đầu, giúp người Việt có thể kiểm soát và điều khiển hoàn toàn quy trình huấn luyện và mô hình [3]. Tuy nhiên, hầu hết các mô hình được xây dựng dựa trên cách tiếp cận chuyển giao và tinh chỉnh từ các mô hình nền tảng của nước ngoài, do Việt Nam thiếu cơ sở hạ tầng tính toán hiệu năng cao (HPC) và GPU quy mô lớn cần thiết để huấn luyện các mô hình AI lớn. Sự phụ thuộc vào các dịch vụ đám mây nước ngoài cho việc huấn luyện quy mô lớn làm tăng chi phí và tiềm ẩn rủi ro về an ninh dữ liệu. Chưa có nghĩa vụ trực tiếp của nhà cung cấp mô hình đối với bên tích hợp, như nghĩa vụ cung cấp thẻ mô hình, thông tin cần cho phân loại và cho xử lý sự cố. Chưa có quy tắc đầy đủ về mô hình mã nguồn mở.

Thứ tư, kiểm thử và đánh giá

Về tiêu chí và phương pháp đánh giá hiệu suất mô hình AI, quá trình phát triển mô hình AI tại Việt Nam thường tuân theo quy trình chuẩn quốc tế, trong đó sau khi mô hình được đào tạo sẽ trải qua quá trình xác thực và kiểm tra nghiêm ngặt để đánh giá hiệu suất, bao gồm việc sử dụng các tập dữ liệu riêng biệt để đánh giá mức độ tổng quát hóa của mô hình đối với dữ liệu mới chưa từng thấy. Tuy nhiên, thách thức lớn nhất trong việc kiểm tra độ tin cậy và dự đoán hành vi của mô hình AI là việc xã hội chưa đủ năng lực dự đoán các hành vi phát sinh từ mô hình phức tạp, nếu không được kiểm soát đúng mức có thể khiến con người mất quyền kiểm soát chính công nghệ mình tạo ra. Điều này đặc biệt quan trọng trong bối cảnh chỉ trong vài năm mô hình học sâu đã tiến từ khả năng nhận diện hình ảnh cơ bản đến việc tự tạo nội dung, mô phỏng cảm xúc, đưa ra quyết định phức tạp, tạo ra loạt rủi ro khó dự đoán từ thiên lệch dữ liệu, đe dọa quyền riêng tư đến các hệ thống tự động hóa không còn minh bạch với con người. Về thực trạng năng lực kiểm định và giám sát trong thực tiễn, việc áp dụng cơ chế hậu kiểm cho tất cả các mức độ rủi ro có thể không phù hợp do AI có thể tạo ra những rủi ro bất ngờ, đây là lĩnh vực mới trong việc quản lý nhà nước yêu cầu nguồn nhân lực chất lượng cao khi tội phạm AI ngày càng phát triển nhanh. Thực tế này phản ánh rõ ràng khoảng cách giữa yêu cầu pháp lý và năng lực thực thi trong việc kiểm định và giám sát mô hình AI tại Việt Nam. Phần lớn việc kiểm thử hiện do chính doanh nghiệp thực hiện. Năng lực kiểm thử độc lập trong nước và tổ chức đánh giá sự phù hợp chuyên về AI đều còn rất hạn chế. Vai trò cụ thể của bên triển khai trong đánh giá sự phù hợp chưa được xác định. Chưa có tiêu chuẩn quốc gia về phương pháp kiểm thử, đánh giá AI làm thước đo cho Điều 13.

Thứ năm, tích hợp và triển khai

Thực trạng tích hợp mô hình AI vào các ứng dụng, phần mềm và hệ thống vật lý tại Việt Nam hiện đang trải qua giai đoạn phát triển tích cực với nhiều thành tựu đáng ghi nhận, nhưng cũng đối mặt với không ít thách thức về hạ tầng, nguồn nhân lực

và khả năng triển khai quy mô lớn. Sự tích hợp này đang diễn ra trên nhiều lĩnh vực từ doanh nghiệp [108], y tế [48], ứng dụng AI bản địa [126], pháp lý [17], giao thông thông minh [111], hệ thống tự động hóa công nghiệp [71]. Về thách thức trong tích hợp AI, phần lớn các doanh nghiệp vừa và nhỏ không có đội ngũ chuyên trách về AI hoặc chuyên gia dữ liệu giàu kinh nghiệm, dẫn đến sự phụ thuộc vào các giải pháp bên ngoài hoặc đơn vị tư vấn, điều này thường gây ra sự chậm trễ trong triển khai và hạn chế khả năng tùy biến các công cụ AI theo nhu cầu thực tế. Ngoài ra, hạ tầng công nghệ không đáp ứng được yêu cầu mở rộng, việc tích hợp AI vào các hệ thống cũ hiện có là một quá trình phức tạp và tốn nhiều thời gian. Trong nhiều ngành công nghiệp truyền thống, các doanh nghiệp vẫn đang trong giai đoạn chuyển đổi từ vận hành thủ công sang số hóa. Chưa có yêu cầu kiểm chứng giám sát thực chất ngoài thời gian chuyển tiếp. Danh mục rủi ro cao là danh mục đóng và chưa bao gồm lĩnh vực tuyển dụng. Chưa có quy định riêng về an toàn khi mua sắm công đối với hệ thống AI trong khi bên triển khai thường là ngân hàng, bệnh viện, cơ quan nhà nước, những chủ thể ít có chuyên môn kỹ thuật sâu về AI.

Thứ sáu, quản lý và bảo trì

Về nguyên tắc chung của việc giám sát mô hình AI sau triển khai, các hệ thống AI cần được giám sát và bảo trì liên tục để bảo đảm tiếp tục hoạt động như mong đợi, bao gồm việc theo dõi hiệu suất của hệ thống, xác định sự suy giảm hoặc bất thường và đào tạo lại mô hình khi cần. Về hiện tượng độ trôi của mô hình và tầm quan trọng của việc phát hiện sớm, thế giới luôn thay đổi và dữ liệu cũng vậy, mô hình AI cần được giám sát liên tục để đảm bảo hiệu suất không bị suy giảm theo thời gian do sự thay đổi của dữ liệu đầu vào là hiện tượng model drift [241]. Về các loại độ trôi cần giám sát, thường có hai nguyên nhân dẫn đến sự suy giảm của mô hình: data drift [182], concept drift [208] hoặc đôi khi là cả hai. Báo cáo của Microsoft chỉ ra rằng các mô hình học máy có thể mất tới 40% độ chính xác trong vòng một năm nếu data drift bị xem nhẹ [148]. Về phương pháp phát hiện độ trôi của dữ liệu, biểu hiện trực tiếp và rõ ràng nhất của Model Drift là độ chính xác dự đoán giảm dần theo thời gian, nhưng cần có cả kết quả dự đoán của model và dữ liệu thực tế (ground truth) [342]. Về quy trình cập nhật và huấn luyện lại mô hình khi phát hiện độ trôi, cần huấn luyện lại mô hình trên tập dữ liệu mới và giữ nguyên tất cả các yếu tố khác như hyperparameters [155], thuật toán, features. Có ba phương pháp chính để huấn luyện lại mô hình: full retraining, incremental learning [262], train mô hình với bộ siêu tham số mới cần được tinh chỉnh lại. Khi triển khai AI vào bảo mật an toàn thông tin, doanh nghiệp cần định kỳ kiểm tra hiệu suất của AI để phát hiện các lỗi như quá khớp (overfitting) [40] hoặc model drift và kết hợp giám sát của con người. Vấn đề pháp lý đặt ra là nhà cung cấp chưa có

nghĩa vụ thông báo cho bên triển khai khi có thay đổi đáng kể hoặc bản cập nhật. Chưa có quy định bắt buộc dùng dữ liệu sự cố làm căn cứ rà soát chính sách.

Thứ bảy, phát triển ứng dụng cuối

Về phát triển giao diện chatbot AI, Việt Nam đã có những bước tiến đáng kể với sự tham gia của nhiều doanh nghiệp công nghệ lớn như VNPT SmartBot của VNPT AI, nền tảng FPT.AI của FPT [39], cho thấy quy mô ứng dụng và khả năng cạnh tranh quốc tế của các sản phẩm AI Việt Nam.

Về thách thức trong phát triển chatbot AI tại Việt Nam, một trong những vấn đề chung là hiểu đúng ngữ cảnh và ý nghĩa sâu xa của câu hỏi hoặc yêu cầu của người dùng. Nếu không được tích hợp trình duyệt web, chatbot cũng không thể truy cập thông tin thời gian thực, vì thế một câu trả lời nghe rất mạch lạc không đồng nghĩa với chính xác, trong nhiều trường hợp chatbot còn có thể tạo ra thông tin sai mà vẫn thể hiện rất tự tin, thường được gọi là hallucination [340].

Về phát triển ứng dụng di động tích hợp AI, trong nửa đầu năm 2025 người Việt dành 283 triệu giờ để sử dụng các ứng dụng AI tạo sinh, tương đương mỗi ngày thị trường Việt Nam ghi nhận gần 1,6 triệu giờ và hơn 41 triệu phiên dùng ứng dụng AI, trung bình mỗi người Việt dành 2,83 giờ và 75 phiên truy cập app AI [244]. Điều này cho thấy sự gia tăng mạnh mẽ trong việc chấp nhận và sử dụng các ứng dụng AI tại Việt Nam. Đây là giai đoạn người dân trực tiếp chịu tác động qua trợ lý ảo, nền tảng nội dung và các công cụ tạo hình ảnh, giọng nói, video. Lừa đảo sử dụng giọng nói và hình ảnh giả mạo bằng AI cũng gia tăng từ năm 2023 đến 2025 trong khi quyền được giải thích một quyết định cụ thể bởi AI, quyền phản đối và quyền yêu cầu con người xem xét lại chưa được quy định. Người bị thiệt hại không có công cụ tiếp cận nhật ký và tài liệu kỹ thuật trong tố tụng dân sự.

Thứ tám, chấm dứt vận hành

Hoạt động kỹ thuật của giai đoạn này gồm ngừng cung cấp dịch vụ dựa trên hệ thống AI, xóa hoặc lưu trữ mô hình và dữ liệu liên quan, thông báo cho người dùng về việc chấm dứt dịch vụ, và xử lý các nghĩa vụ còn tồn đọng liên quan đến hợp đồng, dữ liệu và trách nhiệm đối với thiệt hại đã phát sinh trước thời điểm chấm dứt. Quan hệ pháp luật phát sinh khi chấm dứt vận hành là quan hệ giữa bên chấm dứt vận hành hệ thống và người dùng đang sử dụng dịch vụ về nghĩa vụ thông báo trước và bảo đảm quyền tiếp tục tiếp cận dữ liệu cá nhân hoặc dữ liệu do người dùng tạo ra trong quá trình sử dụng; quan hệ giữa bên chấm dứt vận hành và cơ quan quản lý nhà nước về nghĩa vụ cập nhật thông tin trên Cổng thông tin điện tử một cửa về AI và Cơ sở dữ liệu quốc gia về AI (xóa hoặc cập nhật trạng thái đăng ký của hệ thống); quan hệ về lưu trữ hồ sơ và dữ liệu sau khi chấm dứt vận hành phục vụ mục đích giải quyết tranh chấp

hoặc điều tra sự cố có thể phát sinh sau thời điểm chấm dứt liên quan đến hành vi của hệ thống khi còn vận hành; trong trường hợp chấm dứt vận hành do quyết định bắt buộc của cơ quan có thẩm quyền, quan hệ pháp luật hành chính về thi hành quyết định và quyền khiếu nại của tổ chức bị áp dụng biện pháp. Chủ thể chịu nghĩa vụ là nhà cung cấp hoặc bên triển khai quyết định chấm dứt vận hành (nghĩa vụ thông báo trước cho người dùng trong thời hạn hợp lý, nghĩa vụ xử lý dữ liệu cá nhân theo đúng quy định về lưu trữ hoặc xóa dữ liệu khi chấm dứt xử lý, nghĩa vụ tiếp tục chịu trách nhiệm đối với thiệt hại đã phát sinh trước thời điểm chấm dứt); trong trường hợp đặc thù khi tổ chức vận hành chấm dứt tồn tại (giải thể, phá sản) trước khi hệ thống AI được chính thức ngừng vận hành, phát sinh câu hỏi pháp lý về chủ thể kế thừa nghĩa vụ liên quan đến hệ thống AI đã triển khai. Quyền, lợi ích có nguy cơ bị xâm phạm gồm quyền của người dùng đối với dữ liệu cá nhân và dữ liệu do mình tạo ra trong quá trình sử dụng hệ thống có nguy cơ bị ảnh hưởng khi không có quy định rõ về việc dữ liệu đó được xử lý ra sao sau khi dịch vụ chấm dứt (xóa, chuyển giao, tiếp tục lưu trữ); quyền của người dùng đang trong quá trình sử dụng dịch vụ có tính chất liên tục (như hệ thống AI hỗ trợ theo dõi sức khỏe hoặc tài chính cá nhân dài hạn) trước nguy cơ gián đoạn đột ngột không có thời gian chuyển tiếp hợp lý; lợi ích của cơ quan quản lý nhà nước cùng người dùng bị thiệt hại trước đó trong việc bảo đảm khả năng truy vết và quy trách nhiệm đối với hành vi của hệ thống AI đã ngừng vận hành nhưng gây hậu quả được phát hiện sau thời điểm chấm dứt.

3.1.2. Thực trạng nội dung khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Từ ba tiêu chí xác định các thành phần của khung pháp lý ở Chương 2, kết hợp áp dụng phương pháp phân tích chức năng của Baldwin, Cave và Lodge trong *Understanding Regulation* [321] và phân tích nguồn pháp luật của Xanthaki trong *Drafting Legislation* [222] vào Việt Nam cho thấy bảy tầng là cần thiết để một khung pháp lý AI vận hành với tất cả chức năng điều tiết. Nếu gộp 2 tầng bất kỳ để có tổng 6 tầng, ví dụ gộp tầng 1 vào tầng 2 sẽ làm mất vai trò giới hạn của nền tảng hiến định. Gộp tầng 4 với tầng 5 sẽ xóa nhòa ranh giới về trách nhiệm giải trình. Nếu tách 1 tầng bất kỳ thành 2 tầng để có tổng 8 tầng, ví dụ tách tầng 2 thành luật và văn bản dưới luật về bản chất không đem lại chức năng mới.

3.1.2.1. Tầng thứ nhất - Nền tảng hiến định và văn kiện quốc tế

Thứ nhất, tầng nền tảng hiến định của khung pháp lý AI Việt Nam dựa trên Hiến pháp 2013 với các điều khoản bảo vệ quyền con người có liên quan trực tiếp đến AI. Điều 20 bảo đảm quyền bất khả xâm phạm về thân thể, Điều 21 bảo vệ đời sống riêng tư và bí mật cá nhân, Điều 25 bảo đảm tự do ngôn luận, Điều 26 về bình đẳng

giới và Điều 40 về quyền hưởng thụ các giá trị văn hóa đều tạo ra giới hạn hiến định mà không có yêu cầu công nghệ AI nào được vượt qua.

Tầng này tương đối vững chắc về mặt hình thức. Hiến pháp 2013 có hệ thống quyền con người tương đối đầy đủ, đã được nhiều quốc gia ghi nhận. Tuy nhiên, chưa có cơ chế giải thích Hiến pháp có giá trị ràng buộc, không có Tòa án Hiến pháp hay Hội đồng Bảo hiến dẫn đến các nguyên tắc hiến định về quyền con người liên quan đến AI khó được viện dẫn trực tiếp trong tranh chấp tư pháp. Thực tiễn xét xử Việt Nam hầu như không có án lệ viện dẫn điều khoản hiến định về quyền riêng tư để phán xét hành vi AI, khoảng trống này sẽ trở nên nghiêm trọng hơn khi tác nhân AI ngày càng phổ biến.

Thứ hai, điều ước quốc tế ràng buộc có nội dung liên quan đến AI:

Một là, Công ước Hà Nội - tên gọi tắt của Công ước Liên Hợp Quốc về chống tội phạm mạng được Đại hội đồng LHQ thông qua ngày 24/12/2024 và được mở ký tại Hà Nội ngày 25-26/10/2025 là văn kiện pháp lý đa phương đầu tiên do LHQ chủ trì xây dựng về tội phạm mạng với 9 chương, 71 điều khoản. Việt Nam là nước đăng cai tổ chức lễ mở ký với 72 quốc gia ký kết và đã ký Công ước. Nhìn từ góc độ kiến trúc khung pháp lý AI, Công ước Hà Nội không phải văn kiện điều chỉnh AI thuần túy nhưng có ba giao điểm quan trọng với tầng điều ước quốc tế của khung pháp lý AI Việt Nam. (i) Nghĩa vụ hình sự hóa tội phạm sử dụng AI: Công ước xác lập nghĩa vụ các quốc gia thành viên hình sự hóa các hành vi phạm tội mạng bao gồm hành vi lợi dụng AI để giả mạo hình ảnh, giọng nói, danh tính; sử dụng mã độc và dữ liệu đánh cắp để lừa đảo bằng AI tổng hợp. Đây là nghĩa vụ điều ước quốc tế bổ sung và ràng buộc hơn so với các quy định hình sự hiện hành của Việt Nam về tội phạm công nghệ. (ii) Chuẩn chứng cứ điện tử trong tố tụng AI: Công ước quy định các tiêu chuẩn và thủ tục thu thập, bảo quản và sử dụng chứng cứ điện tử trong điều tra tội phạm mạng bao gồm dữ liệu từ hệ thống AI, nhật ký hoạt động AI và nội dung AI tổng hợp. Khi Công ước được phê chuẩn, Việt Nam có nghĩa vụ điều chỉnh pháp luật tố tụng để đáp ứng chuẩn này. (iii) Hợp tác quốc tế 24/7 về tội phạm AI xuyên biên giới: Mạng lưới hợp tác thực thi pháp luật 24/7 theo Công ước tạo ra cơ chế pháp lý để Việt Nam yêu cầu bảo toàn chứng cứ điện tử từ máy chủ đặt ở nước ngoài trong thời gian ngắn, khắc phục một phần khoảng trống thẩm quyền đối với tội phạm AI xuyên biên giới mà tầng VBQPPL hiện hành chưa giải quyết. Việc Việt Nam ký Công ước Hà Nội là bước tiến tích cực trong tầng 1 của khung pháp lý AI. Tuy nhiên, hệ thống pháp luật trong nước chưa được điều chỉnh để thực thi nghĩa vụ Công ước, mỗi ngày chậm nội luật hóa là mỗi ngày nạn nhân tội phạm AI xuyên biên giới thiếu cơ chế pháp lý bảo vệ hiệu quả. Ngoài ra, tầng này chưa có cam kết quốc tế riêng về AI.

Hai là, Điều ước về thương mại và thương mại điện tử có nội dung liên quan đến AI, tiêu biểu là:

(1) Hiệp định Đối tác Toàn diện và Tiến bộ xuyên Thái Bình Dương (CPTPP, 2018) với điều khoản liên quan AI: Chương 14 (Thương mại điện tử) cho phép việc lưu chuyển thông tin xuyên biên giới bằng các phương tiện điện tử, bao gồm cả thông tin cá nhân (Điều 14.11); cấm yêu cầu pháp nhân sử dụng hoặc đặt hệ thống máy chủ trong nước (Điều 14.13); bảo vệ mã nguồn (Điều 14.17); không đánh thuế hải quan đối với các giao dịch điện tử (Điều 14.3). Tất cả các điều khoản này trực tiếp ảnh hưởng đến môi trường phát triển và triển khai AI xuyên biên giới.

(2) Hiệp định Thương mại Tự do Việt Nam - EU (EVFTA, 2020) ở Chương 8 (Dịch vụ, đầu tư và mua sắm chính phủ) và Điều khoản về thương mại điện tử tạo nghĩa vụ không phân biệt đối xử và minh bạch trong quy định kỹ thuật số, bao gồm quy định về AI; Điều khoản BVDLCN tương thích với GDPR của EU. Đặc biệt, Tuyên bố chung về BVDLCN kèm theo EVFTA xác lập nguyên tắc "mức độ bảo vệ tương đương" tạo áp lực để Việt Nam nâng chuẩn BVDLCN trong huấn luyện AI.

(3) Hiệp định Đối tác Kinh tế Toàn diện Khu vực (RCEP, 2022) ở Chương 12 (Thương mại điện tử) quy định tương tự CPTPP nhưng linh hoạt hơn, đặc biệt Điều 12.15 (tự do luồng dữ liệu với ngoại lệ tự phán quyết về chính sách công hợp pháp). Tác động với AI tạo khung tự do luồng dữ liệu phục vụ huấn luyện AI xuyên biên giới với các đối tác ASEAN, Trung Quốc, Nhật Bản, Hàn Quốc, Úc, New Zealand.

(4) Hiệp định Thương mại Tự do ASEAN - Úc - New Zealand (AANZFTA) và Nghị định thư thứ hai (2023) Điều khoản thương mại điện tử cập nhật tạo nghĩa vụ tương thích về khuôn khổ pháp lý số, bao gồm điều khoản về AI, với Úc và New Zealand.

Việt Nam đã ký kết 17 FTA [122], trong đó các FTA thế hệ mới đều có chương về thương mại điện tử với điều khoản liên quan AI như FTA ASEAN (AFTA), FTA ASEAN - Trung Quốc, FTA ASEAN - Nhật Bản, FTA ASEAN - Ấn Độ, FTA Việt Nam - Nhật Bản (VJEPA), FTA Việt Nam - Chile, FTA Việt Nam - Israel, FTA Việt Nam - UAE…

Ba là, Đối tác Chiến lược Toàn diện có điều khoản về AI/công nghệ như:

(1) Đối tác Chiến lược Toàn diện Việt Nam - Hoa Kỳ (tháng 9/2023) Tuyên bố chung ghi nhận tiềm năng của Việt Nam trong chuỗi cung ứng bán dẫn và AI; khởi động Semiconductor Partnership theo CHIPS and Science Act; NVIDIA đối tác với FPT, Viettel và VinGroup triển khai AI; Microsoft - Trusting Social phát triển giải pháp AI tạo sinh cho thị trường Việt Nam. Đây là cam kết chính trị quan trọng nhất về hợp tác AI song phương gắn với đầu tư thực chất.

(2) Đối tác Chiến lược Toàn diện Việt Nam - Nhật Bản (nâng cấp 2023) Hợp tác công nghệ số và AI trong khuôn khổ đối tác mới, Nhật Bản là quốc gia có AI Governance Framework phát triển cao và là đối tác hợp tác xây dựng năng lực AI tiềm năng nhất cho Việt Nam trong khuôn khổ ASEAN - Nhật Bản.

(3) Đối tác Chiến lược Toàn diện Việt Nam - Hàn Quốc: Khuôn khổ Hợp tác CNTT Việt Nam - Hàn Quốc (ITCP); Diễn đàn Hợp tác Kỹ thuật số Việt Nam - Hàn Quốc (2024) với AI là chủ đề trọng tâm; Năm 2026 ký 73 Biên bản ghi nhớ (MOU) trong chuyến thăm của Tổng thống Lee Jae-myung, bao gồm SK Innovation và SK Telecom ký MOU với NIC về xây dựng trung tâm dữ liệu AI.

(4) Đối tác Chiến lược Toàn diện Việt Nam - Singapore (nâng cấp 2025): MOU về hợp tác kinh tế số (2022 và cập nhật 2023); MOU giữa A*STAR (Singapore) và Trung tâm Đổi mới Sáng tạo Quốc gia Việt Nam (NIC, 2023) hợp tác AI và công nghệ; Singapore là đối tác MRA tiềm năng nhất cho VN-AI Verify trong khuôn khổ ASEAN.

(5) Đối tác Chiến lược Toàn diện Việt Nam - Trung Quốc: Hợp tác công nghệ số và AI trong khuôn khổ "Cộng đồng chia sẻ tương lai" giữa hai nước bao gồm hợp tác về hạ tầng số, 5G và ứng dụng AI.

(6) Đối tác Chiến lược Toàn diện Việt Nam - Úc và Việt Nam - Pháp: Bao gồm điều khoản hợp tác công nghệ số và đổi mới sáng tạo, trong đó AI ngày càng trở thành nội dung cụ thể trong các chương trình hợp tác.

Tầng 1 thiết lập giới hạn tuyệt đối và nguyên tắc tối thượng mà không tầng nào ở dưới được vi phạm. Không giống Tầng 2 (tạo ra quyền và nghĩa vụ cụ thể), Tầng 1 thiết lập quy tắc của quy tắc, xác định điều gì có thể được quy định ở các tầng dưới và điều gì bị cấm hoàn toàn. Điều 14 khoản 2 Hiến pháp 2013 là ví dụ tiêu biểu, quy định này không nói gì về AI nhưng xác lập rằng mọi hạn chế quyền con người, bao gồm hạn chế bởi AI giám sát phải có căn cứ luật định và tỷ lệ với mục tiêu chính đáng.

3.1.2.2. Tầng thứ hai - Văn bản quy phạm pháp luật chuyên biệt và liên quan

Đây là tầng được xây dựng tích cực nhất và đã đạt tiến bộ đáng kể, đặc biệt từ năm 2024 đến nay, bao gồm:

Thứ nhất, văn bản quy phạm pháp luật chuyên biệt

Một là, Luật Trí tuệ nhân tạo 2025 (có hiệu lực từ 01/3/2026) là văn bản trung tâm của tầng này. Từ góc độ kiến trúc khung pháp lý đa tầng, Luật AI 2025 có một số đặc điểm đáng ghi nhận. Định nghĩa chức năng về "hệ thống AI" tương thích với chuẩn mực quốc tế và có khả năng bao quát công nghệ tương lai. Kiến trúc phân loại rủi ro ba tầng (cao, trung bình, thấp) với cơ chế ủy quyền linh hoạt cho Quyết định Thủ tướng Chính phủ bảo đảm danh mục có thể cập nhật mà không cần sửa luật. Triết lý kết hợp điều chỉnh rủi ro với thúc đẩy hệ sinh thái (Chương V về hỗ trợ phát triển hệ sinh thái

AI) phản ánh nhận thức đúng về chức năng kép của khung pháp lý AI. Cơ chế sandbox thử nghiệm có kiểm soát và điều khoản chuyển tiếp được quy định. Chuyên gia tư vấn chính sách (ký hiệu CG-CS-01) đánh giá Luật AI 2025 là một đạo luật vừa mang tính "bảo vệ" vừa mang tính "kiến tạo", cung cấp cho các nhà khoa học, startup và doanh nghiệp một hành lang pháp lý rõ ràng để yên tâm đầu tư, sáng tạo, đồng thời trang bị cho cơ quan quản lý những công cụ sắc bén để chặn đứng các hệ lụy tiêu cực [Xem Phụ lục 5].

Hai là, Thông tư số 05/2026/TT-BKHCN ngày 10/3/2026 ban hành Khung đạo đức AI quốc gia đặt yêu cầu áp dụng các biện pháp bảo mật nhằm phòng ngừa các nguy cơ như tấn công, chiếm quyền điều khiển, đầu độc dữ liệu, rò rỉ thông tin. Danh sách các nguy cơ được liệt kê cụ thể theo loại tấn công, từ tấn công đầu vào đến đầu độc dữ liệu cho thấy nhận thức kỹ thuật được cải thiện đáng kể, dù vẫn ở mức nguyên tắc chứ không phải tiêu chuẩn kỹ thuật có thể kiểm thử. Khung đạo đức quốc gia yêu cầu xem xét đầy đủ tác động đối với các nhóm dễ bị tổn thương như trẻ em, người cao tuổi, người khuyết tật và các nhóm yếu thế, một điều khoản bảo vệ chủ động ít gặp trong văn kiện tương đương ở Đông Nam Á. Về gắn nhãn nội dung AI, nhà cung cấp và bên triển khai có trách nhiệm thông báo, gắn nhãn hoặc áp dụng biện pháp phù hợp để người tiếp nhận nhận biết nội dung do AI tạo ra hoặc chỉnh sửa.

Thông tư 05/2026/TT-BKHCN áp dụng đối với cơ quan quản lý nhà nước, tổ chức, cá nhân với vai trò nhà phát triển, nhà cung cấp, bên triển khai hoặc người sử dụng hệ thống AI phục vụ quản lý nhà nước hoặc cung cấp dịch vụ công, biến đây thành văn kiện ràng buộc về đạo đức đối với toàn bộ hoạt động AI khu vực công. Các cơ quan cần công khai thông tin phù hợp, tạo điều kiện để người dân được biết, được giải thích và phản ánh khi cần thiết; đồng thời phải tham vấn các bên liên quan, chú trọng nhóm yếu thế và thường xuyên đánh giá tác động đạo đức của hệ thống AI trong quá trình vận hành. Yêu cầu đánh giá tác động đạo đức định kỳ là yêu cầu hậu kiểm bổ sung cho yêu cầu tiền kiểm trước triển khai ở nhóm phát triển, tạo ra vòng kiểm soát liên tục.

Cơ chế tự đánh giá thông qua mẫu phiếu kèm theo Thông tư 05/2026/TT-BKHCN là cơ chế kiểm định nội bộ đơn giản nhưng có thể triển khai ngay. Phụ lục của Thông tư cung cấp hướng dẫn cho người sử dụng, nhà phát triển và CQNN về cách áp dụng Khung đạo đức; ngoài ra, có mẫu phiếu tự đánh giá tuân thủ để các bên tự kiểm tra mức độ đáp ứng các nguyên tắc đạo đức. Cơ chế tự đánh giá phân tầng theo nhóm chủ thể (người dùng, nhà phát triển, CQNN) là thiết kế thực tiễn và phù hợp. Điểm còn thiếu là tiêu chí đánh giá đạt/không đạt cho từng yêu cầu cụ thể và chưa có quy trình xem xét kết quả tự đánh giá của CQNN bởi cơ quan cấp trên.

Chuyên gia CG-DN-02 khẳng định trong quá trình phỏng vấn rằng "Việt Nam hiện chưa có tổ chức nào đồng thời đáp ứng ba yêu cầu: năng lực kỹ thuật để đánh giá model phức tạp, khung pháp lý đủ rõ để làm cơ sở đánh giá và tính độc lập để tránh xung đột lợi ích. VinAI và một số đơn vị nghiên cứu có năng lực kỹ thuật nhưng chưa có tư cách pháp lý kiểm định. Cơ quan nhà nước có tư cách pháp lý nhưng thiếu năng lực kỹ thuật" [Xem Phụ lục 5].

Khung đạo đức quốc gia yêu cầu tổ chức, cá nhân tăng cường hợp tác quốc tế, tham gia sáng kiến, tiêu chuẩn, bộ quy tắc ứng xử về đạo đức AI; tận dụng sáng kiến khu vực tư nhân phục vụ lợi ích công theo điều kiện Việt Nam. Cụm từ "theo điều kiện Việt Nam" là điều khoản chủ quyền quan trọng, tương thích quốc tế có chọn lọc thay vì sao chép y nguyên. Đặc biệt, cơ chế rà soát, cập nhật định kỳ ba năm một lần hoặc khi có thay đổi lớn về công nghệ, pháp luật và thực tiễn quản lý là cơ chế thích ứng có tính ràng buộc thể chế, tạo nghĩa vụ cho Bộ KH&CN định kỳ xem xét lại.

Ba là, Nghị định 142/2026/NĐ-CP (có hiệu lực từ 01/5/2026) là văn bản hướng dẫn kỹ thuật quan trọng của Luật AI 2025. Nghị định quy định chi tiết về phân loại hệ thống AI theo mức độ rủi ro với cơ chế tự phân loại của nhà cung cấp; nghĩa vụ gắn nhãn bắt buộc cho âm thanh, hình ảnh và video mô phỏng người thật; báo cáo sự cố AI nghiêm trọng trong 72 giờ hoặc 05 ngày tuỳ tính chất; yêu cầu đánh giá tác động trước khi cơ quan nhà nước triển khai AI trong dịch vụ công; nguyên tắc không trùng lặp với pháp luật chuyên ngành trong kiểm định và phân loại. Từ góc độ kiến trúc, Nghị định 142 thể hiện một số lựa chọn đúng hướng như nguyên tắc không trùng lặp tránh tạo hai lớp yêu cầu song song; cơ chế ngoại lệ rõ ràng cho hệ thống nội bộ và thử nghiệm giảm gánh nặng tuân thủ không cần thiết và điều khoản chuyển tiếp 60 ngày có kế hoạch cụ thể. Tuy nhiên, rà soát toàn văn Luật AI 2025 và Nghị định 142 cho thấy hoàn toàn không có quy định về nghĩa vụ thông báo trước khi chấm dứt vận hành hệ thống AI, xử lý dữ liệu, thời hạn lưu trữ hồ sơ sau chấm dứt và cơ chế kế thừa nghĩa vụ.

Bốn là, Quyết định số 33/2026/QĐ-TTg ban hành danh mục hệ thống AI có rủi ro cao dẫn chiếu nhất quán Điều 9 Luật AI 2025 và khoản 1, 2 Điều 8 Nghị định 142, không tạo mâu thuẫn thuật ngữ. Doanh nghiệp có thể tự tra cứu Danh mục để xác định nghĩa vụ của mình, khắc phục trực tiếp tình trạng không biết mình có thuộc hệ thống AI rủi ro cao không trong giai đoạn đầu vận hành Nghị định 142. Danh mục tạo tiền đề rõ ràng cho Tầng 3 (kiểm định theo từng lĩnh vực) và Tầng 4 (hướng dẫn ngành theo sáu lĩnh vực đã xác định) nhưng T3, T4 chưa vận hành để hiện thực hóa liên kết này. Quyết định 33 chưa công bố chu kỳ rà soát định kỳ, chưa có mục riêng cho tác nhân AI xuyên lĩnh vực.

Thứ hai, văn bản quy phạm pháp luật liên quan:

Một là, Luật số 131/2025/QH15 sửa đổi, bổ sung một số điều của Luật SHTT tại Khoản 5 Điều 1 bổ sung quy định đột phá cho phép tổ chức, cá nhân được sử dụng "văn bản và dữ liệu về đối tượng quyền SHTT đã được công bố hợp pháp và công chúng được phép tiếp cận để phục vụ mục đích nghiên cứu khoa học, thử nghiệm, huấn luyện hệ thống AI" [83]. Đây là lần đầu tiên pháp luật Việt Nam thừa nhận rõ ràng quyền sử dụng dữ liệu có bản quyền để huấn luyện AI, giải quyết một trong những tranh cãi pháp lý lớn nhất trong phát triển AI toàn cầu. Quy định này tạo ra ngoại lệ hợp pháp cho phép các nhà phát triển AI sử dụng tác phẩm có bản quyền mà không cần xin phép hoặc trả phí, miễn là đáp ứng hai điều kiện quan trọng. Điều kiện thứ nhất là mục đích sử dụng phải là "nghiên cứu khoa học, thử nghiệm, huấn luyện hệ thống AI" - tức là giới hạn trong phạm vi phát triển công nghệ AI, không bao gồm các mục đích thương mại trực tiếp khác. Điều kiện thứ hai là "việc sử dụng này không ảnh hưởng bất hợp lý đến quyền và lợi ích hợp pháp của tác giả, chủ sở hữu quyền SHTT". Đây là điều khoản cân bằng quan trọng, bảo vệ quyền lợi của chủ sở hữu bản quyền trong khi vẫn cho phép đổi mới sáng tạo. Tuy nhiên, thuật ngữ "ảnh hưởng bất hợp lý" còn mơ hồ, cần có hướng dẫn chi tiết về các trường hợp cụ thể nào được coi là "bất hợp lý", ví dụ: liệu AI tạo ra nội dung cạnh tranh trực tiếp với tác phẩm gốc có được coi là "bất hợp lý" hay không.

Đặc biệt, đối với "văn bản và dữ liệu là đối tượng được bảo hộ quyền tác giả và quyền liên quan", Luật quy định thêm yêu cầu "việc sử dụng văn bản và dữ liệu theo quy định tại khoản này còn phải thực hiện theo quy định của Chính phủ". Điều này cho thấy Quốc hội nhận thức được tính nhạy cảm của vấn đề quyền tác giả trong huấn luyện AI và giao cho Chính phủ nhiệm vụ ban hành quy định chi tiết, có thể bao gồm cơ chế opt-out (cho phép tác giả từ chối việc sử dụng tác phẩm của mình để huấn luyện AI), cơ chế bồi thường hợp lý, hoặc các điều kiện kỹ thuật cụ thể.

Quy định này phản ánh xu hướng toàn cầu về Text and Data Mining (TDM) exception - ngoại lệ cho phép khai thác văn bản và dữ liệu phục vụ nghiên cứu và phát triển AI, tương tự như Article 4 của EU Directive on Copyright in the Digital Single Market. Tuy nhiên, quy định của Việt Nam có phạm vi hẹp hơn khi chỉ cho phép đối với dữ liệu "đã được công bố hợp pháp và công chúng được phép tiếp cận", không bao gồm dữ liệu riêng tư hay dữ liệu được bảo vệ bởi biện pháp kỹ thuật.

Về quyền SHTT đối với sản phẩm do AI tạo ra, Khoản 2 Điều 1 bổ sung khoản 5 vào Điều 6 Luật SHTT, quy định: "Căn cứ vào quy định của Luật này, Chính phủ quy định việc phát sinh, xác lập quyền sở hữu trí tuệ theo quy định tại các khoản 1, 2, 3 và 4 Điều này trong trường hợp đối tượng quyền SHTT được tạo ra có sử dụng hệ

thống AI" [83]. Đây là quy định mang tính đột phá nhưng cũng rất thận trọng của Việt Nam về vấn đề quyền SHTT đối với sản phẩm AI.

Điều 6 Luật SHTT quy định bốn nguyên tắc cơ bản về phát sinh quyền SHTT: (1) quyền tác giả phát sinh khi tác phẩm được thể hiện dưới hình thức vật chất nhất định; (2) quyền đối với đối tượng quyền sở hữu công nghiệp phát sinh từ thời điểm được cấp văn bằng bảo hộ; (3) quyền đối với giống cây trồng phát sinh từ thời điểm được cấp Bằng bảo hộ giống cây trồng; (4) quyền đối với bí mật kinh doanh phát sinh khi đáp ứng điều kiện bí mật kinh doanh. Khoản 5 mới bổ sung không thay đổi các nguyên tắc này, mà ủy quyền cho Chính phủ quy định chi tiết việc phát sinh, xác lập quyền SHTT khi có sự tham gia của AI.

Quy định này đặt ra nhiều câu hỏi quan trọng sẽ cần Chính phủ giải đáp: Ai là chủ sở hữu quyền tác giả đối với tác phẩm do AI tạo ra - người lập trình AI, người sử dụng AI, hay không ai? Liệu AI có thể được coi là "tác giả" hay "người sáng chế"? Mức độ can thiệp của con người như thế nào thì đủ để được coi là tác giả? Nếu AI tạo ra phát minh, ai có quyền đăng ký sáng chế? Những câu hỏi này phản ánh tranh luận toàn cầu về AI và SHTT, với các quốc gia có cách tiếp cận khác nhau: Mỹ hiện không công nhận bản quyền cho tác phẩm hoàn toàn do AI tạo ra; Anh cho phép bản quyền với "người sắp xếp việc tạo ra tác phẩm" làm chủ sở hữu; còn nhiều quốc gia khác vẫn đang trong quá trình nghiên cứu.

Việc Luật Việt Nam giao cho Chính phủ quy định chi tiết thể hiện sự thận trọng hợp lý, vì đây là vấn đề phức tạp, đang phát triển nhanh chóng, cần có sự linh hoạt để điều chỉnh theo thực tiễn và xu hướng quốc tế. Tuy nhiên, điều này cũng khiến các doanh nghiệp, nhà sáng tạo và nhà đầu tư khó khăn trong việc xác định quyền lợi của mình trong giai đoạn chờ Nghị định hướng dẫn.

Hai là, Luật Bảo vệ dữ liệu cá nhân 2025 nâng bảo vệ dữ liệu lên cấp luật là thành phần thiết yếu của khung pháp lý AI vì rất nhiều hệ thống AI hoạt động dựa trên dữ liệu cá nhân. Các quy định về quyền phản đối xử lý tự động và quyền không bị đưa ra quyết định tự động gây hậu quả pháp lý quan trọng mà không có sự xem xét của con người là tiến bộ thực chất bảo vệ quyền con người trong môi trường AI.

Nghị định 356/2025/NĐ-CP ngày 31/12/2025 quy định chi tiết một số điều và biện pháp thi hành Luật Bảo vệ dữ liệu cá nhân, đặc biệt là Điều 10 về BVDLCN trong hệ thống AI và vũ trụ ảo, là văn bản pháp luật quan trọng tạo ra cầu nối giữa BVDLCN và phát triển, ứng dụng AI. Khoản 1 Điều 10 Nghị định quy định rõ ràng: "Các tổ chức, cá nhân được quyền sử dụng dữ liệu cá nhân để nghiên cứu, phát triển các thuật toán tự học, hệ thống AI và các hệ thống tự động khác nhưng cần bảo đảm tuân thủ các quy định về bảo vệ dữ liệu cá nhân" [26]. Đây là quy định mang tính đột phá, lần đầu tiên

công nhận rõ ràng quyền sử dụng dữ liệu cá nhân cho mục đích phát triển AI, giải quyết một trong những vướng mắc pháp lý lớn nhất trong nghiên cứu và phát triển AI tại Việt Nam.

Quy định này tạo ra cơ sở pháp lý cho các nhà nghiên cứu, doanh nghiệp AI được sử dụng dữ liệu cá nhân để huấn luyện thuật toán máy học, phát triển mô hình AI mà không vi phạm pháp luật BVDLCN. Tuy nhiên, quyền này không phải là tuyệt đối mà đi kèm với nghĩa vụ "tuân thủ các quy định về bảo vệ dữ liệu cá nhân". Điều này có nghĩa là việc sử dụng dữ liệu cá nhân để phát triển AI vẫn phải tuân thủ các nguyên tắc cơ bản như: (1) hợp pháp, chính đáng, minh bạch; (2) đúng mục đích; (3) chính xác; (4) lưu trữ có thời hạn; (5) bảo đảm an toàn, an ninh dữ liệu; và (6) trách nhiệm giải trình theo Luật Bảo vệ dữ liệu cá nhân 2025.

Điểm đáng chú ý là Nghị định không yêu cầu bắt buộc phải có sự đồng ý của chủ thể dữ liệu cho việc sử dụng dữ liệu cá nhân để nghiên cứu, phát triển AI, miễn là tuân thủ các quy định về bảo vệ dữ liệu. Đây là cách tiếp cận cân bằng giữa bảo vệ quyền riêng tư và thúc đẩy đổi mới sáng tạo, tương tự như cách tiếp cận của EU trong GDPR về xử lý dữ liệu cho mục đích nghiên cứu khoa học. Tuy nhiên, Nghị định không làm rõ các trường hợp cụ thể nào cần có sự đồng ý và trường hợp nào không.

Về tiêu chuẩn kỹ thuật, Khoản 5 điểm a Điều 10 Nghị định 356/2025/NĐ-CP ngày 31/12/2025 yêu cầu các tổ chức, cá nhân "nghiên cứu, xây dựng và triển khai hệ thống đáp ứng các tiêu chuẩn an ninh mạng, tiêu chuẩn bảo vệ dữ liệu toàn diện cho các hệ thống AI, đặc biệt chú trọng các yếu tố: bảo mật thông tin, độ tin cậy của thuật toán, tính ổn định hệ thống và khả năng phòng chống tấn công mạng" [26]. Đây là quy định quan trọng về tiêu chuẩn kỹ thuật trong phát triển AI, nhấn mạnh bốn yếu tố cốt lõi: bảo mật thông tin, độ tin cậy của thuật toán, tính ổn định hệ thống và khả năng phòng chống tấn công mạng.

Yêu cầu về "độ tin cậy của thuật toán" đặc biệt quan trọng, thể hiện sự nhận thức về vấn đề thiên lệch, công bằng và độ chính xác trong AI. Tuy nhiên, Nghị định không cụ thể hóa các tiêu chuẩn này là gì, không đưa ra phương pháp đánh giá cụ thể, cũng không tham chiếu đến các tiêu chuẩn quốc tế như ISO/IEC 42001 về hệ thống quản lý AI, ISO/IEC 23894 về đánh giá rủi ro AI, hay NIST AI Risk Management Framework. Điều này có thể gây khó khăn cho doanh nghiệp trong việc xác định cụ thể cần tuân thủ tiêu chuẩn nào.

Về cơ chế giám sát, Khoản 5 điểm b Điều 10 Nghị định 356/2025/NĐ-CP ngày 31/12/2025 thiết lập "cơ chế giám sát hoạt động của hệ thống AI trên hai phương diện: giám sát của CQNN có thẩm quyền; trách nhiệm giải trình đối với chủ thể dữ liệu cá nhân của bên kiểm soát dữ liệu cá nhân, bên kiểm soát và xử lý dữ liệu cá nhân" [26]. Đây là cơ chế giám sát kép, kết hợp giám sát nhà nước và trách nhiệm giải trình với

chủ thể dữ liệu, tạo ra hệ thống checks and balances để bảo vệ quyền của cá nhân. Đây là cơ chế giám sát toàn diện, kết hợp giám sát từ bên ngoài và từ bên trong.

Khoản 5 điểm c yêu cầu phát triển hệ thống giám sát và cảnh báo sớm các nguy cơ an ninh mạng - đây là yêu cầu về giám sát kỹ thuật tự động để phát hiện sớm các rủi ro. Khoản 5đ quy định thực hiện đánh giá tuân thủ các quy định về BVDLCN định kỳ 01 năm/lần đảm bảo hệ thống AI luôn tuân thủ các quy định về bảo vệ dữ liệu trong suốt quá trình vận hành.

Ba là, Điểm g khoản 2 Điều 7 Luật An ninh mạng số 116/2025/QH15 nghiêm cấm: "Sử dụng AI hoặc công nghệ mới để giả mạo video, hình ảnh, giọng nói của người khác trái quy định của pháp luật; tạo lập, đăng tải, phát tán thông tin quy định tại khoản 1 Điều này" [87]. Đây là quy định đầu tiên trong pháp luật Việt Nam cấm rõ ràng việc sử dụng AI để tạo deepfake, thể hiện trách nhiệm pháp lý hình sự và hành chính đối với hành vi lạm dụng AI. Quy định này có ba yếu tố cấu thành quan trọng. Thứ nhất**,** về công nghệ: "sử dụng AI hoặc công nghệ mới" - quy định này không giới hạn chỉ ở AI mà mở rộng cho "công nghệ mới", thể hiện tính dự báo để bao quát cả các công nghệ tương lai có khả năng tạo nội dung giả mạo. Điều này quan trọng vì công nghệ deepfake không chỉ dựa vào AI mà có thể kết hợp nhiều công nghệ khác như computer graphics, computer vision, hay các kỹ thuật xử lý đa phương tiện tiên tiến. Thứ hai, về đối tượng bị giả mạo: "video, hình ảnh, giọng nói của người khác" - quy định này bảo vệ ba dạng biểu hiện nhân thân chính: hình ảnh khuôn mặt và cơ thể (video, hình ảnh) và đặc điểm sinh trắc học giọng nói (voice). Việc liệt kê cụ thể ba dạng này cho thấy sự nhận thức rõ ràng về các công nghệ deepfake phổ biến hiện nay: face swap (hoán đổi khuôn mặt), face reenactment (tái hiện biểu cảm khuôn mặt) và voice cloning (nhân bản giọng nói). Tuy nhiên, quy định chưa đề cập đến các dạng giả mạo khác như text-based deepfake (giả mạo văn bản, ví dụ AI tạo email giả mạo phong cách viết của một người cụ thể) hoặc gesture/body movement deepfake (giả mạo cử chỉ, chuyển động cơ thể). Thứ ba, về tính chất vi phạm: "trái quy định của pháp luật" - đây là điều khoản quan trọng nhất, phân biệt giữa sử dụng AI hợp pháp và bất hợp pháp để tạo nội dung giả mạo. Không phải mọi hành vi sử dụng AI để tạo video, hình ảnh, giọng nói giả mạo đều bị cấm, mà chỉ những hành vi "trái quy định của pháp luật". Điều này mở ra không gian hợp pháp cho các mục đích chính đáng như: (1) giáo dục và nghiên cứu khoa học; (2) nghệ thuật và giải trí (ví dụ: phim ảnh, quảng cáo có sự đồng ý của người được mô phỏng); (3) satire và parody (châm biếm và bắt chước theo phong cách hài hước) - các thể loại được bảo vệ bởi quyền tự do ngôn luận ở nhiều quốc gia; (4) các mục đích y tế (ví dụ: tạo giọng nói tổng hợp cho người mất khả năng nói); (5) accessibility (khả năng tiếp cận, ví dụ: tạo video mô phỏng người thật để hỗ trợ người khuyết tật).

Tuy nhiên, cụm từ "trái quy định của pháp luật" còn mơ hồ và cần được làm rõ. Các "quy định của pháp luật" ở đây có thể bao gồm: Luật SHTT (quyền chân dung, quyền hình ảnh); Bộ luật Dân sự [85](quyền nhân thân, danh dự, nhân phẩm, uy tín); Luật Bảo vệ dữ liệu cá nhân (dữ liệu sinh trắc học); Bộ luật Hình sự [86] (tội vu khống, tội làm nhục người khác, tội lừa đảo chiếm đoạt tài sản); và các luật chuyên ngành khác. Việc thiếu hướng dẫn cụ thể về trường hợp nào là "trái quy định của pháp luật" có thể dẫn đến áp dụng pháp luật không thống nhất. Trường hợp nào được coi là hợp pháp? Ví dụ, một công ty quảng cáo sử dụng AI để tái tạo hình ảnh của một người nổi tiếng đã qua đời (với sự đồng ý của gia đình) để xuất hiện trong quảng cáo có vi phạm không? Một nhà làm phim sử dụng deepfake để "trẻ hóa" diễn viên có cần giấy phép đặc biệt không? Một người dùng tạo video deepfake của chính mình để bảo vệ quyền riêng tư khi tham gia video call có vi phạm không?

Về yêu cầu đồng ý, Luật không quy định rõ ràng việc có cần sự đồng ý của người được mô phỏng hay không. Trong khi các nước như Mỹ (bang California) đã ban hành AB 602 (2019) cấm tạo deepfake khiêu dâm mà không có sự đồng ý [138] và AB 730 (2019) cấm deepfake chính trị trong khoảng thời gian gần bầu cử [139], Việt Nam chưa có quy định cụ thể tương tự.

Về trách nhiệm của nền tảng, các nền tảng mạng xã hội, dịch vụ chia sẻ video có trách nhiệm gì trong việc phát hiện, gỡ bỏ deepfake, họ có bị coi là đồng phạm nếu không kiểm duyệt nội dung deepfake vi phạm pháp luật không. Luật An ninh mạng 2025 đã quy định trách nhiệm của doanh nghiệp cung cấp dịch vụ trên mạng, nhưng chưa cụ thể cho trường hợp deepfake.

Về chế tài xử phạt, Luật quy định hành vi bị nghiêm cấm nhưng chưa quy định cụ thể mức phạt hành chính hoặc khung hình phạt. Việc xử lý vi phạm sẽ phụ thuộc vào Nghị định về xử phạt vi phạm hành chính trong lĩnh vực an ninh mạng và các điều luật hình sự liên quan (tội vu khống, tội làm nhục người khác, tội lừa đảo, v.v.).

Về chứng cứ và phương pháp phát hiện, làm thế nào để chứng minh một video là deepfake trong tố tụng hình sự hoặc hành chính? Ai có thẩm quyền giám định? Các phương pháp kỹ thuật phát hiện deepfake có được chấp nhận làm chứng cứ không? Những câu hỏi này cần được giải đáp để đảm bảo tính khả thi của quy định.

Quy định còn cấm "tạo lập, đăng tải, phát tán thông tin quy định tại khoản 1 Điều này". Khoản 1 Điều 7 liệt kê các thông tin bị cấm trên không gian mạng, bao gồm: thông tin chống Nhà nước, gây rối an ninh, xuyên tạc lịch sử, phá hoại khối đại đoàn kết, xúc phạm dân tộc, tôn giáo, kích động bạo lực, tuyên truyền chiến tranh, khiêu dâm, mại dâm, tệ nạn xã hội, tội phạm, thông tin giả mạo gây hoang mang dư luận, xâm phạm an ninh quốc gia, trật tự an toàn xã hội. Điều này có nghĩa là việc sử

dụng AI để tạo deepfake nhằm phát tán các loại thông tin trên sẽ bị nghiêm cấm và xử lý nghiêm khắc hơn.

Quy định này đặc biệt quan trọng trong bối cảnh công nghệ deepfake ngày càng tinh vi và dễ tiếp cận. Theo báo cáo của các tổ chức quốc tế, số lượng nội dung deepfake sẽ tăng 900% mỗi năm [253], số lượng video deepfake được chia sẻ tăng từ 500.000 năm 2023 lên 8 triệu năm 2025 [263], với các trường hợp sử dụng deepfake để: (1) tạo video khiêu dâm giả mạo người nổi tiếng hoặc người thường (chiếm đa số các deepfake hiện nay); (2) lừa đảo tài chính (ví dụ: CEO fraud - giả mạo giám đốc công ty để yêu cầu chuyển tiền [278]); (3) can thiệp bầu cử (tạo video giả mạo chính trị gia); (4) tống tiền, uy hiếp; (5) phá hoại danh tiếng cá nhân và tổ chức.

Luật An ninh mạng số 116/2025/QH15 không quy định cụ thể về cơ chế giám sát các hệ thống AI tạo deepfake. Tuy nhiên, với tư cách là luật về an ninh mạng, Luật trao cho các cơ quan an ninh mạng (đặc biệt là Bộ Công an và các lực lượng an ninh) thẩm quyền giám sát, phát hiện, ngăn chặn và xử lý các hành vi vi phạm trên không gian mạng, bao gồm cả việc sử dụng AI để tạo nội dung giả mạo. Luật An ninh mạng 2025 đã quy định các biện pháp bảo vệ an ninh mạng như giám sát, phân tích, đánh giá, cảnh báo nguy cơ đe dọa an ninh mạng; phòng ngừa, phát hiện, ngăn chặn, xử lý hành vi vi phạm an ninh mạng; yêu cầu doanh nghiệp cung cấp thông tin, dữ liệu phục vụ điều tra; thu thập, khai thác, xử lý thông tin, dữ liệu về an ninh mạng. Các biện pháp này có thể được áp dụng để giám sát việc sử dụng AI tạo deepfake.

Tuy nhiên, giám sát deepfake đặt ra thách thức kỹ thuật lớn. Thứ nhất, khối lượng nội dung khổng lồ trên internet khiến việc giám sát thủ công không khả thi. Thứ hai, công nghệ deepfake ngày càng tinh vi, khó phân biệt với nội dung thật bằng mắt thường. Thứ ba, deepfake có thể được tạo ra và phát tán nhanh chóng trên nhiều nền tảng khác nhau. Do đó, cơ chế giám sát hiệu quả cần kết hợp công nghệ AI phát hiện deepfake tự động; hợp tác với các nền tảng mạng xã hội, dịch vụ chia sẻ video; cơ chế báo cáo từ người dùng; giáo dục nâng cao nhận thức cộng đồng về deepfake.

Một số hành vi liên quan đến AI trong lĩnh vực dữ liệu cá nhân đã bị xử phạt như Nghị định 330/2026/NĐ-CP xử phạt hành vi dùng AI hoặc deepfake để giả mạo dữ liệu sinh trắc học nhằm xác thực tài khoản trái phép.

Bốn là, các luật nền bao gồm Bộ luật Dân sự 2015, Bộ luật Hình sự 2015 (sửa đổi) tạo thành hệ thống pháp lý nền tảng có thể áp dụng cho AI trong các tình huống chuyên ngành. Tuy nhiên, các văn bản này đều được xây dựng cho đối tượng khác và áp dụng cho AI như hệ quả không hoàn toàn có chủ đích tạo ra nhiều khoảng trống và mâu thuẫn chưa được xử lý, đặc biệt trong phân định trách nhiệm dân sự và hình sự khi AI agents tự chủ gây hậu quả pháp lý.

Xét theo tính đầy đủ, tầng này còn ba khoảng trống: (i) thiếu văn bản chuyên ngành cho y tế, giáo dục và giao thông, (ii) chưa có nghị định xử phạt chuyên biệt về AI, (iii) việc hướng dẫn ngoại lệ về khai thác văn bản và dữ liệu cho huấn luyện AI vẫn đang chờ Chính phủ.

Xét theo tính thống nhất và đồng bộ, các văn bản có hiệu lực lệch nhau qua bốn mốc. Hai hệ phân loại rủi ro theo Luật AI và Luật BVDLCN mới được nối với nhau ở mức hồ sơ. Xét theo tính thích ứng, đã có cơ chế sửa Danh mục và cơ chế phân loại lại nhưng chưa có điều khoản đánh giá định kỳ.

3.1.2.3. Tầng thứ ba - Tiêu chuẩn kỹ thuật có giá trị pháp lý

Hệ thống pháp luật về tiêu chuẩn và quy chuẩn kỹ thuật tại Việt Nam được điều chỉnh bởi Luật Tiêu chuẩn và Quy chuẩn kỹ thuật số 68/2006/QH11 được Quốc hội thông qua ngày 29/6/2006, có hiệu lực từ 01/01/2007 và được sửa đổi, bổ sung một số điều theo Luật số 35/2018/QH14 ngày 13/6/2018, Luật số 70/2025/QH15 ngày 14/6/2025 [84]. Luật này thiết lập khung pháp lý tổng thể cho việc xây dựng, ban hành và áp dụng tiêu chuẩn, quy chuẩn kỹ thuật trong mọi lĩnh vực, bao gồm cả công nghệ thông tin và AI. Văn bản hợp nhất 56/VBHN-VPQH 2025 Điều 3 định nghĩa các khái niệm cơ bản, trong đó Khoản 1 quy định *"Tiêu chuẩn* là quy định về đặc tính kỹ thuật và yêu cầu quản lý dùng làm chuẩn để phân loại, đánh giá sản phẩm, hàng hóa, dịch vụ, quá trình, môi trường và các đối tượng khác trong hoạt động kinh tế - xã hội". Định nghĩa này hoàn toàn có thể áp dụng cho hệ thống AI khi coi AI là một loại sản phẩm hoặc dịch vụ công nghệ. Khoản 17 định nghĩa "Tiêu chuẩn quốc gia là tiêu chuẩn được CQNN có thẩm quyền công bố theo trình tự, thủ tục quy định tại Luật này". Tiêu chuẩn quốc gia ở Việt Nam được ký hiệu là TCVN (Tiêu chuẩn Việt Nam). Theo khoản 2 Điều 10a, tiêu chuẩn quốc gia có tính chất tự nguyện áp dụng: "Tiêu chuẩn quốc gia được áp dụng trên cơ sở tự nguyện. Toàn bộ hoặc một phần tiêu chuẩn quốc gia, tiêu chuẩn quốc tế, tiêu chuẩn khu vực, tiêu chuẩn nước ngoài trở thành bắt buộc áp dụng khi được viện dẫn trong quy chuẩn kỹ thuật, văn bản quy phạm pháp luật". Điều này có nghĩa là các doanh nghiệp AI có thể tự nguyện áp dụng TCVN để nâng cao chất lượng và uy tín, nhưng không bị bắt buộc trừ khi TCVN đó được viện dẫn trong QCVN hoặc văn bản pháp luật khác. Khoản 2 Điều 3 định nghĩa *"Quy chuẩn kỹ thuật* là quy định về mức giới hạn của đặc tính kỹ thuật và yêu cầu quản lý mà sản phẩm, hàng hóa, dịch vụ, quá trình, môi trường và các đối tượng khác trong hoạt động kinh tế - xã hội phải tuân thủ để bảo đảm an toàn vệ sinh, sức khỏe con người; bảo vệ động vật, thực vật, môi trường; bảo vệ lợi ích và an ninh quốc gia, quyền lợi của người tiêu dùng và các yêu cầu thiết yếu khác; bảo đảm tuân thủ cam kết trong các điều ước quốc tế mà nước Cộng hòa xã hội chủ nghĩa Việt Nam là thành viên và không tạo rào cản không cần thiết cho thương mại.

Quy chuẩn kỹ thuật do CQNN, người có thẩm quyền ban hành dưới dạng văn bản quy phạm pháp luật để bắt buộc áp dụng". Điều 26 Quy chuẩn kỹ thuật quốc gia ký hiệu là QCVN. Khác với TCVN, QCVN mang tính bắt buộc áp dụng theo quy định tại Điều 26a khoản 2: "Quy chuẩn kỹ thuật được áp dụng bắt buộc, là cơ sở cho hoạt động đánh giá sự phù hợp".

Đến thời điểm tháng 9/2026, Việt Nam đã ban hành chính thức 8 TCVN về AI. Đó là TCVN 13902:2023 CNTT - AI - Các khái niệm và thuật ngữ AI; TCVN 13903: 2023 CNTT - AI - Tổng quan về tính đáng tin cậy trong AI; TCVN 14199-1:2024 AI - Quy trình vòng đời và yêu cầu chất lượng - Phần 1: Mô hình Meta chất lượng; TCVN 14199-2:2024 AI - Quy trình vòng đời và yêu cầu chất lượng - Phần 2: Độ bền vững; TCVN 14465:2025 CNTT - AI - Khung quản lý quá trình phân tích dữ liệu lớn; TCVN ISO/IEC 42001:2025 CNTT - AI - Hệ thống quản lý; TCVN 14365:2025 CNTT - AI - Tổng quan về các phương pháp tính toán cho các hệ thống AI và TCVN 14364:2025 CNTT - AI - Khung cho hệ thống AI sử dụng học máy [125]. Đây là điểm khởi đầu cần thiết nhưng chưa đủ để thực thi các yêu cầu pháp lý cụ thể.

Các tiêu chuẩn quốc tế về AI đang được nhiều tổ chức tại Việt Nam tham khảo bao gồm:

Tiêu chuẩn ISO/IEC về AI: ISO/IEC JTC 1/SC 42 [233] là Ủy ban kỹ thuật phụ trách xây dựng các tiêu chuẩn quốc tế về AI. Từ năm 2021 đến tháng 9/2026, 145 tiêu chuẩn quan trọng đã được ban hành [234]. Việt Nam đã có những bước đi rất nhanh chóng trong việc nội hóa các tiêu chuẩn ISO/IEC, đặc biệt là TCVN ISO/IEC 42001:2025. Đây là công cụ pháp lý - kỹ thuật quan trọng nhất để các doanh nghiệp trong nước có thể đạt được chứng chỉ quốc tế, giúp sản phẩm AI của Việt Nam dễ dàng thâm nhập vào thị trường EU (tuân thủ AI Act) và Mỹ.

Tiêu chuẩn IEEE về AI: Các tiêu chuẩn của Viện Kỹ sư Điện và Điện tử (IEEE) như IEEE 7000 series về AI Ethics, IEEE 2846-2022 về đánh giá thiên kiến trong hệ thống AI, IEEE 7001-2021 về minh bạch của các hệ thống tự động [335]. IEEE là một trong những tổ chức dẫn đầu thế giới về xây dựng tiêu chuẩn kỹ thuật và các tiêu chuẩn của IEEE thường được cộng đồng công nghệ quốc tế chấp nhận rộng rãi. Hiện nay, các cơ quan quản lý tại Việt Nam đang khuyến khích doanh nghiệp tự nguyện áp dụng IEEE 7000 series như một bộ quy tắc ứng xử đạo đức trước khi các văn bản quy phạm pháp luật bắt buộc được ban hành.

Tuy nhiên, việc tham chiếu đến các tiêu chuẩn quốc tế này gặp một số hạn chế. Thứ nhất, các tiêu chuẩn quốc tế thường được viết bằng tiếng Anh và đòi hỏi trình độ chuyên môn cao để hiểu và áp dụng đúng, khiến nhiều doanh nghiệp nhỏ và vừa tại Việt Nam gặp khó khăn. Thứ hai, các tiêu chuẩn quốc tế có thể không hoàn toàn phù

hợp với điều kiện cụ thể của Việt Nam về mặt pháp lý, văn hóa, xã hội. Thứ ba, việc áp dụng tiêu chuẩn quốc tế thường tốn kém do phải mua bản quyền tiêu chuẩn, thuê tư vấn và có thể phải chứng nhận bởi các tổ chức quốc tế.

Về thách thức trong xây dựng tiêu chuẩn kỹ thuật AI:

Một là, tốc độ phát triển cực nhanh của công nghệ: Trong khi quy trình xây dựng tiêu chuẩn quốc gia theo Luật Tiêu chuẩn và Quy chuẩn kỹ thuật thường mất từ mười hai đến hai mươi bốn tháng, công nghệ AI có thể thay đổi căn bản chỉ trong vài tháng. Ví dụ, từ khi GPT-3 ra đời năm 2020 đến GPT-4 năm 2023, năng lực của mô hình ngôn ngữ lớn đã tăng lên gấp nhiều lần. Nếu xây dựng tiêu chuẩn quá chi tiết dựa trên công nghệ hiện tại, tiêu chuẩn đó có thể trở nên lỗi thời ngay khi được ban hành.

Hai là, thách thức về năng lực kỹ thuật: Việc xây dựng tiêu chuẩn kỹ thuật về AI đòi hỏi sự kết hợp giữa kiến thức chuyên sâu về công nghệ AI, hiểu biết về pháp luật và chính sách và kinh nghiệm thực tiễn trong triển khai AI. Hiện tại, Việt Nam đang thiếu nguồn nhân lực có đủ ba yếu tố này. Nhiều chuyên gia AI có kiến thức kỹ thuật tốt nhưng thiếu hiểu biết về các vấn đề pháp lý, đạo đức và xã hội. Ngược lại, nhiều cán bộ làm công tác tiêu chuẩn hóa có kinh nghiệm về quy trình xây dựng tiêu chuẩn nhưng thiếu hiểu biết sâu về công nghệ AI.

Ba là, thách thức về cân bằng giữa yêu cầu kỹ thuật và khả năng tuân thủ: Tiêu chuẩn kỹ thuật về AI cần đủ nghiêm ngặt để đảm bảo an toàn và chất lượng, nhưng cũng phải khả thi để các doanh nghiệp, đặc biệt là SME có thể tuân thủ được. Nếu tiêu chuẩn quá cao sẽ tạo ra rào cản gia nhập thị trường và kìm hãm đổi mới sáng tạo. Nếu tiêu chuẩn quá thấp sẽ không đạt được mục đích bảo vệ người dùng và lợi ích công cộng.

Đây là tầng yếu nhất và thiếu hụt nhất trong toàn bộ khung pháp lý Việt Nam, khoảng cách giữa yêu cầu pháp lý của Luật AI 2025 và Nghị định 142 với hệ thống tiêu chuẩn kỹ thuật quốc gia là khá lớn. Tiêu chuẩn còn thiếu là TCVN về kiểm định hệ thống AI trong khi Nghị định 142 yêu cầu đánh giá sự phù hợp cho hệ thống rủi ro cao nhưng không có tiêu chuẩn nào định nghĩa "sự phù hợp" là gì và đo lường theo phương pháp nào. Chưa có TCVN về AI agents và khả năng kiểm soát trong khi thực tiễn AI agents đang bùng nổ. Chưa có TCVN về gắn nhãn nội dung AI tổng hợp trong khi Nghị định 142 yêu cầu mà không xác định tiêu chuẩn kỹ thuật tương ứng. Chưa có TCVN về đánh giá công bằng thuật toán trong khi thuật toán tín dụng, tuyển dụng và dịch vụ công đang tạo ra rủi ro phân biệt đối xử thực tiễn. Giai đoạn chấm dứt vận hành làm phát sinh nhu cầu về tiêu chuẩn về xóa dữ liệu an toàn, tiêu chuẩn về định dạng lưu trữ nhật ký vận hành để phục vụ tranh chấp phát sinh sau chấm dứt và tiêu chuẩn về chuyển giao dữ liệu cho dịch vụ thay thế (nếu người dùng lựa chọn). Hệ quả thực tiễn của khoảng trống này là Nghị định 142 quy định nhà cung cấp tự phân loại rủi

ro và chịu trách nhiệm trước pháp luật về tính chính xác, trung thực của kết quả phân loại nhưng chưa có tiêu chuẩn nào hướng dẫn. Điều này khiến doanh nghiệp bị ràng buộc trách nhiệm nhưng không biết cụ thể phải làm gì để tuân thủ. Khoảng trống trở nên rõ ràng và cấp bách hơn sau khi Quyết định 33/2026/QĐ-TTg được ban hành, chưa có TCVN chuyên biệt cho sáu lĩnh vực Danh mục xác định là rủi ro cao. Quyết định 33 gián tiếp làm bộc lộ rõ hơn khoảng cách giữa T2 và T3. Danh mục xác định lĩnh vực giao thông vận tải có 31 hệ thống rủi ro cao (bao gồm hệ thống điều khiển tín hiệu đường sắt, đường bộ, hàng không) nhưng chưa có TCVN xác định phương pháp kiểm định cho các hệ thống này.

3.1.2.4. Tầng thứ tư - Luật mềm và hướng dẫn ngành

Thứ nhất, khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam có thể tham chiếu giá trị từ các văn bản luật mềm quốc tế sau:

Một là, Văn kiện trong khuôn khổ LHQ bao gồm Khuyến nghị về Đạo đức Trí tuệ nhân tạo của UNESCO (2021), Nghị quyết LHQ về AI vì sự phát triển bền vững (2024), Nghị quyết LHQ A/RES/78/311 Quy trình đa phương về quản trị AI (2024).

Hai là, Văn kiện trong khuôn khổ OECD và G20 bao gồm Nguyên tắc AI của OECD (2019, cập nhật 2024), Nguyên tắc AI G20 (2019).

Ba là, Văn kiện trong khuôn khổ G7 như Tiến trình AI Hiroshima (2023-2024).

Bốn là, Văn kiện trong khuôn khổ ASEAN bao gồm:

(1) Hướng dẫn ASEAN về Quản trị và Đạo đức AI (2024) Khung quản trị AI khu vực với nguyên tắc về minh bạch, công bằng, trách nhiệm giải trình và bảo mật tương thích với nguyên tắc OECD và UNESCO. Các quốc gia thành viên (bao gồm Việt Nam) được khuyến khích tích hợp vào pháp luật quốc gia.

(2) Hướng dẫn ASEAN mở rộng về Quản trị và Đạo đức AI - AI Tạo sinh (2025) Bổ sung 9 lĩnh vực trọng tâm đặc thù cho AI tạo sinh: nguồn gốc nội dung, báo cáo sự cố, an ninh mạng chống prompt injection, yêu cầu minh bạch cho hệ thống GenAI. Đặc biệt có phần về ngôn ngữ tiếng Việt trong phát triển AI tổng quát.

(3) Tuyên bố Hà Nội về Hợp tác Kỹ thuật số ASEAN (2026) do Việt Nam đăng cai. Nội dung về AI: xác nhận AI là "tầng thông minh" của chuyển đổi kỹ thuật số ASEAN; kêu gọi hài hòa hóa chính sách AI khu vực; thúc đẩy hợp tác an toàn AI.

(4) Tuyên bố thành lập ASEAN AI Safety Network (2025) Mạng lưới khu vực nhằm tăng cường năng lực AI, sẵn sàng pháp lý và biện pháp bảo vệ. Đặc biệt liên quan đến phối hợp ASEAN trong phòng chống deepfake, fake news và thao túng thông tin bằng AI.

(5) Lộ trình AI có trách nhiệm ASEAN (2025-2030) xác định ưu tiên hành động tập thể cho quản trị AI khu vực đến 2030 trong đó Việt Nam tham gia xây dựng và cam kết thực hiện.

Thứ hai, tầng luật mềm trong nước hiện còn mỏng và thiếu hệ thống. Một là Quyết định 1290/QĐ-BKHCN ngày 11/6/2024 của Bộ KH&CN về một số nguyên tắc phát triển AI có trách nhiệm - soft law đầu tiên của Việt Nam về AI, tuy nhiên còn ở mức chung chung và chưa phát triển thành hệ thống hướng dẫn ngành dọc chi tiết. Hai là, Công văn số 557/BKHCN-CĐSQG ngày 31/3/2025 của Bộ KH&CN bản chất là tài liệu hướng dẫn nghiệp vụ chuyên ngành hành chính. Nội dung công văn hướng dẫn công chức, viên chức và người lao động trong cơ quan nhà nước sử dụng trợ lý ảo - chatbot AI phục vụ công việc. Văn bản thiết lập các giới hạn hành vi tự quản như: tuyệt đối không nạp dữ liệu bí mật nhà nước, thông tin cá nhân lên chatbot công cộng; bắt buộc kiểm chứng kết quả từ AI và cấm phụ thuộc hoàn toàn vào quyết định tự động của máy móc. Ba là, Công văn số 2250/BGDĐT-GDPT ngày 12/5/2025 về việc triển khai tài liệu Hướng dẫn sử dụng Ứng dụng AI trong dạy và học do Viện Khoa học Giáo dục Việt Nam biên soạn. Bốn là, Quyết định 2557/QĐ-BKHCN ngày 25/5/2026 của Bộ KH&CN ban hành hướng dẫn về liêm chính khoa học và đạo đức nghề nghiệp trong nghiên cứu khoa học và phát triển công nghệ.

Ngày 30/9/2026, Ngân hàng Nhà nước ban hành Thông tư quy định về an toàn, quản lý rủi ro và điều kiện triển khai đối với việc ứng dụng hệ thống AI trong ngành ngân hàng. Chưa có bằng chứng công khai về việc các Bộ chuyên ngành (Bộ Y tế, Bộ GTVT, Bộ Dân tộc và Tôn giáo) đã khởi động soạn thảo hướng dẫn ứng dụng và quản trị AI đối với lĩnh vực trong thẩm quyền phụ trách. Lộ trình chuyển tiếp trong Quyết định 33 tạo ra áp lực thời gian cụ thể (khác biệt theo ngành: y tế/giáo dục/tài chính đến 01/9/2027; các lĩnh vực còn lại đến 01/3/2027) mà T4 phải đáp ứng, là cơ sở định lượng thời gian mới để đánh giá tiến độ xây dựng T4 trong các nghiên cứu tiếp theo.

Cần nhìn nhận thực chất và khẩn trương bù đắp cho tầng này mới ở mức sơ khai vì luật mềm là tầng có thể cập nhật nhanh nhất theo công nghệ, thiếu tầng này đồng nghĩa với thiếu cơ chế phản ứng nhanh; thiếu hướng dẫn ngành dọc khiến luật khung áp dụng một cách đồng đều cho các tình huống rủi ro rất khác nhau dễ vi phạm nguyên tắc tỷ lệ. Chưa có bộ hướng dẫn áp dụng theo ngành để doanh nghiệp và bên triển khai biết cách đáp ứng nghĩa vụ trong từng bối cảnh. Tầng này vì vậy chưa làm được chức năng chuyển hóa nghĩa vụ chung thành thực hành.

3.1.2.5. Tầng thứ năm - Bộ quy tắc ứng xử và tự quản ngành

Tầng bộ quy tắc ứng xử và tự quản ngành AI Việt Nam còn ở giai đoạn sơ khai, thiếu cả về số lượng lẫn chất lượng.

Thực trạng hiện có Hiệp hội Phần mềm và Dịch vụ CNTT Việt Nam (VINASA) thành lập Ủy ban Đạo đức AI năm 2024 [133] đã bắt đầu quan tâm đến quản trị AI nhưng chưa ban hành bộ quy tắc ứng xử AI ngành nào có tính ràng buộc thực chất đối với thành viên. Đáng chú ý là Bộ Quy tắc sử dụng AI có trách

nhiệm phiên bản 1.0 (DCC 1.0) bản chất là công cụ tự quản ngành do Hội Truyền thông số Việt Nam (VDCA) và Viện Nghiên cứu Chính sách và Phát triển Truyền thông (IPS) phối hợp công bố ngày 30/09/2025 [127]. Nội dung Bộ quy tắc đưa ra 6 nguyên tắc cốt lõi và 29 hướng dẫn cụ thể cho người dùng sáng tạo nội dung, cùng 5 lưu ý đặc thù cho quảng cáo trực tuyến. DCC 1.0 đóng vai trò là "kim chỉ nam" giúp các cá nhân, tòa soạn báo chí, và nhà sản xuất nội dung số tự phân định ranh giới giữa sáng tạo hợp pháp và hành vi vi phạm đạo đức truyền thông khi sử dụng AI.

Ngoài ra, Quyết định số 423/QĐ-BVHTTDL ngày 5/3/2026 ban hành Bộ Quy tắc ứng xử văn hóa trên môi trường số bản chất là Bộ quy tắc ứng xử xã hội trên không gian số do Bộ Văn hoá, Thể thao và Du lịch phát động. Nội dung định hình văn hóa ứng xử lành mạnh, yêu cầu người dùng và các nhà cung cấp dịch vụ mạng xã hội tự quản lý tài khoản, không sử dụng các công cụ công nghệ mới (như AI tạo sinh) để tạo tin giả, giả mạo tài khoản cá nhân hoặc bạo lực mạng.

Tại Diễn đàn Báo chí toàn quốc lần thứ III năm 2026, Hội Nhà báo Việt Nam đã công bố 10 quy tắc sử dụng AI trong báo chí: AI chỉ là công cụ kỹ thuật hỗ trợ, không thay thế nhà báo; không xem AI là nguồn tin; không đăng tải nguyên bản nội dung AI tạo ra; tuyệt đối không tạo hoặc phát tán thông tin giả; không sử dụng AI để tạo Deepfake hoặc giả mạo; bảo vệ nguồn tin và dữ liệu cá nhân; tôn trọng bản quyền; minh bạch khi sử dụng AI; nội dung nhạy cảm phải kiểm chứng nhiều lớp; tuân thủ pháp luật và định hướng [73].

Có thể thấy tầng bộ quy tắc ứng xử chưa đầy đủ đặc biệt nghiêm trọng trong bối cảnh bất cân xứng thông tin cao khi doanh nghiệp AI hiểu hệ thống của mình sâu hơn bất kỳ cơ quan quản lý nào và bộ quy tắc ứng xử là cơ chế duy nhất có thể chuyển kiến thức kỹ thuật đó thành cam kết điều chỉnh hành vi có thể kiểm chứng.

3.1.2.6. Tầng thứ sáu - Án lệ, lẽ công bằng và tập quán pháp

Tầng án lệ và lẽ công bằng trong khung pháp lý AI Việt Nam đang trong giai đoạn hình thành ban đầu, với nhiều khoảng trống pháp lý chưa được án lệ lấp đầy.

Về thực trạng án lệ, tính đến tháng 8/2026, chưa có án lệ chính thức nào được TAND Tối cao công bố liên quan trực tiếp đến AI. Các vụ tranh chấp liên quan đến AI đã xảy ra (đặc biệt trong lĩnh vực thương mại điện tử và lừa đảo sử dụng deepfake) nhưng chưa tạo ra án lệ có giá trị hướng dẫn. Sự thiếu vắng quy định này tạo ra sự tương phản rõ rệt so với các quốc gia phát triển đã có hệ thống pháp lý hoặc án lệ rõ ràng. Tại EU, Tòa án Công lý Châu Âu (Court of Justice of the European Union - CJEU) đã có nhiều án lệ quan trọng về hoạt động scraping, điển hình là Ryanair Ltd v. PR Aviation BV [174], từ đó thiết lập các nguyên tắc pháp lý về quyền và nghĩa vụ của các bên liên quan. Tại Hoa Kỳ, vụ án tiêu biểu hiQ Labs v. LinkedIn năm 2022 [362] đã tạo ra tiền lệ quan trọng về quyền thu thập dữ liệu công khai từ các nền tảng mạng

xã hội, khẳng định rằng việc thu thập dữ liệu công khai không vi phạm Computer Fraud and Abuse Act. Bên cạnh đó, tình huống tổ chức vận hành chấm dứt tồn tại (giải thể, phá sản) trước khi hệ thống AI được chính thức ngừng vận hành tạo ra một loại tranh chấp mới về kế thừa nghĩa vụ pháp lý liên quan đến hệ thống AI có cấu trúc tương tự nghĩa vụ hậu mãi trong pháp luật bảo vệ người tiêu dùng nhưng có đặc thù riêng do tính liên tục và tích lũy dữ liệu của AI.

Nguyên nhân do tòa án thường xử lý tranh chấp AI thông qua các quy định dân sự, hình sự chung mà không phát triển lý luận đặc thù cho AI; thiếu thẩm phán được đào tạo về kỹ thuật AI để xét xử các vụ việc phức tạp và đương sự thường không đủ năng lực pháp lý để đưa vụ việc AI ra tòa.

Về lẽ công bằng trong thực tiễn xét xử AI, khoản 2 Điều 6 Bộ luật Dân sự 2015 cho phép áp dụng lẽ công bằng khi không có quy định pháp luật hay tập quán. Đây là công cụ quan trọng để tòa án giải quyết tranh chấp AI khi phân bổ trách nhiệm giữa các bên trong chuỗi giá trị AI khi văn bản pháp luật chưa có quy định cụ thể. Tuy nhiên, thực tiễn xét xử Việt Nam hiếm khi viện dẫn lẽ công bằng một cách chủ động, xu hướng là "không có quy định thì không xử" thay vì "không có quy định thì áp dụng lẽ công bằng".

Về tập quán pháp AI, tập quán thực hành trong cộng đồng AI Việt Nam đang dần hình thành nhưng chưa đủ ổn định và phổ biến để được Nhà nước thừa nhận như nguồn pháp luật bổ sung. Một số thực hành đang trở thành chuẩn mực không chính thức (ví dụ: các nền tảng thương mại điện tử lớn tự nguyện thông báo khi dùng AI trong đề xuất giá) nhưng chưa được pháp luật công nhận.

Ba nhóm tranh chấp AI đang chờ án lệ lấp đầy là trách nhiệm dân sự khi tác nhân AI tự chủ gây thiệt hại tài chính (đặc biệt nghiêm trọng với lừa đảo deepfake, tổn thất hàng tỷ đồng/vụ); quyền tác giả đối với nội dung do AI tạo sinh; nghĩa vụ trung thực của doanh nghiệp khi người dùng không biết đang tương tác với AI và bị thiệt hại. Ba nhóm tranh chấp này không có quy phạm thực định rõ ràng và sẽ chỉ được giải quyết nhất quán khi có án lệ hoặc giải thích pháp luật chính thức từ TAND Tối cao, cho thấy cần có hướng dẫn thống nhất trước khi các bản án đầu tiên xuất hiện.

Việc Quyết định 33 xếp "tố tụng" vào nhóm sáu lĩnh vực rủi ro cao là một dữ kiện mới có ý nghĩa gián tiếp quan trọng cho T6, xác nhận ở cấp độ văn bản hành pháp cao nhất rằng AI trong lĩnh vực tố tụng được coi là có rủi ro đáng kể nhưng chưa có bằng chứng về việc điều này đã thúc đẩy hoạt động cụ thể của TAND Tối cao.

3.1.2.7. Tầng thứ bảy - Thiết chế thực thi và cơ chế giám sát

Đây là tầng quyết định hiệu lực thực tiễn của toàn bộ khung pháp lý và cũng là tầng thứ hai sau tiêu chuẩn kỹ thuật đang có vấn đề cơ cấu nghiêm trọng.

Cơ quan chủ trì Bộ KH&CN được Luật AI 2025 giao là đầu mối quản lý nhà nước về AI. Cổng thông tin điện tử một cửa về AI theo Nghị định 142 là bước xây dựng hạ tầng quản lý quan trọng. Cơ sở dữ liệu quốc gia về hệ thống AI tạo ra hạ tầng thông tin cho quản lý AI. Hiện chưa có cơ chế cập nhật trạng thái "đã chấm dứt vận hành" cho hệ thống AI đã đăng ký, cho thấy công cụ giám sát hiện có chủ yếu được thiết kế cho giai đoạn đăng ký - vận hành mà chưa tính đến khâu đóng hồ sơ khi hệ thống ngừng hoạt động, một thiếu sót về thiết kế vòng đời của chính công cụ quản lý.

Về vấn đề phân tán thẩm quyền, thực tiễn quản lý AI phân tán giữa nhiều cơ quan như Bộ Công an (an ninh mạng và tội phạm AI), NHNN (AI trong tài chính), Bộ Y tế (AI trong y tế), Bộ GD&ĐT (AI trong giáo dục)… mà chưa có cơ chế điều phối liên ngành thường trực có thực quyền.

Nghị định 142 xây dựng mô hình quản lý theo phân công thẩm quyền đa cơ quan, trong đó Bộ KH&CN giữ vai trò đầu mối điều phối. Điều 41 giao Bộ KH&CN quản lý và vận hành Cổng thông tin điện tử một cửa về AI và Cơ sở dữ liệu quốc gia về hệ thống AI, đây là hạ tầng số của toàn bộ hệ thống giám sát. Điều 42 phân công Bộ Công an chịu trách nhiệm xây dựng Trung tâm dữ liệu quốc gia và bảo đảm an ninh mạng; Điều 43 quy định các bộ ngành phối hợp theo chức năng. Điều 44 phân cấp thẩm quyền cấp phép thử nghiệm sandbox theo phạm vi địa lý. UBND cấp tỉnh cho thử nghiệm trong phạm vi một tỉnh, Bộ chuyên ngành cho thử nghiệm liên tỉnh trong lĩnh vực mình quản lý, Bộ KH&CN cho các trường hợp phức tạp liên ngành. Cơ chế phân cấp này có tính thực tiễn cao, tận dụng được chuyên môn ngành dọc và năng lực quản lý địa phương. Tuy nhiên, sự phân tán thẩm quyền cũng tiềm ẩn nguy cơ thiếu nhất quán trong áp dụng pháp luật, đặc biệt với các hệ thống AI hoạt động xuyên lĩnh vực. Một khung pháp lý AI dù tốt đến đâu cũng vô nghĩa nếu thiếu cơ quan có thẩm quyền, năng lực và nguồn lực để thực thi. OECD, UN và EU đều nhấn mạnh tính độc lập và năng lực chuyên môn của cơ quan quản lý AI là điều kiện tiên quyết. Việt Nam có thể học hỏi từ kinh nghiệm quốc tế nhằm nâng cao hiệu quả triển khai Luật AI trên thực tế.

Nguyên tắc không trùng lặp trong Nghị định 142 xử lý vấn đề phân loại và kiểm định nhưng không giải quyết vấn đề điều phối phản ứng khi sự cố AI xảy ra. Khoảng trống điều phối đặc biệt lộ rõ trong xử lý lừa đảo deepfake khi Bộ Công an điều tra tội phạm, Bộ KHCN xử lý nội dung, NHNN cảnh báo tổ chức tín dụng tạo ra ba luồng song song nhưng không phối hợp với nhau, không có cơ quan tổng hợp chung.

Về năng lực kiểm định, chưa có tổ chức kiểm định AI nào tại Việt Nam hiện đủ cả ba điều kiện năng lực kỹ thuật, tính độc lập pháp lý và quyền truy cập dữ liệu vận hành. Yêu cầu kiểm định trong Luật AI 2025 đứng trước nguy cơ biến thành kiểm định hồ sơ, kiểm tra tài liệu kỹ thuật mà không kiểm tra hành vi thực tế của hệ thống AI.

Về công cụ giám sát, chưa có hệ thống giám sát thuật toán liên tục; công cụ phân tích dữ liệu sự cố AI có tính hệ thống và cơ chế phát hiện sớm khi AI học được hành vi có hại sau khi triển khai.

Như vậy, nhìn tổng thể toàn bộ các tầng, mối quan hệ giữa bảy tầng của khung pháp lý không phải quan hệ một chiều áp đặt từ trên xuống mà là mạng lưới tương tác lẫn nhau, tập trung vào hai chiều chính. Chiều phụ thuộc lấy nền tảng làm gốc, trong đó các quy định ngành, tiêu chuẩn kỹ thuật hay các cơ quan xử phạt (các tầng từ 3 đến 7) không thể tự nhiên hoạt động. Tất cả đều phải dựa vào "xương sống" là Hiến pháp (Tầng 1) và Luật (Tầng 2) để có quyền lực và cơ sở pháp lý hợp lệ. Chiều phản hồi lấy thực tiễn để sửa luật thể hiện ở các cơ quan giải quyết tranh chấp và thực thi (Tầng 6, Tầng 7) khi làm việc sẽ thu thập các khó khăn, sự cố từ thực tế. Những dữ liệu này được dùng để "báo cáo" ngược lên, giúp Nhà nước sửa đổi, cập nhật Luật (Tầng 2) cho phù hợp. Ngược lại, khi Luật thay đổi, các hướng dẫn bên dưới cũng bắt buộc phải cập nhật theo.

Tổng hợp lại, khung pháp lý đạt yêu cầu tối thiểu ở cả 5 tiêu chí. Tuy vậy, khung pháp lý chưa đáp ứng đầy đủ các tiêu chí, đặc biệt là tính đầy đủ và cân đối. Ba tầng chịu lực là tầng 1, tầng 2 và tầng 7 đều đạt từ mức tối thiểu trở lên, trong đó tầng 2 đạt mức cơ bản đầy đủ. Tầng 5 và tầng 6 vẫn chưa có, nên khung pháp lý chưa đầy đủ và cân đối. Sự mất cân đối nằm ở chỗ những tầng mà các tầng khác phụ thuộc vào, tức tầng tiêu chuẩn và tầng thiết chế, lại phát triển chậm nhất.

3.2. THỰC TRẠNG HOẠT ĐỘNG XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM

3.2.1. Về chủ thể xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

3.2.1.1. Nhóm chủ thể có thẩm quyền quyết định

Thứ nhất, cơ quan lập pháp

Về ưu điểm, Quốc hội đã hoàn thành đúng tiến độ việc thông qua Luật AI 2025, Luật CNCNS 2025, Luật BVDLCN 2025 và Luật Ban hành VBQPPL 2025 trong bối cảnh áp lực chính trị mạnh từ Nghị quyết 57-NQ/TW, thể hiện năng lực xử lý khối lượng công việc lập pháp lớn trong thời gian ngắn. Nguyên nhân của ưu điểm này nằm ở quy trình lập pháp đã được chuẩn hóa theo Luật Ban hành VBQPPL 2025 cung cấp khung thủ tục rõ ràng để cơ quan lập pháp vận hành hiệu quả và ý chí chính trị thống nhất từ định hướng của Đảng tạo ra sự đồng thuận cao trong quá trình biểu quyết, giảm thời gian tranh luận kéo dài như thường thấy đối với các đạo luật có tính nhạy cảm chính trị cao hơn.

Về hạn chế, Quốc hội chủ yếu thực hiện chức năng thẩm tra và thông qua văn bản đã được cơ quan hành pháp soạn thảo hoàn chỉnh, chưa tham gia vào giai

đoạn thiết kế kiến trúc từ sớm. Nguyên nhân của hạn chế này xuất phát từ mô hình phân công giữa lập pháp và hành pháp tại Việt Nam theo truyền thống đặt trọng tâm soạn thảo dự án luật vào cơ quan hành pháp (Chính phủ trình dự án luật), trong khi Quốc hội đóng vai trò thẩm tra và quyết định thông qua. Mô hình phân công này phù hợp với đa số lĩnh vực lập pháp nhưng bộc lộ hạn chế khi áp dụng cho lĩnh vực đòi hỏi thiết kế kiến trúc kỹ thuật phức tạp ngay từ giai đoạn đầu như AI, nơi các quyết định kiến trúc (mô hình phân loại rủi ro, phạm vi định nghĩa) cần được cân nhắc trước khi văn bản được soạn thảo chi tiết chứ không phải sau khi đã hoàn chỉnh. Nguyên nhân thứ hai là Quốc hội chưa có Tiểu ban chuyên trách về AI trong Ủy ban Khoa học, Công nghệ và Môi trường với năng lực kỹ thuật đủ để tham gia thiết kế kiến trúc từ sớm. Quốc hội hoàn toàn có thẩm quyền tham gia sớm hơn nếu có cơ cấu tổ chức phù hợp.

Thứ hai, cơ quan hành pháp (Chính phủ và Bộ KH&CN)

Về ưu điểm, Bộ KH&CN đã hoàn thành vai trò đầu mối tổ chức ban soạn thảo, phối hợp lấy ý kiến các bộ ngành, trình Chính phủ và Quốc hội trong tiến độ được yêu cầu theo phân công của Chính phủ với ba sản phẩm cụ thể: Luật AI 2025, Nghị định 142/2026/NĐ-CP và Quyết định 33/2026/QĐ-TTg về Danh mục hệ thống AI rủi ro cao thể hiện năng lực tổ chức triển khai một chuỗi văn bản có tính liên kết logic trong thời gian tương đối ngắn (từ tháng 12/2025 đến tháng 6/2026). Nguyên nhân của ưu điểm này là việc Chính phủ xác lập rõ vai trò đầu mối duy nhất cho một Bộ chủ trì (thay vì phân tán giữa nhiều Bộ ngang quyền), tạo ra tính nhất quán trong định hướng thiết kế xuyên suốt cả ba văn bản.

Về hạn chế, sự thiếu hụt về năng lực công nghệ-pháp lý đặc thù như khả năng kết hợp hiểu biết kỹ thuật AI đủ sâu với kiến thức pháp lý để thiết kế quy phạm có tính khả thi kỹ thuật. Toàn bộ 15 chuyên gia phỏng vấn xác nhận thiếu hụt này là vấn đề cơ cấu quan trọng nhất trong hệ thống chủ thể xây dựng, trong đó đáng chú ý là chuyên gia CG-DN-04 so sánh ở Hàn Quốc phiên điều trần kỹ thuật diễn ra với câu hỏi rất cụ thể về tính khả thi của từng yêu cầu pháp lý dự kiến. Khác với Việt Nam quá trình tham vấn chủ yếu là đăng dự thảo và chờ phản hồi qua cổng điện tử trong 60 ngày [Xem Phụ lục 5]. Điều đáng lưu ý, thời gian có hiệu lực của Nghị định 142 (01/5/2026) chậm hai tháng, Quyết định 33 (15/8/2026) chậm năm tháng so với Luật AI 2025 (01/3/2026) không hoàn toàn tuân thủ yêu cầu cùng thời điểm có hiệu lực của Khoản 2 Điều 53 Luật Ban hành VBQPPL 2025. Đây là bằng chứng cho thấy thách thức trong việc thực hiện yêu cầu mới của Luật Ban hành VBQPPL 2025 về chuẩn bị văn bản hướng dẫn song song với quá trình soạn thảo luật. Về phạm vi thẩm quyền theo thời gian của Bộ KH&CN với tư cách đầu mối theo Chương VII Nghị định 142 hiện có

thẩm quyền quản lý hệ thống AI đang vận hành, nhưng chưa rõ thẩm quyền và trách nhiệm đối với hồ sơ, dữ liệu của hệ thống đã chấm dứt vận hành.

Nguyên nhân của hạn chế này có tính khách quan, hoạt động xây dựng khung pháp lý AI đòi hỏi đồng thời hai loại tri thức khác biệt về bản chất (pháp lý và kỹ thuật) mà hệ thống đào tạo và tuyển dụng công chức Việt Nam chưa hình thành như một chuẩn mực nghề nghiệp riêng biệt. Chức danh "chuyên viên pháp lý công nghệ" chưa tồn tại trong danh mục vị trí việc làm là khoảng trống về cơ cấu đào tạo và tuyển dụng cần thời gian để xây dựng.

Thứ ba, cơ quan tư pháp

Về ưu điểm, hệ thống tòa án đã có cơ chế án lệ vận hành ổn định trong các lĩnh vực khác theo cơ chế công bố của TAND tối cao và căn cứ pháp lý về lẽ công bằng tại Điều 6 Bộ luật Dân sự 2015 tạo ra hạ tầng thể chế sẵn có thể sử dụng ngay mà không cần xây dựng cơ chế mới từ đầu. Nguyên nhân của ưu điểm này là việc Việt Nam đã trải qua một giai đoạn cải cách tư pháp trước đó (thừa nhận án lệ, thừa nhận lẽ công bằng) độc lập với và có trước làn sóng lập pháp AI. AI được hưởng lợi từ hạ tầng thể chế đã có sẵn thay vì phải xây dựng đồng thời cả hạ tầng án lệ lẫn nội dung án lệ AI cụ thể.

Về hạn chế, chưa có án lệ nào liên quan đến AI được công bố. Nguyên nhân của hạn chế này khác biệt với nguyên nhân đã nêu ở hai chủ thể trên, về tính chủ động của cơ chế án lệ tại Việt Nam theo thiết kế hiện hành vận hành theo phương thức thụ động, chọn lọc từ các bản án đã có sẵn để công bố thành án lệ, không có cơ chế chủ động nhận diện và ưu tiên phát triển án lệ cho lĩnh vực mới nổi có tính cấp bách như AI. Nguyên nhân sâu xa hơn là hệ thống tòa án chưa có chương trình đào tạo chuyên biệt về kỹ thuật AI cho thẩm phán và chưa có cơ chế tư vấn kỹ thuật độc lập trong xét xử khiến ngay cả khi vụ việc AI được đưa ra xét xử, khả năng phát triển lý luận pháp lý đặc thù thay vì chỉ áp dụng quy định chung bị hạn chế bởi thiếu công cụ hỗ trợ chuyên môn.

3.2.1.2. Nhóm chủ thể tham gia chính thức

Thứ nhất, cơ quan thẩm định pháp lý (Bộ Tư pháp)

Về ưu điểm, Bộ Tư pháp đã thực hiện đầy đủ và đúng thẩm quyền chức năng thẩm định được giao, kiểm tra tính hợp hiến, hợp pháp và tính thống nhất của Luật AI 2025, Nghị định 142, Quyết định 33 với hệ thống VBQPPL hiện hành. Nguyên nhân của ưu điểm này là Bộ Tư pháp có chức năng thẩm định được quy định rõ ràng và ổn định trong Luật Ban hành VBQPPL 2025, không phụ thuộc vào đặc thù riêng của từng lĩnh vực được thẩm định. Đây là năng lực thể chế chung, có thể áp dụng nhất quán cho AI như cho bất kỳ lĩnh vực nào khác.

Về hạn chế, phạm vi thẩm định dừng lại ở tầng VBQPPL, chưa mở rộng sang đánh giá kiến trúc tổng thể bảy tầng. Nguyên nhân của hạn chế này là hạn chế về phạm vi thẩm quyền được giao, chức năng thẩm định theo quy định pháp luật hiện hành được thiết kế để kiểm tra tính hợp pháp của từng văn bản đơn lẻ đối chiếu với hệ thống pháp luật hiện hành, không bao gồm nhiệm vụ đánh giá tính đầy đủ của kiến trúc đa tầng (một khái niệm mới mà luận án đề xuất tại Chương 2, chưa được phản ánh trong bất kỳ quy định thẩm quyền hiện hành nào).

Thứ hai, cơ quan/tổ chức được trao thẩm quyền ban hành tiêu chuẩn kỹ thuật hoặc công nhận bộ quy tắc ứng xử

Về ưu điểm, Uỷ ban Tiêu chuẩn Đo lường Chất lượng Quốc gia thuộc Bộ KHCN đã ban hành được 8 TCVN nền tảng, thể hiện năng lực kỹ thuật cơ bản trong việc nội luật hóa tiêu chuẩn quốc tế. Nguyên nhân của ưu điểm này là Việt Nam đã là thành viên tổ chức tiêu chuẩn hóa quốc tế ISO/IEC từ trước, có quy trình và năng lực kỹ thuật nội luật hóa tiêu chuẩn quốc tế đã được thiết lập cho nhiều lĩnh vực khác trước khi áp dụng cho AI, tương tự trường hợp cơ quan tư pháp, đây là việc tận dụng hạ tầng thể chế đã có sẵn.

Về hạn chế, chưa có Ban kỹ thuật thường trực chuyên về AI (TC/AI) trong Uỷ ban Tiêu chuẩn Đo lường Chất lượng Quốc gia và quy trình ban hành TCVN thông thường kéo dài 2-3 năm không tương xứng với tốc độ cần thiết cho lĩnh vực AI. Nguyên nhân của hạn chế này có hai lớp: nguyên nhân trực tiếp là quy trình ban hành TCVN được thiết kế theo chuẩn mực chung cho mọi lĩnh vực kỹ thuật (không phân biệt tốc độ thay đổi công nghệ của từng lĩnh vực), trong khi đặc tính công nghệ AI có chu kỳ đổi mới 12-18 tháng đòi hỏi quy trình nhanh hơn đáng kể; nguyên nhân sâu xa hơn là chưa có quyết định phân bổ nguồn lực và thẩm quyền riêng để thành lập Ban kỹ thuật chuyên trách AI. Về cơ chế công nhận bộ quy tắc ứng xử, hiện chưa có bất kỳ cơ quan nào được trao thẩm quyền cụ thể thực hiện chức năng này đối với lĩnh vực AI.

3.2.1.3. Nhóm chủ thể có ảnh hưởng đến quá trình xây dựng khung pháp lý

Thứ nhất, tổ chức quốc tế liên chính phủ và cơ chế hợp tác đa phương

Về ưu điểm, Việt Nam đã tận dụng hiệu quả kênh ảnh hưởng từ UNESCO và từ khuôn khổ ASEAN với vai trò chủ động khi đăng cai Hội nghị Bộ trưởng Kỹ thuật số ASEAN lần thứ sáu tại Hà Nội (01/2026) cùng việc ký Công ước Hà Nội. Nguyên nhân của ưu điểm này là Việt Nam có truyền thống tham gia tích cực vào các cơ chế đa phương khu vực (đặc biệt ASEAN) từ trước khi vấn đề AI trở nên cấp bách, tạo ra kênh ảnh hưởng đã được thiết lập sẵn mà lĩnh vực AI có thể sử dụng ngay.

Về hạn chế, việc tiếp nhận ảnh hưởng từ tổ chức quốc tế liên chính phủ chủ yếu dừng ở cấp độ nguyên tắc chung (như các nguyên tắc đạo đức AI của UNESCO), chưa có bằng chứng về việc các khuyến nghị này được diễn giải và cụ thể hóa có hệ thống thành nội dung quy phạm cụ thể. Nguyên nhân của hạn chế này phù hợp với đặc trưng đã phân tích bản chất "khuyến nghị mềm" của tổ chức liên chính phủ đòi hỏi một quá trình diễn giải và cụ thể hóa chủ động từ phía chủ thể trực tiếp xây dựng khung pháp lý. Quá trình này chỉ có thể thực hiện tốt khi có năng lực công nghệ pháp lý đủ mạnh, tạo ra một chuỗi nguyên nhân liên hoàn: thiếu năng lực công nghệ pháp lý ở chủ thể trực tiếp dẫn đến khả năng hấp thụ hạn chế đối với ảnh hưởng từ tổ chức quốc tế liên chính phủ.

Thứ hai, tổ chức tiêu chuẩn hóa kỹ thuật quốc tế và khu vực

Về ưu điểm, việc Việt Nam nội luật hóa TCVN từ ISO/IEC là bằng chứng cho thấy kênh ảnh hưởng từ tổ chức tiêu chuẩn hóa kỹ thuật quốc tế đã phát huy tác dụng ở mức độ cơ bản. Nguyên nhân của ưu điểm này là do bản chất kỹ thuật tiêu chuẩn quốc tế có khả năng chuyển hóa trực tiếp với chi phí thấp hơn nhiều so với khuyến nghị nguyên tắc, không đòi hỏi quá trình diễn giải phức tạp như đối với ảnh hưởng từ tổ chức liên chính phủ.

Về hạn chế, mức độ ảnh hưởng còn rất hạn chế so với tiềm năng, chỉ 8 tiêu chuẩn AI đã được ISO/IEC SC 42 ban hành được nội luật hóa và chưa có thỏa thuận công nhận lẫn nhau (MRA) nào với bất kỳ đối tác quốc tế nào về kết quả kiểm định AI. Nguyên nhân của hạn chế này trùng khớp trực tiếp với nguyên nhân đã phân tích về Uỷ ban Tiêu chuẩn Đo lường Chất lượng Quốc gia cho thấy hạn chế của chủ thể trực tiếp (thiếu Ban kỹ thuật chuyên trách, quy trình ban hành chậm) giới hạn khả năng hấp thụ ảnh hưởng từ chủ thể có ảnh hưởng tương ứng.

Thứ ba, mạng lưới chuyên gia độc lập và cộng đồng học thuật xuyên quốc gia

Về ưu điểm, cộng đồng học thuật pháp lý trong nước đã phát triển nhanh về số lượng công trình nghiên cứu liên quan đến pháp luật AI trong giai đoạn 2022-2026, phản ánh sự lan tỏa của thảo luận học thuật quốc tế vào Việt Nam. Nguyên nhân của ưu điểm này là làn sóng nghiên cứu quốc tế về pháp luật AI đã phát triển sôi động từ năm 2021 và các kênh trao đổi học thuật quốc tế (hội thảo, xuất bản, hợp tác nghiên cứu) tạo điều kiện để tri thức đó lan tỏa tương đối nhanh vào cộng đồng học thuật trong nước.

Về hạn chế, chưa có cơ chế chính thức kết nối nghiên cứu học thuật với hoạt động soạn thảo pháp luật. Nguyên nhân của hạn chế này nằm ở phía chủ thể trực tiếp xây dựng, thiếu một cơ chế thể chế hóa (như Ban cố vấn nghiên cứu AI hoặc nghĩa vụ tóm tắt nghiên cứu) để kết nối có hệ thống hai phía.

Thứ tư, chủ thể thị trường tạo ra chuẩn mực thực tiễn dẫn dắt

Về ưu điểm, một số doanh nghiệp công nghệ lớn trong nước (như VinAI, FPT AI, Viettel AI) đã xây dựng chính sách AI nội bộ và tham gia các mạng lưới quốc tế (như việc FPT tham gia AI Verify Foundation của Singapore), tạo ra nguồn kinh nghiệm thực tiễn có thể tham chiếu khi xây dựng chuẩn mực chính thức. Nguyên nhân của ưu điểm này là các doanh nghiệp công nghệ lớn có động lực kinh tế trực tiếp (nhu cầu tiếp cận thị trường quốc tế, nhu cầu quản trị rủi ro nội bộ) để chủ động xây dựng thực hành tốt trước khi có yêu cầu pháp lý bắt buộc như một động lực tồn tại độc lập với hoạt động của chủ thể trực tiếp xây dựng khung pháp lý.

Về hạn chế, ảnh hưởng của chủ thể thị trường có nguy cơ tạo ra bất đối xứng đại diện trong quá trình tham vấn. Bằng chứng từ chuyên gia DN-02: "Trong các hội thảo tham vấn về Luật AI 2025 mà tôi được mời, phần lớn người tham gia là doanh nghiệp lớn. Startup AI, freelance developer... thiếu kênh tham vấn phù hợp". Nguyên nhân của hạn chế này có tính cơ cấu, doanh nghiệp lớn có nguồn lực tổ chức (bộ phận pháp chế chuyên trách, quan hệ công tác thường xuyên với cơ quan quản lý) để duy trì kênh ảnh hưởng liên tục, trong khi doanh nghiệp nhỏ và startup không có nguồn lực tương đương. Đây là sự bất đối xứng về nguồn lực tự nhiên của thị trường, chỉ có thể được cân bằng thông qua thiết kế chủ động từ phía chủ thể trực tiếp, không thể kỳ vọng thị trường tự điều chỉnh.

Phân tích thực trạng cho thấy phần lớn hạn chế của chủ thể tham gia và có ảnh hưởng (Nhóm thứ hai, ba) có nguyên nhân trực tiếp từ hạn chế tương ứng của chủ thể quyết định (Nhóm thứ nhất). Cụ thể, mức độ ảnh hưởng hạn chế của tổ chức tiêu chuẩn hóa quốc tế có nguyên nhân trực tiếp từ hạn chế năng lực của Uỷ ban Tiêu chuẩn Đo lường Chất lượng Quốc gia; mức độ hấp thụ hạn chế từ tổ chức liên chính phủ có nguyên nhân từ khoảng trống năng lực công nghệ-pháp lý ở cơ quan hành pháp; tình trạng thiếu kết nối với mạng lưới học thuật có nguyên nhân từ thiếu cơ chế thể chế hóa ở phía chủ thể trực tiếp.

3.2.2. Về nội dung xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Thứ nhất, kiến trúc phân tầng của khung pháp lý

Thực tiễn xây dựng khung pháp lý AI Việt Nam cho thấy kiến trúc phân tầng gần như không tồn tại dưới dạng tài liệu công khai. Không có tài liệu kiến trúc chính thức nào giải thích tại sao Việt Nam chọn ba tầng rủi ro (thay vì bốn như EU); tại sao Bộ KH&CN là đầu mối (thay vì cơ quan AI chuyên biệt); quan hệ kiến trúc giữa Luật AI 2025 và các luật liên quan là gì; vị trí của luật mềm, bộ quy tắc ứng xử, án lệ trong kiến trúc tổng thể được thiết kế như thế nào.

Rà soát hồ sơ pháp lý công khai của Luật AI 2025 và Nghị định 142/2026/NĐ-CP cho thấy sự vắng mặt của Tài liệu kiến trúc pháp lý AI quốc gia. Chuyên gia CS-02 phân tích hậu quả kinh tế của thiếu tài liệu kiến trúc tạo ra chi phí giao dịch rất cao cho nhà đầu tư nước ngoài. Họ phải thuê luật sư để đọc cả Luật AI, Luật BVDLCN, Luật CNCNS và các nghị định và tự xây dựng bản đồ nghĩa vụ. Chuyên gia CG-DN-04 xác nhận bộ phận pháp chế công ty mất ba tuần chỉ để xác định hệ thống nghĩa vụ áp dụng cho sản phẩm AI của mình. Điều này tạo ra bất lợi so với các công ty ở Hàn Quốc hay Singapore, nơi có hướng dẫn tuân thủ một cửa rõ ràng [Xem Phụ lục 5].

Điều quan trọng cần nhấn mạnh là thiếu tài liệu kiến trúc không có nghĩa là thiếu tư duy kiến trúc, các lựa chọn thiết kế trong Luật AI 2025 như phân loại ba tầng rủi ro, định nghĩa chức năng và nguyên tắc không trùng lặp với pháp luật chuyên ngành cho thấy có tư duy kiến trúc. Vấn đề là tư duy đó được thực hiện trong nội bộ nhóm soạn thảo mà không được tài liệu hóa và công bố dẫn đến cộng đồng chỉ có thể phản biện từng điều khoản mà không thể phản biện lựa chọn kiến trúc và khi cần cập nhật, người soạn thảo sau không biết logic của người soạn thảo trước.

Thứ hai, Việt Nam đã có bước tiến quan trọng từ việc ban hành các định hướng chiến lược mang tính nguyên tắc (như Chiến lược quốc gia về AI đến năm 2030, tầm nhìn đến năm 2045 theo Quyết định 1671/QĐ-TTg) sang việc thể chế hóa một phương pháp luận quản trị hiện đại, tiệm cận với chuẩn mực quốc tế. Phương pháp luận quản trị được lựa chọn nhất quán là tiếp cận dựa trên rủi ro kết hợp với nguyên tắc lấy con người làm trung tâm. Hệ thống pháp lý Việt Nam phân loại AI thành các cấp độ rủi ro (cao, trung bình, thấp) nhằm tối ưu hóa nguồn lực quản lý và tránh tạo gánh nặng tuân thủ không cần thiết cho doanh nghiệp. Đồng thời, tư duy lập pháp đã dịch chuyển rõ rệt từ thái cực "kìm hãm vì rủi ro" sang tư duy "quản lý để phát triển", vừa xác lập hàng rào kỹ thuật bảo vệ quyền con người, an toàn dữ liệu và an ninh quốc gia, vừa tạo hành lang pháp lý vượt trội để thu hút đầu tư và thúc đẩy đổi mới sáng tạo.

So với dự thảo, phương pháp tiếp cận đã thay đổi. Dự thảo trình Quốc hội có bốn mức rủi ro, trong đó hệ thống có rủi ro không chấp nhận được bị cấm phát triển, cung cấp, triển khai hoặc sử dụng dưới mọi hình thức. Luật được thông qua chỉ còn ba mức. Phần cấm chuyển thành hành vi bị cấm xác định theo mục đích. Điểm b khoản 2 Điều 7 chỉ cấm việc giả mạo, mô phỏng người thật để lừa dối có chủ đích, có hệ thống và gây tổn hại nghiêm trọng. Ngưỡng cấm vì vậy cao và chưa có loại hệ thống nào bị cấm theo bản chất của nó.

Khoản 2 Điều 4 Luật AI đòi con người duy trì kiểm soát đối với mọi quyết định của hệ thống. Thế nhưng điểm b khoản 2 Điều 8 Nghị định 142 lại dùng chính "sự giám sát thực chất của con người" làm căn cứ để loại một hệ thống ra khỏi

nhóm rủi ro cao. Một nguyên tắc vốn đặt ra nghĩa vụ đã trở thành căn cứ miễn trừ. Thêm vào đó, van an toàn tại khoản 5 Điều 8 Nghị định 142, cho phép vẫn đưa hệ thống đã được loại trừ vào Danh mục, chỉ áp dụng khi có nguy cơ gây thiệt hại đến quốc phòng, an ninh, trật tự, an toàn xã hội. Van này không áp dụng khi thiệt hại rơi vào quyền của cá nhân.

Thứ ba, cơ chế thử nghiệm có kiểm soát tại Việt Nam hiện đang trong quá trình chuyển giao từ mô hình thí điểm ngành sang thể chế hóa diện rộng. Ban đầu, Sandbox được ứng dụng rải rác trong lĩnh vực tài chính - ngân hàng (Fintech), giao thông thông minh hay dịch vụ công điện tử. Đến nay, khung pháp lý chuyên ngành về AI đã chính thức mở đường cho việc thành lập không gian thử nghiệm công nghệ cho các hệ thống AI có rủi ro tiềm ẩn hoặc chưa có quy định điều chỉnh trực tiếp. Dù vậy, thực trạng triển khai hoạt động này vẫn đối mặt với không ít thách thức như năng lực thẩm định và giám sát kỹ thuật của các cơ quan quản lý chuyên ngành còn hạn chế; cơ chế miễn trừ trách nhiệm pháp lý có điều kiện cho doanh nghiệp tham gia thử nghiệm chưa thực sự hoàn thiện; quy trình cấp phép thử nghiệm còn nhiều thủ tục hành chính, khiến thời gian đưa sản phẩm AI thử nghiệm ra thị trường chưa đạt tốc độ kỳ vọng. Quyền của người tham gia thử nghiệm chưa đủ do chưa có yêu cầu đồng ý có thông tin, chưa có quyền rút khỏi thử nghiệm và chưa rõ quy định miễn, loại trừ trách nhiệm có loại cả nghĩa vụ bồi thường cho người tham gia bị thiệt hại hay không. Kết quả thử nghiệm chỉ được dùng để miễn, giảm nghĩa vụ cho từng tổ chức, chưa được tổng hợp để sửa quy định chung. Thẩm quyền chấp thuận chia theo cấp độ dữ liệu và địa bàn, cộng thêm cơ chế thử nghiệm chuyên ngành, nên tiêu chí thẩm định dễ thiếu thống nhất. Chưa có cơ chế công khai danh sách các hồ sơ được chấp thuận hoặc bị từ chối.

Thứ tư, hệ thống quyền và nghĩa vụ của các chủ thể đã được phân vai rõ ở cấp hệ thống. Tuy nhiên, nhà cung cấp mô hình không có nghĩa vụ trực tiếp nào. Yêu cầu có đầu mối tại Việt Nam chỉ áp dụng với nhà cung cấp hệ thống rủi ro cao. Quyền của người bị ảnh hưởng nằm rải rác ở bốn văn bản: Luật AI chỉ bảo đảm thông tin ở mức mô tả chức năng (điểm e khoản 1 và điểm đ khoản 2 Điều 14), Luật BVDLCN ghi nhận các quyền được biết, đồng ý, yêu cầu xóa, hạn chế xử lý, phản đối xử lý và yêu cầu bồi thường (khoản 1 Điều 4), Nghị định 356/2025/NĐ-CP yêu cầu bên kiểm soát dữ liệu phải thông báo về việc xử lý tự động, giải thích nguyên tắc hoạt động của thuật toán và cho chủ thể dữ liệu lựa chọn không tham gia. Thông tư 53 quy định ngân hàng phải giải thích quyết định tự động và cho phép con người rà soát khi có khiếu nại. Bên thứ ba không được dùng dữ liệu của đơn vị để huấn luyện mô hình nếu chưa có đồng ý bằng văn bản.

Thứ năm, nhiều loại công cụ bảo đảm tuân thủ được sử dụng đồng thời: kiểm tra theo mức rủi ro (khoản 5 Điều 10); yêu cầu phân loại lại hoặc tạm dừng sử dụng (khoản 6 Điều 10); báo cáo sự cố (Điều 12; Điều 19 Nghị định 142); đánh giá sự phù hợp, có công khai kết quả; thanh tra, có nghĩa vụ cung cấp hồ sơ, nhật ký, dữ liệu huấn luyện và công khai kết luận (khoản 3, khoản 4 Điều 28). Đây là kết quả của việc kế thừa mô hình chế tài hành chính đã có tiền lệ ổn định trong hệ thống pháp luật Việt Nam (Luật Xử lý vi phạm hành chính) kết hợp với việc tham chiếu kỹ thuật giám sát sự cố có tiền lệ trong pháp luật quản lý rủi ro nói chung. Về hạn chế, chưa có chế tài hành chính riêng cho AI, chưa bao gồm công cụ suy đoán tuân thủ qua tiêu chuẩn kỹ thuật. Về cơ chế khắc phục hậu quả, chưa có quyết định về nguyên tắc phân bổ rủi ro công bằng khi nhiều chủ thể cùng liên quan đến thiệt hại mà không thể quy trách nhiệm rõ ràng cho một bên.

So với dự thảo, cơ chế khắc phục đã bị thu hẹp: Dự thảo coi thiệt hại do hệ thống rủi ro cao gây ra là thiệt hại do nguồn nguy hiểm cao độ. Nhà cung cấp và bên triển khai liên đới bồi thường, theo mức độ kiểm soát thực tế, khả năng dự báo và biện pháp quản lý rủi ro của mỗi bên. Dự thảo buộc cung cấp hồ sơ kỹ thuật, nhật ký và dữ liệu huấn luyện cả khi có yêu cầu của Tòa án và Trọng tài. Dự thảo đặt mức phạt tối đa 2 tỷ đồng với tổ chức, 1 tỷ đồng với cá nhân, và đến 2% doanh thu đối với vi phạm nghiêm trọng. Luật được thông qua chỉ buộc bên triển khai bồi thường không phụ thuộc lỗi. Bên triển khai chỉ được yêu cầu hoàn trả nếu có thỏa thuận giữa các bên (khoản 2 Điều 29). Nghĩa vụ cung cấp thông tin chỉ phát sinh trong thanh tra, kiểm tra (khoản 3 Điều 28).

Thứ sáu, cơ chế rà soát và cập nhật định kỳ là khâu mới nhất đang được định hình trong chu trình quản trị AI tại Việt Nam. Một số đối tượng đã có cơ chế rà soát: Chiến lược được rà soát ít nhất 03 năm một lần (khoản 1 Điều 19). Khung đạo đức được rà soát, cập nhật định kỳ (khoản 2 Điều 26). Nghị định 142 phân công trách nhiệm rà soát, đề xuất sửa đổi (khoản 2 Điều 7), cho phép đưa vào Danh mục dựa trên bằng chứng (khoản 5 Điều 8), và quy định trường hợp phải phân loại lại khi Danh mục thay đổi, với thời gian chuyển tiếp 12 tháng (Điều 11). Đồng thời, sự ra đời của các thiết chế như Quỹ Phát triển AI Quốc gia giúp Nhà nước chủ động nguồn lực để điều chỉnh hạ tầng và hỗ trợ tuân thủ. Mặc dù vậy, bản thân Luật AI không có điều khoản đánh giá định kỳ. Các nguồn dữ liệu sau không bắt buộc phải được dùng làm căn cứ rà soát: báo cáo sự cố (Điều 19), báo cáo thử nghiệm (Điều 25), kết luận thanh tra (khoản 4 Điều 28) và khiếu nại. Hoạt động rà soát thích ứng hiện vẫn chịu áp lực lớn do năng lực công cụ kiểm toán thuật toán và hệ thống cảnh báo sớm rủi ro AI ở cấp độ quốc gia mới ở giai đoạn đầu xây dựng, chưa đáp ứng

đầy đủ yêu cầu quản trị thời gian thực đối với các tác nhân AI tự chủ ngày càng phức tạp.

3.2.3. Về quy trình xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Tiến trình xây dựng khung pháp lý AI Việt Nam từ 2017 đến 2026 là quá trình phản ứng tích lũy theo áp lực từ ba phía: áp lực chính trị từ các văn kiện của Đảng và Nhà nước; áp lực thực tiễn từ hệ sinh thái AI phát triển nhanh và rủi ro AI ngày càng hiện hữu và áp lực hội nhập từ xu hướng lập pháp AI toàn cầu. Sự tương tác giữa ba áp lực này tạo ra bốn giai đoạn có tính chất và trọng tâm khác nhau, từ nhận diện AI là ưu tiên công nghệ đến hoàn thành khung VBQPPL cơ bản trong gần một thập kỷ. Mỗi giai đoạn có một mốc đột phá tạo điều kiện cho giai đoạn tiếp theo.

Giai đoạn 1 - Chẩn đoán nhu cầu điều chỉnh:

Nhà nước lần đầu tiên đặt AI vào tầm nhìn chính sách nhưng trong khuôn khổ rộng hơn của CMCN lần thứ tư, AI chưa được nhận diện như đối tượng cần điều chỉnh pháp lý độc lập mà được tiếp cận chủ yếu như công nghệ cần ứng dụng và phát triển.

Bắt đầu từ Chỉ thị số 16/CT-TTg ngày 04/5/2017 của Thủ tướng Chính phủ về "Tăng cường năng lực tiếp cận cuộc Cách mạng công nghiệp lần thứ 4" là văn bản đầu tiên đặt AI vào định hướng chính sách quốc gia. Tiếp đó, Bộ Chính trị ban hành Nghị quyết số 52-NQ/TW ngày 27/9/2019 [10] về một số chủ trương, chính sách chủ động tham gia cuộc CMCN lần thứ tư là văn kiện Đảng quan trọng nhất của giai đoạn này. Nghị quyết yêu cầu "hoàn thiện pháp luật tạo thuận lợi cho quá trình chuyển đổi số quốc gia và phát triển kinh tế số", đây là lần đầu tiên một nghị quyết của Bộ Chính trị đặt ra yêu cầu hoàn thiện pháp luật trong bối cảnh AI và công nghệ số, dù còn chung chung. Nghị quyết 57-NQ/TW [11] đã đặt nền móng cho nhận thức về tầm quan trọng của KHCN, ĐMST và chuyển đổi số, tạo tiền đề chính trị cho giai đoạn tiếp theo. Quyết định 127/QĐ-TTg ngày 26/01/2021 ban hành Chiến lược quốc gia đầu tiên về nghiên cứu, phát triển và ứng dụng AI đặt ra yêu cầu "hoàn thiện thể chế, chính sách" cho AI như một trong các giải pháp trọng tâm là tín hiệu chính thức đầu tiên cho nhu cầu xây dựng khung pháp lý AI. Các văn bản này xác định nhu cầu phải có thể chế cho AI ở mức tổng quát, nhận diện chính sách đặt AI vào danh sách ưu tiên công nghệ, chứ không phải xây dựng quy phạm điều chỉnh hành vi.

Việc chẩn đoán ở mức cụ thể chỉ xuất hiện rõ trong hồ sơ dự án luật năm 2025. Tuy nhiên, nó được thực hiện đồng thời với việc soạn thảo, chứ không làm tiền đề cho việc soạn thảo, và có ba hạn chế. Hạn chế thứ nhất là chẩn đoán chưa đi đến tiêu chí điều chỉnh. Trong báo cáo thẩm tra, Ủy ban Khoa học, Công nghệ và Môi trường nhận xét dự thảo chưa có tiêu chí định lượng hoặc định tính để xác định

và phân loại rủi ro, chưa có biện pháp, công cụ và phương pháp đánh giá rủi ro. Hạn chế thứ hai là cơ sở bằng chứng còn mỏng. Không có thống kê chính thức riêng về sự cố và thiệt hại do AI. Các số liệu được dùng chủ yếu là số liệu chung về lừa đảo trực tuyến. Hạn chế thứ ba là chẩn đoán chưa được tổ chức theo vòng đời hệ thống. Vì vậy, những nhu cầu điều chỉnh ở hai đầu vòng đời, tức giai đoạn nghiên cứu và giai đoạn chấm dứt vận hành, gần như không được nhận diện.

Đối chiếu với tiêu chí căn cứ thực tiễn, giai đoạn chẩn đoán đạt yêu cầu ở mức định hướng, chưa đạt ở mức xác định vấn đề chính sách cụ thể.

Giai đoạn 2 - Thiết kế kiến trúc tổng thể:

Nhà nước chuyển từ đặt AI trong bối cảnh CMCN 4.0 sang xây dựng chiến lược AI độc lập với mục tiêu, nguồn lực và lộ trình cụ thể. Mặt tích cực của giai đoạn này, theo cơ quan chủ trì soạn thảo, là Việt Nam chọn cách tiếp cận kết hợp có chọn lọc: an toàn cao hơn mức của Hàn Quốc nhưng ít thủ tục hơn châu Âu, phát triển mạnh mẽ như Nhật Bản. Lựa chọn này đóng vai trò định hướng cho kiến trúc, nhưng chưa được cụ thể hóa thành bản đồ phân bổ nội dung giữa các tầng. Không có tài liệu nào thể hiện một quyết định trước khi soạn thảo về việc nội dung nào được đặt trong luật, nội dung nào được giao cho nghị định, tiêu chuẩn, hướng dẫn hay tự điều chỉnh và lộ trình ban hành các văn bản. Kiến trúc của khung pháp lý được hình thành ngay trong quá trình soạn thảo. Ba nhóm bằng chứng cho thấy điều này.

Nhóm bằng chứng thứ nhất là vị trí của các quy định về AI trong hệ thống văn bản thay đổi chỉ trong vài tháng. Luật CNCNS được Quốc hội thông qua tháng 6/2025 đã có một chương riêng về AI. Đến tháng 12/2025, Điều 33 Luật AI bãi bỏ Chương IV cùng một số khoản liên quan của Luật CNCNS. Một nội dung vừa được đặt vào luật này rồi phải chuyển sang luật khác ngay trong cùng năm. Đây là dấu hiệu rõ của việc thiếu một bản thiết kế kiến trúc có trước.

Nhóm bằng chứng thứ hai là cấu trúc của dự thảo luật thay đổi lớn qua từng vòng. Bản dự thảo lấy ý kiến tháng 10/2025 gồm 9 chương, 70 điều. Bản trình Quốc hội ngày 21/11/2025 còn 8 chương, 36 điều. Bản được thông qua còn 35 điều. Số điều giảm gần một nửa cho thấy phần lớn việc phân bổ nội dung giữa luật và văn bản dưới luật đã được quyết định trong lúc chỉnh lý dự thảo. Câu hỏi về kiến trúc cũng chỉ được đặt ra muộn, ở bước thẩm tra. Cơ quan thẩm tra đánh giá việc xây dựng theo hướng luật khung là phù hợp, nhưng lưu ý khả năng phải sửa luật trong thời gian ngắn. Cơ quan thẩm tra cũng đề nghị rà soát sự đồng bộ với BLDS, Luật Chất lượng sản phẩm, hàng hóa và Luật Tiêu chuẩn và quy chuẩn kỹ thuật.

Nhóm bằng chứng thứ ba là trình tự ban hành các văn bản định hướng và hiệu lực của các văn bản bị lệch nhau. Chiến lược quốc gia mới về AI được phê duyệt ngày 28/8/2026, thay thế Quyết định 127/QĐ-TTg. Như vậy, văn bản định

hướng chiến lược được ban hành sau khi luật và nghị định đã có hiệu lực. Các văn bản thuộc khung pháp lý lần lượt có hiệu lực ở những thời điểm khác nhau: Luật ngày 01/3/2026, Nghị định 142 ngày 01/5/2026, Quyết định 33 ngày 15/8/2026.

Đối chiếu với tiêu chí tính khoa học trong thiết kế kiến trúc, đây là giai đoạn yếu nhất của quy trình, cũng là nguyên nhân trực tiếp của tình trạng mất cân đối giữa các tầng đã nêu ở mục 3.1.2.

Giai đoạn 3 - Tạo lập các thành tố:

Giai đoạn tạo lập diễn ra rất nhanh và tập trung vào khoảng một năm. Trong chưa đầy một năm, Bộ KH&CN đã chủ trì soạn thảo, sửa đổi và trình thông qua 10 luật. Luật AI được thông qua năm 2025. Sau đó, các văn bản quy định chi tiết và hướng dẫn được ban hành liên tiếp trong năm 2026: Thông tư 05/2026/TT-BKHCN về Khung đạo đức tháng 3, Nghị định 142 tháng 4, Quyết định 804/QĐ-TTg về danh mục bộ dữ liệu tháng 5, Quyết định 33 về Danh mục hệ thống rủi ro cao tháng 6, và Thông tư 53/2026/TT-NHNN tháng 9. Tốc độ này cho thấy năng lực huy động rất lớn của bộ máy, lý giải vì sao giai đoạn chẩn đoán và giai đoạn thiết kế bị nén vào chính quá trình soạn thảo.

Việc tham vấn đã được tổ chức. Dự thảo luật được đăng tải lấy ý kiến rộng rãi trên Cổng thông tin điện tử của Bộ. Ủy ban Khoa học, Công nghệ và Môi trường của Quốc hội phối hợp với Bộ Khoa học và Công nghệ và VCCI tổ chức hội thảo góp ý. Qua các vòng tham vấn, nhiều nội dung thực chất đã thay đổi. Bản dự thảo lấy ý kiến có danh mục hệ thống bị cấm, bao gồm chấm điểm tín nhiệm xã hội diện rộng bởi cơ quan nhà nước, nhận dạng sinh trắc học từ xa theo thời gian thực tại nơi công cộng và xây dựng cơ sở dữ liệu nhận dạng khuôn mặt bằng cách thu thập tràn lan. Luật được thông qua chỉ còn ba mức rủi ro, giảm so với bốn mức của dự thảo trước đó. Cơ quan chủ trì giải thích định hướng của Luật là chỉ điều chỉnh đầu ra và hành vi sử dụng AI, không cản trở đổi mới sáng tạo.

Hạn chế lớn nhất của giai đoạn này nằm ở phạm vi. Hoạt động tạo lập tập trung gần như hoàn toàn vào tầng VBQPPL. Các thành tố ở tầng tiêu chuẩn kỹ thuật, tầng tự điều chỉnh và tầng án lệ chưa được tạo lập tương ứng. Tầng thiết chế mới được giao thẩm quyền, chưa đủ năng lực vận hành. Mỗi văn bản được xây dựng theo trình tự riêng của nó, không theo một lộ trình chung cho cả khung pháp lý. Đây là hệ quả trực tiếp của việc thiếu giai đoạn thiết kế kiến trúc.

Giai đoạn 4 - Vận hành, đánh giá và thích ứng

Giai đoạn này mới bắt đầu. Do đó, luận án chỉ đánh giá được việc các điều kiện cho vận hành đã được chuẩn bị đến đâu, chưa đánh giá được hiệu quả. Luật có hiệu lực từ ngày 01/3/2026. Hệ thống đã hoạt động trước đó được chuyển tiếp 18 tháng trong lĩnh vực y tế, giáo dục, tài chính và 12 tháng trong các lĩnh vực khác.

Điều đó có nghĩa là phần lớn nghĩa vụ đối với hệ thống đang vận hành chỉ thực sự được áp dụng từ ngày 01/3/2027 và ngày 01/9/2027. Trong 60 ngày kể từ khi Nghị định 142 có hiệu lực, tổ chức thuộc diện chuyển tiếp phải thông báo trên Cổng thông tin một cửa, kèm kế hoạch chuyển đổi hoặc phương án tuân thủ. Tuy nhiên, ngày 28/8/2026, Cổng Thông tin một cửa về AI mới chính thức được vận hành, còn sơ sài trong giao diện với tham vọng quản lý tập trung việc đăng ký tham gia sandbox, báo cáo sự cố nghiêm trọng, phân loại rủi ro, báo cáo đánh giá tác động và công khai kết quả. Mục Khung pháp lý AI trên Cổng chỉ xác nhận quy định chính thức bởi Luật AI và các văn bản hướng dẫn, thể hiện quan điểm đồng nhất khung pháp lý với khung pháp luật, chưa bao gồm các tầng khác của mô hình khung pháp lý đầy đủ như luận án đề xuất. Chưa có nguồn độc lập ngoài truyền thông nhà nước đánh giá về tính ổn định kỹ thuật, khả năng chịu tải, hay cơ chế bảo mật dữ liệu cá nhân, doanh nghiệp khai báo.

Đã xuất hiện một số vòng thích ứng nhỏ. Vòng rõ nhất là Thông tư 53/2026/TT-NHNN. Văn bản này cụ thể hóa những điểm mà Danh mục hệ thống rủi ro cao còn để ngỏ trong lĩnh vực ngân hàng. Thông tư xác định hệ thống tự động cấp tín dụng là rủi ro cao khi khoản cấp tín dụng từ 50 triệu đồng trở lên, và đặt ngưỡng riêng cho hệ thống tự động thanh toán. Đây là một vòng nối tiếp đúng chức năng: văn bản chuyên ngành bổ sung tiêu chí còn thiếu ở văn bản chung.

Ở cấp chiến lược, cơ chế đánh giá đã được thiết lập. Bộ Khoa học và Công nghệ chịu trách nhiệm theo dõi, đo lường, đánh giá và báo cáo Thủ tướng hằng năm về việc thực hiện Chiến lược. Ở cấp khung pháp lý thì chưa có cơ chế đánh giá tương ứng. Việc rà soát Danh mục mới được giao trách nhiệm cho các cơ quan, chưa có chu kỳ và chưa có nguồn dữ liệu bắt buộc. Báo cáo sự cố, kết quả thử nghiệm có kiểm soát và kết luận thanh tra chưa được nối với việc sửa đổi quy định.

3.2.4. Về kỹ thuật xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Từ phân tích kinh nghiệm quốc tế, áp dụng khung đánh giá sáu kỹ thuật vào thực tiễn xây dựng Luật AI 2025, Nghị định 142/2026/NĐ-CP và các văn bản liên quan cho thấy mức độ vận dụng không đồng đều, một số kỹ thuật được vận dụng đúng hướng, nhiều kỹ thuật còn thiếu chiều sâu.

3.2.4.1. Kỹ thuật 1 - RIA:

Kỹ thuật AI-RIA là kỹ thuật thể hiện rõ nhất khoảng cách giữa tuân thủ hình thức và thực chất. Hồ sơ dự án Luật AI 2025 có Báo cáo đánh giá tác động chính sách theo đúng yêu cầu của Luật Ban hành VBQPPL 2025 là thành quả thể hiện sự trưởng thành trong quy trình lập pháp và cần được ghi nhận như điều kiện cần của RIA chất

lượng. Điều đáng chú ý là từ điều kiện cần đến điều kiện đủ còn khoảng cách đáng kể khi đối chiếu với chuẩn mực OECD.

Chưa có cơ sở xác nhận công khai về mức độ đầy đủ của ba chiều phân tích đặc thù AI trong Báo cáo đánh giá tác động chính sách. Về chiều phân tích kịch bản công nghệ, không có bằng chứng về đánh giá văn bản dự thảo trong ba kịch bản tốc độ phát triển AI khác nhau, đặc biệt kịch bản tác nhân AI lan rộng trong 18 tháng thay vì 5 năm, vốn là kịch bản đang diễn ra trong thực tế. Về chiều phân tích chi phí tuân thủ theo quy mô doanh nghiệp, không có ước lượng công khai so sánh gánh nặng tuân thủ giữa doanh nghiệp lớn, SME và startup AI dẫn đến việc một số ngưỡng trong Nghị định 142, như yêu cầu báo cáo sự cố trong vòng 72 giờ hoặc 05 ngày, được đặt ra theo tham khảo quốc tế mà thiếu cơ sở định lượng từ thực tiễn Việt Nam. Về chiều đánh giá năng lực thể chế, không có bằng chứng về việc Luật AI 2025 được xây dựng kèm theo đánh giá liệu Việt Nam có đủ tổ chức kiểm định AI đủ năng lực để thực thi yêu cầu kiểm định không - một câu hỏi mà chuyên gia CS-05 cho rằng chưa được trả lời chính thức trong hồ sơ soạn thảo [Xem Phụ lục 5].

3.2.4.2. Kỹ thuật 2 - Định nghĩa chức năng

Trong số sáu kỹ thuật được xem xét, kỹ thuật định nghĩa chức năng là kỹ thuật được vận dụng tốt nhất so với thực hành quốc tế, đồng thời là bằng chứng rõ ràng nhất cho thấy quá trình xây dựng Luật AI 2025 đã học hỏi có chọn lọc từ kinh nghiệm quốc tế. Luật AI 2025 định nghĩa "hệ thống AI" theo tiêu chí hành vi và chức năng, tham chiếu cách tiếp cận của OECD AI Principles 2024. Lựa chọn kỹ thuật này bảo đảm định nghĩa có khả năng bao phủ các thế hệ công nghệ AI mới mà không cần sửa đổi văn bản - một ưu điểm quan trọng trong lĩnh vực mà chu kỳ đổi mới công nghệ chỉ kéo dài 12-18 tháng.

Tuy vậy, kỹ thuật định nghĩa chức năng chưa được áp dụng nhất quán trong toàn bộ kiến trúc của văn bản, đặc biệt đối với AI agents thực thể đang bùng nổ trong thực tiễn ứng dụng AI từ 2024-2026. Trong khi OECD AI Principles 2024 đã lần đầu tiên chính thức phân biệt "AI system" và "agentic AI" theo đặc trưng hành vi là tự chủ thực thi chuỗi hành động không cần phê duyệt từng bước của con người, Nghị định 142/2026/NĐ-CP vẫn chưa có điều khoản xác định khi nào một hệ thống AI trở thành "AI agent" và hàm ý pháp lý khác biệt của loại hệ thống đó. Bên cạnh khoảng trống về AI agents, việc thiếu cơ chế diễn giải định nghĩa chính thức cũng là một hạn chế đáng kể khi AI agents và các thế hệ AI mới tiếp tục xuất hiện và đặt câu hỏi về phạm vi áp dụng của định nghĩa hiện hành, không có thiết chế nào được giao thẩm quyền diễn giải kịp thời mà không phải sửa luật, trong khi Hàn Quốc đã giải quyết vấn đề này bằng cơ chế diễn giải chính thức của Ủy ban AI Quốc gia với quy trình 30 ngày.

3.2.4.3. Kỹ thuật 3 - Ủy quyền linh hoạt phân tầng

Kỹ thuật ủy quyền linh hoạt phân tầng đạt mức trung bình, cao hơn hầu hết các kỹ thuật còn lại, nhưng cũng ẩn chứa sự mất cân đối nội tại đáng chú ý. Ở tầng VBQPPL, cơ chế Quyết định Thủ tướng cập nhật danh mục AI rủi ro cao theo quy trình rút gọn là vận dụng đúng kỹ thuật này, cho phép cập nhật danh mục mà không cần sửa Luật AI 2025, bảo đảm khả năng phản ứng với phát triển công nghệ mà không làm bất ổn tầng văn bản cốt lõi. Tương tự, nguyên tắc không trùng lặp với pháp luật chuyên ngành trong Nghị định 142 thể hiện kỹ thuật phân tầng ủy quyền theo ngành hợp lý, tránh tình trạng doanh nghiệp phải tuân thủ song song hai lớp yêu cầu có khả năng mâu thuẫn nhau.

Song song với những điểm đúng hướng đó, ba khoảng trống quan trọng phản ánh tầng dưới của kiến trúc ủy quyền chưa được hoàn thiện. Đáng kể nhất là sự thiếu vắng hoàn toàn của điều khoản rà soát bắt buộc, trong khi EU AI Act Điều 112 quy định nghĩa vụ rà soát và báo cáo Nghị viện và Hội đồng châu Âu ba năm một lần như một cơ chế thích ứng thể chế hóa, cả Luật AI 2025 lẫn Nghị định 142 đều không có điều khoản tương đương khiến tính thích ứng phụ thuộc hoàn toàn vào ý chí hành chính. Liên quan đến tầng tiêu chuẩn kỹ thuật, ủy quyền cập nhật TCVN AI vẫn phải trải qua quy trình ban hành thông thường kéo dài 2-3 năm, thay vì có quy trình rút gọn 12 tháng tương tự cơ chế tiêu chuẩn hài hòa của EU. Hơn nữa, không có cơ chế ủy quyền cho nhóm làm việc đồng thiết kế cập nhật hướng dẫn kỹ thuật theo mô hình Singapore nơi nhóm thường trực 30+ doanh nghiệp có thể cập nhật AI Verify ba lần/năm mà chỉ cần IMDA phê duyệt trong 30 ngày.

3.2.4.4. Kỹ thuật 4 - Phân định nghĩa vụ theo chuỗi giá trị

Kỹ thuật phân định nghĩa vụ theo chuỗi giá trị là kỹ thuật thể hiện rõ nhất khoảng cách giữa nhận thức nguyên tắc và chi tiết vận hành. Nghị định 142 đã phân biệt "nhà cung cấp" và "bên triển khai" với phân công nghĩa vụ cơ bản, cho thấy nhận thức về mô hình phân định chuỗi giá trị là một bước tiến quan trọng so với cách tiếp cận "tổ chức, cá nhân liên quan đến AI có nghĩa vụ chung" vốn không giải quyết được "many hands problem" trong chuỗi AI phức tạp. Tuy nhiên, so với mô hình nhà cung cấp-nhà triển khai đầy đủ của EU AI Act, khung phân định của Nghị định 142 vẫn thiếu hai cơ chế vận hành quan trọng.

Khoảng trống thứ nhất là thiếu quy tắc "vượt phạm vi kiểm định trở thành nhà cung cấp" khi doanh nghiệp Việt Nam tinh chỉnh mô hình AI nước ngoài, câu hỏi về nghĩa vụ pháp lý của phần sửa đổi đó vẫn chưa có câu trả lời rõ ràng trong văn bản hiện hành, dẫn đến khoảng trống trách nhiệm trong tình huống phổ biến nhất của hệ sinh thái AI nội địa. Khoảng trống thứ hai, nghiêm trọng hơn từ góc độ thực tiễn, là thiếu cơ chế xử lý nhà cung cấp mô hình nền tảng nước ngoài vắng mặt pháp lý tại

Việt Nam khi OpenAI, Google hay Anthropic không có hiện diện pháp lý và không thực hiện nghĩa vụ nhà cung cấp, nhà triển khai Việt Nam phải chịu trách nhiệm đến mức nào và theo cơ chế gì đang gây ra tranh chấp thực tiễn chưa có căn cứ pháp lý giải quyết nhất quán. Khoảng trống về kế thừa nghĩa vụ khi tổ chức vận hành chấm dứt tồn tại trước khi hệ thống AI được đóng là một biến thể mới của vấn đề phân định trách nhiệm, mở rộng phạm vi của Kỹ thuật 4 từ phân định trách nhiệm giữa các chủ thể đang hoạt động đồng thời sang phân định trách nhiệm theo trục thời gian, bao gồm cả tình huống một chủ thể không còn tồn tại.

3.2.4.5. Kỹ thuật 5 - Miễn trừ tỷ lệ

Kỹ thuật miễn trừ tỷ lệ là kỹ thuật còn yếu ảnh hưởng trực tiếp đến khả năng đổi mới của hệ sinh thái AI Việt Nam. Điều quan trọng cần ghi nhận là Nghị định 142 đã có các ngoại lệ cho hệ thống AI nội bộ không tác động ra ngoài tổ chức, hệ thống trong giai đoạn thử nghiệm môi trường kiểm soát, và hệ thống chỉ tạo khuyến nghị tham khảo, đây là các miễn trừ hợp lý theo tiêu chí mục đích sử dụng. Tuy nhiên, miễn trừ theo mục đích sử dụng và miễn trừ theo quy mô doanh nghiệp là hai thiết kế khác nhau về bản chất và trong khi miễn trừ thứ nhất đã có, miễn trừ thứ hai lại thiếu vắng hoàn toàn.

Hệ quả thực tiễn của khoảng trống này được chuyên gia CG-DN-03 mô tả cụ thể: một startup AI với 12 nhân viên và doanh thu chưa đến 5 tỷ VNĐ/năm phải tuân thủ cùng yêu cầu với doanh nghiệp lớn, trong khi chi phí kiểm định AI bên thứ ba ước tính 200-500 triệu VNĐ/hệ thống tương đương 4-10% doanh thu hàng năm. So sánh với mô hình ba tầng miễn trừ của Hàn Quốc nơi startup dưới ba năm tuổi được miễn hoàn toàn kiểm định bên thứ ba sau khi cộng đồng startup phản hồi và thiết kế lại điều khoản, sự thiếu vắng của ngưỡng định lượng cụ thể theo quy mô doanh nghiệp trong khung pháp lý AI Việt Nam thiếu cơ chế bảo vệ tương đương cho hệ sinh thái startup AI trong nước [Xem Phụ lục 5].

Gắn liền với đó là sự thiếu vắng của VN-AI Verify công cụ tự đánh giá mà nếu có, startup sẽ có con đường chứng minh tuân thủ thay thế kiểm định bên thứ ba đầy đủ. Hai khoảng trống này cần được giải quyết đồng thời vì thiết kế ngưỡng mà không có công cụ tự chứng minh tuân thủ là thiếu giải pháp vận hành cho ngưỡng đó.

3.2.4.6. Kỹ thuật 6 - luật hóa tiêu chuẩn

Kỹ thuật pháp luật hóa tiêu chuẩn kỹ thuật là kỹ thuật có khoảng cách lớn nhất so với thực hành quốc tế, đồng thời là kỹ thuật phụ thuộc nhiều nhất vào điều kiện tiên quyết bên ngoài VBQPPL. Nhận thức về tầm quan trọng của tầng tiêu chuẩn kỹ thuật trong Nghị định 142 là có điều khoản về kiểm định, đánh giá sự phù hợp và watermarking đều viện dẫn "tiêu chuẩn kỹ thuật" song viện dẫn đó không đi kèm với hai yếu tố kỹ thuật thiết yếu để tạo ra hiệu lực thực chất.

Yếu tố thiếu thứ nhất là chỉ định TCVN cụ thể trong điều khoản viện dẫn. Viện dẫn "tiêu chuẩn kỹ thuật" theo kiểu chung chung mà không chỉ rõ TCVN nào dẫn đến tình huống mà chuyên gia CS-03 mô tả doanh nghiệp tuân thủ TCVN ISO/IEC 42001 không biết liệu điều đó có nghĩa là đã đáp ứng yêu cầu quản lý rủi ro AI của Nghị định 142 hay không, vì không có văn bản nào khẳng định điều đó, phủ nhận chính mục tiêu của kỹ thuật luật hóa tiêu chuẩn là tạo ra sự chắc chắn pháp lý. Yếu tố thiếu thứ hai là cơ chế suy đoán tuân thủ. Trong khi EU AI Act Điều 40 thiết kế cơ chế tiêu chuẩn hài hoà rõ ràng, tuân thủ tiêu chuẩn trong danh sách công bố trên Official Journal tạo ra suy đoán phù hợp chính thức và Texas TRAIGA cung cấp biện hộ khẳng định mạnh hơn cho doanh nghiệp tuân thủ NIST AI RMF, Nghị định 142 chưa có cơ chế tương đương ở bất kỳ cấp độ nào.

3.2.5. Đánh giá việc xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Tiêu chí 1, tính có căn cứ thực tiễn và dựa trên bằng chứng đạt tối thiểu ở chiều quy trình và đạt một phần ở ba chiều còn lại. Điểm chung của cả bốn chiều là chẩn đoán đúng ở tầm vấn đề nhưng chưa chuyển được thành căn cứ cho các lựa chọn cụ thể. Thứ nhất, không có công bố chính thức về đánh giá tác động pháp lý (*RIA*) được thực hiện trước khi soạn thảo Luật AI 2025, tức là không có tài liệu công khai phân tích chi phí - lợi ích của các lựa chọn chính sách khác nhau và không có phân tích thực trạng rủi ro AI cụ thể tại Việt Nam làm căn cứ thiết kế.

Thứ hai, về căn cứ thực tiễn trong nước, trong khi Việt Nam đã chứng kiến nhiều vụ việc AI nghiêm trọng (lừa đảo deepfake, quyết định tín dụng thiên lệch, vi phạm dữ liệu cá nhân), chưa có thống kê chính thức riêng về sự cố và thiệt hại do AI. Chưa có bằng chứng công khai cho thấy dữ liệu từ các vụ việc này được hệ thống hóa và sử dụng một cách chính thức làm căn cứ thiết kế quy phạm.

Thứ ba, về tham khảo quốc tế, Luật AI 2025 cho thấy ảnh hưởng rõ ràng của EU AI Act (đặc biệt trong kiến trúc phân loại rủi ro và định nghĩa) nhưng không có tài liệu giải thích tại sao lựa chọn một số yếu tố từ EU và không lựa chọn các yếu tố khác, cũng như tại sao không học từ Hàn Quốc hay Singapore nhiều hơn dù bối cảnh tương đồng hơn.

Tiêu chí 2, tính khoa học trong thiết kế kiến trúc chưa đạt ở chiều quy trình, vì không có bước thiết kế kiến trúc có sản phẩm thành văn trước khi soạn thảo. Ở ba chiều còn lại, tiêu chí này chỉ đạt một phần, hạn chế của mỗi chiều đều là hệ quả của sự thiếu vắng nói trên.

Thứ nhất, không có tài liệu kiến trúc chính thức giải thích logic nội tại của toàn bộ hệ thống, tại sao lựa chọn cơ chế danh mục theo QĐ TTg; quan hệ của Luật AI 2025 với các luật liên quan như thế nào.

Thứ hai, không có phân tích khoảng trống chính thức xác định những thành phần còn thiếu của khung pháp lý như tiêu chuẩn kỹ thuật, thiết chế điều phối xuyên ngành, cơ chế thích ứng.

Thứ ba, chưa phân biệt rõ ràng giữa hệ thống AI thụ động và tác nhân AI trong kiến trúc điều chỉnh.

Tiêu chí 3, tính tham vấn thực chất và bao trùm đạt tối thiểu ở chiều chủ thể và chiều quy trình, đạt một phần ở chiều nội dung, và chưa đạt ở chiều kỹ thuật. Còn có những hạn chế đáng kể như: *(i)* chưa có bằng chứng công khai về tham vấn kỹ thuật chuyên sâu với doanh nghiệp AI - loại tham vấn để thu thập ý kiến về tính khả thi kỹ thuật của từng yêu cầu pháp lý; *(ii)* chưa có bằng chứng về cơ chế tiếp cận tiếng nói của người tiêu dùng và các nhóm dễ tổn thương bị ảnh hưởng bởi AI, đặc biệt người lao động có nguy cơ bị thay thế bởi AI, người cao tuổi, người khuyết tật; *(iii)* chưa có báo cáo phản hồi ý kiến công bố lý do tiếp nhận hoặc không tiếp nhận các ý kiến cụ thể.

Tiêu chí 4, tính minh bạch và trách nhiệm giải trình của quy trình đạt tối thiểu ở chiều nội dung và đạt một phần ở ba chiều còn lại. Các dự thảo và báo cáo thẩm tra đã được công khai, và định hướng chung của các thay đổi đã được cơ quan chủ trì giải thích. Tuy nhiên, còn ba khoảng trống: Những thay đổi lớn chưa được giải trình riêng như việc bỏ tầng rủi ro không chấp nhận được, thu hẹp trách nhiệm liên đới và bỏ nghĩa vụ cung cấp chứng cứ cho Tòa án. Chưa xác định ai chịu trách nhiệm khi nghĩa vụ đã có hiệu lực mà thiết chế thi hành chưa sẵn sàng. Ở chiều kỹ thuật, chưa có văn bản hợp nhất làm căn cứ chính thức và chưa có bản dịch chính thức. Tóm lại, hoạt động xây dựng minh bạch về kết quả, nhưng chưa minh bạch về lý do của các lựa chọn.

Tiêu chí 5, tính bền vững và khả năng cải tiến liên tục:

Luật Ban hành VBQPPL 2025 cung cấp một số công cụ hữu ích như cơ chế nghị quyết Chính phủ cho vấn đề cấp bách; quy trình rút gọn mở rộng; quy định khi VBQPPL hết hiệu lực thì văn bản chi tiết vẫn có thể tiếp tục hiệu lực bảo đảm tính kế thừa. Tuy nhiên, Luật AI 2025 và Nghị định 142 vẫn thiếu điều khoản rà soát bắt buộc, cơ chế thu thập, phân tích và dự báo các xu hướng, tín hiệu và biến động tiềm ẩn trong môi trường bên ngoài và vòng phản hồi từ dữ liệu sự cố đến cập nhật chính sách.

Không có văn bản nào quy định trách nhiệm của cơ quan nhà nước trong việc theo dõi xu hướng phát triển AI và đánh giá hàm ý pháp lý. Không có cơ chế học hỏi thể chế hóa từ thực tiễn vận hành, báo cáo sự cố tự nguyện, đánh giá hậu kỳ, phản hồi chính sách từ sandbox.

Hệ quả khung pháp lý AI Việt Nam đã được trang bị một số công cụ để thích ứng, nhưng hệ thống chủ thể xây dựng chưa có năng lực để duy trì sự thích ứng đó.

Về các điều kiện bảo đảm xây dựng khung pháp lý, có thể đánh giá sơ bộ như sau:

Điều kiện về ý chí chính trị và định hướng chiến lược là điều kiện mạnh nhất, thể hiện ở Nghị quyết 57-NQ/TW năm 2024 xác định thể chế là đột phá; Nghị quyết 66-NQ/TW năm 2025 đặt yêu cầu đổi mới công tác xây dựng và thi hành pháp luật; Quyết định 1671/QĐ-TTg ngày 28/8/2026 phê duyệt Chiến lược quốc gia về AI đến năm 2030, tầm nhìn 2045. Ban Chỉ đạo Trung ương về phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số đã được thành lập.

Điều kiện về năng lực thể chế và nhân lực đạt tối thiểu một phần, là điều kiện yếu nhất cùng với tài chính. Do tổ chức đánh giá sự phù hợp chuyên về AI chưa hình thành. Chưa có chương trình bồi dưỡng có hệ thống về AI cho thẩm phán và thanh tra viên.

Điều kiện về hạ tầng pháp lý và kỹ thuật còn lệch nhau. Hạ tầng pháp lý nền tương đối đầy đủ: Luật Dữ liệu số 60/2024/QH15; Luật Bảo vệ dữ liệu cá nhân số 91/2025/QH15; Luật Ban hành VBQPPL số 64/2025/QH15, có trình tự rút gọn; Thông tư 13/2026/TT-BKHCN về trình tự rút gọn xây dựng TCVN. Trong khi hạ tầng kỹ thuật phục vụ kiểm thử và đánh giá sự phù hợp còn yếu. Năng lực vận hành thực tế của các nền tảng quản lý AI cần được kiểm chứng độc lập.

Điều kiện về sự tham gia của các bên liên quan đạt tối thiểu. Kênh chính thức đã vận hành nhưng năng lực tham gia của khu vực ngoài nhà nước còn hạn chế: chưa có bằng chứng công khai về tham vấn kỹ thuật chuyên sâu; hiệp hội ngành chưa đủ năng lực tự điều chỉnh.

Điều kiện về nguồn lực tài chính có cam kết vĩ mô mạnh nhưng phân bổ cho khâu thể chế chưa rõ. Nghị quyết 57-NQ/TW đặt mục tiêu bố trí ít nhất 3% tổng chi ngân sách hằng năm cho khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số. Nghị định 142 quy định phiếu hỗ trợ tối đa 50% chi phí cho doanh nghiệp (khoản 3 Điều 26). Tuy vậy, chưa xác định được dòng kinh phí riêng cho xây dựng tiêu chuẩn về AI, cho hạ tầng kiểm thử và cho đánh giá tác động pháp luật về AI.

3.2.6. Nguyên nhân của các hạn chế trong xây dựng khung pháp lý

Thứ nhất, nguyên nhân của hạn chế về chủ thể xây dựng

Một là, việc phân vai giữa các cơ quan chủ yếu dựa vào cấu trúc bộ máy sẵn có, chưa dựa trên đánh giá năng lực. Cơ quan đầu mối vì vậy phải gánh khối lượng vượt quá khả năng chuẩn bị trong thời gian ngắn.

Hai là, chưa có cơ chế phối hợp có hiệu lực pháp lý giữa cơ quan đầu mối về AI, cơ quan quản lý hạ tầng và dữ liệu, các bộ quản lý ngành và các chủ thể trong

chuỗi rủi ro như nền tảng số. Ngành ngân hàng cho thấy cơ chế phối hợp song phương làm được, nhưng nó chưa được nhân rộng thành một cơ chế chung.

Ba là, kênh tham gia chính thức còn hạn chế đối với tổ chức xã hội, giới nghiên cứu và nhóm chịu tác động.

Bốn là, chưa đầu tư vào năng lực duy trì. Chưa có thiết chế thường trực để lưu giữ tri thức qua các vòng xây dựng khung pháp lý. Chưa có chính sách quốc gia về đào tạo nhân lực liên ngành pháp lý-công nghệ.

Thứ hai, nguyên nhân của hạn chế về nội dung xây dựng

Một là, chưa có đánh giá công khai về tác động đối với người chịu ảnh hưởng. Thiếu cơ chế bắt buộc công bố căn cứ lựa chọn nội dung: đối với văn bản ủy quyền cấp Quyết định Thủ tướng (như Quyết định 33/2026/QĐ-TTg), không có quy định pháp luật nào bắt buộc phải công bố đồng thời tài liệu thuyết minh căn cứ lựa chọn nội dung, khác với yêu cầu thuyết minh chi tiết áp dụng cho dự án luật. Thiếu dữ liệu thực tiễn làm căn cứ định lượng cho quyết định: nhiều quyết định nội dung phức tạp đòi hỏi dữ liệu thực tiễn về chi phí tuân thủ và mô hình tinh chỉnh phổ biến tại Việt Nam mà quá trình soạn thảo chưa có bằng chứng thu thập một cách hệ thống.

Hai là, nghĩa vụ được đặt chủ yếu lên bên triển khai trong nước, vì đây là chủ thể dễ quản lý nhất. Nó không được đặt theo khả năng kiểm soát và thông tin của từng bên.

Ba là, các luật có liên quan như Luật BVDLCN, Luật SHTT sửa đổi và Luật AI được xây dựng song song mà không có thiết kế chung. Hệ quả là các chế độ pháp lý chỉ được nối với nhau ở mức hồ sơ.

Thứ ba, nguyên nhân của hạn chế về quy trình xây dựng

Một là, sự vắng mặt của một mô hình quy trình được thiết kế lại có chủ đích cho đối tượng công nghệ thay đổi nhanh. Quy trình lập pháp Việt Nam hiện hành được thiết kế phù hợp cho đối tượng điều chỉnh có tốc độ thay đổi chậm (nơi mô hình tuần tự, tham vấn một vòng và thiếu kênh dữ liệu thời gian thực không tạo ra hậu quả nghiêm trọng). AI với chu kỳ đổi mới 12-18 tháng bộc lộ rõ những hạn chế vốn tồn tại tiềm ẩn trong quy trình lập pháp truyền thống mà các lĩnh vực điều chỉnh chậm thay đổi trước đây chưa từng làm lộ ra.

Hai là, yêu cầu về tốc độ được đặt lên trên yêu cầu về trình tự. Định hướng thể chế phải đi trước, cùng với mục tiêu tháo gỡ điểm nghẽn trong thời gian ngắn đã khiến giai đoạn chẩn đoán và giai đoạn thiết kế bị nén vào quá trình soạn thảo.

Ba là, không bổ sung bước thiết kế kiến trúc đầu tiên. Thiếu bước kiểm tra mức độ sẵn sàng của cơ chế vận hành trước khi quyết định thời điểm có hiệu lực của VBQPPL và không dự liệu phương thức tiếp nhận thay thế trong thời gian Cổng

thông tin một cửa chưa vận hành chính thức. Giai đoạn đánh giá và thích ứng vì vậy chưa có dữ liệu để vận hành.

Thứ tư, nguyên nhân của hạn chế về kỹ thuật lập pháp

Nhận thức hẹp đồng nhất "xây dựng khung pháp lý" với "ban hành văn bản quy phạm pháp luật": mức độ vận dụng sáu kỹ thuật giảm dần một cách nhất quán từ kỹ thuật thuần túy pháp lý-văn bản (định nghĩa chức năng, đạt mức cao nhất) đến kỹ thuật đòi hỏi điều kiện kỹ thuật bên ngoài văn bản (miễn trừ tỷ lệ, pháp luật hóa tiêu chuẩn, đạt mức thấp nhất). Mô hình này phản ánh việc nguồn lực và sự chú ý của quá trình xây dựng tập trung vào hoạt động soạn thảo văn bản, trong khi các kỹ thuật đòi hỏi xây dựng đồng thời tầng luật mềm, tiêu chuẩn kỹ thuật và công cụ tự đánh giá vốn không được nhận thức đầy đủ là một phần không thể tách rời của "xây dựng khung pháp lý" nhận được ít đầu tư tương xứng.

3.2.7. Tình huống điển hình: Phản ứng pháp lý liên ngành trước làn sóng lừa đảo deepfake

3.2.7.1. Lý do lựa chọn và phương pháp luận tình huống điển hình

Luận án thiết kế tình huống đơn có bốn đơn vị lồng ghép tương ứng với bốn chiều xây dựng khung pháp lý, để phân tích quá trình phản ứng pháp lý liên ngành đối với một rủi ro AI xác định, từ khi rủi ro được nhận diện đến tháng 9/2026 đồng thời so sánh với Trung Quốc, Liên minh châu Âu và Singapore. Tình huống lừa đảo bằng deepfake được lựa chọn vì là tình huống then chốt để kiểm định lý thuyết và bao trùm cả thời kỳ trước và sau khi Luật AI có hiệu lực. Nhờ đó, luận án quan sát được cùng một rủi ro qua hai trạng thái của khung pháp lý. Kết quả cho phép khái quát hóa về mặt phân tích đối với các cơ chế được nêu trong bốn mệnh đề, cung cấp bằng chứng hỗ trợ kiểm nghiệm tính đúng đắn của mô hình khung pháp lý trong điều kiện thực tiễn Việt Nam và xác nhận rằng các khoảng trống là hậu quả của thiếu tư duy kiến trúc hệ thống trong xây dựng khung pháp lý. Kết quả không cho phép khái quát hóa thống kê sang mọi rủi ro AI, sang các hệ thống pháp luật khác hay sang hiệu quả thi hành trước khi thời hạn chuyển tiếp kết thúc năm 2027.

3.2.7.2. Mô tả tình huống

Thứ nhất, về quy mô và phạm vi: Từ 2023 đến hết năm 2025, Việt Nam chứng kiến làn sóng lừa đảo sử dụng công nghệ deepfake và AI voice cloning ở quy mô chưa từng có. Theo số liệu của Cục An toàn Thông tin (Bộ TT&TT cũ), trong năm 2024 hệ thống tiếp nhận thông tin cảnh báo do Cục quản lý và vận hành đã ghi nhận 5.058 phản ánh trường hợp lừa đảo trực tuyến do người dùng internet gửi về. Trong đó, có 190 trường hợp phản ánh được tiếp nhận thông qua Trang cảnh báo an toàn thông tin Việt Nam (canhbao.khonggianmang.vn); 4.868 trường hợp phản ánh qua tổng đài 156/5656 [64]. Cứ khoảng 555 người thì có 1 người là nạn nhân lừa đảo,

tương đương 0,18%, thiệt hại hơn 6.000 tỷ đồng trong 11 tháng đầu năm 2025 [68]. Theo Sáng kiến Toàn cầu Chống Tội phạm Có tổ chức Xuyên quốc gia (GI-TOC), Việt Nam đã ghi nhận mức tăng cao nhất trong khu vực về các vụ lừa đảo liên quan đến deepfake (25,3%) [215], cho thấy mối đe dọa này đang ảnh hưởng nghiêm trọng làm suy yếu an ninh mạng bằng cách đánh cắp danh tính và lừa dối cả người dùng lẫn hệ thống xác thực.

Đặc điểm kỹ thuật của phương thức lừa đảo là kẻ tấn công thu thập ảnh và video của nạn nhân mục tiêu (từ mạng xã hội, công khai), sử dụng công cụ deepfake (DeepFaceLab, OpenAI, Zao hoặc các mô hình nguồn mở) để tạo video và giọng nói tổng hợp của người thân, sếp, hoặc cán bộ nhà nước liên lạc qua Zalo Video Call, Facebook Messgenger hoặc điện thoại để yêu cầu chuyển tiền khẩn cấp.

Đối tượng bị hại có phổ rộng đáng chú ý, không chỉ người cao tuổi thiếu kỹ năng số mà cả người trung niên có học thức, thậm chí cán bộ nhà nước và doanh nhân. Điều này phản ánh mức độ tinh vi ngày càng cao của công nghệ deepfake và giọng nói AI tổng hợp, đến mức người thân quen vẫn khó phân biệt với giao tiếp thực.

Thứ hai, về các vụ việc điển hình có thể kể đến:

Vụ 1 - Lừa đảo bằng deepfake giả mạo người thân (tháng 7/2023): Một người phụ nữ tên L tại thành phố Hồ Chí Minh nhận được tin nhắn của bạn thân mượn 10 triệu vì có việc cần gấp, chuyển vào tài khoản người khác. Nghi ngờ facebook của bạn bị hack nên chị đã gọi video call để kiểm chứng. Thắc mắc thấy mặt bạn hơi mờ thì được giải thích sóng yếu nên chị L yên tâm chuyển khoản 10 triệu đồng theo số tài khoản bạn yêu cầu. Sau đó, chị L gọi điện thoại trực tiếp cho bạn thì mới biết mình đã sập bẫy deepfake [96].

Vụ 2 - Lừa đảo giả mạo lãnh đạo bằng AI voice cloning (tháng 8/2024): Một cán bộ tại TP. Hà Nội nhận được một cuộc gọi từ người tự nhận là Bí thư Thành ủy TP. Hồ Chí Minh với mục đích vay tiền. Giọng nói trong điện thoại rất giống đồng chí Bí thư Nguyễn Văn Nên, nhưng khi xác minh số tài khoản thì người nhận lại không đúng tên. Bản thân Bí thư và những người làm việc xung quanh thường xuyên nhận được tin tức về những vụ lừa đảo tương tự. Không chỉ có lãnh đạo mà ngay cả các cán bộ trong Thành ủy TP. Hồ Chí Minh cũng liên tiếp bị những kẻ xấu lợi dụng hình ảnh để tạo ra những câu chuyện nhằm lừa đảo [124].

Vụ 3 - Deepfake CEO lừa đảo đầu tư (tháng 12/2024): Ông Nguyễn Duy Hưng, người đứng đầu HĐQT của Công ty CP Chứng khoán SSI, đã trở thành nạn nhân của việc sử dụng công nghệ Deepfake AI để giả mạo hình ảnh của ông. Kẻ gian đã thực hiện các livestream giả mạo trên các trang Zalo, Tiktok, và Telegram nhằm mục đích lừa đảo các nhà đầu tư. Họ cũng đã tạo một tài khoản mang tên Nguyễn Duy Hưng trên Zalo để phát biểu ý kiến và đưa ra tư vấn đầu tư. Qua Zalo, nhóm lừa đảo này lại

hướng dẫn người dùng sang các nhóm kín trên Telegram. Tại đây, chúng tiếp tục sử dụng công nghệ AI để tạo ra hình ảnh và video giả mạo ông Nguyễn Duy Hưng, củng cố việc tư vấn đầu tư chứng khoán và thuyết phục các nhà đầu tư mắc bẫy [123].

Thứ ba, phân tích tình huống theo bốn chiều

Ở chiều chủ thể, phản ứng pháp lý do cơ quan gần nhất với thiệt hại là Ngân hàng Nhà nước dẫn dắt trong phạm vi ngành bằng công cụ của mình. Ở cấp ngành, đã có cơ chế phối hợp. Ngân hàng Nhà nước đang mở rộng hệ thống giám sát và phòng ngừa rủi ro SIMO [62], tăng cường trao đổi dữ liệu và phối hợp truy vết với các đơn vị chức năng của Bộ Công an theo quy chế phối hợp giữa hai cơ quan. Vì vậy, việc thiếu điều phối xảy ra ở hai điểm giao nhau: giữa ngành ngân hàng với các chủ thể khác trong chuỗi lừa đảo, nhất là nền tảng số; giữa pháp luật chuyên ngành với cơ quan đầu mối quản lý nhà nước về AI.

Ở chiều nội dung, Luật chỉ cấm việc giả mạo, mô phỏng người thật để lừa dối khi hành vi đó có chủ đích, có hệ thống và gây tổn hại nghiêm trọng. Nghĩa vụ gắn nhãn chỉ áp dụng cho hai trường hợp: nội dung mô phỏng người thật và nội dung tái hiện sự kiện thực tế. Khung pháp lý đã có lớp phòng ngừa ở khâu xác thực và giao dịch, nhưng vẫn thiếu lớp khắc phục. Khi công cụ AI tạo sinh được sử dụng thuộc về nhà cung cấp nước ngoài không có hiện diện pháp lý tại Việt Nam (một tình huống thực tế rất phổ biến với các công cụ deepfake nguồn mở hoặc dịch vụ nước ngoài), cơ chế thực thi nghĩa vụ hành chính đối với chủ thể này gần như không khả thi trên thực tế. Thiếu quy tắc ai chịu tổn thất khi các bên đã hoặc chưa làm đủ nghĩa vụ; quyền tiếp cận chứng cứ của nạn nhân. Câu hỏi ai chịu trách nhiệm khi AI thực hiện giao dịch có sai sót cũng là vấn đề được giới chuyên môn nêu ra, nhưng chưa có lời giải rõ ràng. BLHS hiện hành chưa có tình tiết định khung tăng nặng đặc thù cho hành vi sử dụng công nghệ AI/deepfake làm phương tiện phạm tội. Về trách nhiệm và quyền của nạn nhân, nạn nhân có quyền yêu cầu bồi thường thiệt hại theo quy định về trách nhiệm bồi thường thiệt hại ngoài hợp đồng (Bộ luật Dân sự 2015) đối với đối tượng lừa đảo (khi được xác định danh tính và có tài sản để thi hành án - một điều kiện thực tế thường khó đáp ứng trong loại tội phạm này), có quyền yêu cầu cơ quan điều tra áp dụng biện pháp phong tỏa, thu hồi tài sản đối với tài khoản nhận tiền theo quy định của Bộ luật Tố tụng hình sự 2015. Nạn nhân không có nghĩa vụ pháp lý phải biết trước về công nghệ deepfake, không có quy định pháp luật hiện hành nào đặt ra nghĩa vụ cho công dân phải có kiến thức nhận diện deepfake như điều kiện để được bảo vệ pháp lý. Trong trường hợp đối tượng lừa đảo không bị bắt giữ hoặc không có tài sản để thi hành án (tình huống phổ biến với tội phạm công nghệ cao xuyên biên giới) và không có chủ thể nào khác trong chuỗi (ngân hàng, nhà cung cấp công cụ AI) được xác định có lỗi đủ

rõ ràng để chịu trách nhiệm bồi thường theo quy định hiện hành, nạn nhân có nguy cơ hoàn toàn không có kênh khắc phục hiệu quả.

Ở chiều quy trình, nghĩa vụ đánh dấu và gắn nhãn có hiệu lực từ tháng 3 và tháng 5/2026. Chế tài đối với hành vi dùng deepfake giả mạo sinh trắc học được quy định tại Nghị định 330/2026/NĐ-CP, ban hành ngày 19/8/2026. Chế tài chuyên biệt đối với vi phạm nghĩa vụ gắn nhãn theo pháp luật về AI vẫn chưa hoàn thiện. Đây là độ trễ đúng như mệnh đề dự báo, phản ứng đi sau thiệt hại và thiếu đồng bộ ở tầng pháp luật chung về AI. Nhưng ở tầng ngân hàng, chuỗi văn bản từ năm 2024 đến 2026 cho thấy một vòng thích ứng có trình tự, công bố dữ liệu vận hành từ SIMO làm căn cứ và điểm yếu được cơ quan công an chỉ ra ở nhiều hệ thống, đối tượng chỉ cần chèn một bức ảnh tĩnh hoặc đoạn video ngắn là qua được bước xác thực. Văn bản tiếp theo nâng yêu cầu đúng vào điểm yếu đó. Từ yêu cầu xác thực sinh trắc học năm 2024, chuỗi văn bản chuyển sang chuẩn chống giả mạo ISO 30107 năm 2026, sang chứng nhận bắt buộc dự kiến năm 2027.

Ở chiều kỹ thuật, việc đánh dấu và gắn nhãn nội dung tổng hợp ở tầng pháp luật chung về AI dựa trên hướng dẫn kỹ thuật tham chiếu nhưng chưa có tiêu chuẩn quốc gia để làm thước đo chung. Ở tầng chuyên ngành, Ngân hàng Nhà nước đã dẫn chiếu bắt buộc một tiêu chuẩn quốc tế về chống giả mạo sinh trắc học trong thông tư, và đang chuyển sang yêu cầu chứng nhận.

Tình huống điển hình này xác nhận bằng dữ liệu chính thống rằng hạn chế không tồn tại ở một chiều đơn lẻ trong xây dựng khung pháp lý, củng cố luận điểm trung tâm của luận án rằng hoàn thiện khung pháp lý AI đòi hỏi can thiệp đồng thời vào nhiều tầng và nhiều chiều xây dựng.

3.2.7.3. So sánh phản ứng pháp lý của các quốc gia khác

Trung Quốc đối mặt với làn sóng lừa đảo deepfake tương tự từ năm 2022 và đã phản ứng bằng hai VBQPPL đặc thù: Quy định về Quản lý Dịch vụ Deep Synthesis trên Thông tin Internet (2023) và Quy định về Quản lý Dịch vụ AI Tổng hợp (2023). Cả hai quy định đều thiết lập định nghĩa kỹ thuật-pháp lý rõ ràng; yêu cầu watermarking bắt buộc với tiêu chuẩn kỹ thuật cụ thể; nghĩa vụ xác thực danh tính thực của người dùng và mức phạt cụ thể khi vi phạm.

Singapore thông qua Đạo luật Chống Tội phạm Trực tuyến (OCHA) vào tháng 7 năm 2023, cho phép chính phủ ban hành chỉ thị cho các nền tảng trực tuyến để ngăn chặn các tài khoản hoặc nội dung có khả năng liên quan đến lừa đảo tiếp cận người dùng Singapore. Các chương trình giáo dục công cộng về truyền thông kỹ thuật số và kiến thức thông tin, an ninh mạng và các nỗ lực chống lừa đảo phải đủ mạnh để bảo vệ và duy trì một xã hội kỹ thuật số an toàn. Một số chương trình bao gồm chiến dịch an

ninh mạng quốc gia "Kẻ thù vô hình" của Cơ quan An ninh mạng Singapore (CSA), và chiến dịch chống lừa đảo "Tôi có thể hành động chống lại lừa đảo" của Lực lượng Cảnh sát Singapore và Hội đồng Phòng chống Tội phạm Quốc gia. Văn phòng Giáo dục Công cộng về Lừa đảo (SPEO) cũng được thành lập vào năm 2023 để thúc đẩy các nỗ lực giáo dục công cộng chống lừa đảo và mở rộng phạm vi tiếp cận [310]. Năm 2025, Singapore tiến hành sửa đổi Bộ luật Hình sự để làm rõ điều khoản về tội phạm sử dụng công nghệ AI. Theo Điều 377BE(5) của Bộ luật Hình sự, các "deepfake" khiêu dâm trong đó hình ảnh hoặc bản ghi âm bị chỉnh sửa để mô tả nạn nhân. Việc sở hữu hoặc truy cập, hoặc phân phối hoặc đe dọa phân phối các "deepfake" khiêu dâm như vậy mà không có sự đồng ý đã là một hành vi phạm tội. Tuy nhiên, AI tạo sinh hiện nay cho phép tạo ra các hình ảnh hoặc video hoàn toàn tổng hợp về một người, mà không cần sử dụng hình ảnh hoặc bản ghi âm có sẵn. Điều khoản 50(a) và (b) của Dự luật cập nhật điều khoản liên quan để đón đầu sự phát triển này [275]. Cơ quan tiền tệ Singapore (MAS) ban hành tài liệu "Rủi ro mạng liên quan đến Deepfake" nhằm mục đích nâng cao nhận thức của các tổ chức tài chính bằng cách cung cấp một cái nhìn tổng quan toàn diện về các mối đe dọa và rủi ro mới nổi do deepfake gây ra, tác động đối với lĩnh vực tài chính và các biện pháp giảm thiểu khả thi mà các tổ chức tài chính có thể thực hiện để giải quyết những rủi ro này [269].

EU AI Act (Điều 50) yêu cầu nhà cung cấp hệ thống AI tạo ra nội dung âm thanh, hình ảnh và video tổng hợp phải đảm bảo nội dung được đánh dấu theo phương thức có thể phát hiện bằng máy. Quy định này tạo ra nghĩa vụ pháp lý ràng buộc với tiêu chuẩn kỹ thuật rõ ràng; cơ sở pháp lý cho kiểm định và xử phạt và áp dụng cho mọi nhà cung cấp AI tạo ra nội dung cho người dùng EU, bất kể trụ sở ở đâu.

Phân tích phản ứng pháp lý của ba khu vực trên cho phép xác định ba yếu tố tối thiểu mà một khung pháp lý hiệu quả với deepfake/synthetic media đều có, trong khi Việt Nam hiện chưa đáp ứng đủ ba yếu tố này: định nghĩa pháp lý rõ ràng về deepfake, synthetic media và ngưỡng kỹ thuật phân biệt; nghĩa vụ gắn nhãn bắt buộc với tiêu chuẩn kỹ thuật cụ thể; cơ chế trách nhiệm nền tảng rõ ràng đối với việc phát tán nội dung AI tổng hợp gây hại.

Tổng kết Chương 3

Việt Nam đã và đang chuẩn bị những điều kiện thiết yếu cho quá trình nghiên cứu, ứng dụng và phát triển AI phục vụ chuyển đổi số. Trên nền tảng các nội dung lý luận cơ bản ở Chương 2, Chương 3 phân tích, đánh giá thực trạng phát triển, ứng dụng AI; thực trạng khung pháp lý và xây dựng khung pháp lý cho phát triển, ứng

dụng AI ở Việt Nam, qua đó thấy được nhiều dấu hiệu tích cực. Tuy nhiên, mức hoàn thành khung pháp lý tổng thể chỉ đạt yêu cầu cơ bản, cơ cấu giữa các tầng còn mất cân đối; việc xây dựng khung pháp lý vẫn chưa thực sự được bảo đảm và phát huy so với yêu cầu đặt ra, còn nhiều hạn chế, do nhiều nguyên nhân khác nhau. Về chủ thể: việc phân vai giữa các cơ quan chưa dựa trên đánh giá năng lực, chưa có sự phối hợp giữa các chủ thể, kênh tham gia chính thức còn hạn chế, thiếu chuyên gia liên ngành pháp lý-công nghệ; về nội dung: chưa có đánh giá công khai về tác động đối với người chịu ảnh hưởng, thiếu cơ chế bắt buộc công bố căn cứ lựa chọn nội dung, thiếu dữ liệu thực tiễn làm căn cứ định lượng cho quyết định, nghĩa vụ không được đặt theo khả năng kiểm soát và thông tin của từng bên, các luật có liên quan được xây dựng song song mà không có thiết kế chung; về quy trình: mô hình quy trình chưa được thiết kế lại có chủ đích cho đối tượng công nghệ thay đổi nhanh, yêu cầu về tốc độ được đặt lên trên yêu cầu về trình tự, không bổ sung bước thiết kế kiến trúc đầu tiên; về kỹ thuật: nhận thức hẹp về khung pháp lý chỉ là VBQPPL. Tình huống điển hình về phản ứng pháp lý liên ngành trước làn sóng lừa đảo bằng deepfake bổ sung bằng chứng hỗ trợ kiểm nghiệm, cho thấy hạn chế tồn tại ở cả bốn chiều của việc xây dựng khung pháp lý.

Đây là cơ sở để luận án xác lập một số quan điểm chỉ đạo và các giải pháp nhằm thúc đẩy xây dựng khung pháp lý theo hướng thực chất và hiệu quả hơn ở Chương 4.

Chương 4
QUAN ĐIỂM, GIẢI PHÁP XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM

4.1. QUAN ĐIỂM XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM

Các quan điểm xây dựng, hoàn thiện khung pháp lý cho phát triển và ứng dụng AI ở Việt Nam thể hiện sự nhất quán, toàn diện và khoa học, xuất phát từ chủ trương của Đảng về chủ động tham gia cuộc CMCN lần thứ tư, hướng tới mục tiêu phát huy lợi thế công nghệ, bảo đảm tính hợp hiến và đồng bộ của hệ thống pháp luật, phù hợp với điều kiện đất nước và hội nhập quốc tế, đồng thời luôn đặt quốc phòng, an ninh và lợi ích quốc gia lên hàng đầu, tạo nền tảng vững chắc để Việt Nam phát triển bền vững trong kỷ nguyên số.

4.1.1. Bảo đảm thể chế đầy đủ quan điểm, chủ trương của Đảng và Nhà nước về thúc đẩy phát triển khoa học, công nghệ và trí tuệ nhân tạo

Cở sở hình thành quan điểm này bắt nguồn từ quyết định pháp luật trong thể chế chính trị Việt Nam phải trả lời trước ý chí chính trị đã được xác lập của Đảng và Nhà nước. Đây là tính đặc thù cấu trúc chính trị xuất phát từ nguyên tắc hiến định về vai trò lãnh đạo của Đảng Cộng sản Việt Nam đối với Nhà nước và xã hội (Điều 4 Hiến pháp 2013), không thể bị thay thế bởi các tiêu chí khoa học, quốc tế hay pháp lý thuần túy.

Trong hệ thống chính trị Việt Nam, quan điểm và chủ trương của Đảng là định hướng chính trị có giá trị dẫn dắt toàn bộ hoạt động lập pháp, hành pháp và tư pháp. Xây dựng khung pháp lý cho phát triển, ứng dụng AI không phải ngoại lệ, quá trình xây dựng phải bám sát và thể chế hóa đầy đủ định hướng của Đảng và Nhà nước, không được đi lệch hoặc thu hẹp phạm vi. Với ý nghĩa đảm bảo tính thống nhất chính trị và tầm nhìn chiến lược của Đảng, khung pháp lý phải cụ thể hóa các yêu cầu về đẩy mạnh chuyển đổi số, phát triển CNCNS và các định hướng từ các Nghị quyết của Bộ Chính trị (các Nghị quyết số 57 [11], 59 [12], 66 [13], 68 [14] năm 2025 về khoa học, công nghệ, đổi mới sáng tạo, chuyển đổi số, phát triển công nghiệp số, an ninh mạng và thể chế).

Đặc biệt, Nghị quyết 57-NQ/TW ngày 22/12/2024 của Bộ Chính trị về đột phá phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia [11] là văn kiện Đảng mạnh mẽ nhất và cập nhật nhất về AI. Nghị quyết xác định: "Phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia là đột phá quan trọng hàng đầu, là động lực chính để […] đưa đất nước phát triển bứt phá, giàu mạnh trong kỷ nguyên mới". Nghị quyết 57 yêu cầu thể chế là điều kiện tiên quyết, cần hoàn thiện và đi trước một bước. Đổi mới tư duy xây dựng pháp luật bảo đảm yêu cầu quản lý và

khuyến khích đổi mới sáng tạo, loại bỏ tư duy không quản được thì cấm. Hai yêu cầu này đặt ra tiêu chuẩn kép cho việc xây dựng khung pháp lý cho phát triển, ứng dụng AI: vừa đầy đủ không bỏ sót tầng nào trong cấu trúc đa tầng vừa thông thoáng không tạo rào cản không cần thiết cho đổi mới. Chủ trương của Đảng không chỉ được thể chế hóa qua văn bản quy phạm pháp luật mà phải được phản ánh trong toàn bộ bảy tầng của khung pháp lý. Nếu Đảng xác định từng bước làm chủ công nghệ AI là ưu tiên quốc gia nhưng hướng dẫn ngành dọc, bộ quy tắc ứng xử và tiêu chuẩn kỹ thuật hỗ trợ phát triển AI còn thiếu, thể chế hóa mới là hình thức, chưa phải thực chất.

Văn kiện Đại hội XIV của Đảng nhấn mạnh "khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số phải thực sự trở thành động lực then chốt của tăng trưởng" [38, tr.31]. Cụ thể hơn, Đảng đề ra nhiệm vụ "ưu tiên tiếp thu, chuyển giao, ứng dụng sáng tạo, làm chủ các công nghệ tiên tiến, công nghệ chiến lược, công nghệ lõi, công nghệ nguồn, công nghệ số, đặc biệt là trong các lĩnh vực công nghiệp điện tử, chip bán dẫn, trí tuệ nhân tạo…" [38, tr.107]. Như vậy, khung pháp lý về AI không đơn thuần là công cụ quản lý mà là biểu hiện của quyết tâm chính trị cao trong việc nắm bắt thời cơ của cuộc cách mạng công nghệ để đưa đất nước phát triển nhanh và bền vững.

4.1.2. Bảo đảm tính khoa học, thống nhất trong hệ thống pháp luật và đáp ứng yêu cầu thực tiễn

Căn cứ hình thành quan điểm này độc lập với việc một quyết định có phù hợp ý chí chính trị hay không, quyết định đó còn phải trả lời trước chuẩn mực về chất lượng tri thức, có dựa trên bằng chứng, có nhất quán logic nội tại, có khả thi kỹ thuật, có đáp ứng thực tiễn hay không. Một quyết định hoàn toàn phù hợp ý chí chính trị vẫn có thể sai về mặt khoa học nếu thiếu bằng chứng hoặc mâu thuẫn logic, do vậy hai loại đánh giá này không thể thay thế cho nhau.

Thứ nhất, tính khoa học trong xây dựng khung pháp lý không chỉ là yêu cầu hình thức, đây là điều kiện sống còn để khung pháp lý cho phát triển, ứng dụng AI có thể vận hành hiệu quả trong thực tiễn. Một khung pháp lý thiếu cơ sở khoa học sẽ không thể điều chỉnh được hành vi của đối tượng điều chỉnh có hiểu biết kỹ thuật sâu hơn cơ quan ban hành quy định.

Thứ hai, tính thống nhất đòi hỏi toàn bộ bảy tầng của khung pháp lý phải nhất quán với nhau về nội dung và logic, phải nhất quán với phần còn lại của hệ thống pháp luật Việt Nam. Một là, nhất quán theo chiều dọc thể hiện ở chỗ tầng văn bản quy phạm pháp luật phải không mâu thuẫn với Hiến pháp; nghị định phải không vượt quá phạm vi ủy quyền của luật; tiêu chuẩn kỹ thuật phải được ban hành trong phạm vi thẩm quyền và nhất quán với yêu cầu pháp lý. Đặc biệt, khi luật mềm và bộ quy tắc ứng xử phát triển, phải nhất quán với nguyên tắc đã đặt ra trong văn bản quy phạm pháp luật, không được phép có bộ quy tắc ứng xử ngành cho phép điều mà luật cấm. Hai là, nhất

quán theo chiều ngang thể hiện ở việc Luật AI 2025 phải không mâu thuẫn với Luật Bảo vệ dữ liệu cá nhân 2025, Luật An ninh mạng 2025, Luật Sở hữu trí tuệ sửa đổi 2025 là những văn bản điều chỉnh các khía cạnh khác nhau của cùng một đối tượng. Nghị định 142/2026/NĐ-CP đã thể hiện ý thức về vấn đề này qua nguyên tắc không trùng lặp với pháp luật chuyên ngành, nhưng các điểm giao thoa và khả năng mâu thuẫn trong tương lai cần được theo dõi và xử lý có hệ thống.

Thứ ba, tính thực tiễn đòi hỏi khung pháp lý AI phải được thiết kế phù hợp với điều kiện đặc thù của Việt Nam, không phải sao chép nguyên mô hình nước ngoài. Một là, phù hợp với năng lực thể chế: Yêu cầu pháp lý không được vượt quá năng lực thực thi thực tế của cơ quan nhà nước và năng lực tuân thủ thực tế của doanh nghiệp trong từng giai đoạn. Đây là lý do kiến trúc hai giai đoạn, yêu cầu cơ bản ngay và yêu cầu nâng cao theo lộ trình phù hợp hơn với thực tiễn Việt Nam so với kiến trúc đặt toàn bộ yêu cầu cùng lúc. Hai là, phù hợp với đặc điểm hệ sinh thái: Hệ sinh thái AI Việt Nam đặc trưng bởi: tỷ lệ SME và startup chiếm đa số; mức độ phụ thuộc cao vào AI nước ngoài (đặc biệt là nền tảng của Big Tech Mỹ và Trung Quốc) và năng lực tự phát triển mô hình nền tảng còn hạn chế. Khung pháp lý phải phản ánh những đặc điểm này: cơ chế miễn trừ tỷ lệ cho SME, cơ chế quản lý AI xuyên biên giới và khuyến khích phát triển năng lực AI nội địa. Ba là, phù hợp với trình tự ưu tiên theo thực tiễn rủi ro: Nguồn lực thể chế hạn chế đòi hỏi phân bổ có trọng tâm, tập trung vào rủi ro đang hiện hữu và có quy mô lớn (lừa đảo deepfake, thuật toán tín dụng thiên lệch, vi phạm dữ liệu trong huấn luyện AI) hơn là rủi ro lý thuyết. Thứ tự ưu tiên xây dựng khung pháp lý phải phản ánh thứ tự ưu tiên rủi ro thực tiễn tại Việt Nam. Bốn là, phù hợp với văn hóa pháp lý: Truyền thống civil law của Việt Nam ưa thích quy tắc rõ ràng hơn nguyên tắc linh hoạt; văn hóa tuân thủ còn đang hình thành. Điều này có nghĩa là mô hình luật mềm thuần túy như Nhật Bản chưa phù hợp với Việt Nam, cần kết hợp luật cứng ràng buộc với luật mềm hướng dẫn theo lộ trình phát triển văn hóa tuân thủ.

4.1.3. Bảo đảm xây dựng đồng bộ, cân đối giữa các tầng của khung pháp lý

Quan điểm này đòi hỏi các tầng của khung pháp lý về AI phải phát triển tương xứng với vai trò của từng tầng trong cơ chế điều chỉnh và phải được tạo lập theo một lộ trình chung. Nội hàm của quan điểm gồm bốn yêu cầu: Một là, đồng bộ về thời gian. Một nghĩa vụ chỉ nên có hiệu lực khi văn bản quy định chi tiết, chế tài, tiêu chuẩn làm thước đo và thiết chế thi hành nghĩa vụ đó đã sẵn sàng. Hai là, cân đối về mức độ phát triển. Những tầng mà nhiều tầng khác phụ thuộc vào, trước hết là tầng tiêu chuẩn kỹ thuật và tầng thiết chế, phải được ưu tiên. Không để tầng văn bản quy phạm phát triển vượt xa phần còn lại của khung. Ba là, đặt đúng vị trí trong thứ bậc. Nội dung hạn chế quyền, đặt nghĩa vụ mới hoặc quy định chế tài phải nằm

ở tầng luật. Nội dung kỹ thuật, có tính thay đổi nhanh, được đặt ở các tầng có khả năng cập nhật linh hoạt. Bốn là, thừa nhận đúng mức các nguồn bổ trợ. Án lệ, tập quán, lẽ công bằng và bộ quy tắc ứng xử của ngành có vị trí riêng trong khung, không bị coi là phần phụ.

Quan điểm này có căn cứ chính trị trực tiếp, đó là Kết luận số 09-KL/TW ngày 10/3/2026 của Bộ Chính trị xác định cấu trúc hệ thống pháp luật theo thứ bậc các nguồn luật. Cấu trúc này gồm hệ thống văn bản quy phạm pháp luật và các nguồn bổ trợ là án lệ, tập quán, lẽ công bằng. Kết luận cũng yêu cầu mở rộng phạm vi áp dụng và xác định rõ giá trị pháp lý của các nguồn bổ trợ. Kết luận còn yêu cầu nghiên cứu bổ sung những nội dung phải được quy định trong văn bản luật, liên quan đến hạn chế quyền, đặt ra nghĩa vụ mới và biện pháp chế tài đối với người dân, doanh nghiệp. Điểm tiến bộ nổi bật của Kết luận này là chuyển từ tư duy xây dựng pháp luật theo từng văn bản đơn lẻ sang tư duy thiết kế hệ thống pháp luật như một chỉnh thể có cấu trúc.

4.1.4. Bảo đảm phù hợp với pháp luật quốc tế và xu hướng chung của các quốc gia trong bối cảnh hội nhập

Căn cứ cho quan điểm này độc lập với cả ý chí chính trị trong nước lẫn chất lượng khoa học nội tại, quyết định pháp luật của một quốc gia trong bối cảnh hội nhập còn phải trả lời trước hệ quy chiếu bên ngoài có tương thích với chuẩn mực và xu hướng chung của cộng đồng quốc tế hay không.

Hội nhập quốc tế về AI có những đặc điểm khác biệt so với hội nhập trong các lĩnh vực thương mại hay đầu tư truyền thống. Một là, không có điều ước quốc tế ràng buộc về AI, không như WTO trong thương mại hay Công ước Vienna trong pháp luật điều ước, lĩnh vực AI chưa có văn kiện pháp lý quốc tế ràng buộc nào. Hai là, cạnh tranh địa-chính trị về AI đang tạo ra cạnh tranh chuẩn mực giữa mô hình EU, mô hình Mỹ và mô hình Trung Quốc buộc Việt Nam phải định vị trong không gian cạnh tranh này. Ba là, tính tương thích quốc tế của khung pháp lý AI ảnh hưởng trực tiếp đến khả năng tiếp cận thị trường quốc tế của doanh nghiệp AI Việt Nam và khả năng thu hút đầu tư nước ngoài. Từ đặc điểm này, quan điểm hội nhập trong xây dựng khung pháp lý AI phải là "tương thích có chọn lọc", tham chiếu có phân tích đối với chuẩn mực quốc tế và học có chọn lọc từ kinh nghiệm các quốc gia phù hợp với bối cảnh Việt Nam.

Tương thích quốc tế không có nghĩa là từ bỏ chủ quyền chính sách. Việt Nam cần duy trì quyền thiết kế khung pháp lý AI phù hợp với lợi ích quốc gia và định hướng phát triển, điều này đặc biệt quan trọng trong bối cảnh cạnh tranh địa-chính trị AI gay gắt. Bảo vệ chủ quyền chính sách trong xây dựng khung pháp lý AI bao gồm: chủ quyền dữ liệu quy định về lưu trữ, xử lý và chuyển dữ liệu người dùng Việt Nam phải

phản ánh lợi ích quốc gia trong bảo vệ thông tin; chủ quyền công nghệ khuyến khích phát triển năng lực AI nội địa, không phụ thuộc hoàn toàn vào hạ tầng AI nước ngoài và chủ quyền thông tin kiểm soát AI có khả năng tác động đến dư luận và an ninh chính trị theo phương thức phù hợp với Hiến pháp và các cam kết quốc tế. Quan điểm chủ quyền pháp lý không phải là bảo hộ cực đoan hay đóng cửa mà là định vị chiến lược hội nhập có nguyên tắc, tích hợp vào hệ sinh thái AI toàn cầu nhưng bảo vệ không gian chính sách để thiết kế khung pháp lý phù hợp với lợi ích quốc gia và con đường phát triển đặc thù.

Quan điểm hội nhập không chỉ là tiếp nhận chuẩn mực quốc tế mà phải bao gồm chủ động tham gia định hình chuẩn mực đó. Khoảng trống lớn trong xây dựng khung pháp lý AI quốc tế là thiếu quan điểm của các quốc gia đang phát triển, phần lớn chuẩn mực hiện nay phản ánh lợi ích và bối cảnh của EU, Mỹ và một số quốc gia phát triển châu Á. Việt Nam có cơ hội và trách nhiệm đóng góp quan điểm quốc gia đang phát triển vào các diễn đàn quốc tế như OECD.AI, ISO/IEC SC 42, UNESCO và ASEAN Working Group on AI Governance để chuẩn mực toàn cầu phản ánh đầy đủ hơn thực tiễn và lợi ích của các quốc gia như Việt Nam. Đây không phải nhiệm vụ thứ yếu mà là một trong những lợi ích chiến lược của hội nhập AI.

4.1.5. Bảo đảm an ninh quốc phòng, trật tự công cộng, lợi ích cộng đồng và bảo vệ quyền con người

Độc lập với ba nguồn trên, tồn tại một nhóm giá trị mà không quyết định pháp luật nào được phép vi phạm, đó là an ninh quốc gia ở mức tồn vong, trật tự công cộng cơ bản, và quyền con người ở nội hàm cốt lõi không thể tước bỏ.

Bảo vệ quyền con người là điều kiện tiên quyết của việc phát triển AI có trách nhiệm. Công nghệ AI chỉ được xã hội chấp nhận và phát triển bền vững khi người dân tin tưởng rằng AI phục vụ, không phải xâm phạm. Quan điểm này phải được phản ánh nhất quán trong từng tầng của khung pháp lý AI, từ tầng hiến định đến tầng tiêu chuẩn kỹ thuật, từ tầng văn bản quy phạm đến tầng bộ quy tắc ứng xử.

AI đang trở thành nhân tố có tính quyết định trong an ninh quốc gia và quốc phòng từ nhận diện mối đe dọa, phân tích tình báo đến hệ thống vũ khí tự chủ. Quan điểm thứ tư đặt ra yêu cầu khung pháp lý AI phải bảo vệ lợi ích an ninh quốc gia theo hai chiều. Một là, chiều phòng thủ: Khung pháp lý phải ngăn chặn AI bị sử dụng như vũ khí chống lại an ninh quốc gia như thao túng thông tin quy mô lớn bằng AI tạo sinh, tấn công hạ tầng quan trọng qua hệ thống AI và gián điệp kỹ thuật số thông qua thu thập dữ liệu AI. Quy định về AI trong hạ tầng thông tin quan trọng quốc gia, kiểm soát AI từ các chủ thể nước ngoài có nguy cơ ảnh hưởng đến an ninh và yêu cầu đánh giá an ninh cho AI agents trong các hệ thống nhạy cảm là những nội dung cần thiết. Hai là, chiều chủ động: Khung pháp lý phải tạo điều kiện cho Việt Nam phát triển và ứng

dụng AI phục vụ an ninh và quốc phòng theo cách có trách nhiệm bao gồm cơ chế phân loại rõ ràng về AI được sử dụng trong bối cảnh quân sự hay an ninh (loại trừ khỏi phạm vi áp dụng của Luật AI 2025 hoặc điều chỉnh bởi quy định chuyên biệt) và bảo vệ bí mật quốc gia trong quá trình kiểm định và công khai thông tin về hệ thống AI nhạy cảm. Điểm cân bằng quan trọng là yêu cầu an ninh quốc gia không được được sử dụng như lý do biện minh vô hạn để hạn chế quyền tự do và bảo mật cá nhân. Mọi giới hạn quyền tự do cá nhân nhân danh an ninh quốc gia phải đáp ứng tiêu chuẩn tỷ lệ, cần thiết và có cơ sở pháp lý rõ ràng.

Tội phạm sử dụng AI (lừa đảo deepfake, lừa đảo voice cloning, phát tán thông tin sai sự thật quy mô lớn) đang đặt ra thách thức trật tự công cộng nghiêm trọng tại Việt Nam. Quan điểm thứ tư đòi hỏi khung pháp lý phải bảo đảm công cụ pháp lý đủ mạnh và đủ cập nhật để phòng chống tội phạm AI. Vì vậy, pháp luật tố tụng cần được điều chỉnh để thừa nhận và quy định giá trị chứng cứ của bằng chứng kỹ thuật số liên quan đến AI, đặc biệt trong bối cảnh deepfake làm cho bằng chứng hình ảnh và âm thanh không còn là bằng chứng tuyệt đối. Thiết chế thực thi pháp luật cần được tăng cường năng lực kỹ thuật về AI để có thể điều tra, truy tố và phòng ngừa tội phạm AI hiệu quả.

Lợi ích cộng đồng trong bối cảnh AI bao gồm việc bảo đảm AI phục vụ lợi ích của toàn xã hội, không chỉ của những người có điều kiện tiếp cận và hưởng lợi từ AI. Quan điểm này đặt ra ba yêu cầu cụ thể: ngăn chặn bất bình đẳng do AI, bảo vệ nhóm dễ tổn thương và bảo đảm tiếp cận bình đẳng với lợi ích AI.

4.2. GIẢI PHÁP XÂY DỰNG KHUNG PHÁP LÝ CHO PHÁT TRIỂN, ỨNG DỤNG TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM

4.2.1. Về phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Giải pháp cho phát triển, ứng dụng AI ở Việt Nam chịu sự chi phối bởi nguyên tắc safety-by-design và privacy-by-design (can thiệp sớm trong vòng đời có chi phí thấp hơn can thiệp muộn) và bởi chính đặc tính bản chất "tính dựa trên dữ liệu" của AI vốn là gốc rễ của bốn vấn đề pháp lý cốt lõi (chất lượng dữ liệu, thiên lệch, quyền riêng tư, khó xác định trách nhiệm). Trong điều kiện Việt Nam xác nhận khoảng trống pháp lý phân bố không đều theo vòng đời mà tập trung ở hai đầu (R&D/thu thập dữ liệu và chấm dứt vận hành chưa được quy định) trong khi giai đoạn giữa đã nhận được sự chú ý pháp lý nhiều hơn qua Nghị định 142 và Quyết định 33.

Nhìn từ góc độ vòng đời AI xác định tại mỗi giai đoạn phát triển và ứng dụng AI, khung pháp lý cần tạo ra điều kiện gì, kiểm soát rủi ro gì và phân định trách nhiệm như thế nào. Tiếp cận theo vòng đời bổ sung cho kiến trúc bảy tầng xác định hệ thống

quy phạm cần xây dựng, vòng đời AI xác định chủ thể và thời điểm mà từng quy phạm phát huy tác dụng.

4.2.1.1. Nghiên cứu và phát triển (R&D)

Giai đoạn R&D là nơi các rủi ro AI được định hình như thiên lệch trong dữ liệu huấn luyện, việc lựa chọn kiến trúc mô hình có hàm ý đạo đức và hướng nghiên cứu có tác động xã hội dài hạn. Can thiệp pháp lý ở giai đoạn này theo nguyên tắc thiết kế an toàn và trách nhiệm vào AI từ giai đoạn R&D thay vì bổ sung sau khi sản phẩm đã hoàn thiện có chi phí thấp hơn và hiệu quả cao hơn can thiệp ở các giai đoạn sau.

Khung pháp lý cần tạo ra ba điều kiện cho R&D AI có trách nhiệm. Điều kiện thứ nhất là hành lang pháp lý an toàn cho nghiên cứu AI nguy cơ cao, cơ chế sandbox nghiên cứu cho phép thử nghiệm công nghệ AI mới trong môi trường kiểm soát với miễn trừ tạm thời khỏi một số yêu cầu thực thi, từ cơ chế thử nghiệm có kiểm soát của Nghị định 142 nhưng mở rộng sang phạm vi nghiên cứu học thuật. Điều kiện này đặc biệt quan trọng cho nghiên cứu về AI tạo sinh, tác nhân AI tự chủ và AI trong lĩnh vực y sinh học.

Điều kiện thứ hai là chuẩn đạo đức nghiên cứu AI, Quyết định 2557/QĐ-BKHCN ngày 25/5/2026 đã tạo nền tảng với Điều 16 về liêm chính khoa học khi sử dụng AI, khung pháp lý cần mở rộng từ đạo đức nghiên cứu sang yêu cầu đánh giá tác động xã hội đối với dự án nghiên cứu AI nguy cơ cao trước khi được phê duyệt tài trợ. Cơ quan đánh giá là Hội đồng Đạo đức AI Nghiên cứu Quốc gia do Viện Hàn lâm KH&CN chủ trì, với thành phần gồm chuyên gia kỹ thuật, luật học và đạo đức học.

Điều kiện thứ ba là khuyến khích R&D AI an toàn và có trách nhiệm thông qua chính sách thuế và đầu tư, ưu đãi thuế thu nhập doanh nghiệp cho chi phí nghiên cứu AI có tích hợp quy trình đánh giá rủi ro và đạo đức; ưu tiên trong xét duyệt đề tài nghiên cứu nhà nước cho dự án AI gắn với giải pháp bảo đảm an toàn và minh bạch và cơ chế SHTT linh hoạt cho kết quả nghiên cứu AI của tổ chức KH&CN công lập để tạo điều kiện thương mại hóa.

4.2.1.2. Thu thập và xử lý dữ liệu

Dữ liệu là nguyên liệu nền tảng của AI từ chất lượng, nguồn gốc và phương thức thu thập dữ liệu quyết định cả hiệu suất kỹ thuật lẫn tính công bằng và hợp pháp của hệ thống AI.

Khung pháp lý cho giai đoạn này cần giải quyết đồng thời ba vấn đề chưa được tháo gỡ rõ ràng trong Luật AI 2025 và Luật BVDLCN 2025. Vấn đề thứ nhất là căn cứ pháp lý cho việc sử dụng dữ liệu đã thu thập hợp pháp vào mục đích huấn luyện AI. Cần có hướng dẫn chính thức xác định nguyên tắc đối với việc đồng ý thu thập dữ liệu phải tường minh bao gồm hoặc loại trừ mục đích huấn luyện AI; dữ liệu ẩn danh hóa

đủ tiêu chuẩn TCVN tương ứng được miễn yêu cầu đồng ý huấn luyện AI và dữ liệu công khai có thể dùng huấn luyện AI nhưng phải ghi nhận nguồn trong hồ sơ kỹ thuật.

Vấn đề thứ hai là yêu cầu về chất lượng và tài liệu hóa bộ dữ liệu huấn luyện tương thích với TCVN ISO/IEC (data quality for AI) khi được ban hành. Cho đến khi TCVN có hiệu lực, hướng dẫn ngành dọc cho từng lĩnh vực ứng dụng AI cần xác định yêu cầu tối thiểu về tài liệu hóa nguồn gốc, quy mô, phân phối và phương pháp làm sạch bộ dữ liệu.

Vấn đề thứ ba là quản lý dữ liệu xuyên biên giới trong huấn luyện AI khi mô hình AI được huấn luyện trên dữ liệu từ nhiều quốc gia hoặc trên nền tảng điện toán đám mây đặt tại nước ngoài. Nguyên tắc áp dụng lãnh thổ mở rộng trong Nghị định 142 cần được hướng dẫn cụ thể, dữ liệu người dùng Việt Nam được dùng để huấn luyện AI ở nước ngoài thuộc phạm vi điều chỉnh của pháp luật Việt Nam về bảo vệ dữ liệu, bất kể nơi xử lý.

Ngoài ra, cần xây dựng các bộ dữ liệu thuộc Danh mục tại Quyết định 804/QĐ-TTg, ưu tiên dữ liệu tiếng dân tộc thiểu số và dữ liệu vùng ít số hóa. Mỗi bộ dữ liệu có hồ sơ mô tả nguồn gốc và căn cứ pháp lý của việc thu thập: sự đồng ý, ngoại lệ tại khoản 5 Điều 7 Luật SHTT hoặc dữ liệu mở. Vận hành môi trường khai thác dữ liệu an toàn theo Điểm c khoản 3 Điều 35 Nghị định 142 cho phép huấn luyện mô hình trên hạ tầng của cơ quan quản lý dữ liệu mà không chuyển quyền sở hữu dữ liệu. Nhờ đó, doanh nghiệp có thể huấn luyện bằng dữ liệu y tế, giáo dục mà không phải chuyển dữ liệu ra nước ngoài theo Điều 20 Luật BVDLCN.

4.2.1.3. Xây dựng và huấn luyện mô hình

Giai đoạn xây dựng và huấn luyện mô hình là trọng tâm của nhóm nghĩa vụ nhà cung cấp trong Luật AI 2025, đặc biệt đối với hệ thống AI rủi ro cao và mô hình AI đa năng. Khung pháp lý ở giai đoạn này cần cân bằng giữa yêu cầu minh bạch kỹ thuật đủ để cho phép giám sát và trách nhiệm giải trình với bảo vệ bí mật thương mại hợp pháp của doanh nghiệp phát triển AI.

Ba yêu cầu pháp lý cốt lõi cần được cụ thể hóa trong hướng dẫn thực hiện Luật AI 2025. Yêu cầu thứ nhất là hồ sơ kỹ thuật tối thiểu bắt buộc mà mọi nhà cung cấp hệ thống AI rủi ro cao phải lập và duy trì: mô tả kiến trúc và nguyên tắc vận hành (không nhất thiết phải công bố trọng số mô hình); nguồn gốc, quy mô và đặc điểm bộ dữ liệu huấn luyện; kết quả đánh giá hiệu năng trên các nhóm dân số khác nhau; rủi ro đã được nhận diện và biện pháp giảm thiểu và điều kiện sử dụng an toàn. Hồ sơ kỹ thuật không được công khai toàn bộ mà được cung cấp cho cơ quan quản lý theo yêu cầu và cho nhà triển khai theo nghĩa vụ hợp đồng.

Yêu cầu thứ hai là kiểm thử thiên lệch bắt buộc trước triển khai cho AI rủi ro cao trong các lĩnh vực có tác động ở phạm vi rộng: thuật toán tín dụng phải kiểm thử

thiên lệch theo giới tính, dân tộc và địa lý; AI tuyển dụng phải kiểm thử thiên lệch theo giới tính và tuổi; AI trong tố tụng phải kiểm thử thiên lệch theo nhóm xã hội.

Yêu cầu thứ ba là quy trình an toàn cho mô hình AI đa năng - loại mô hình đặc biệt nguy hiểm vì tính năng đa dụng làm tăng độ khó khi dự đoán rủi ro trong từng ngữ cảnh triển khai cụ thể. Tham chiếu Bộ nguyên tắc Hiroshima Process, khung pháp lý cần yêu cầu nhà cung cấp mô hình nền tảng công bố báo cáo về khả năng và giới hạn đã biết trước khi mô hình được triển khai ở Việt Nam, bất kể mô hình đó được phát triển ở đâu.

4.2.1.4. Kiểm thử và đánh giá

Giai đoạn kiểm thử và đánh giá là giai đoạn tạo ra bằng chứng về sự tuân thủ và an toàn của hệ thống AI trước khi triển khai. Ba giải pháp ưu tiên cho giai đoạn này bao gồm: Thứ nhất, ban hành TCVN về kiểm định AI (giai đoạn 1 của Lộ trình TCVN AI đề xuất) để tạo phương pháp kiểm định có thể đo lường và so sánh. Thứ hai, phát triển năng lực kiểm định AI tại ba tổ chức kiểm định trong nước theo lộ trình hai giai đoạn: Giai đoạn 1 (2027-2028) đào tạo kỹ năng kiểm định AI cho Viện Công nghệ Thông tin, Viện KH&CN và một tổ chức kiểm định thương mại; Giai đoạn 2 (2029-2030) ít nhất một trong ba tổ chức được công nhận thông qua các cơ quan thành viên của Tổ chức Công nhận toàn cầu (Global ACI). Phát triển thị trường dịch vụ kiểm thử, kiểm toán AI và sản phẩm bảo hiểm trách nhiệm dân sự cho hệ thống AI. Thứ ba, thiết kế quy trình kiểm định phân tầng theo mức rủi ro để tránh áp dụng quy trình đồng nhất cho tất cả. AI rủi ro thấp tự đánh giá qua VN-AI Verify với kiểm tra mẫu 10-15%; AI rủi ro trung bình kiểm định hồ sơ kỹ thuật bởi tổ chức được công nhận; AI rủi ro cao và tác nhân AI tự chủ kiểm định kỹ thuật thực chất với kiểm thử thực địa.

4.2.1.5. Tích hợp và triển khai

Giai đoạn tích hợp và triển khai là nơi nhà cung cấp mô hình nhường trách nhiệm phần lớn cho nhà triển khai. Khung pháp lý cho giai đoạn này cần giải quyết ba vấn đề đặc thù. Vấn đề thứ nhất là thông báo và đồng ý của người dùng: mọi hệ thống AI tương tác với người dùng phải thông báo rõ ràng bản chất AI của hệ thống theo hình thức phù hợp với đặc điểm của nhóm người dùng, thông báo cho người cao tuổi khác với thông báo cho người dùng kỹ thuật; thông báo trong quyết định tín dụng khác với thông báo trong chatbot giải trí. Hướng dẫn ngành dọc cần xác định tiêu chuẩn thông báo cụ thể cho từng lĩnh vực.

Vấn đề thứ hai là quản lý rủi ro trong quá trình tích hợp, khi AI được tích hợp vào hệ thống hiện có (như tích hợp AI vào phần mềm quản lý bệnh viện hoặc hệ thống ngân hàng lõi), rủi ro tích hợp có thể khác và cao hơn rủi ro của AI đứng độc lập. Yêu cầu đánh giá rủi ro tích hợp trước khi triển khai trong môi trường sản xuất phân biệt với kiểm định mô hình AI cần được quy định rõ trong hướng dẫn ngành. Triển khai các dự

án ứng dụng mẫu trong y tế, giáo dục, ngân hàng và dịch vụ công. Thiết kế của các dự án này có sẵn cơ chế giám sát thực chất của con người, ghi nhật ký, kênh tiếp nhận khiếu nại và phương án dự phòng, sau đó công bố mô hình để nhân rộng. Ở khu vực công, lồng yêu cầu về tài liệu kỹ thuật, quyền kiểm tra và điều khoản chấm dứt vào hồ sơ mời thầu của từng dự án, trong khuôn khổ mua sắm đổi mới sáng tạo.

Vấn đề thứ ba là quyền từ chối và quyền được giải thích tại điểm triển khai, người dùng có quyền từ chối nhận dịch vụ từ AI và yêu cầu phục vụ bởi con người trong các tình huống có ảnh hưởng quan trọng đến quyền lợi và có quyền nhận giải thích về cơ sở của quyết định AI ảnh hưởng bất lợi đến mình. Hai quyền này cần được quy định cụ thể về hình thức thực hiện, thời hạn phản hồi và cơ chế khiếu nại.

4.2.1.6. Quản lý và bảo trì

Khung pháp lý cho giai đoạn này cần thiết lập bốn cơ chế vận hành. Cơ chế thứ nhất là giám sát hiệu năng liên tục cho AI rủi ro cao, nhà triển khai phải có hệ thống theo dõi các chỉ số hiệu năng quan trọng theo thời gian thực, với ngưỡng cảnh báo kích hoạt rà soát khi hiệu năng giảm xuống dưới mức đã xác định. Cơ chế này là yêu cầu kỹ thuật bắt buộc được xác định trong TCVN tương ứng.

Cơ chế thứ hai là báo cáo sự cố AI có phân tích chính sách. Nghị định 142 đã quy định nghĩa vụ báo cáo sự cố nghiêm trọng, nhưng cần bổ sung yêu cầu về phân tích nguyên nhân gốc rễ trong 30 ngày và chia sẻ bài học với cộng đồng qua AMIS theo cách ẩn danh hóa thông tin doanh nghiệp. Hệ thống báo cáo sự cố là vòng phản hồi học tập của toàn ngành.

Cơ chế thứ ba là quản lý cập nhật và tái kiểm định khi mô hình AI được cập nhật đáng kể trong quá trình vận hành, tương tự quy tắc vượt phạm vi kiểm định trong phân định chuỗi giá trị nhưng áp dụng cho thay đổi theo thời gian thay vì thay đổi bởi chủ thể khác.

Cơ chế thứ tư là quản lý vòng đời AI và lưu trữ hồ sơ quy định thời hạn lưu trữ nhật ký vận hành AI, hồ sơ quyết định và dữ liệu đào tạo cho mục đích kiểm toán và giải quyết tranh chấp, đặc biệt quan trọng vì hầu hết tranh chấp AI phát sinh sau khi hệ thống đã vận hành một thời gian.

4.2.1.7. Phát triển ứng dụng cuối

Giai đoạn phát triển ứng dụng AI cho người dùng cuối là giai đoạn mà chuỗi giá trị AI đến gần nhất với người chịu tác động cuối cùng. Khung pháp lý cần giải quyết ba vấn đề đặc thù: thứ nhất, bảo vệ nhóm người dùng dễ tổn thương, ứng dụng AI hướng đến người cao tuổi, trẻ em hoặc người có khuyết tật phải tuân thủ yêu cầu bổ sung về thiết kế giao diện và cơ chế bảo vệ. Cụ thể, ứng dụng AI cho người cao tuổi không được thiết kế cơ chế gây nghiện hoặc khai thác tâm lý; ứng dụng AI cho trẻ em phải có đánh giá tác động trước triển khai và không thu thập dữ liệu hành vi

không cần thiết; và ứng dụng AI hỗ trợ người khuyết tật phải được thiết kế theo tiêu chuẩn dễ tiếp cận.

Thứ hai, đào tạo đội ngũ vận hành ở bên triển khai về giám sát và can thiệp đối với hệ thống AI, để khâu giám sát của con người có thực chất. Ví dụ là cán bộ tín dụng ở ngân hàng, bác sĩ dùng công cụ hỗ trợ chẩn đoán, công chức xử lý hồ sơ. Triển khai chương trình nâng cao nhận thức cho người dân về nhận diện nội dung giả mạo do AI tạo ra và lừa đảo trực tuyến, ưu tiên người cao tuổi và người dân tộc thiểu số.

Thứ ba, trách nhiệm xã hội của nhà phát triển ứng dụng AI được thể chế hóa qua cơ chế báo cáo tác động xã hội hàng năm cho ứng dụng AI có trên một triệu người dùng tại Việt Nam. Báo cáo tác động là tài liệu có cấu trúc chuẩn xác định bởi hướng dẫn của Cục Công nghệ số và AI thuộc Bộ KHCN, bao gồm dữ liệu về khiếu nại của người dùng, sự cố đã xảy ra, biện pháp khắc phục và kết quả kiểm toán thiên lệch. Báo cáo này được nộp lên AMIS và một phần được công khai tóm tắt tạo ra áp lực minh bạch từ đầu thay vì chỉ dựa vào quy trình hành chính. Hỗ trợ hợp tác kỹ thuật giữa ngân hàng, doanh nghiệp viễn thông và nền tảng để phát hiện cuộc gọi, video giả mạo trong giao dịch.

4.2.1.8. Chấm dứt vận hành

Giai đoạn chấm dứt vận hành đánh dấu điểm kết thúc của vòng đời AI nhưng không đồng nghĩa với việc chấm dứt các rủi ro pháp lý và kỹ thuật liên đới. Việc xác lập cơ chế pháp lý cho hoạt động chấm dứt vận hành AI dựa trên ba trụ cột cốt lõi: quản trị rủi ro theo chu kỳ sống theo ISO/IEC 22989:2022 (TCVN 13902:2023), trách nhiệm giải trình hậu vận hành, lý thuyết vòng đời dữ liệu và an toàn thông tin [192]. Kinh nghiệm quản trị từ EU AI Act đã chuẩn hóa khái niệm "rút khỏi thị trường" (withdrawal) và "thu hồi" (recall) như những nghĩa vụ pháp lý bắt buộc đối với nhà cung cấp khi hệ thống AI rủi ro cao không còn đáp ứng các tiêu chuẩn tuân thủ. Đồng thời, NIST AI RMF đã xác lập chức năng "Retirement" (Ngừng vận hành) như một phần bắt buộc, yêu cầu lập kế hoạch chuyển giao an toàn và thông báo minh bạch cho các bên liên quan. Các mô hình này chứng minh rằng chấm dứt vận hành phải được quy phạm hóa thành một quy trình kỹ thuật - hành chính mang tính bắt buộc, không thể để doanh nghiệp tự quyết định theo chính sách nội bộ.

Thứ nhất, thể chế hóa căn cứ, thẩm quyền và quy trình chấm dứt vận hành hệ thống AI. Cần bổ sung vào các văn bản hướng dẫn thi hành Luật AI các quy định chi tiết phân định rõ hai hình thức chấm dứt vận hành: chấm dứt vận hành bắt buộc (thẩm quyền hành chính) và chấm dứt vận hành tự nguyện (chủ thể chủ động).

Thứ hai, ban hành các Tiêu chuẩn quốc gia (TCVN) hướng dẫn kỹ thuật chấm dứt vận hành đối với việc xử lý dữ liệu, vô hiệu hoá mô hình và cơ chế bảo

lưu nhật ký hệ thống nhằm ngăn chặn nguy cơ "rác dữ liệu" và rò rỉ thông tin sau khi dừng hệ thống.

Thứ ba, xác lập nguyên tắc trách nhiệm pháp lý kéo dài để bảo vệ quyền và lợi ích hợp pháp của bên thứ ba bị thiệt hại, pháp luật cần khẳng định rõ ràng nguyên tắc: Việc chấm dứt vận hành hệ thống AI không làm miễn trừ trách nhiệm dân sự, hành chính của nhà cung cấp hoặc bên triển khai đối với các thiệt hại phát sinh từ các hành vi, quyết định sai trái mà AI đã thực hiện trong thời gian còn hoạt động. Trường hợp doanh nghiệp sở hữu AI giải thể hoặc phá sản đồng thời với việc chấm dứt vận hành hệ thống, cần thể chế hóa cơ chế trích lập Quỹ dự phòng rủi ro AI hoặc nghĩa vụ tham gia Bảo hiểm trách nhiệm sản phẩm AI bắt buộc để bảo đảm nguồn tài chính bồi thường cho nạn nhân.

Thứ tư, bắt buộc xây dựng Phương án dự phòng trong khu vực công và dịch vụ thiết yếu, pháp luật cần quy định điều kiện tiên quyết để cấp phép vận hành là phải có Phương án chuyển giao duy trì hoạt động bằng con người hoặc hệ thống truyền thống. Với các hệ thống phục vụ dịch vụ thiết yếu thuộc Danh mục rủi ro cao, một ngày vận hành là chu kỳ nghiệp vụ cơ bản. Gián đoạn quá một ngày thường gây thiệt hại trực tiếp cho người dân, khi giao dịch không được xử lý hay hồ sơ không được giải quyết. Mốc 24 giờ cũng ngắn hơn mốc 72 giờ báo cáo sơ bộ tại Nghị định 142, theo quan điểm rằng dịch vụ cho người dân phải được khôi phục trước khi thủ tục báo cáo hoàn tất. Khi có quyết định chấm dứt vận hành hệ thống AI (do sự cố kỹ thuật hoặc mệnh lệnh hành chính), phương án dự phòng phải được kích hoạt ngay lập tức trong thời gian không quá 24 giờ để đảm bảo tính liên tục của dịch vụ công và trật tự an toàn xã hội, tránh tình trạng đứt gãy quản lý.

4.2.2. Về các tầng của khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

4.2.2.1. Tầng thứ nhất

Tầng nền tảng hiến định của khung pháp lý AI Việt Nam với Hiến pháp 2013 tương đối vững chắc về hình thức. Vấn đề là thiếu cơ chế kích hoạt các nguyên tắc hiến định trong thực tiễn điều chỉnh AI, do đó các giải pháp cần được thực hiện đồng thời:

Thứ nhất, Ủy ban Thường vụ Quốc hội xem xét ban hành Nghị quyết giải thích riêng đối với Khoản 2 Điều 14 Hiến pháp 2013, làm rõ phương pháp luận kiểm tra tính cần thiết của biện pháp pháp luật hạn chế quyền con người, quyền công dân khi biện pháp đó được thực hiện thông qua công nghệ AI (như hệ thống nhận diện khuôn mặt tại nơi công cộng, hệ thống chấm điểm rủi ro tự động trong lĩnh vực tư pháp hình sự, hệ thống giám sát hành vi trên không gian mạng). Nội dung giải thích đề xuất tập trung xác lập ba bậc kiểm tra (mục tiêu hợp pháp - tính phù hợp - tính cần thiết tối thiểu) như phương pháp luận áp dụng thống nhất, được trình bày như phương pháp giải thích một

điều khoản cụ thể đã xác định. Đây là kiến nghị dài hạn có điều kiện, chỉ đặt ra khi đã có bằng chứng về cách hiểu khác nhau trong áp dụng và quy trình giải thích Hiến pháp đã được quy định đầy đủ, vì đến nay chưa có tiền lệ áp dụng thẩm quyền này.

Thứ hai, thể chế hóa kiểm tra tính hợp hiến trong xây dựng văn bản pháp lý AI. Bộ Tư pháp trong chức năng thẩm định văn bản quy phạm pháp luật cần bổ sung tiêu chí kiểm tra tính hợp hiến về quyền con người trong môi trường AI. Đây là cơ chế lọc từ sớm tránh ban hành những quy định sau này bị tòa án bác bỏ hoặc áp dụng không nhất quán.

Thứ ba, nội luật hóa Công ước Hà Nội theo kiến trúc tổng thể. Sau khi Quốc hội phê chuẩn Công ước Hà Nội, cần xây dựng lộ trình nội luật hóa theo phương pháp kiến trúc tổng thể: Bộ Tư pháp thực hiện rà soát tuân thủ toàn diện xác định ba nhóm điều khoản (đã được pháp luật trong nước đáp ứng; cần bổ sung vào VBQPPL hiện có và cần văn bản mới). Ưu tiên nội luật hóa ba nghĩa vụ trực tiếp liên quan đến AI: hình sự hóa tội phạm sử dụng AI tổng hợp; chuẩn chứng cứ điện tử AI trong tố tụng và cơ quan đầu mối hợp tác quốc tế điều tra tội phạm AI xuyên biên giới.

Khi khung pháp lý trong nước đã ổn định, Chính phủ có thể đánh giá khả năng tham gia các văn kiện quốc tế về AI có giá trị ràng buộc, khi những văn kiện này mở cho các quốc gia ngoài khu vực ký kết ban đầu. Đây là bước cuối của lộ trình, vì cam kết quốc tế chỉ có ý nghĩa khi các tầng bên dưới đã có khả năng thực hiện.

4.2.2.2. Tầng thứ hai

Đây là tầng đã có nền tảng tốt nhất nhưng vẫn còn một số khoảng trống quan trọng cần lấp đầy.

Thứ nhất, bổ sung vào Nghị định sửa đổi Nghị định 142/2026/NĐ-CP định nghĩa tác nhân AI, dẫn chiếu TCVN 13902:2023, cùng quy định về điểm kiểm soát phê duyệt và giám sát thực chất đối với tác nhân thuộc nhóm rủi ro cao. Thang năm mức tự động hóa hành động được công bố dưới dạng hướng dẫn kỹ thuật và sau đó là tiêu chuẩn quốc gia, phản ánh các mức tự động hóa của TCVN 13902. Hệ quả pháp lý được xác định theo tổ hợp giữa mức tự động hóa và loại hành động. Giám sát của con người áp dụng từ mức 2. Kiểm soát chặt hơn áp dụng từ mức 4 đối với hành động có hệ quả pháp lý, hoặc ở mọi mức trong lĩnh vực thuộc Danh mục. Mức 5 chỉ được phép trong môi trường thử nghiệm có kiểm soát khi hành động có hệ quả pháp lý.

Thứ hai, quy định chuỗi trách nhiệm pháp lý theo mô hình nhà cung cấp-nhà triển khai có điều chỉnh cho Việt Nam. Nghị định hướng dẫn cần phân định rõ ba nhóm nghĩa vụ theo vị trí trong chuỗi giá trị. Nhà cung cấp mô hình nền tảng có nghĩa vụ công bố hồ sơ mô tả năng lực, giới hạn đã biết, dữ liệu huấn luyện (không phải bí mật thương mại), rủi ro đã được kiểm tra và điều khoản sử dụng có trách nhiệm. Nhà tích

hợp và triển khai có nghĩa vụ kiểm thử trong ngữ cảnh triển khai cụ thể, cấu hình an toàn và thông báo người dùng về bản chất AI theo hình thức phù hợp. Người dùng chuyên nghiệp có nghĩa vụ sử dụng đúng mục đích đã được kiểm định và báo cáo sự cố. Ngoài ra, hai quy tắc cần bổ sung vào Nghị định sửa đổi: (i) Vượt phạm vi kiểm định trở thành nhà cung cấp: Khi bên triển khai sửa đổi hệ thống AI vượt ra ngoài phạm vi sử dụng được kiểm định ban đầu, bên đó phải chịu nghĩa vụ nhà cung cấp đối với phần sửa đổi. (ii) Trách nhiệm bổ sung khi nhà cung cấp nước ngoài vắng mặt: Khi nhà cung cấp mô hình nền tảng không có hiện diện pháp lý tại Việt Nam và không thực hiện nghĩa vụ nhà cung cấp, bên triển khai Việt Nam phải thực hiện nghĩa vụ kiểm định và tài liệu hóa bổ sung tương ứng.

Về cơ chế bồi thường, cần nghiên cứu khả năng bổ sung quy định về quỹ bồi thường AI theo mô hình các ngành rủi ro cao như hàng không và dược phẩm để bảo đảm nạn nhân có thể nhận bồi thường ngay cả khi không thể chứng minh lỗi cụ thể của chủ thể nào trong chuỗi. Bổ sung quy định nghĩa vụ thông báo trước, xử lý dữ liệu sau chấm dứt, thời hạn lưu trữ hồ sơ, và quy tắc kế thừa nghĩa vụ khi tổ chức vận hành chấm dứt tồn tại.

Thứ ba, bổ sung nguyên tắc áp dụng lãnh thổ mở rộng đối với AI xuyên biên giới. Cần bổ sung vào nghị định hướng dẫn nguyên tắc mọi hệ thống AI cung cấp dịch vụ cho người dùng tại Việt Nam đều thuộc phạm vi điều chỉnh của Luật AI 2025, bất kể nhà cung cấp đặt trụ sở ở đâu. Kết hợp với cơ chế thực thi thực tiễn: (i) yêu cầu nhà cung cấp AI nước ngoài chỉ định đại diện pháp lý tại Việt Nam khi đạt ngưỡng người dùng xác định; (ii) cơ chế hợp tác thực thi với cơ quan quản lý AI các nước đối tác; (iii) biện pháp kỹ thuật (địa chỉ IP, giao thức mạng) hỗ trợ thực thi khi nhà cung cấp không tuân thủ.

Thứ tư, kiến nghị hoàn thiện văn bản dưới luật, cụ thể là sửa đổi, bổ sung hướng dẫn thi hành Luật BVDLCN 2025 và Nghị định 142 (đối với nội dung quyền riêng tư tại Điều 21 Hiến pháp 2013 liên quan đến AI tạo sinh) và kiến nghị bổ sung định nghĩa "phân biệt đối xử gián tiếp qua thuật toán" vào Luật AI 2025 sửa đổi (đối với nội dung quyền bình đẳng tại Điều 16 Hiến pháp 2013).

Thứ năm, cập nhật Bộ luật Hình sự về tội phạm sử dụng AI cân nhắc theo hai hướng. Một là, bổ sung tình tiết tăng nặng đặc thù liên quan đến AI vào các tội danh hiện có: sử dụng AI tạo sinh để giả mạo giọng nói, khuôn mặt hoặc nội dung của người khác; sử dụng AI để tự động hóa hành vi lừa đảo hoặc tấn công quy mô lớn không thể thực hiện bằng con người. Tình tiết này phản ánh thực tiễn rằng AI cho phép một người phạm tội ảnh hưởng đến hàng nghìn nạn nhân đồng thời, mức độ nguy hiểm xã hội cao hơn rõ rệt so với hành vi cùng loại không dùng AI. Hai là, tội danh mới về vi

phạm quy trình kiểm định AI. Khi nhà cung cấp hoặc nhà triển khai cố ý bỏ qua hoặc làm giả kết quả kiểm định AI rủi ro cao gây ra hậu quả nghiêm trọng cho người dùng là hành vi có cấu thành riêng biệt không trùng với tội danh gian lận thông thường.

4.2.2.3. Tầng thứ ba

Thứ nhất, ban hành lộ trình TCVN về AI giai đoạn 2027-2032 với danh sách ưu tiên và thời hạn xác định. Bộ KH&CN (qua Ủy ban Tiêu chuẩn Đo lường Chất lượng Quốc gia) cần công bố lộ trình TCVN AI chia thành ba giai đoạn.

Giai đoạn 1 (2027-2028) Năm TCVN cấp bách trực tiếp hỗ trợ thực thi Luật AI 2025 và Nghị định 142: (1) TCVN về kiểm định và đánh giá hệ thống AI (tham chiếu ISO/IEC TS 42119-2:2025) là điều kiện tiên quyết để yêu cầu kiểm định AI rủi ro cao trong Luật AI 2025 có thể thực thi thực chất. (2) TCVN về tác nhân AI và khả năng kiểm soát cần thiết để thực thi định nghĩa tác nhân AI đề xuất bổ sung và bốn yêu cầu kỹ thuật tối thiểu. (3) TCVN về gắn nhãn và nhận diện nội dung AI tổng hợp (tham chiếu C2PA Technical Specification và ISO/IEC đang trong giai đoạn xây dựng) là công cụ kỹ thuật để thực thi nghĩa vụ gắn nhãn bắt buộc tại Điều 18 Nghị định 142. (4) TCVN về đánh giá tác động hệ thống cần thiết để thực hiện yêu cầu đánh giá tác động trước khi cơ quan nhà nước triển khai AI trong dịch vụ công theo Nghị định 142. (5) TCVN về xóa dữ liệu an toàn và định dạng lưu trữ nhật ký vận hành phục vụ giải quyết tranh chấp sau khi chấm dứt vận hành hệ thống AI.

Giai đoạn 2 (2029-2030) Bốn TCVN mở rộng nền tảng điều chỉnh chất lượng và công bằng: (1) TCVN về đánh giá công bằng và kiểm soát thiên lệch thuật toán (tham chiếu ISO/IEC TS 12791:2024). (2) TCVN về an toàn hệ thống AI trong hạ tầng quan trọng quốc gia. (3) TCVN về minh bạch và giải thích được của hệ thống AI (tham chiếu ISO/IEC 12792:2025). (4) TCVN về quản lý dữ liệu huấn luyện AI và tài liệu hóa bộ dữ liệu (tham chiếu ISO/IEC 8183:2023).

Giai đoạn 3 (2031-2032) Hệ thống TCVN chuyên ngành cho ba lĩnh vực ưu tiên: TCVN AI trong tài chính - ngân hàng; TCVN AI trong y tế và thiết bị y tế; TCVN AI trong dịch vụ công và quản lý hành chính.

Thứ hai, thiết lập Ban kỹ thuật thường trực chuyên về AI trong Uỷ ban TCĐLCLQG. Ban kỹ thuật AI (Technical Committee on AI - TC/AI) được thành lập với thành phần bao gồm kỹ sư AI từ doanh nghiệp, chuyên gia pháp lý, học giả và đại diện tổ chức người tiêu dùng. Nhiệm vụ theo dõi phát triển ISO/IEC về AI và đề xuất nội luật hóa kịp thời; xây dựng TCVN đặc thù Việt Nam cho vấn đề không được tiêu chuẩn quốc tế bao phủ đủ; ban hành tiêu chuẩn tạm thời trong vòng 90 ngày khi có nhu cầu thực tiễn cấp bách, tránh khoảng trống kỹ thuật kéo dài trong khi TCVN chính thức đang được xây dựng.

Thứ ba, thiết lập cơ chế đàm phán thỏa thuận công nhận lẫn nhau về kết quả kiểm định AI với IMDA Singapore trước tiên, Hàn Quốc và trong tương lai với EU (khi Việt Nam đã xây dựng đủ năng lực). MRA giúp doanh nghiệp Việt Nam tiếp cận thị trường quốc tế mà không cần kiểm định lại và tạo ra áp lực nâng cao chất lượng TCVN nội địa để đạt chuẩn công nhận lẫn nhau.

Có hai chuyên gia CG-CS-01 và CG-DN-02 đề xuất giải pháp đạo đức sinh thái và tác động môi trường: Mặc dù Điều 4 Luật AI có khuyến khích phát triển "AI xanh", nhưng mối liên hệ giữa AI và phát triển bền vững vẫn là một khoảng trống chính sách. Các hệ thống siêu máy tính huấn luyện AI tiêu thụ năng lượng khổng lồ. Chúng ta cần bổ sung các tiêu chí đánh giá cụ thể về mức độ phát thải carbon, rác thải điện tử và mức tiêu thụ tài nguyên của các trung tâm dữ liệu AI [Xem Phụ lục 5].

4.2.2.4. Tầng thứ tư

Thứ nhất, thể chế hóa giá trị pháp lý của hướng dẫn ngành trong VBQPPL. Từ góc độ luật học, thách thức của tầng luật mềm là địa vị pháp lý của hướng dẫn. Nếu hướng dẫn ngành không có giá trị pháp lý ràng buộc, tuân thủ sẽ không bảo đảm; nếu hướng dẫn ngành được ban hành như VBQPPL đầy đủ, mất đi tính linh hoạt và tốc độ cập nhật. Giải pháp là thiết kế cơ chế pháp lý trung gian cho hướng dẫn ngành. Nghị định 142 sửa đổi bổ sung điều khoản: "Bộ trưởng chuyên ngành ban hành hướng dẫn áp dụng Luật AI 2025 và Nghị định này trong lĩnh vực thuộc phạm vi quản lý nhà nước. Tổ chức, doanh nghiệp không tuân thủ hướng dẫn phải giải thích bằng văn bản cách thức đạt được mục tiêu bảo vệ tương đương với cơ quan quản lý nhà nước có thẩm quyền khi được yêu cầu". Điều khoản này thể chế hóa cơ chế tuân thủ hoặc giải thích trong VBQPPL cấp nghị định tạo ra giá trị pháp lý ràng buộc thủ tục cho hướng dẫn ngành mà không biến hướng dẫn thành VBQPPL đầy đủ. Hậu quả pháp lý của việc không tuân thủ và không có giải thích hợp lý là yếu tố bất lợi trong xử lý vi phạm. Những hướng dẫn đã ổn định qua một chu kỳ áp dụng được chuyển thành tiêu chuẩn quốc gia theo lộ trình của tầng 3. Hướng dẫn và tiêu chuẩn vì vậy là hai nấc của cùng một lộ trình: hướng dẫn đi trước để thử nghiệm yêu cầu, tiêu chuẩn đi sau để ổn định yêu cầu.

Thứ hai, ban hành hướng dẫn ngành dọc theo cấu trúc ba lớp cho các lĩnh vực ưu tiên. Học từ mô hình ba lớp của Singapore (nguyên tắc → kỳ vọng cụ thể → ví dụ thực hành) kết hợp bổ sung yêu cầu quy trình chuyển tiếp khi chấm dứt dịch vụ AI liên tục, các bộ ngành cần ưu tiên ban hành hướng dẫn AI theo thứ tự:

Bộ Y tế ban hành hướng dẫn AI trong y tế và thiết bị y tế (trong 18 tháng): Lớp nguyên tắc: AI trong y tế phải đặt an toàn bệnh nhân lên trên hết; bác sĩ là người ra quyết định cuối cùng. Lớp kỳ vọng cụ thể: AI chẩn đoán hình ảnh phải được xác nhận lâm sàng trước triển khai; phần mềm AI y tế phải có hướng dẫn sử dụng rõ ràng về giới

hạn và điều kiện áp dụng. Lớp ví dụ thực hành: trường hợp điển hình từ bệnh viện Bạch Mai, Chợ Rẫy đã triển khai AI chẩn đoán.

Hướng dẫn về AI trong khu vực công gồm 2 phần, mỗi phần thuộc thẩm quyền của một bộ. Phần thuộc Bộ KH&CN với tư cách cơ quan đầu mối, tiếp tục chủ trì hướng dẫn chung về sử dụng AI trong cơ quan nhà nước, gồm phương pháp và mẫu đánh giá tác động theo Điều 20 Nghị định 142; tiêu chí xác định hệ thống có liên quan đến quyền con người theo khoản 7 Điều 20; yêu cầu kỹ thuật về giám sát của con người; việc công khai báo cáo trên Cổng thông tin một cửa. Hướng dẫn này được ban hành dưới hình thức thông tư của Bộ trưởng Bộ KH&CN, hoặc nội dung này được bổ sung vào nghị định sửa đổi Nghị định 142. Phần thuộc Bộ Nội vụ ban hành thông tư hoặc văn bản hướng dẫn chuyên môn trong phạm vi công vụ và cải cách hành chính: trách nhiệm công vụ của người có thẩm quyền khi sử dụng kết quả của hệ thống AI làm căn cứ ra quyết định hành chính, bao gồm yêu cầu ghi nhận lý do khi chấp nhận hoặc bác bỏ kết quả của hệ thống; chuẩn năng lực và nội dung bồi dưỡng về sử dụng AI cho cán bộ, công chức, viên chức; việc đưa tiêu chí về sử dụng AI an toàn, minh bạch vào bộ chỉ số cải cách hành chính và vào phương pháp đo lường sự hài lòng của người dân; việc rà soát vị trí việc làm khi một phần nhiệm vụ được AI hỗ trợ.

Thứ ba, thiết lập cơ chế tuân thủ hoặc giải thích chính thức trong các hướng dẫn ngành: đối với mỗi kỳ vọng trong hướng dẫn, doanh nghiệp có thể (i) tuân thủ trực tiếp; hoặc (ii) áp dụng giải pháp tương đương và giải thích công khai tại sao giải pháp của họ đạt được mục tiêu tương tự. Cơ quan quản lý có 30 ngày để phản hồi và xác nhận hoặc từ chối giải pháp thay thế. Cơ chế này biến luật mềm từ "khuyến nghị" thành "nghĩa vụ mềm có áp lực giải trình" phù hợp với giai đoạn xây dựng văn hóa tuân thủ AI tại Việt Nam.

4.2.2.5. Tầng thứ năm

Thứ nhất, xây dựng VN-AI Verify theo mô hình đồng thiết kế. Lưu ý VN-AI Verify là công cụ hỗ trợ tự đánh giá và chuẩn hóa hồ sơ; việc sử dụng công cụ không mặc nhiên thay thế nghĩa vụ pháp lý, tạo chứng nhận sự phù hợp, chuyển gánh nặng chứng minh hoặc loại trừ trách nhiệm nếu chưa có căn cứ luật định và cơ chế đánh giá độc lập. VN-AI Verify là bộ công cụ tự đánh giá AI tiếng Việt cho doanh nghiệp, có bằng chứng thực nghiệm trực tiếp từ quá trình phát triển Khung kiểm thử xác minh AI của Singapore (AI Verify Framework) quy trình gần nhất về mặt thiết kế với đề xuất của luận án. Theo đó, Cơ quan Phát triển Truyền thông Thông tin (IMDA) phát triển AI Verify, một khung kiểm thử quản trị AI và bộ công cụ phần mềm từ năm 2021 với sự tham vấn của các công ty thuộc nhiều quy mô và lĩnh vực khác nhau, bao gồm AWS, DBS, Google, Meta, Microsoft, Singapore Airlines, NCS/LTA, Standard

Chartered, UCARE.AI và X0PA. Tháng 5/2022, AI Verify đã được ra mắt dưới dạng Sản phẩm Khả thi Tối thiểu (MVP) để thử nghiệm quốc tế và nhận phản hồi từ IMDA và PDPC. Vào ngày 7/6/2023, AI Verify đã được công khai mã nguồn trên GitHub với sự tham gia của hơn 50 công ty, bao gồm Dell, Hitachi và IBM. Đến năm 2025, khung kiểm thử AI Verify được cập nhật nâng cấp để giải quyết các rủi ro do AI thế hệ mới (Gen AI) gây ra. Với bản cập nhật này, các công ty hiện có thể áp dụng khung kiểm thử AI Verify cho cả các trường hợp sử dụng AI truyền thống và Gen AI [230].

So sánh điều kiện thực tiễn cho thấy Việt Nam có các yếu tố tương đồng đủ để chuyển giao. VINASA và VAIA đóng vai trò kết nối doanh nghiệp AI. Từ 2023, NIC (Trung tâm Đổi mới Sáng tạo Quốc gia Việt Nam) đã ký Biên bản hợp tác về đổi mới sáng tạo với Trung tâm Công nghệ và Sản xuất tiên tiến (ARTC) thuộc Cơ quan Khoa học, Công nghệ và Nghiên cứu Singpore (A*STAR), tạo kênh trao đổi kinh nghiệm trực tiếp [42]. Vì AI Verify Toolkit đã là mã nguồn mở, Việt Nam có thể điều chỉnh thay vì xây dựng từ đầu. Điểm khác biệt cần lưu ý là hệ sinh thái startup AI Việt Nam nhỏ hơn Singapore, đòi hỏi thiết kế module đơn giản hóa hơn cho doanh nghiệp siêu nhỏ. Đề xuất VN-AI Verify gồm ba module độc lập, mỗi module phản ánh điều khoản cụ thể trong Luật AI 2025 và Nghị định 142 [Xem Phụ lục 2]:

Module A Phân loại rủi ro: Chuỗi câu hỏi phân nhánh theo Điều 9 Luật AI 2025 và Điều 6 Nghị định 142, kết thúc bằng phân loại có căn cứ (rủi ro cao/trung bình/thấp) kèm danh sách nghĩa vụ tương ứng. Module này thay thế trực tiếp nghĩa vụ "tự phân loại" của nhà cung cấp và được AMIS ghi nhận là hoàn thành thủ tục thông báo.

Module B Kiểm tra tuân thủ nghĩa vụ: Danh sách kiểm tra (checklist) cấu trúc theo từng điều khoản gồm watermarking (Điều 18), báo cáo sự cố (Điều 22), hồ sơ kỹ thuật (Điều 14), nghĩa vụ minh bạch (Điều 16). Mỗi câu hỏi kèm hướng dẫn kỹ thuật và ví dụ minh họa bằng tiếng Việt, điều chỉnh cho ba ngữ cảnh đặc thù: AI trong tiếng Việt (NLP/LLM), AI tín dụng tiêu dùng và AI hỗ trợ quyết định y tế.

Module C Đánh giá rủi ro thiên lệch: Bộ câu hỏi kiểm thử thiên lệch thuật toán theo nhóm nhân khẩu học phù hợp với dân số Việt Nam (giới tính, dân tộc, địa lý đô thị - nông thôn), tham chiếu ISO/IEC về phương pháp đánh giá thiên lệch AI. Module này là tùy chọn cho AI rủi ro thấp, bắt buộc cho AI rủi ro cao trong tín dụng, tuyển dụng và dịch vụ công.

Lộ trình thực hiện đối với Việt Nam là Bộ KH&CN phối hợp VINASA và VAIA thành lập Nhóm phát triển VN-AI Verify (30+ doanh nghiệp gồm cả doanh nghiệp lớn, SME, startup); kiểm thử thực tiễn trong 12 tháng với nhóm doanh nghiệp tình nguyện; ra mắt phiên bản 1.0 dưới dạng mã nguồn mở miễn phí; cập nhật hàng năm dựa trên phản hồi sử dụng thực tế. VN-AI Verify phải kiểm thử cả khả năng xử lý

tiếng Việt (phát hiện thiên lệch ngôn ngữ, xử lý đặc thù văn hóa) và phải phản ánh yêu cầu từng nguyên tắc trong OECD AI Principles và Luật AI 2025 cụ thể.

Thứ hai, hỗ trợ phát triển bộ quy tắc ứng xử AI theo ngành với kiểm toán độc lập: VINASA, VAIA và các hiệp hội ngành được hỗ trợ xây dựng bộ quy tắc ứng xử AI ngành theo khung của Bộ KH&CN. Điều kiện để bộ quy tắc được công nhận chính thức: xây dựng qua quy trình đa chủ thể bao gồm đại diện người dùng cuối; có tiêu chí tuân thủ cụ thể và có thể kiểm tra; có cơ chế kiểm toán độc lập định kỳ bởi bên thứ ba được công nhận. Doanh nghiệp tham gia và tuân thủ bộ quy tắc được công nhận có thể sử dụng chứng nhận này như bằng chứng tuân thủ trong quy trình xét duyệt sandbox và trong tranh chấp pháp lý. Bộ quy tắc được công bố trên Cổng thông tin một cửa về AI để Bộ KH&CN tiếp nhận và theo dõi. Sau khi có đánh giá về mức độ tuân thủ thực tế, việc ghi nhận bộ quy tắc ứng xử và hệ quả pháp lý được quy định trong lần sửa đổi tiếp theo của Nghị định 142.

4.2.2.6. Tầng thứ sáu

Tầng thứ sáu cần đi trước các tranh chấp, vì đây là tầng duy nhất mà sản phẩm phụ thuộc vào việc có vụ việc thực tế.

Thứ nhất, TAND Tối cao chủ động rà soát, lựa chọn và công bố án lệ từ các bản án cụ thể đã được xét xử liên quan đến AI theo đúng cơ chế án lệ hiện hành (Nghị quyết 04/2019/NQ-HĐTP và Nghị quyết số 02/2025/NQ-HĐTP). Đây là cơ chế phát triển pháp luật gắn với vụ việc cụ thể đã được xét xử, khác về bản chất với việc ban hành trước một danh mục nguyên tắc trừu tượng chưa qua kiểm nghiệm thực tiễn xét xử. Ưu điểm về thẩm quyền của cách tiếp cận này là án lệ chỉ có giá trị ràng buộc trong phạm vi tình huống pháp lý tương tự với vụ việc đã xét xử, không mang tính khái quát hóa vượt quá những gì đã thực sự được kiểm nghiệm qua tố tụng, phù hợp đúng bản chất của thẩm quyền tư pháp (áp dụng và giải thích pháp luật trong từng vụ việc cụ thể).

TAND Tối cao cần xây dựng chương trình chủ động phát triển án lệ AI theo ba bước. Bước 1, Tổng hợp và phân loại: Rà soát toàn bộ bản án và quyết định có liên quan đến AI trong hệ thống tòa án; phân loại theo bốn nhóm tranh chấp AI ưu tiên (trách nhiệm dân sự tác nhân AI; quyền tác giả nội dung AI; lừa đảo sử dụng AI tổng hợp; kế thừa nghĩa vụ khi tổ chức vận hành chấm dứt tồn tại); xác định vụ việc có tiềm năng phát triển thành án lệ. Bước 2, Đào tạo thẩm phán: Chương trình đào tạo 40 giờ về kỹ thuật AI cơ bản cho thẩm phán tòa án các cấp, tập trung vào các vụ việc AI phức tạp; cơ chế tư vấn kỹ thuật độc lập cho thẩm phán khi xét xử vụ việc AI đòi hỏi hiểu biết kỹ thuật sâu. Bước 3, Công bố án lệ: TAND Tối cao ưu tiên xem xét công bố án lệ về bốn nhóm tranh chấp AI ưu tiên trong giai đoạn 2027-2028 tạo ra tầng điều chỉnh quan trọng lấp đầy khoảng trống trong giai đoạn khung văn bản chưa hoàn thiện.

Thứ hai, TAND Tối cao có thẩm quyền hướng dẫn áp dụng pháp luật, làm rõ cách áp dụng thống nhất quy phạm đã có sẵn vào các loại tình huống cụ thể theo Điều 30 Luật Tổ chức TAND. Đề xuất Hội đồng Thẩm phán ban hành Nghị quyết hướng dẫn tòa án các cấp về trình tự và phương pháp xác định khi giải quyết tranh chấp liên quan đến AI gây thiệt hại, cụ thể gồm bốn nội dung: (i) hướng dẫn tòa án thực hiện đúng trình tự Điều 6 Bộ luật Dân sự 2015 trước khi viện dẫn lẽ công bằng, phải xem xét khả năng áp dụng tương tự Điều 601 (nguồn nguy hiểm cao độ) đối với hệ thống AI có đặc điểm tương đồng về mức độ rủi ro khó kiểm soát, chỉ khi không thể áp dụng tương tự mới chuyển sang lẽ công bằng; (ii) hướng dẫn về việc vận dụng các quy định hiện hành của Bộ luật Tố tụng dân sự 2015 (Điều 91, 92) về nghĩa vụ chứng minh và các trường hợp đã được pháp luật thừa nhận cho phép tòa án yêu cầu đương sự cung cấp chứng cứ thuộc phạm vi kiểm soát của mình, áp dụng cho tình huống chứng cứ kỹ thuật về hệ thống AI (như nhật ký vận hành, hồ sơ kỹ thuật) chỉ do bên vận hành AI nắm giữ, trong phạm vi mà pháp luật tố tụng hiện hành đã cho phép, không mở rộng vượt quá; (iii) hướng dẫn về việc thu thập ý kiến chuyên gia kỹ thuật độc lập trong quá trình tố tụng theo cơ chế giám định tư pháp hoặc người có kiến thức chuyên môn đã được Bộ luật Tố tụng dân sự 2015 quy định, để hỗ trợ thẩm phán đánh giá chứng cứ kỹ thuật phức tạp; (iv) hướng dẫn suy đoán trách nhiệm theo lợi ích: Chủ thể hưởng lợi kinh tế chủ yếu từ hệ thống AI bị suy đoán có trách nhiệm bồi thường thiệt hại gây ra bởi hệ thống đó, đặt gánh nặng chứng minh không có lỗi lên phía có lợi thế thông tin hơn (nhà cung cấp/nhà triển khai) thay vì phía yếu thế hơn (nạn nhân).

Thứ ba, thiết kế quy tắc chứng cứ AI trong tố tụng. Bộ luật Tố tụng dân sự và Bộ luật Tố tụng hình sự chưa có quy tắc đặc thù cho chứng cứ AI. Dữ liệu từ hệ thống AI, kết quả phân tích AI và nội dung AI tạo sinh đang được xử lý theo quy tắc chứng cứ thông thường không phù hợp. Cần nghiên cứu và đề xuất bổ sung vào hai bộ luật tố tụng: (i) Điều kiện chấp nhận chứng cứ AI bao gồm tiêu chuẩn về tính toàn vẹn dữ liệu, khả năng kiểm chứng và không bị giả mạo; (ii) Phân bổ gánh nặng chứng minh khi chứng cứ AI bị phản đối: bên sử dụng chứng cứ AI chứng minh tính xác thực hay bên phản đối chứng minh sự giả mạo; (iii) Giá trị chứng minh của kết quả phân tích AI là chứng cứ bổ trợ cần được chuyên gia xác nhận trong vụ việc có quy mô lớn.

4.2.2.7. Tầng thứ bảy

Thứ nhất, thiết lập cơ chế điều phối AI xuyên ngành có thẩm quyền ràng buộc điều chỉnh theo một trong hai phương án sau:

Phương án 1 Kiện toàn Ủy ban Quốc gia về Chuyển đổi số: Bổ sung chức năng điều phối AI chuyên biệt vào Ủy ban Quốc gia về Chuyển đổi số đang có, thành lập Tiểu ban AI trực thuộc với Tổ thư ký kỹ thuật chuyên trách, ngân sách độc lập và thẩm

quyền phán quyết xung đột thẩm quyền AI. Đây là phương án ít tốn kém về thể chế nhất vì tận dụng cơ cấu đã có.

Phương án 2 Quyết định Thủ tướng thành lập Ban Chỉ đạo quốc gia về AI: Thủ tướng ban hành Quyết định thành lập Ban Chỉ đạo quốc gia về AI với Phó Thủ tướng làm Trưởng ban, tương tự mô hình Ban Chỉ đạo về Đề án 06 (Cơ sở dữ liệu quốc gia về dân cư). Đây là mô hình đã có tiền lệ trong nước, không tạo ra cơ quan mới nhưng tạo ra cơ chế điều phối có thực quyền.

Giữa hai phương án, Phương án 2 phù hợp hơn về mặt pháp lý vì tạo ra cơ chế điều phối độc lập với các Bộ chuyên ngành, tránh tình trạng Bộ KH&CN vừa là đầu mối vừa là trọng tài trong xung đột với Bộ khác.

Thứ hai, xây dựng năng lực công nghệ-pháp lý trong đơn vị đầu mối hiện có: Củng cố Cục Công nghệ số và AI trong Bộ KH&CN với chỉ tiêu biên chế có kỹ sư AI được giao chính thức; chương trình đào tạo công nghệ-pháp lý bắt buộc cho cán bộ đơn vị này; ngân sách hoạt động độc lập cho chức năng quản lý AI; và thẩm quyền ban hành văn bản hướng dẫn kỹ thuật.

Bên cạnh đó, cần thể chế hóa vị trí chuyên trách AI. Bộ Nội vụ ban hành vị trí chức danh Chuyên viên Pháp lý Công nghệ AI trong danh mục vị trí việc làm, với tiêu chí tuyển dụng và khung năng lực xác định, áp dụng cho Bộ KH&CN và các Bộ chuyên ngành có chức năng quản lý AI. Đây là giải pháp tạo ra năng lực công nghệ-pháp lý có hệ thống mà không đòi hỏi thành lập đơn vị mới.

Nghị định 142 sửa đổi cần bổ sung chương về phân định thẩm quyền với quy tắc rõ ràng: *(i)* Nguyên tắc thẩm quyền chính: Ban Chỉ đạo quốc gia về AI là cơ quan có thẩm quyền chính đối với hệ thống AI rủi ro cao đa ngành; các Bộ chuyên ngành có thẩm quyền đối với AI trong lĩnh vực mình quản lý; *(ii)* Quy tắc xung đột thẩm quyền: Khi cùng một hệ thống AI thuộc phạm vi quản lý của nhiều Bộ, nguyên tắc thẩm quyền ưu tiên theo ngành có rủi ro cao hơn áp dụng; *(iii)* Cơ chế giải quyết xung đột: Ban Chỉ đạo quốc gia về AI có thẩm quyền phán quyết xung đột thẩm quyền giữa các Bộ trong thời hạn 30 ngày kể từ ngày phát sinh xung đột.

Thứ ba, xây dựng lộ trình kiểm định AI ba giai đoạn:

Giai đoạn 1 (2027-2028) VN-AI Verify và kiểm định hồ sơ: Doanh nghiệp sử dụng VN-AI Verify để tự chứng nhận; cơ quan quản lý kiểm tra mẫu 10-15% doanh nghiệp mỗi năm; tổ chức kiểm định bên thứ ba được công nhận theo tiêu chí năng lực và độc lập.

Giai đoạn 2 (2029-2030) Kiểm định kỹ thuật thực chất: Ít nhất ba tổ chức kiểm định AI độc lập đủ năng lực được công nhận; phối hợp kỹ thuật với IMDA Singapore, BSI Anh và TÜV SÜD Đức để chuyển giao phương pháp kiểm định.

Giai đoạn 3 (2031-2032) Công nhận lẫn nhau: Đàm phán MRA với ASEAN, Hàn Quốc và Singapore; doanh nghiệp Việt Nam có thể sử dụng kết quả kiểm định trong nước để tiếp cận thị trường đối tác mà không cần kiểm định lại.

Thứ tư, thiết lập hệ thống báo cáo sự cố AI bắt buộc có phân tích chính sách: Cục Công nghệ số và AI thuộc Bộ KHCN cần bổ sung tổng hợp và phân tích dữ liệu sự cố định kỳ hàng quý; phát hành Báo cáo sự cố AI ẩn danh hóa thông tin doanh nghiệp; cơ chế phản hồi chính sách khi xu hướng sự cố chỉ ra khoảng trống pháp lý, Bộ KHCN đề xuất cập nhật với Ban Chỉ đạo quốc gia về AI trong 90 ngày. Bổ sung yêu cầu kỹ thuật cho Cơ sở dữ liệu quốc gia về AI về cơ chế cập nhật trạng thái "đã chấm dứt vận hành" với thời hạn xác định kể từ ngày chấm dứt.

Nghị định 142 đã thiết lập Cổng thông tin điện tử một cửa về AI và Cơ sở dữ liệu quốc gia về hệ thống AI với ý nghĩa là giao diện hành chính tiếp nhận thông báo, cấp mã định danh, báo cáo định kỳ. Bộ KH&CN cần bảo đảm Cổng thông tin một cửa về AI vận hành đầy đủ các chức năng mà Luật AI đã quy định trước ngày 01/3/2027. Kiến nghị Cục Công nghệ số và AI của Bộ KHCN vận hành Hệ thống thông tin quản lý AI (AMIS) là tầng phân tích và phản hồi chính sách đặt phía sau Cổng, biến dữ liệu hành chính thành thông tin ra quyết định chính sách [Xem Phụ lục 4]. Hai hệ thống bổ trợ nhau và cần được thiết kế để kết nối liền mạch với nhau. AMIS đồng thời là hệ thống đầu mối kỹ thuật 24/7 kết nối với mạng lưới Công ước Hà Nội khi được phê chuẩn, tiếp nhận và chuyển tiếp yêu cầu bảo toàn chứng cứ điện tử AI trong điều tra tội phạm xuyên biên giới.

Nghị định 142 sửa đổi cần bổ sung điều khoản để tránh tranh chấp pháp lý về sau: *"Kết quả phân tích, cảnh báo tự động của Hệ thống thông tin quản lý AI có giá trị tham khảo, hỗ trợ ra quyết định quản lý nhà nước, không có giá trị làm căn cứ pháp lý trực tiếp để áp dụng biện pháp xử lý đối với tổ chức, doanh nghiệp mà không qua quy trình xem xét, xác minh theo quy định của pháp luật về xử lý vi phạm hành chính hoặc pháp luật liên quan".* Điều khoản này bảo đảm tính chất công cụ hỗ trợ, không thay thế quyết định của con người của AMIS nhất quán với nguyên tắc kiểm soát có ý nghĩa của con người đã xác lập xuyên suốt luận án đối với chính các hệ thống AI được điều chỉnh, áp dụng ngược lại cho công cụ AI (nếu AMIS sử dụng kỹ thuật AI trong phân tích) do Nhà nước vận hành.

Thứ năm, thiết kế cơ chế thẩm quyền lãnh thổ mở rộng và hợp tác tư pháp AI. Điều 2 Luật AI 2025 về phạm vi áp dụng cần được hướng dẫn chi tiết về thẩm quyền lãnh thổ mở rộng. Hệ thống AI phát triển và vận hành ở nước ngoài nhưng tác động đến người dùng Việt Nam thuộc thẩm quyền điều chỉnh của pháp luật Việt Nam về nghĩa vụ bảo vệ người dùng và trách nhiệm bồi thường. Cơ chế thực thi thẩm quyền lãnh thổ mở rộng: (i) Yêu cầu chỉ định đại diện pháp lý tại Việt Nam cho nhà cung cấp

AI nước ngoài vượt ngưỡng số người dùng; (ii) Cơ chế tống đạt tư pháp AI xuyên biên giới trong khuôn khổ Công ước Hà Nội khi được phê chuẩn; (iii) Biện pháp cưỡng chế gián tiếp đối với nhà cung cấp nước ngoài không tuân thủ bao gồm yêu cầu app store và nền tảng phân phối gỡ bỏ sản phẩm vi phạm.

4.2.3. Về các giải pháp cụ thể cho việc xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

4.2.3.1. Giải pháp về chủ thể xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Thứ nhất, xây dựng năng lực công nghệ-pháp lý trong hệ thống nhà nước:

Chương trình đào tạo công nghệ-pháp lý AI do Bộ KH&CN phối hợp Bộ Tư pháp và cơ sở đào tạo (trường Đại học Luật - Đại học Quốc gia Hà Nội, Học viện Tư pháp, trường Đại học Luật Hà Nội) thiết kế và vận hành chương trình 6 tháng cho cán bộ xây dựng pháp luật AI. Nội dung kỹ thuật AI ở mức đủ để hiểu hành vi và giới hạn hệ thống AI (không cần thành kỹ sư); pháp luật so sánh về AI; phân tích tác động pháp lý và kỹ thuật soạn thảo quy phạm công nghệ. Mục tiêu định lượng đến năm 2028, mỗi Bộ quản lý AI trong lĩnh vực mình có ít nhất năm cán bộ có chứng chỉ công nghệ-pháp lý AI; đến năm 2030, có 100 chuyên gia công nghệ-pháp lý trên toàn quốc.

Thiết lập cơ chế biệt phái kỹ sư AI và chuyên gia kỹ thuật từ khu vực tư nhân sang làm việc trong cơ quan nhà nước trong thời hạn 1-2 năm, với doanh nghiệp tiếp tục trả lương; cơ quan nhà nước cung cấp môi trường làm việc và cơ hội đóng góp thực chất vào xây dựng chính sách; cơ chế quản lý xung đột lợi ích rõ ràng. Cơ chế này cho phép cơ quan quản lý AI có ngay nguồn nhân lực kỹ thuật AI chất lượng cao mà không cần chờ chu kỳ đào tạo dài hạn.

Hợp tác quốc tế xây dựng năng lực thông qua ký kết thỏa thuận kỹ thuật với OECD AI Policy Observatory (phương pháp phân tích chính sách AI), NIST Hoa Kỳ (AI Risk Management Framework) và IMDA Singapore (phát triển VN-AI Verify).

Thứ hai, tăng cường vai trò lập pháp chuyên sâu của Quốc hội: Đề xuất thành lập Tiểu ban AI trong Ủy ban Khoa học, Công nghệ và Môi trường của Quốc hội học từ mô hình Tiểu ban Quốc hội đặc biệt về AI của Hàn Quốc. Chức năng của Tiểu ban AI là tham gia thiết kế kiến trúc các dự án luật liên quan đến AI quan trọng ngay từ giai đoạn chẩn đoán (không chỉ thẩm tra văn bản đã soạn thảo xong); tổ chức phiên điều trần kỹ thuật thường niên với chuyên gia AI trong và ngoài nước; giám sát thực thi Luật AI 2025 dựa trên báo cáo thường niên từ Chính phủ.

Thứ ba, thể chế hóa Hội đồng tư vấn AI quốc gia bao gồm đại diện doanh nghiệp AI (cân bằng giữa doanh nghiệp lớn, SME và startup), chuyên gia học thuật, đại diện người dùng và tổ chức xã hội. Hội đồng có quyền đề xuất chính sách chính thức và nhận phản hồi bắt buộc từ Ban Chỉ đạo quốc gia về AI trong 90 ngày. Họp ít nhất

mỗi quý với Ban Chỉ đạo quốc gia về AI; kết quả thảo luận được công bố công khai. Cơ chế đặc biệt đảm bảo đại diện của cộng đồng startup và SME là những đối tượng bị ảnh hưởng nhiều nhất bởi gánh nặng tuân thủ nhưng thường thiếu tiếng nói nhất.

Thứ tư, tăng cường đại diện quốc tế chủ động. Việt Nam cần tham gia chủ động trong các diễn đàn quốc tế về chuẩn mực AI như đăng ký tham gia ISO/IEC JTC 1/SC 42 Ủy ban kỹ thuật về AI; tham gia tích cực ASEAN Working Group on AI Governance với vai trò đề xuất nghị trình; đóng góp nghiên cứu về quản trị AI đối với các nước đang phát triển trong OECD AI Policy Observatory.

Thứ năm, thể chế hóa thẩm định kỹ thuật bắt buộc trong quy trình lập pháp AI. Sửa đổi Nghị định hướng dẫn Luật Ban hành VBQPPL 2025 để bổ sung yêu cầu: dự thảo VBQPPL AI có yêu cầu kỹ thuật phải kèm theo Báo cáo thẩm định kỹ thuật được thực hiện bởi đơn vị độc lập với nhóm soạn thảo, xác nhận mỗi yêu cầu kỹ thuật có thể thực hiện bằng công nghệ hiện tại không; chi phí thực hiện ước tính theo ba tầng quy mô doanh nghiệp; có giải pháp kỹ thuật thay thế nào đạt cùng mục tiêu bảo vệ với chi phí thấp hơn không. Báo cáo thẩm định kỹ thuật được công bố công khai cùng dự thảo văn bản cho phép cộng đồng kỹ thuật phản biện.

Thứ sáu, bổ sung vào nhiệm vụ của đơn vị đầu mối (Cục Công nghệ số và AI thuộc Bộ KH&CN) giám sát trong toàn bộ vòng đời, bao gồm cả giai đoạn hậu chấm dứt, tránh tình trạng thẩm quyền mặc định chỉ giới hạn ở giai đoạn hệ thống đang vận hành.

4.2.3.2. Giải pháp về nội dung xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Nội dung 1: Kiến trúc phân tầng của khung pháp lý

Thứ nhất, thể chế hóa Chương trình Nghiên cứu Chính sách AI Quốc gia. Bộ KH&CN thành lập và cấp ngân sách ổn định cho Chương trình Nghiên cứu Chính sách AI Quốc gia, giao cho các cơ sở nghiên cứu độc lập thực hiện theo hợp đồng. Nhiệm vụ trọng tâm của Chương trình gồm: lập hồ sơ chẩn đoán theo vòng đời hệ thống; tổ chức đánh giá sau ban hành ngay sau khi thời hạn chuyển tiếp kết thúc; nghiên cứu so sánh phục vụ lần sửa đổi Luật AI. Ưu tiên nghiên cứu: thực trạng rủi ro AI tại Việt Nam với dữ liệu định lượng (hàng năm); tác động AI đến thị trường lao động theo ngành (hai năm một lần); tác động phân phối của thuật toán AI theo nhóm xã hội (ba năm một lần) và chi phí tuân thủ của hệ sinh thái doanh nghiệp AI Việt Nam (khi ban hành văn bản quan trọng). Kết quả nghiên cứu phải được công bố công khai và sử dụng bắt buộc trong hồ sơ dự án luật.

Thứ hai, xây dựng và duy trì Tài liệu kiến trúc pháp lý AI quốc gia. Cục Công nghệ số và AI thuộc Bộ KHCN chịu trách nhiệm xây dựng và cập nhật Tài liệu kiến trúc pháp lý AI quốc gia - văn bản giải thích công khai sơ đồ quan hệ giữa tất cả văn

bản pháp lý AI (luật, nghị định, thông tư, tiêu chuẩn, hướng dẫn); logic ủy quyền và phân tầng; khoảng trống còn lại và lộ trình bổ sung với thời hạn xác định. Tài liệu được Ban Chỉ đạo quốc gia về AI phê duyệt, công bố công khai và cập nhật khi có thay đổi đáng kể trong khung pháp lý. Đây là công cụ bảo đảm tính nhất quán nội tại của khung pháp lý khi được xây dựng dần theo thời gian bởi nhiều chủ thể khác nhau.

Thứ ba, thể chế hóa phân tích khoảng trống bảy tầng hàng năm. Cục Công nghệ số và AI thuộc Bộ KHCN thực hiện phân tích khoảng trống bảy tầng hàng năm, đánh giá mức độ hoàn thành của từng tầng trong khung pháp lý AI so với một năm trước và so với mục tiêu lộ trình. Kết quả phân tích là đầu vào trực tiếp cho lập kế hoạch xây dựng pháp luật AI hàng năm của Bộ KH&CN và cho chương trình làm việc của Ban Chỉ đạo quốc gia về AI. Kết quả phân tích phải được công bố công khai tạo áp lực xã hội thúc đẩy lấp đầy khoảng trống.

Thứ tư, thể chế hóa quy trình rà soát tuân thủ điều ước quốc tế như một bước bắt buộc trong soạn thảo văn bản pháp lý AI. Cụ thể, trước khi ban hành bất kỳ văn bản AI quan trọng nào, cơ quan soạn thảo phải thực hiện rà soát hai chiều: *(i)* văn bản dự thảo có xung đột với điều ước quốc tế Việt Nam đã tham gia không (CPTPP, EVFTA, RCEP, Công ước quyền con người); *(ii)* có nghĩa vụ từ điều ước nào chưa được phản ánh trong dự thảo không. Quy trình này do Bộ Tư pháp (Vụ Pháp luật quốc tế) thực hiện phối hợp với Bộ Ngoại giao, kết quả được đưa vào hồ sơ thẩm định và công bố cùng văn bản.

Nội dung 2: Nguyên tắc và phương pháp điều chỉnh

Bên cạnh hành vi bị cấm theo mục đích, bổ sung nhóm hệ thống bị cấm theo bản chất, ví dụ hệ thống chấm điểm xã hội diện rộng đối với cá nhân mà trước đây dự thảo Luật đã từng xếp vào tầng rủi ro không chấp nhận được.

Trả nguyên tắc kiểm soát của con người về đúng chức năng. Ngoại lệ chỉ áp dụng khi việc giám sát kiểm chứng được: người giám sát có đủ thông tin, có thẩm quyền ghi đè, có nhật ký can thiệp, như Nghị định 142 đã đặt ra cho hệ thống đang chuyển tiếp. Mở rộng van an toàn sang trường hợp thiệt hại đến quyền con người.

Nội dung 3: Sandbox

Bảo vệ người tham gia bằng cách bổ sung yêu cầu người tham gia ở cấp độ 2 và 3 phải đồng ý bằng văn bản trên cơ sở được thông tin đầy đủ; bổ sung quyền rút khỏi thử nghiệm. Nêu rõ quy định miễn, loại trừ trách nhiệm không loại trừ nghĩa vụ bồi thường cho người tham gia bị thiệt hại.

Đưa kết quả thử nghiệm vào cơ chế rà soát. Cơ quan chấp thuận có nghĩa vụ định kỳ tổng hợp kết quả thử nghiệm gửi Bộ KH&CN. Bộ công bố báo cáo tổng hợp, đã ẩn danh, trên Cổng thông tin một cửa. Ban hành bộ tiêu chí thẩm định thống nhất giữa các cơ quan chấp thuận và công khai danh sách giấy xác nhận đã cấp.

Nội dung 4: Hệ thống quyền và nghĩa vụ của các chủ thể

Người bị ảnh hưởng có quyền được giải thích quyết định cá nhân và quyền yêu cầu con người xem xét lại quyết định của hệ thống rủi ro cao, theo hướng nâng mô hình của Thông tư 53 thành quy định chung.

Nghĩa vụ thông tin của nhà cung cấp mô hình khi tích hợp vào hệ thống rủi ro cao được phân tầng: nhẹ với mô hình mã nguồn mở không kèm hỗ trợ thương mại, đầy đủ với mô hình thương mại. Nhà cung cấp mô hình hoặc dịch vụ AI nước ngoài có quy mô đáng kể phải có đầu mối liên hệ tại Việt Nam.

Nội dung 5: Cơ chế bảo đảm tuân thủ và khắc phục hậu quả

(i) Nghị định xử phạt chuyên biệt bao gồm hành vi vi phạm: không thông báo phân loại hoặc phân loại sai, không đánh dấu hoặc gắn nhãn, không đánh giá sự phù hợp, không báo cáo sự cố. Mức phạt tương xứng với mức rủi ro. Nguyên tắc áp dụng khi một hành vi đồng thời vi phạm Nghị định 330, để không xử phạt hai lần.

(ii) Khi không có thỏa thuận, bên triển khai vẫn được hoàn trả theo mức độ đóng góp vào thiệt hại của từng bên. Thỏa thuận loại trừ hoàn toàn quyền hoàn trả không có hiệu lực đối với doanh nghiệp nhỏ và vừa và cơ quan nhà nước.

(iii) Khôi phục nghĩa vụ cung cấp chứng cứ hồ sơ, nhật ký theo yêu cầu của Tòa án và Trọng tài như dự thảo.

(iv) Bổ sung kênh tiếp nhận phản ánh của người bị ảnh hưởng trên Cổng thông tin một cửa.

Nội dung 6: cơ chế rà soát và cập nhật

Công khai báo cáo đánh giá thi hành Luật AI định kỳ, gợi ý 03 năm cho đồng bộ với chu kỳ Chiến lược. Chu kỳ rà soát Danh mục cố định, gợi ý hằng năm, kèm danh mục căn cứ dữ liệu bắt buộc. Bắt buộc dùng báo cáo sự cố, tổng hợp kết quả thử nghiệm, kết luận thanh tra và khiếu nại làm căn cứ của hồ sơ rà soát.

4.2.3.3. Giải pháp về quy trình xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Thứ nhất, chuẩn hóa quy trình bốn giai đoạn cho văn bản AI quan trọng. Mọi văn bản pháp lý AI quan trọng (luật, nghị định, tiêu chuẩn kỹ thuật) phải trải qua bốn giai đoạn với sản phẩm cụ thể: Giai đoạn 1 (Chẩn đoán) Báo cáo chẩn đoán với bằng chứng định lượng về vấn đề chính sách; Giai đoạn 2 (Thiết kế kiến trúc) Tài liệu kiến trúc giải thích lựa chọn mô hình và phân tầng văn bản; Giai đoạn 3 (Soạn thảo và tham vấn) Dự thảo văn bản kèm Báo cáo RIA và Báo cáo phản hồi ý kiến tham vấn; Giai đoạn 4 (Ban hành và chuẩn bị thực thi) Văn bản ban hành kèm lộ trình áp dụng, hướng dẫn thực hiện và VN-AI Verify (nếu áp dụng).

Thứ hai, nâng cao chất lượng tham vấn bằng bốn biện pháp bổ sung. Bên cạnh quy trình tham vấn bắt buộc theo Luật Ban hành VBQPPL 2025, cần bổ sung: tổ chức

ít nhất hai phiên tham vấn kỹ thuật chuyên sâu với kỹ sư AI về tính khả thi của từng yêu cầu pháp lý dự kiến trước khi đăng tải dự thảo công khai; tổ chức phiên điều trần riêng với cộng đồng startup và SME về gánh nặng tuân thủ và đề xuất cơ chế miễn trừ tỷ lệ; bắt buộc ban hành Báo cáo phản hồi ý kiến tham vấn giải thích lý do tiếp nhận hoặc không tiếp nhận từng nhóm ý kiến quan trọng; đối với văn bản AI quan trọng, tổ chức một vòng tham vấn quốc tế với cơ quan quản lý AI của Hàn Quốc, Singapore và EU để phát hiện sớm rủi ro xung đột chuẩn mực. Khi được hỏi về nội dung cần sửa đổi trong quá trình xây dựng khung pháp lý AI, chuyên gia CG-CS-02 và CG-NC-03 đồng tình lựa chọn giải pháp quan trọng nhất chính là tăng thời gian nghiên cứu góp ý, phản hồi dự thảo luật, nghị định đồng thời mở rộng đối tượng được tiếp cận với hồ sơ dự thảo [Xem Phụ lục 5].

Thứ ba, vận hành cơ chế phản ứng nhanh bằng nghị quyết Chính phủ. Luật Ban hành VBQPPL 2025 đã trao cho Chính phủ quyền ban hành nghị quyết để giải quyết vấn đề cấp bách. Ban Chỉ đạo quốc gia về AI được giao thẩm quyền đề xuất Chính phủ ban hành nghị quyết về Biện pháp tạm thời khi xuất hiện rủi ro AI khẩn cấp với thời hạn hiệu lực 6 tháng, đủ thời gian để xây dựng văn bản QPPL chính thức qua quy trình đầy đủ. Quy trình là Ban Chỉ đạo quốc gia về AI họp khẩn trong 48 giờ đề xuất Biện pháp tạm thời trình Thủ tướng phê duyệt, Chính phủ ban hành nghị quyết trong tối đa 30 ngày từ khi nhận diện rủi ro. Cơ chế này tránh tình trạng khoảng trống pháp lý kéo dài như đã xảy ra với lừa đảo bằng deepfake giai đoạn 2023-2025.

Thứ tư, thể chế hóa đánh giá hậu kỳ bắt buộc. Cục Công nghệ số và AI thuộc Bộ KHCN ban hành quy định bắt buộc thực hiện đánh giá hậu kỳ ba năm sau khi văn bản AI quan trọng có hiệu lực, bao gồm đánh giá mức độ tuân thủ thực tế (không chỉ đánh giá hình thức); phân tích dữ liệu sự cố được báo cáo; khảo sát đối tượng điều chỉnh về tính khả thi và gánh nặng tuân thủ; và so sánh với mục tiêu chính sách đã đặt ra trong RIA. Kết quả đánh giá hậu kỳ phải được công bố công khai và trở thành đầu vào bắt buộc cho lần sửa đổi văn bản tiếp theo.

Các giải pháp trên chỉ phát huy tác dụng khi bốn giai đoạn được nối với nhau bằng những điểm chuyển tiếp rõ ràng. Mỗi điểm chuyển tiếp gắn với một sản phẩm phải có trước khi được bước sang giai đoạn sau: hồ sơ chẩn đoán để chuyển sang thiết kế; bản thiết kế kiến trúc được phê duyệt để chuyển sang soạn thảo; gói dự thảo kèm bảng theo dõi thay đổi, bảng đối chiếu liên chế độ và báo cáo đánh giá mức độ sẵn sàng để chuyển sang thông qua và ban hành; báo cáo đánh giá sau ban hành để khởi động vòng tiếp theo. Các sản phẩm này được công bố trên Cổng thông tin một cửa về AI hoặc cổng thông tin của cơ quan chủ trì. Như vậy, việc giám sát quy trình có thể dựa vào văn bản, không phải dựa vào suy đoán.

Áp dụng vào thực tế trước mắt, giai đoạn từ nay đến khi kết thúc thời hạn chuyển tiếp nên tập trung vào ba việc: xác lập trạng thái gốc, thiết lập vòng phản hồi, lập hồ sơ chẩn đoán theo vòng đời cho lần sửa đổi Luật. Bản thiết kế kiến trúc cho lần sửa đổi đó nên được hoàn thành sau đợt đánh giá năm 2027 và sau lần cập nhật Chiến lược, trước khi giao soạn thảo dự án Luật sửa đổi. Theo trình tự này, lần sửa đổi tới sẽ là vòng đầu tiên của khung pháp lý về AI đi qua đủ bốn giai đoạn theo đúng thứ tự chức năng của chúng.

4.2.3.4. Giải pháp về kỹ thuật xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam

Kỹ thuật 1, nâng cấp AI-RIA lên ba chiều đặc thù

Hạn chế cốt lõi của AI-RIA trong quá trình soạn thảo Luật AI 2025 là thiếu ba chiều phân tích đặc thù của lĩnh vực AI mà chuẩn mực OECD Better Regulation Framework đòi hỏi. Từ nhận diện đó, giải pháp cần bổ sung ba chiều còn thiếu vào hướng dẫn thực hiện.

Chiều thứ nhất, phân tích kịch bản công nghệ, yêu cầu văn bản dự thảo phải được đánh giá trong ba kịch bản tốc độ phát triển AI khác nhau trong vòng năm năm: chậm với chu kỳ đổi mới 4-5 năm, vừa với 2-3 năm và nhanh với 12-18 tháng. Với mỗi kịch bản, cần xác định điều khoản nào sẽ lỗi thời đầu tiên và cơ chế ủy quyền nào cần được thiết kế từ đầu để cho phép cập nhật không cần sửa luật. Chiều thứ hai, phân tích chi phí tuân thủ theo ba tầng quy mô, yêu cầu ước lượng chi phí thực sự bao gồm chi phí ban đầu, chi phí vận hành hàng năm và chi phí thích ứng khi văn bản cập nhật cho doanh nghiệp lớn trên 500 nhân viên, SME từ 11-499 nhân viên và micro/startup dưới 10 nhân viên hoặc dưới ba năm với kết quả trực tiếp làm căn cứ thiết kế ngưỡng miễn trừ tỷ lệ. Chiều thứ ba, đánh giá năng lực thể chế thực tế, yêu cầu trước khi quy định bất kỳ nghĩa vụ kiểm định hay đánh giá sự phù hợp nào phải xác định được Việt Nam hiện có bao nhiêu tổ chức đủ năng lực thực hiện, nếu thiếu thì lộ trình xây dựng năng lực cần bao nhiêu thời gian, và trong thời gian đó cơ chế chuyển tiếp nào được thiết kế sẵn.

Để ba chiều bổ sung có căn cứ định lượng, Cục Công nghệ số và AI thuộc Bộ KHCN cần xây dựng cơ sở dữ liệu chi phí tuân thủ AI qua khảo sát 500 doanh nghiệp trong năm 2027, tạo ra căn cứ thực địa cho ngưỡng miễn trừ tỷ lệ thay vì đặt ngưỡng dựa trên phán đoán hành chính. Đây chính là lý do Kỹ thuật 1 là điều kiện tiên quyết cho Kỹ thuật 5, không có dữ liệu chi phí tuân thủ từ RIA đầy đủ, ngưỡng miễn trừ tỷ lệ sẽ không có cơ sở thực tiễn vững chắc.

Kỹ thuật 2, hoàn thiện định nghĩa chức năng và thiết lập cơ chế diễn giải

Kỹ thuật định nghĩa chức năng đã được vận dụng tốt nhất trong số sáu kỹ thuật, song hai khoảng trống còn lại cần được khắc phục để bảo đảm tính nhất quán của toàn

bộ kiến trúc. Khoảng trống thứ nhất là chưa mở rộng kỹ thuật này sang tác nhân AI - thực thể đang bùng nổ trong thực tiễn nhưng chưa có định nghĩa pháp lý đặc thù trong hệ thống văn bản hiện hành.

Trong kỳ sửa đổi Nghị định 142 hoặc nghị định bổ sung, cần bổ sung định nghĩa chức năng theo hướng: Tác nhân AI là hệ thống AI có khả năng tự chủ nhận thức môi trường, lập kế hoạch và thực thi chuỗi hành động để đạt mục tiêu xác định mà không cần phê duyệt của người vận hành cho từng hành động đơn lẻ trong chuỗi. Định nghĩa này nhất quán với cách tiếp cận chức năng đã sử dụng cho "hệ thống AI" trong Luật AI và tương thích với định nghĩa agentic AI mà OECD AI Principles 2024 xác lập, qua đó bảo đảm tính nhất quán ngôn ngữ pháp lý quốc tế. Kèm theo định nghĩa là thang phân loại mức độ tự chủ với hàm ý pháp lý khác nhau theo từng mức, gắn với yêu cầu giám sát có ý nghĩa của con người khi mức tự chủ tăng cao.

Khoảng trống thứ hai là thiếu cơ chế diễn giải định nghĩa chính thức, vấn đề sẽ ngày càng trở nên cấp bách khi công nghệ AI tiếp tục phát triển và các thế hệ AI mới đặt câu hỏi về phạm vi áp dụng của định nghĩa hiện hành. Ban Chỉ đạo quốc gia về AI nên được giao thẩm quyền ban hành diễn giải chính thức theo quy trình rút gọn: doanh nghiệp hoặc cơ quan quản lý đề nghị trong 15 ngày, tham vấn kỹ thuật, ban hành diễn giải trong 30 ngày và công bố công khai. Diễn giải này có giá trị tham chiếu pháp lý cho đến khi có sửa đổi chính thức qua VBQPPL, tránh tình trạng khoảng trống diễn giải kéo dài từng làm các nhà cung cấp Trung Quốc sử dụng mô hình khuếch tán phải chờ CAC phát hành thông báo giải thích bổ sung trong nhiều tháng.

Để bảo đảm tính nhất quán của kỹ thuật này trong tương lai, cần xây dựng bộ công cụ kỹ thuật lập pháp riêng cho AI. Kỹ thuật lập pháp thông thường theo Luật Ban hành VBQPPL 2025 không đủ để xử lý bốn đặc điểm phân biệt của AI so với đối tượng điều chỉnh truyền thống: tốc độ thay đổi công nghệ vượt chu kỳ lập pháp; đặc điểm "hộp đen" khiến yêu cầu pháp lý khó kiểm tra kỹ thuật; chuỗi chủ thể phức tạp không có mô hình tương tự trong pháp luật hiện hành; và rủi ro tích lũy theo thời gian do hệ thống AI tiếp tục học không ngừng nghỉ. Bộ công cụ kỹ thuật lập pháp AI là tập hợp phương pháp, quy trình kiểm tra và mẫu soạn thảo giúp cán bộ soạn thảo xử lý bốn đặc điểm trên một cách có hệ thống [Xem Phụ lục 3]. Bộ công cụ được ban hành dưới dạng Hướng dẫn kỹ thuật của Bộ Tư pháp, ban đầu áp dụng khuyến nghị, dần dần nâng lên bắt buộc sau khi đã được kiểm nghiệm qua ít nhất hai kỳ sửa đổi văn bản AI. Đào tạo sử dụng bộ công cụ được tích hợp vào chương trình đào tạo cán bộ pháp chế của Học viện Tư pháp và Vụ Pháp luật các Bộ, là điều kiện năng lực đầu vào cho cán bộ tham gia soạn thảo văn bản AI quan trọng.

Kỹ thuật 3, bổ sung đầy đủ ủy quyền linh hoạt phân tầng

Ủy quyền linh hoạt phân tầng là kỹ thuật có mức vận dụng không đồng đều nhất, đúng hướng ở tầng VBQPPL nhưng gần như thiếu vắng ở tầng tiêu chuẩn kỹ thuật và tầng luật mềm, vì vậy giải pháp cần bù đắp theo ba hướng bổ sung nhau.

Hướng thứ nhất, bổ sung điều khoản rà soát bắt buộc khi Luật AI 2025 được sửa đổi lần đầu. Chính phủ có nghĩa vụ thực hiện đánh giá toàn diện hiệu quả thực thi sau mỗi năm năm và trình báo cáo Quốc hội kèm đề xuất sửa đổi nếu rà soát xác định bất cập cơ cấu, với báo cáo phải dựa trên dữ liệu thực tiễn từ AMIS. Điều khoản này biến rà soát từ tùy ý thành nghĩa vụ pháp lý học trực tiếp từ Điều 112 EU AI Act, vốn là cơ chế thích ứng thể chế hóa quan trọng nhất của khung pháp lý AI châu Âu. Kể từ khi EU AI Act 2024/1689/EU có hiệu lực, Ủy ban châu Âu đã bắt đầu quy trình xây dựng Danh sách tiêu chuẩn hài hòa theo Điều 40 giao CEN/CENELEC và ETSI xây dựng các tiêu chuẩn hài hòa và công bố trên Official Journal of the EU. Kinh nghiệm thực tiễn cho thấy ngay từ khi Danh sách dự kiến được thảo luận công khai (trước khi chính thức công bố), doanh nghiệp AI châu Âu đã bắt đầu điều chỉnh kế hoạch tuân thủ theo hướng đầu tư vào các tiêu chuẩn dự kiến được đưa vào Danh sách. Điều này cho thấy hiệu ứng định hướng thị trường của cơ chế Danh sách tiêu chuẩn hài hòa xuất hiện ngay cả trước khi cơ chế có hiệu lực pháp lý chính thức, một hiệu ứng quan trọng mà Thông tư Danh mục TCVN AI của Việt Nam có thể khai thác.

Hướng thứ hai, thiết kế ủy quyền TCVN AI theo quy trình rút gọn 12 tháng, bổ sung vào Nghị định 142 hoặc ban hành Quyết định Bộ trưởng KH&CN. Bộ KH&CN ban hành và cập nhật hàng năm Danh mục TCVN AI có hiệu lực, tuân thủ TCVN trong Danh mục được suy đoán là tuân thủ yêu cầu kỹ thuật tương ứng tại Nghị định này. Danh mục TCVN cập nhật qua Thông tư Bộ trưởng với chu kỳ 12 tháng không phải sửa Nghị định và sử dụng kỹ thuật viện dẫn không ngày học từ Nhật Bản. Áp dụng rộng rãi trong pháp luật về tiêu chuẩn kỹ thuật, METI (Bộ Kinh tế, Thương mại và Công nghiệp) áp dụng kỹ thuật viện dẫn không ngày thực tế trong hướng dẫn AI (2024), cho phép hướng dẫn tự động cập nhật khi OECD AI Principles được sửa đổi mà không cần ban hành văn bản mới. Đây là bằng chứng thực nghiệm cho tính khả thi của kỹ thuật viện dẫn không ngày trong hệ thống pháp luật thành văn. Ở Việt Nam, ghi "TCVN ISO/IEC 42001 phiên bản hiện hành" thay vì ghi số phiên bản cụ thể, bảo đảm khi Uỷ ban TCĐLCLQG ban hành phiên bản mới, văn bản tự động áp dụng mà không cần sửa đổi. Cơ chế kết hợp Kỹ thuật 3 và Kỹ thuật 6 trong một bước đơn giản, giải pháp này vừa giải quyết ủy quyền linh hoạt vừa đặt nền tảng cho luật hóa tiêu chuẩn.

Hướng thứ ba, thí điểm ủy quyền đồng thiết kế hướng dẫn ngành, khởi đầu với NHNN thành lập Nhóm làm việc AI Ngân hàng gồm đại diện NHNN, 8-10 tổ chức tài chính và chuyên gia kỹ thuật AI độc lập để cùng soạn nội dung hướng dẫn AI tài chính

theo mô hình Singapore. NHNN đặt ra nguyên tắc tối thiểu không thể thỏa hiệp, nhóm làm việc soạn nội dung chi tiết, NHNN phê duyệt trong 30 ngày và nếu không có phản hồi coi như chấp thuận. Cơ chế này chuyển quyền khởi xướng về phía khu vực tư nhân có hiểu biết kỹ thuật sâu hơn, trong khi nhà nước giữ quyền kiểm soát qua phê duyệt. Sau 18 tháng thí điểm, đánh giá hiệu quả và nhân rộng sang y tế và dịch vụ công theo cùng mô hình.

Kỹ thuật 4, hoàn thiện phân định nghĩa vụ theo chuỗi giá trị

Xuất phát từ hai khoảng trống đã phân tích, giải pháp kỹ thuật 4 hướng vào việc hoàn thiện chi tiết vận hành của khung phân định đã có.

Về quy tắc "vượt phạm vi kiểm định trở thành nhà cung cấp", khi sửa đổi Nghị định 142 cần bổ sung nội dung: bên triển khai thực hiện sửa đổi đáng kể hệ thống AI rủi ro cao vượt ra ngoài phạm vi sử dụng được kiểm định hoặc mô tả trong tài liệu kỹ thuật của nhà cung cấp phải thực hiện các nghĩa vụ của nhà cung cấp đối với phần sửa đổi đó. Quy tắc này giải quyết trực tiếp tình huống phổ biến nhất của hệ sinh thái AI Việt Nam là tinh chỉnh mô hình nước ngoài.

Về cơ chế xử lý nhà cung cấp nước ngoài vắng mặt, học từ Điều 28 EU AI Act, cần bổ sung quy định khi nhà cung cấp mô hình nền tảng không có hiện diện pháp lý tại Việt Nam và không thực hiện nghĩa vụ nhà cung cấp theo Nghị định, bên triển khai Việt Nam phải thực hiện bổ sung các nghĩa vụ kiểm định và tài liệu hóa tối thiểu tương ứng. Quy định này không yêu cầu nhà triển khai làm điều không thể về mặt kỹ thuật mà chỉ yêu cầu tài liệu hóa những gì họ có thể biết và kiểm soát về mô hình đang triển khai, gắn với nghĩa vụ công khai thông tin về giới hạn đã biết của mô hình.

Song hành với hai quy tắc trên, Cục Công nghệ số và AI thuộc Bộ KHCN duy trì Danh sách nền tảng AI được xác nhận tuân thủ, các nền tảng trải qua kiểm định độc lập và đăng ký vào Danh sách để bên triển khai SME sử dụng nền tảng trong Danh sách chỉ cần tự đánh giá phần tích hợp và triển khai mà không cần kiểm định toàn bộ hệ thống. Cơ chế này học từ kỹ thuật "chứng nhận nền tảng giảm gánh nặng triển khai" của Hàn Quốc, đồng thời kết nối trực tiếp với giải pháp Kỹ thuật 5 về miễn trừ tỷ lệ. Bên triển khai SME sử dụng nền tảng trong Danh sách và đạt dưới ngưỡng định lượng của Kỹ thuật 5 được hưởng chế độ tuân thủ đơn giản hóa toàn diện.

Kỹ thuật 5, thiết kế đầy đủ kỹ thuật miễn trừ tỷ lệ

Kỹ thuật miễn trừ tỷ lệ có mối quan hệ phụ thuộc hai chiều, phụ thuộc vào Kỹ thuật 1 vì cần dữ liệu chi phí tuân thủ để xác định ngưỡng có căn cứ, và phụ thuộc vào việc phát triển VN-AI Verify vì không thể có ngưỡng đơn giản hóa mà không có công cụ chứng minh tuân thủ theo chế độ đó. Nhận thức được sự phụ thuộc này, giải pháp được thiết kế theo trình tự: xây dựng cơ sở dữ liệu chi phí tuân thủ trước (2027), phát

triển VN-AI Verify đồng thời (2028), sau đó bổ sung ngưỡng miễn trừ vào văn bản khi đã có đủ căn cứ (2029).

Về ngưỡng miễn trừ, dựa trên kết quả khảo sát chi phí tuân thủ từ Kỹ thuật 1, Nghị định 142 kỳ sửa đổi bổ sung điều khoản cho phép áp dụng chế độ tuân thủ đơn giản hóa khi doanh nghiệp đồng thời đáp ứng hai tiêu chí: doanh thu hàng năm từ sản phẩm, dịch vụ AI dưới ngưỡng được xác định sau khảo sát và số người dùng thường xuyên tại Việt Nam của hệ thống AI dưới ngưỡng tương ứng. Hai tiêu chí cộng dồn tránh được trường hợp doanh nghiệp lớn có doanh thu cao lách qua ngưỡng nhờ ít người dùng. Ngưỡng cụ thể phải là kết quả của khảo sát thực tiễn, đây chính là lý do Kỹ thuật 1 phải đi trước.

Về thể chế hóa VN-AI Verify như công cụ chứng minh tuân thủ trong chế độ đơn giản hóa, cần bổ sung vào Nghị định 142 hai quy định: doanh nghiệp đạt dưới ngưỡng được sử dụng VN-AI Verify thay cho kiểm định bên thứ ba đầy đủ; kết quả VN-AI Verify tạo biện hộ khẳng định (affirmative defense) trong tranh chấp pháp lý khi doanh nghiệp đã thực hiện tự đánh giá đầy đủ và công khai kết quả được suy đoán là đã tuân thủ thiện chí cho đến khi bên khiếu nại chứng minh được ngược lại. Cơ chế này học từ Texas TRAIGA áp dụng cho tuân thủ NIST AI RMF, mạnh hơn "suy đoán tuân thủ" đơn thuần vì đặt gánh nặng chứng minh lên bên khiếu nại, không phải bên bị khiếu nại.

Kỹ thuật 6, hoàn thiện luật hóa tiêu chuẩn kỹ thuật theo lộ trình hai giai đoạn

Kỹ thuật pháp luật hóa tiêu chuẩn phụ thuộc vào điều kiện tiên quyết là có TCVN AI đủ phong phú, do đó giải pháp phải thực hiện theo hai giai đoạn.

Trong giai đoạn 1 từ năm 2027 đến 2028 phát huy ngay với TCVN đã có, Bộ KH&CN ban hành Thông tư chỉ định TCVN AI tạo suy đoán tuân thủ, cụ thể là tuyên bố: hệ thống quản lý AI của tổ chức, doanh nghiệp tuân thủ TCVN ISO/IEC 42001 phiên bản hiện hành được suy đoán là đáp ứng yêu cầu về hệ thống quản lý rủi ro AI tại Điều X Nghị định 142. Thông tư này không cần sửa Nghị định, chỉ cần quyết định chính sách và thẩm quyền ban hành Thông tư cấp Bộ trưởng, tức là có thể thực hiện trong vòng sáu tháng sau khi có ý chí chính trị. Đây là cơ hội tạo ra tác động lập tức nhất trong số tất cả các giải pháp kỹ thuật được đề xuất, vì giải quyết ngay tình trạng "không biết tuân thủ TCVN nào thì đủ" mà chuyên gia DN-04 đã mô tả là chi phí giao dịch lớn nhất trong tuân thủ AI hiện nay [Xem Phụ lục 5].

Trong giai đoạn 2 từ năm 2029 đến 2030 mở rộng theo lộ trình TCVN, khi năm TCVN cấp bách được ban hành theo lộ trình đề xuất tại Mục 4.2.2.3, Bộ KH&CN bổ sung từng TCVN vào Danh mục hàng năm. Mục tiêu đến năm 2030 là mỗi yêu cầu pháp lý quan trọng trong Luật AI 2025 và Nghị định 142 đều có ít nhất một TCVN tương ứng trong Danh mục, doanh nghiệp tuân thủ tiêu chuẩn tương ứng có sự chắc chắn pháp lý hoàn toàn về tuân thủ yêu cầu luật định. Toàn bộ Danh mục áp dụng kỹ

thuật viện dẫn không ngày học từ Nhật Bản, bảo đảm khi Uỷ ban Tiêu chuẩn Đo lường Chất lượng Quốc gia ban hành phiên bản TCVN mới, Thông tư tự động áp dụng mà không cần sửa đổi là cơ chế tạo ra vòng cập nhật liên tục không đòi hỏi hành động lập pháp tại mỗi chu kỳ.

Đặt ba giải pháp cuối cùng (Kỹ thuật 4, 5, 6) trong mối quan hệ với ba giải pháp đầu (Kỹ thuật 1, 2, 3), có thể thấy rõ logic hệ thống. Kỹ thuật 1 tạo ra dữ liệu để Kỹ thuật 5 sử dụng dữ liệu thiết kế ngưỡng; Kỹ thuật 2 mở rộng định nghĩa tác nhân AI tạo cơ sở cho Kỹ thuật 4 phân định nghĩa vụ cho tác nhân AI trong chuỗi giá trị; Kỹ thuật 3 ủy quyền Danh mục TCVN giúp Kỹ thuật 6 luật hóa TCVN trong Danh mục đó.

4.2.3.5. Giải pháp về điều kiện bảo đảm

Thứ nhất, ý chí chính trị và định hướng chiến lược: Đưa vào chương trình hành động thực hiện Nghị quyết 57-NQ/TW và Quyết định 1671/QĐ-TTg các chỉ tiêu riêng, đo được về thể chế AI, chẳng hạn tỷ lệ tầng đạt mức tối thiểu hoặc số tiêu chuẩn quốc gia về AI được công bố theo năm, kèm cơ quan chịu trách nhiệm. Thiết lập chế độ báo cáo định kỳ trước Ban Chỉ đạo Trung ương về tiến độ xây dựng khung.

Thứ hai, năng lực thể chế và nhân lực: Tổ chức bồi dưỡng chuyên đề về AI cho cán bộ soạn thảo, thẩm định, thanh tra và thẩm phán. Đầu tư đào tạo đội ngũ chuyên gia đánh giá sự phù hợp và năng lực công nhận tổ chức đánh giá.

Thứ ba, hạ tầng pháp lý và kỹ thuật: Xây dựng phòng thử nghiệm và kiểm định AI ở cấp quốc gia, hoặc theo mô hình hợp tác công tư, ưu tiên các lĩnh vực có nhiều hệ thống trong Danh mục rủi ro cao. Ban hành bản dịch tiếng Anh chính thức của Luật AI và các văn bản hướng dẫn. Nâng cấp năng lực kỹ thuật và kết nối liên thông giữa Cổng thông tin một cửa với hệ thống tiếp nhận thông báo về dữ liệu cá nhân.

Thứ tư, sự tham gia của các bên liên quan: Thiết lập diễn đàn đối thoại kỹ thuật thường xuyên giữa cơ quan quản lý, doanh nghiệp, giới nghiên cứu và tổ chức xã hội, không cần gắn với một dự thảo cụ thể. Hỗ trợ hiệp hội ngành nâng cao năng lực nghiên cứu chính sách và tự điều chỉnh. Chuẩn bị tài liệu giải thích chính sách bằng ngôn ngữ dễ hiểu, kể cả cho người dân tộc thiểu số và người khuyết tật.

Thứ năm, nguồn lực tài chính: Xác định dòng kinh phí riêng cho xây dựng khung pháp lý về AI trong mức chi cho KHCN, ĐMST và chuyển đổi số, gồm nghiên cứu chính sách, đánh giá tác động, xây dựng tiêu chuẩn và hạ tầng kiểm thử. Bảo đảm nguồn chi cho phiếu hỗ trợ theo Nghị định 142, để quy định này không bị treo. Huy động nguồn thu từ dịch vụ đánh giá sự phù hợp và hỗ trợ kỹ thuật quốc tế. Công khai và đánh giá định kỳ hiệu quả sử dụng các nguồn này.

Nếu ưu tiên theo tác động pháp lý cao và khả thi trong 12-18 tháng, ba giải pháp hàng đầu là Nghị quyết HĐTP TANDTC hướng dẫn tòa án các cấp về trình tự và phương pháp xác định khi giải quyết tranh chấp liên quan đến AI gây thiệt hại (thẩm quyền sẵn có, không cần sửa luật); Thông tư Danh mục TCVN AI (công cụ pháp lý sẵn có); Bổ sung định nghĩa tác nhân AI vào Nghị định 142 sửa đổi (kỹ thuật lập pháp xác định, không cần điều kiện tiên quyết). Ba giải pháp này có thể được triển khai đồng thời và tạo ra thay đổi trong kiến trúc khung pháp lý AI, tạo nền tảng cho các giải pháp đòi hỏi điều kiện tiên quyết trong giai đoạn 2027-2030.

Tổng kết Chương 4

Từ những vấn đề lý luận và thực tiễn xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam, cần quán triệt năm quan điểm cũng như hệ thống giải pháp được đề xuất. Quan điểm thứ nhất, bảo đảm thể chế đầy đủ chủ trương Đảng và Nhà nước về thúc đẩy phát triển khoa học, công nghệ và AI. Quan điểm thứ hai, bảo đảm tính khoa học, thống nhất trong hệ thống pháp luật và đáp ứng yêu cầu thực tiễn. Quan điểm thứ ba, bảo đảm xây dựng đồng bộ, cân đối giữa các tầng của khung pháp lý. Quan điểm thứ tư, bảo đảm phù hợp với pháp luật quốc tế và xu hướng chung của các quốc gia trong bối cảnh hội nhập. Quan điểm thứ năm, bảo đảm an ninh quốc phòng, trật tự công cộng, lợi ích cộng đồng và bảo vệ quyền con người.

Về giải pháp phát triển, ứng dụng AI, luận án đề xuất tám nhóm giải pháp tương ứng với tám giai đoạn của vòng đời hệ thống AI. Về hoàn thiện khung pháp lý, luận án đóng góp bảy nhóm giải pháp phù hợp với bảy tầng. Về giải pháp xây dựng khung pháp lý, Chương 4 kiến nghị theo bốn chiều. Nguyên lý thống nhất xuyên suốt hệ thống giải pháp là khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam không thể hoàn thiện chỉ bằng cách ban hành thêm VBQPPL mà đòi hỏi xây dựng đồng thời tất cả bảy tầng với chất lượng xây dựng được nâng cao qua bốn chiều chủ thể, nội dung, quy trình và kỹ thuật, đặt con người làm trung tâm của mọi lựa chọn kiến trúc và kỹ thuật.

KẾT LUẬN

Trong quá trình xây dựng Nhà nước pháp quyền XHCN kiến tạo phát triển, liêm chính, hành động, phục vụ nhân dân, lấy KHCN, ĐMST và chuyển đổi số làm động lực chính, việc xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam là một yêu cầu khách quan, cấp thiết nhằm hiện thực hóa mục tiêu làm chủ các công nghệ tiên tiến, công nghệ chiến lược, công nghệ lõi, công nghệ nguồn, công nghệ số, góp phần nâng cao chất lượng cuộc sống, vững bước vào kỷ nguyên vươn mình của nước phát triển. Vì bản chất của AI thế hệ mới đang đặt ra những vấn đề pháp lý mà hệ thống pháp luật được thiết kế cho thế giới trước AI không thể giải quyết.

Do luận án nghiên cứu hoạt động xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam trong giai đoạn khung pháp lý vừa được hình thành, các kết luận khoa học rút ra như sau:

Về lý luận, luận án làm rõ hệ thống khái niệm cốt lõi, "khung pháp lý", "phát triển, ứng dụng AI", "khung pháp lý cho phát triển, ứng dụng AI", "xây dựng khung pháp lý cho phát triển, ứng dụng AI". Khung pháp lý về AI được tiếp cận như một hệ thống có cấu trúc gồm hai nhóm thành tố: công cụ điều chỉnh và thiết chế bảo đảm thực hiện. Hai nhóm này được cụ thể hóa thành bảy tầng, từ nền tảng hiến định đến thiết chế thực thi. Hoạt động xây dựng khung pháp lý được phân tích theo bốn chiều là chủ thể, nội dung, quy trình và kỹ thuật, và được đánh giá bằng ba bộ tiêu chí riêng cho quá trình, cho sản phẩm và cho điều kiện bảo đảm.

Tham khảo kinh nghiệm quốc tế từ tám đối tượng (LHQ/UNESCO, OECD, EU, Hoa Kỳ, Trung Quốc, Nhật Bản, Hàn Quốc, Singapore) rút ra bài học kiến trúc có thể điều chỉnh và vận dụng vào bối cảnh Việt Nam. Từ đó, luận án đề xuất mô hình kiến trúc khung pháp lý cho phát triển, ứng dụng AI khả thi cho Việt Nam là kiến trúc phân tầng thích ứng với bảy tầng.

Về thực trạng, luận án nhận định và lý giải có hệ thống những hạn chế nhằm xác định chính xác những gì cần tiếp tục làm: (i) Mức độ hoàn thiện tổng thể của khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam đạt yêu cầu cơ bản, nhưng còn mất cân đối giữa các tầng. (ii) Thực trạng xây dựng khung pháp lý cho phát triển, ứng dụng AI theo bốn chiều: chủ thể thiếu năng lực công nghệ-pháp lý; nội dung thiếu Tài liệu kiến trúc pháp lý AI quốc gia; quy trình bắt nguồn từ mô hình tuần tự; sáu kỹ thuật xây dựng được vận dụng không đồng đều. Tình huống điển hình phản ứng pháp lý liên ngành trước làn sóng lừa đảo bằng deepfake 2023-2025 bổ sung bằng chứng xác nhận hạn chế không tồn tại ở một điểm đơn lẻ mà phân bố xuyên suốt toàn bộ bốn chiều xây dựng khung pháp lý.

Về các giải pháp đặc thù, luận án đặc biệt nhấn mạnh những giải pháp sau: (1) Mô hình VN-AI Verify theo quy trình đồng thiết kế. (2) Hệ thống TCVN AI rút gọn 12 tháng với tiêu chuẩn tạm thời. (3) Bộ công cụ kỹ thuật lập pháp AI. (4) Chương trình phát triển án lệ AI chủ động của TAND Tối cao. (5) Hệ thống thông tin quản lý AI AMIS như hạ tầng vòng phản hồi chính sách.

Luận án đề xuất giải pháp xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam theo bốn thành tố. Về chủ thể, xây dựng năng lực công nghệ-pháp lý trong hệ thống nhà nước; tăng cường vai trò lập pháp chuyên sâu của Quốc hội; thể chế hóa Hội đồng tư vấn AI quốc gia; tăng cường đại diện quốc tế chủ động; thể chế hóa thẩm định kỹ thuật bắt buộc trong quy trình lập pháp AI. Về nội dung, thể chế hóa Chương trình Nghiên cứu Chính sách AI Quốc gia; xây dựng và duy trì Tài liệu kiến trúc pháp lý AI quốc gia; thể chế hóa phân tích khoảng trống bảy tầng hàng năm; thể chế hóa quy trình rà soát tuân thủ điều ước quốc tế. Về quy trình, chuẩn hóa quy trình bốn giai đoạn cho văn bản AI quan trọng; nâng cao chất lượng tham vấn bằng bốn biện pháp bổ sung; vận hành cơ chế phản ứng nhanh bằng nghị quyết Chính phủ; thể chế hóa đánh giá hậu kỳ bắt buộc. Về kỹ thuật, nâng cấp AI-RIA lên ba chiều đặc thù; hoàn thiện định nghĩa chức năng và thiết lập cơ chế diễn giải; bổ sung đầy đủ ủy quyền linh hoạt phân tầng; hoàn thiện phân định nghĩa vụ theo chuỗi giá trị; thiết kế đầy đủ kỹ thuật miễn trừ tỷ lệ; hoàn thiện luật hóa tiêu chuẩn kỹ thuật theo lộ trình.

Nếu thực hiện đầy đủ các giải pháp này, Việt Nam sẽ có một khung pháp lý tương đối hoàn chỉnh cho phát triển và ứng dụng AI; nguồn nhân lực dồi dào và chế độ hỗ trợ mạnh mẽ cho doanh nghiệp; hội nhập sâu vào cộng đồng quốc tế, có mạng lưới công nhận lẫn nhau về chứng nhận AI và đóng góp tích cực vào việc định hình chuẩn mực toàn cầu về AI. Việt Nam cần chuẩn bị đầy đủ điều kiện để có cơ hội nằm trong số các quốc gia dẫn đầu ASEAN về quản trị AI, tạo nền tảng vững chắc cho sự phát triển bền vững của nền kinh tế số và xã hội số trong kỷ nguyên AI.

DANH MỤC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ ĐÃ CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN

1. Nguyễn Quỳnh Trang (2023), "Công nghệ nhận diện khuôn mặt - Những vấn đề đặt ra đối với quyền riêng tư và bảo vệ dữ liệu cá nhân", *Tạp chí Khoa học pháp lý Việt Nam*, số 09(169)/2023, tr.11-22.
2. Nguyễn Quỳnh Trang (2024), "Tác động của trí tuệ nhân tạo đến thương mại quốc tế: hàm ý đối với Việt Nam", *Tạp chí Pháp luật và Phát triển*, số 09/2024, tr.57-70.
3. Nguyễn Quỳnh Trang, Nguyễn Mai Lan (2024), "Ứng dụng trí tuệ nhân tạo trong hoạt động xét xử của Toà án Hoa Kỳ, Trung Quốc và giá trị tham khảo cho Việt Nam", *Tạp chí Nghề luật*, số tháng 8/2024, tr.51-58.
4. Nguyễn Quỳnh Trang, Nguyễn Quỳnh Hương Ly (2026), "Yêu cầu thể chế mới đối với quản trị rủi ro trí tuệ nhân tạo trong cung ứng dịch vụ hành chính công", *Tạp chí Quản lý nhà nước*, số 368 (8/2026), tr.28-33.
5. Nguyen Quynh Trang (2023), "Artificial Intelligence application in migration management", in *International Conference Proceedings: International migration and social integration: Inclusion and entire development*, Vietnam Women's Academy, Hanoi, The Labour Publishing House, Hanoi, p.359-371, ISBN: 978-604-480-673-0.
6. Nguyen Quynh Trang (2025), "Artificial intelligence improves the operational efficiency of global supply chains", in *International Conference Proceedings: Promoting Sustainable and Fair Employment in Global Supply Chains: Policies, Laws and Practices*, Binh Duong University, Ho Chi Minh City, People's Public Security Publishing House, Hanoi, p.633-640, ISBN: 978-604-72-7936-4.

DANH MỤC TÀI LIỆU THAM KHẢO

*** Tài liệu tiếng Việt**

1. Chu Thị Thanh An, Phạm Thị Hiền (2018), "Trí tuệ nhân tạo và những thách thức trong việc xác định trách nhiệm bồi thường thiệt hại ngoài hợp đồng", *Tạp chí Luật học* số 11-2018, tr.3-16.
2. Nguyễn Hải An, Đinh Thị Thảo, Nguyễn Đặng Hồng Thiên (2022), *"Bảo hộ quyền Sở hữu trí tuệ đối với trí tuệ nhân tạo của một số quốc gia trên thế giới và kiến nghị cho Việt Nam"*, *Tạp chí Toà án nhân dân* số 17 năm 2022, tr.30-38
3. Gia Tuấn Anh (2025), *Việt Nam đã có mô hình AI do người Việt làm chủ và kiểm soát huấn luyện*, Báo Thanh niên, tại trang https://thanhnien.vn/viet-nam-da-co-mo-hinh-ai-do-nguoi-viet-lam-chu-va-kiem-soat-huan-luyen-18525061621131588 4.htm, [truy cập ngày 30/12/2025]
4. Lê Lan Anh, & Lê Thị Kim Oanh (2025), "Khung pháp lý để thực hiện các mục tiêu chuyển đổi số trong bối cảnh sự trỗi dậy của quốc gia Việt Nam", *Tạp chí Đổi mới Quản lý và Kinh tế Hoa Kỳ*, 7(09), tr.25-30
5. Nguyễn Thị Quế Anh, Vũ Công Giao và Mai Văn Thắng (đồng chủ biên) (2019), *Trí tuệ nhân tạo với pháp luật và quyền con người*, Nxb. Tư pháp, Hà Nội.
6. Nguyễn Thị Quế Anh (2021), "Một số khía cạnh pháp lý trong lĩnh vực luật tư khi ứng dụng trí tuệ nhân tạo", *Tạp chí Nghiên cứu Lập pháp* số 22 (446), tháng 11/2021, tr.3-14
7. Nguyễn Thị Quế Anh (2022), "Hoàn thiện pháp luật sở hữu trí tuệ Việt Nam trong bối cảnh phát triển trí tuệ nhân tạo", *Tạp chí Luật học* (Đại học Quốc gia Hà Nội) số 3 năm 2022, tr.1-16
8. Nguyễn Thị Quế Anh, Nguyễn Bích Thảo (đồng chủ biên) (2024), *Phát triển trí tuệ nhân tạo có trách nhiệm: Tiếp cận từ khía cạnh đạo đức, xã hội và pháp lý*, NXB Chính trị quốc gia Sự thật, Hà Nội.
9. Đặng Thái Bình, Nguyễn Trọng Luận (2023), "Quyền tác giả đối với tác phẩm được tạo ra từ trí tuệ nhân tạo", Tạp chí Khoa học Pháp lý Việt Nam số 08(168)/2023, tr.61-70.
10. Bộ Chính trị (2019), *Nghị quyết số 52-NQ/TW ngày 27/9/2019 về một số chủ trương, chính sách chủ động tham gia cuộc Cách mạng công nghiệp lần thứ tư*, Hà Nội.
11. Bộ Chính trị (2024), *Nghị quyết số 57-NQ/TW ngày 22/12/2024 về đột phá phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia*, Hà Nội.
12. Bộ Chính trị (2025), *Nghị quyết số 59-NQ/TW ngày 25/1/2025 về hội nhập quốc tế trong tình hình mới*, Hà Nội.

13. Bộ Chính trị (2025), *Nghị quyết số 66-NQ/TW ngày 30/4/2025 về đổi mới công tác xây dựng và thi hành pháp luật*, Hà Nội.
14. Bộ Chính trị (2025), *Nghị quyết số 68-NQ/TW ngày 4/5/2025 về phát triển kinh tế tư nhân*, Hà Nội.
15. Bộ Khoa học và Công nghệ (2024), *Quyết định số 1290/QĐ-BKHCN ngày 11/06/2024 về việc hướng dẫn một số nguyên tắc về nghiên cứu, phát triển các hệ thống AI có trách nhiệm*, Hà Nội.
16. Bộ Khoa học và Công nghệ (2026), *Thông tư số 05/2026/TT-BKHCN ngày 10/3/2026 về Khung đạo đức trí tuệ nhân tạo quốc gia*, Hà Nội.
17. CMC ATI (2025), *Trợ lý ảo CLS tiếp tục gây ấn tượng trên sóng truyền hình*, tại trang https://cmcati.vn/tro-ly-ao-cls-tiep-tuc-gay-an-tuong-tren-song-truyen-hinh/, [truy cập ngày 30/12/2025]
18. Nguyễn Văn Cương (chủ biên) (2021), *Tính ổn định của pháp luật: Nhận thức, thực tiễn và giải pháp đổi mới*, Nhà xuất bản Chính trị quốc gia Sự thật, Hà Nội.
19. Hoa Hữu Cường, Chử Thị Nhuần (2022), "Chiến lược xây dựng đạo đức trí tuệ nhân tạo tại CHLB Đức và một số gợi mở cho Việt Nam", *Tạp chí Nghiên cứu Châu Âu* số 1 năm 2022, tr.58-66.
20. Chính phủ (2020), *Nghị quyết số 50/2020/NQ-CP ngày 17/4/2020 ban hành Chương trình hành động của Chính phủ thực hiện Nghị quyết số 52-NQ/TW ngày 27/9/2019 của Bộ Chính trị về một số chủ trương, chính sách chủ động tham gia cuộc Cách mạng công nghiệp lần thứ tư*, Hà Nội.
21. Chính phủ (2021), *Nghị định số 98/2021/NĐ-CP ngày 8/11/2021 về quản lý trang thiết bị y tế*, Hà Nội.
22. Chính phủ (2023), *Nghị định số 13/2023/NĐ-CP ngày 17/4/2023 về bảo vệ dữ liệu cá nhân*, Hà Nội.
23. Chính phủ (2024), *Nghị định số 182/2024/NĐ-CP ngày 31/12/2024 về thành lập, quản lý và sử dụng Quỹ Hỗ trợ đầu tư*, Hà Nội.
24. Chính phủ (2025), *Nghị định số 169/2025/NĐ-CP ngày 30/6/2025 quy định hoạt động KHCN, ĐMST và sản phẩm, dịch vụ về dữ liệu*, Hà Nội.
25. Chính phủ (2025), *Nghị định số 267/2025/NĐ-CP ngày 14/10/2025 quy định chi tiết và hướng dẫn một số điều Luật KHCN&ĐMST*, Hà Nội.
26. Chính phủ (2025), *Nghị định số 356/2025/NĐ-CP ngày 31/12/2025 quy định chi tiết một số điều và biện pháp thi hành Luật Bảo vệ dữ liệu cá nhân*, Hà Nội.
27. Chính phủ (2025), *Nghị định số 94/2025/NĐ-CP ngày 21/7/2025 quy định về cơ chế thử nghiệm có kiểm soát đối với hoạt động ngân hàng, hoạt động kinh doanh ngoại hối và hoạt động khác trong lĩnh vực ngân hàng*, Hà Nội.

28. Chính phủ (2026), *Nghị định số 142/2026/NĐ-CP ngày 30/4/2026 quy định chi tiết một số điều và biện pháp thi hành Luật Trí tuệ nhân tạo*, Hà Nội.
29. Bùi Quang Chung, Vũ Hữu Đạo (2024), "Nghiên cứu và phát triển thử nghiệm mô hình trợ giảng số sử dụng công nghệ AI và IoT", *Tạp chí Khoa học công nghệ thông tin và truyền thông*, số 1.
30. Ngô Ngọc Diễm, Chu Huyền My (2023), "Cơ sở pháp lý điều chỉnh những vấn đề liên quan đến phần mềm ChatGPT", *Tạp chí Luật sư Việt Nam* số 3, tháng 3-2023, tr.27-30
31. Phạm Thị Diễm (2022), "Tác động của ứng dụng trí tuệ nhân tạo trong cung ứng dịch vụ công ở Việt Nam và hàm ý chính sách", *Tạp chí Khoa học xã hội Việt Nam* số 9(177), tr.42-49
32. Thái Thị Tuyết Dung, Nguyễn Văn Lãm, Nguyễn Văn Dương (2025), "Xây dựng khung pháp lý về trí tuệ nhân tạo: kinh nghiệm của một số quốc gia và gợi mở cho Việt Nam", *Tạp chí Nghiên cứu lập pháp*, số 04(507), tháng 12/2025, tr.50-64
33. Nguyễn Tiến Dũng, Phạm Trung Thiên, Lê Ngọc Dũng, Đỗ Văn Tỉnh, Vũ Xuân Tú, Nguyễn Văn Hiệp, Nguyễn Ngọc Thể (2022), "Nghiên cứu trợ lý ảo ứng dụng trí tuệ nhân tạo", *Tạp chí Khoa học và Công nghệ* - Trường Đại học Kinh tế - Kỹ thuật Công nghiệp số 33 năm 2022, tr.29-35.
34. Trần Việt Dũng (2023), *Pháp luật Việt Nam trước sự phát triển của trí tuệ nhân tạo: nghiên cứu, đánh giá và hoàn thiện,* Đề tài NAFOSTED, Trường Đại học Luật TP Hồ Chí Minh
35. Trần Việt Dũng (2024), "Bàn về phát triển quy phạm đạo đức trong tiến trình xây dựng pháp luật về trí tuệ nhân tạo", Tạp chí khoa học pháp lý Việt Nam, số 09(181)/2024, tr.30-39
36. Nguyễn Ngọc Hồng Dương (2022), "Bảo vệ quyền tác giả đối với tác phẩm của trí tuệ nhân tạo", *Tạp chí Công thương* số 9 năm 2022, tr.28-33.
37. Đảng Cộng sản Việt Nam (2026), *Văn kiện Đại hội đại biểu toàn quốc lần thứ XIV*, NXB Chính trị quốc gia Sự thật, Hà Nội.
38. Đảng Cộng sản Việt Nam (2026), *Văn kiện Đại hội đại biểu toàn quốc lần thứ XIV*, Tập I, NXB Chính trị quốc gia Sự thật, Hà Nội.
39. FPT.AI (2024), *FPT.AI đạt danh hiệu Giải thưởng Nền tảng AI vì Tương lai Doanh nghiệp Số*, tại trang https://fpt.ai/vi/tin-tuc/fpt-ai-duoc-giai-thuong-nen-tang-ai-vi-tuong-lai-doanh-nghiep-so/, [truy cập ngày 30/12/2025]

40. Nguyễn Minh Hải (2025), *Overfitting là gì? Làm thế nào để phát hiện và khắc phục hiện tượng overfitting?*, tại trang https://vnptai.io/vi/blog/detail/overfitting-la-gi, [truy cập ngày 10/3/2026]
41. Nguyễn Thị Hảo, Nguyễn Văn Trọng (2025), "Quản lý nhà nước về phát triển và ứng dụng trí tuệ nhân tạo tại Việt Nam hiện nay", *Tạp chí Quản lý nhà nước*, tại trang https://www.quanlynhanuoc.vn/2025/04/10/quan-ly-nha-nuoc-ve-phat-trien-va-ung-dung-tri-tue-nhan-tao-tai-viet-nam-hien-nay/, [truy cập ngày 22/6/2026]
42. Lưu Hiệp (2023), *Kết nối kinh tế Việt Nam-Singapore*, tại trang https://cand.vn/ket-noi-kinh-te-viet-nam-singapore-post705182.html, [truy cập ngày 22/6/2026]
43. Chu Thị Hoa (2019), "Sandbox - cơ chế thử nghiệm áp dụng trong phạm vi hạn chế - kinh nghiệm quốc tế và một số gợi ý cho Việt Nam", *Tạp chí Nghiên cứu lập pháp*, số 16(391), tr.58-64
44. Chu Thị Hoa (chủ nhiệm) (2025), *Những khía cạnh pháp lý của trí tuệ nhân tạo ở Việt Nam: Lý luận và thực tiễn, Đề tài cấp Bộ*, Viện Chiến lược và Khoa học pháp lý, Bộ Tư pháp, Hà Nội.
45. Nguyễn Thị Hoa (2023), Xác định trách nhiệm bồi thường thiệt hại do trí tuệ nhân tạo gây ra - Kinh nghiệm của Liên minh châu Âu cho Việt Nam", Tạp chí Khoa học Pháp lý Việt Nam số 08(168)/2023, tr.1-10
46. Đặng Xuân Hoan, Nguyễn Quỳnh Nga (2020), "Đổi mới phương thức điều hành của chính phủ trên nền tảng ứng dụng trí tuệ nhân tạo", *Tạp chí Tổ chức nhà nước* số 3/2020, tr.7-14
47. Trần Thị Hoan (2020), *Nghiên cứu ứng dụng một số công cụ kiểm thử có ứng dụng trí tuệ nhân tạo,* Luận văn thạc sĩ ngành Công nghệ thông tin, Đại học Quốc gia Hà Nội, Hà Nội.
48. Nguyễn Hoàng (2021), *Trí tuệ nhân tạo - "trợ lý" đắc lực trong nỗ lực xây dựng Y tế số*, tại trang https://suckhoedoisong.vn/tri-tue-nhan-tao-tro-ly-dac-luc-trong-no-luc-xay-dung-y-te-so-169190694.htm, [truy cập ngày 30/12/2025]
49. Hội đồng Thẩm phán Toà án nhân dân Tối cao (2019), *Nghị quyết số 04/2019/NQ-HĐTP ngày 18/6/2019 về quy trình lựa chọn, công bố và áp dụng án lệ*, Hà Nội.
50. Lê Minh Hùng, Nguyễn Thiện Tâm (2023), "Chủ thể chịu trách nhiệm bồi thường thiệt hại do trí tuệ nhân tạo gây ra", *Tạp chí Khoa học Pháp lý Việt Nam,* số 08(168)/2023, tr.11-26
51. Vũ Văn Hùng, Hồ Kim Hương (2022), "Nghiên cứu tác động của trí tuệ nhân tạo trong quá trình xây dựng thành phố thông minh: Hàm ý chính sách đối với Việt Nam", *Tạp chí Kinh tế và Quản lý*, số 41 (2022), tr.74-82.

52. Mạc Duy Hưng (2020), *Nghiên cứu ứng dụng trí tuệ nhân tạo trong dự báo chất lượng không khí*, Luận án tiến sĩ môi trường, Trường Đại học Bách Khoa Hà Nội, Hà Nội.
53. Nguyễn Thị Hường (2023), *Nghiên cứu ứng dụng trí tuệ nhân tạo (AI) cho khâu chuẩn bị sản xuất trong nhà máy may*, Đề tài cấp Quốc gia, trường Đại học Công nghiệp Dệt May Hà Nội, Hà Nội.
54. Nguyễn Thị Thu Hương (2020), *Kinh nghiệm quốc tế trong xây dựng chính sách phát triển trí tuệ nhân tạo*, Tạp chí Quản lý nhà nước số 2, 2020, tại trang https://www.quanlynhanuoc.vn/2020/05/05/kinh-nghiem-quoc-te-trong-xay-dung-chinh-sach-phat-trien-tri-tue-nhan-tao/ [truy cập ngày 30/12/2023]
55. Vũ Thị Thanh Hương (2022), *Tác động của trí tuệ nhân tạo trong quản trị nguồn nhân lực*, Tạp chí Công thương số 12 năm 2022, tại trang https://tapchicongthuong.vn/bai-viet/tac-dong-cua-tri-tue-nhan-tao-trong-quan-tri-nguon-nhan-luc-97508.htm, [truy cập ngày 30/12/2022]
56. Nguyễn Hoàng Thái Hy, Nguyễn Phượng An (2023), "Trí tuệ nhân tạo và trách nhiệm dân sự: Trường hợp của phần mềm hỗ trợ quyết định và chẩn đoán y tế - kinh nghiệm từ pháp luật cộng hòa Pháp", Tạp chí Khoa học Pháp lý Việt Nam số 08(168)/2023, tr.71-86
57. IPS (2024), *Chính sách phát triển và kiểm soát rủi ro đối với AI*, Hà Nội; IPS (2026), *Thực hành tốt về tích hợp trí tuệ nhân tạo hiệu quả và trách nhiệm trong khu vực công*, Hà Nội; IPS (2026) *Hướng dẫn tích hợp trí tuệ nhân tạo hiệu quả và trách nhiệm trong khu vực công*, Hà Nội
58. Trần Kiên, Hồ Minh Thành (2023), "Trách nhiệm pháp lý liên quan đến trí tuệ nhân tạo trong việc xử lý dữ liệu cá nhân - pháp luật Liên minh châu Âu và gợi mở cho Việt Nam", Tạp chí Khoa học Pháp lý Việt Nam số 08(168)/2023, tr.38-48
59. Nguyễn Thường Lạng, Trương Thị Ánh, Đào Minh Hiền, Trần Bích Liên, Nguyễn Thị Thu Quỳnh, Đỗ Thị Phương Trinh (2022), "Tác động của tự động hóa quy trình bằng trí tuệ nhân tạo (Chatbot) đến trải nghiệm khách hàng ngành bán lẻ thương mại điện tử ở Việt Nam", *Tạp chí Kinh tế và Dự báo*, số 18 năm 2022, tr.33-36
60. Nguyễn Văn Lê (2013), *Nghiên cứu ứng dụng trí tuệ nhân tạo chẩn đoán sự cố tiềm ẩn trong máy biến áp lực - Ứng dụng cho hệ thống điện Việt Nam*, Luận án tiến sĩ kỹ thuật, Đại học Đà Nẵng, Đà Nẵng.
61. Vũ Ngọc Diệu Linh, Quản Thiện Hưng, Kim Thăng Long (2026), Hình sự hoá hành vi sử dụng trí tuệ nhân tạo thao túng thị trường tài chính - Kinh nghiệm quốc tế khuyến nghị cho Việt Nam, *Tạp chí điện tử Luật sư Việt Nam*, tại

trang https://lsvn.vn/hinh-su-hoa-hanh-vi-su-dung-tri-tue-nhan-tao-thao-tung-thi-truong-tai-chinh-kinh-nghiem-quoc-te-va-khuyen-nghi-cho-viet-nam-a168870.html, [truy cập ngày 22/6/2026].

62. Hạ Linh (2026), *Ngân hàng Nhà nước ngăn chặn gần 4.000 tỉ đồng giao dịch rủi ro*, Báo Lao động, tại trang https://laodong.vn/kinh-doanh/ngan-hang-nha-nuoc-ngan-chan-gan-4000-ti-dong-giao-dich-rui-ro-1678531.ldo, [truy cập ngày 1/8/2026].
63. Hồ Đắc Lộc, Huỳnh Châu Duy (2020), "Phát triển trí tuệ nhân tạo tại Việt Nam: Thực trạng và giải pháp", *Tạp chí Khoa học và Công nghệ Việt Nam*, số 1+2 năm 2020, tr.27-31
64. M.H (2024), Cục An toàn thông tin ghi nhận hơn 5.000 phản ánh về lừa đảo trực tuyến, tại trang https://antoanthongtin.vn/tin/cuc-an-toan-thong-tin-ghi-nhan-hon-5000-phan-anh-ve-lua-dao-truc-tuyen, [truy cập ngày 19/5/2026]
65. Nguyễn Đức Minh (2017), "Bàn về khái niệm chính sách pháp luật", *Tạp chí Nhà nước và Pháp luật*, số 3/2017, tr.10-14.
66. MN (2019), "Nguyễn Thanh Thuỷ: Nhân lực là yếu tố cốt lõi đảm bảo sự thành công của phát triển trí tuệ nhân tạo", *Tạp chí Khoa học & Công nghệ Việt Nam*, số 9 năm 2019, tr.12-15
67. Đỗ Giang Nam, Đào Trọng Khôi (2022), "Pháp luật bồi thường thiệt hại ngoài hợp đồng Việt Nam trước thách thức của trí tuệ nhân tạo và công nghệ số mới nổi", *Tạp chí nghiên cứu lập pháp* số 20 (468) - tháng 10/2022, tr.3-13.
68. NCA (2026), An ninh mạng người dùng cá nhân 2025: Giảm lừa đảo, tăng thách thức, tại trang https://nca.org.vn/news/detail/an-ninh-mang-nguoi-dung-ca-nhan-2025-giam-lua-dao-tang-thach-thuc-1767769688440?l=vi, [truy cập ngày 19/5/2026]
69. NCS (2025), *Báo cáo tình hình an ninh mạng Việt Nam năm 2024*, tại trang https://ncsgroup.vn/bao-cao-tinh-hinh-an-ninh-mang-viet-nam-2024/, [truy cập ngày 3/3/2026]
70. NIC, JICA, BCG (2025), *Báo cáo Nền kinh tế AI của Việt Nam*: Cơ hội bứt phá và định hướng chiến lược phát triển, Hà Nội.
71. Chu Thị Ngân (2023), *Robot dạng người thông minh IVASBot*, tại trang https://vast.gov.vn/tin-chi-tiet/-/chi-tiet/robot-dang-nguoi-thong-minh-ivastbot-77878-403.html, [truy cập ngày 1/1/2024]
72. Hồ Thuý Ngọc (2025), Bảo hộ quyền sở hữu trí tuệ đối với các sản phẩm tạo ra từ trí tuệ nhân tạo, Luận án tiến sĩ Luật học, Trường Đại học Luật TP Hồ Chí Minh, TP Hồ Chí Minh.
73. Tiến Nguyễn (2026), *Công bố 10 quy tắc sử dụng AI trong báo chí*, tại trang https://tuoitre.vn/cong-bo-10-quy-tac-su-dung-ai-trong-bao-chi-100260620192011778.htm, [truy cập ngày 11/6/2026]

74. Quách Ánh Nguyệt (2025), *Các khía cạnh về quyền sở hữu trí tuệ theo quy định của Đạo luật Trí tuệ nhân tạo của Liên minh châu Âu và khuyến nghị cho Việt Nam*, Luận văn thạc sĩ Luật học, trường Đại học Luật Hà Nội, Hà Nội.
75. Nguyễn Nhâm (2025), *Giải quyết bài toán dữ liệu và nhân lực cho phát triển AI "Make in Vietnam"*, Báo Điện tử Tiếng nói Việt Nam, tại trang https://vov.vn/cong-nghe/chuyen-doi-so/giai-quyet-bai-toan-du-lieu-va-nhan-luc-cho-phat-trien-ai-make-in-vietnam-post1197265.vov, [truy cập ngày 30/12/2025]
76. Đoàn Văn Nhật (2020), *Trách nhiệm bảo vệ quyền con người trong việc ứng dụng trí tuệ nhân tạo*, Luận văn thạc sĩ Luật học, Đại học Quốc gia Hà Nội, Hà Nội.
77. Phạm Thanh Nhật (2022), "Phát triển và ngăn chặn gian lận trong hoạt động ngân hàng thông qua trí tuệ nhân tạo", *Tạp chí Tài chính* kỳ 1 tháng 3/2022, tr.63-65
78. Đặng Minh Phương, Nguyễn Như Hà (2022), "Nhận biết những tác động của trí tuệ nhân tạo đến pháp luật sáng chế của Liên minh Châu Âu", *Tạp chí Luật học* (Đại học Quốc gia Hà Nội) số 3 năm 2022, tr.30-38
79. Hà Phương (2024), *Tổng Bí thư Tô Lâm trao đổi chuyên đề "Kỷ nguyên phát triển mới - Kỷ nguyên vươn mình của dân tộc Việt Nam"*, tại trang https://www.tapchicongsan.org.vn/web/guest/thong-tin-ly-luan/-/2018/1005102/tong-bi-thu-to-lam-trao-doi-chuyen-de-%E2%80%9Cky-nguyen-phat-trien-moi---ky-nguyen-vuon-minh-cua-dan-toc-viet-nam%E2%80%9D.aspx, [truy cập ngày 20/2/2026]
80. Lê Trang Thu Phương (2025), *Quyền được bảo hộ dữ liệu cá nhân trong bối cảnh phát triển trí tuệ nhân tạo (AI) ở Việt Nam hiện nay*, Luận văn thạc sĩ Luật học, trường Đại học Luật Hà Nội, Hà Nội.
81. Nguyễn Mạnh Quân (2018), "Chuẩn bị cho tương lai của trí tuệ nhân tạo", *Tạp chí Khoa học và Công nghệ Việt Nam* số 6 năm 2018, tr.53-55.
82. Nguyễn Văn Quân (2019), "Một số tác động của trí tuệ nhân tạo tới nghề luật", *Tạp chí Nghiên cứu lập pháp*, số 12(388) tháng 6/2019, tr.12-18
83. Quốc hội (2005), *Luật Sở hữu trí tuệ* (sửa đổi, bổ sung năm 2009, 2019, 2022, 2025), Hà Nội.
84. Quốc hội (2006), *Luật Tiêu chuẩn và Quy chuẩn kỹ thuật* (sửa đổi, bổ sung năm 2018, 2025), Hà Nội.
85. Quốc hội (2015), *Bộ luật Dân sự*, Hà Nội.
86. Quốc hội (2015), *Bộ luật Hình sự* (sửa đổi, bổ sung năm 2017), Hà Nội.
87. Quốc hội (2025), *Luật An ninh mạng*, Hà Nội.
88. Quốc hội (2025), *Luật Bảo vệ dữ liệu cá nhân*, Hà Nội.
89. Quốc hội (2025), *Luật Công nghiệp công nghệ số*, Hà Nội.
90. Quốc hội (2025*), Luật Khoa học, Công nghệ và đổi mới sáng tạo*, Hà Nội.
91. Quốc hội (2025), *Luật Trí tuệ nhân tạo*, Hà Nội.

92. Tào Thị Quyên (chủ nhiệm) (2025), *Nghiên cứu hoàn thiện thể chế phát triển khoa học và công nghệ ở Việt Nam trong bối cảnh mới*, Đề tài cấp Bộ trọng điểm, Học viện Chính trị quốc gia Hồ Chí Minh, Hà Nội.
93. Doãn Ngọc San (chủ nhiệm) (2023), *Nghiên cứu xây dựng hệ thống trí tuệ nhân tạo tích hợp dữ liệu địa chất dầu khí đánh giá triển vọng dầu khí*, Đề tài trọng điểm quốc gia, Trường Đại học Dầu khí Việt Nam - Bộ Giáo dục và Đào tạo, Hà Nội.
94. Bạch Tân Sinh (2017), "Kế hoạch trở thành quốc gia dẫn đầu thế giới về trí tuệ nhân tạo của Trung Quốc", *Tạp chí Khoa học và Công nghệ Việt Nam* số 10 năm 2017, tr.44-46
95. Trần Hồng Sơn (chủ nhiệm) (2024), *Nghiên cứu, xây dựng ban hành tiêu chuẩn Công nghệ thông tin - Trí tuệ nhân tạo - Hướng dẫn quản lý rủi ro*, Đề tài cấp Bộ, Viện Công nghệ số và Chuyển đổi số quốc gia, Hà Nội.
96. T.N (2023), *Bài 2: Lừa đảo tinh vi bằng cuộc gọi video Deepfake*, Chuyên trang Công an TP. Hồ Chí Minh, tại trang https://congan.com.vn/vu-an/canh-giac/bai-2-lua-dao-cuoc-goi-video-deepfake_150557.html, [truy cập ngày 19/5/2026]
97. Tatjana Hörnle (2024), "The Challenges of Human-Robot Interaction for Substantive Criminal Law" in book Human-Robot Interaction in Law and Its Narratives, Cambridge University Press, Cambridge.
98. Lê Minh Tâm (2025), *Những khái niệm cơ bản về cấu trúc hệ thống pháp luật và mối quan hệ giữa các thành tố trong cấu trúc hệ thông pháp luật*, Báo điện tử Pháp luật Việt Nam, tại trang https://baophapluat.vn/gs-ts-le-minh-tam-he-thong-phap-luat-luon-phat-trien-bien-doi-khong-ngung.html, [truy cập ngày 8/5/2026]
99. Trần Minh Tâm (2019), "Tìm hiểu về trí tuệ nhân tạo", *Tạp chí Khoa học - Đại học Văn Lang* số 13, tháng 1- 2019, tr.139-145
100. Phùng Trung Tập (2015), "Phong tục, tập quán và áp dụng tập quán trong công tác xét xử án dân sự", *Tạp chí Nghiên cứu lập pháp* số 5(285), tháng 3/2015
101. Phùng Trung Tập (2020), "Áp dụng lẽ công bằng để giải quyết tranh chấp dân sự", *Tạp chí Nghiên cứu lập pháp* số 1(401), tháng 1/2020, tr.11-16
102. Trần Anh Tú, Đinh Hải Đăng (2024), "Thúc đẩy nghiên cứu phát triển và sử dụng trí tuệ nhân tạo một cách có trách nhiệm ở Việt Nam", *Tạp chí Khoa học & Công nghệ Việt Nam*, số 12, tr.12-15
103. Nguyễn Minh Tuấn, Tống Thị Phương (2024), "Thực trạng nhận thức, áp dụng luật mềm ở Việt Nam và một số khuyến nghị", *Tạp chí Nghiên cứu lập pháp* số 12(491), tháng 6/2024, tr.3-9
104. Trần Minh Tuấn (chủ nhiệm) (2018), *Nghiên cứu, đề xuất định hướng thúc đẩy*

phát triển trí tuệ nhân tạo (AI) ở Việt Nam, Đề tài khoa học cấp Bộ, Viện Chiến lược thông tin và truyền thông - Bộ Thông tin và Truyền thông, Hà Nội.

105. Nguyễn Thanh Tùng (2025), "Điều chỉnh pháp luật về trí tuệ nhân tạo - Kinh nghiệm quốc tế và gợi mở cho Việt Nam từ sự xuất hiện của ChatGPT", *Tạp chí Dân chủ và Pháp luật, số Kỳ 2 (441) tháng* 11/2025
106. Phạm Thanh Tùng, Nguyễn Thủy Nguyên, Nguyễn Tiến Nhật Quang (2021), "Chính sách pháp luật về phát triển trí tuệ nhân tạo của Mỹ và một số bài học kinh nghiệm cho Việt Nam", *Tạp chí Công Thương* số 23, tháng 10 năm 2021, tr.79-85
107. Dương Thị Tươi (2019), "Bảo đảm chất lượng xây dựng pháp luật ở Việt Nam hiện nay", *Tạp chí Lý luận chính trị* số 6-2019, tr.23-29
108. Nguyễn Phương Thanh (2025), *Cuộc trò chuyện với trợ lý trí tuệ nhân tạo AI AVA của MISA - Trợ lý AI Kế toán đa nhiệm*, tại trang https://amis.misa.vn/118960/cuoc-tro-chuyen-voi-tro-ly-tri-tue-nhan-tao-ai-cua-misa-tro-ly-ai-ke-toan-da-nhiem/, [truy cập ngày 30/12/2025]
109. Chu Thị Thắm (chủ nhiệm) (2024), *Nghiên cứu xây dựng tiêu chuẩn về Công nghệ thông tin -Trí tuệ nhân tạo - Quản trị sử dụng trí tuệ nhân tạo trong các tổ chức*, Đề tài cấp Bộ, Viện Công nghệ số và Chuyển đổi số quốc gia, Hà Nội.
110. Trần Quyết Thắng (2026), "Bản chất pháp lý và cơ chế vận hành của khung pháp lý thử nghiệm có kiểm soát - thực trạng và hướng hoàn thiện", *Tạp chí Quản lý nhà nước*, số 360, tr.109-114.
111. Văn Thể, Hoàng Gia (2022), *VinBus ký kết hợp tác triển khai ứng dụng App VinBus*, tại trang https://vinbus.vn/vinbus-ky-ket-hop-tac-trien-khai-ung-dung-app-vinbus, [truy cập ngày 1/1/2024]
112. Thủ tướng Chính phủ (2021), *Quyết định số 127/2021/QĐ-TTg ngày 21/01/2021 ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng AI đến năm 2030*, Hà Nội.
113. Thủ tướng Chính phủ (2026), *Quyết định số 33/2026/QĐ-TTg ngày 30/6/2026 ngày ban hành Danh mục hệ thống trí tuệ nhân tạo có rủi ro cao*, Hà Nội.
114. Hà Quang Thuy, Nguyễn Thanh Thuỷ, Phạm Bảo Sơn, Phan Xuân Hiếu, Trần Trọng Hiếu, Trần Mai Vũ, Trần Quốc Long, Nguyễn Trí Thành, Lý Hoàng Tùng (2020), "Môi trường và giải pháp cho chiến lược quốc gia về trí tuệ nhân tạo của Việt Nam", *Tạp chí Khoa học và Công nghệ Việt Nam* số 8 năm 2020, tr.15-19.
115. Nguyễn Đức Thuỷ (2022), "Quản lý trí tuệ nhân tạo kinh nghiệm quốc tế và gợi mở với Việt Nam", *Tạp chí Công nghệ thông tin và truyền thông* số 8 - tháng 8/2022, tr.68-75

116. Nguyễn Đức Thuỷ (chủ nhiệm) (2023), *Nghiên cứu định hướng chuẩn hoá về trí tuệ nhân tạo ở Việt Nam*, Đề tài khoa học cấp Bộ, Viện công nghiệp phần mềm và nội dung số Việt Nam - Bộ Thông tin và Truyền thông, Hà Nội.
117. Nguyễn Đức Thuỷ (chủ nhiệm) (2023), *Nghiên cứu xây dựng tiêu chuẩn quốc gia về Khung cho hệ thống trí tuệ nhân tạo (AI) sử dụng học máy (ML)*, Đề tài cấp Bộ, Viện Công nghệ số và Chuyển đổi số quốc gia, Hà Nội.
118. Nguyễn Thanh Thuỷ (1999), *Trí tuệ nhân tạo (Các phương pháp giải quyết vấn đề và kỹ thuật xử lý tri thức)*, NXB Giáo dục, Hà Nội
119. Nguyễn Thanh Thuỷ, Hà Quang Thuy, Phan Xuân Hiếu, Nguyễn Trí Thành (2019), "Góp bàn về chiến lược phát triển trí tuệ nhân tạo của Việt Nam", *Tạp chí Khoa học và Công nghệ Việt Nam* số 11 năm 2018, tr.8-11
120. Phạm Thu Thuỷ (2021), "Chiến lược phát triển công nghệ trí tuệ nhân tạo của Nhật Bản hiện nay", *Tạp chí Nghiên cứu Đông Bắc Á* số 4 (242) 4-2021, tr.60-68.
121. Nguyễn Quý Thành Trung (chủ nhiệm) (2024), *Nghiên cứu, xây dựng ban hành tiêu chuẩn về Công nghệ thông tin - Trí tuệ nhân tạo (AI) - Sự thiên vị trong hệ thống AI và AI hỗ trợ ra quyết định*, Đề tài cấp Bộ, Viện Công nghệ số và Chuyển đổi số quốc gia, Hà Nội.
122. Trung tâm WTO và Hội nhập (2026), Tổng hợp các FTA của Việt Nam tính đến tháng 2/2026, tại trang https://trungtamwto.vn/thong-ke/12065-tong-hop-cac-fta-cua-viet-nam-tinh-den-thang-112018, [truy cập ngày 11/5/2026]
123. Thắng Trung (2024), Cách dùng camera trên smartphone để tránh bị lừa bằng AI Deepfake, tại trang https://congan.com.vn/vu-an/canh-giac/cach-dung-camera-tren-smartphone-de-tranh-bi-lua-bang-ai-deepfake_171553.html#google_vignette, [truy cập ngày 19/5/2026]
124. Thắng Trung (2024), Cảnh báo tội phạm lừa đảo qua công nghệ giả giọng nói "như thật", tại trang https://congan.com.vn/vu-an/phong-chong-toi-pham-tren-khong-gian-mang/canh-bao-toi-pham-lua-dao-qua-cong-nghe-gia-giong-noi-nhu-that_165575.html, [truy cập ngày 19/5/2026]
125. Ủy ban Tiêu chuẩn đo lường chất lượng quốc gia (2026), *Tra cứu Tiêu chuẩn quốc gia - TCVN*, tại trang http://tracuu.tcvn.vn/sdomain/front/tieu-chuan-viet-nam, [truy cập ngày 13/9/2026]
126. VBD (2023), *VinBigdata ra mắt "ChatGPT" phiên bản Việt đầu tiên dành cho người dùng cuối*, tại trang https://vinbigdata.com/news/vinbigdata-ra-mat-chatgpt-phien-ban-viet-dau-tien-danh-cho-nguoi-dung-cuoi, [truy cập ngày 1/1/2024]

127. VDCA (2025), Ra mắt Bộ quy tắc sử dụng AI có trách nhiệm - "kim chỉ nam cho truyền thông số, tại trang https://vdca.org.vn/ra-mat-bo-quy-tac-su-dung-ai-co-trach-nhiem-kim-chi-nam-cho-truyen-thong-so, truy cập ngày 30/10/2025]
128. Viettel (2024), *Viettel khai trương trung tâm dữ liệu lớn nhất Việt Nam, công nghệ xanh, sẵn sàng cho phát triển AI*, tại trang https://viettel.com.vn/vi/tin-tuc-va-su-kien/tin-tuc/viettel-khai-truong-trung-tam-du-lieu-lon-nhat-viet-nam-cong-nghe-xanh-san-sang-cho-phat-trien-ai/, [truy cập ngày 20/2/2025]
129. Viện Công nghệ Thông tin - Đại học Quốc gia Hà Nội (2025), *Báo cáo thường niên về Trí tuệ nhân tạo tại Việt Nam*, Hà Nội.
130. Viện Khoa học pháp lý, Bộ Tư pháp (2022), "Các yếu tố để tạo ra một văn bản quy phạm pháp luật có chất lượng tốt", Tham luận tại Hội thảo *Chất lượng của quy định và sự thuận lợi của môi trường kinh doanh*, đăng trên VIB Online.
131. Viện Ngôn ngữ học (2003), *Từ điển Tiếng Việt*, NXB Đà Nẵng, Trung tâm Từ điển học, Hà Nội - Đà Nẵng, tr.1034
132. Trịnh Tiến Việt (2021), "Hình phạt áp dụng đối với thực thể trí tuệ nhân tạo (AI): Từ khoa học viễn tưởng đến viễn cảnh tương lai của Bộ luật Hình sự Việt Nam trong thời gian tới"*, Tạp chí Toà án nhân dân* số 8 năm 2021, tr.19-22.
133. VINASA (2024), *VINASA ra mắt Uỷ ban Đạo đức trí tuệ nhân tạo*, tại trang https://vinasa.org.vn/vinasa/4/3076/4211/18739/Tin-VINASA/VINASA-ra-mat-Uy-ban-Dao-duc-tri-tue-nhan-tao.aspx, [truy cập ngày 10/2/2026]
134. Võ Khánh Vinh (2025), "Đổi mới tư duy xây dựng pháp luật trong kỷ nguyên phát triển mới của đất nước", Tham luận tại Hội thảo khoa học "Vai trò của pháp luật trong việc thực hiện bốn đột phá chiến lược đưa đất nước phát triển trong kỷ nguyên mới", ngày 25/11/2025; đăng trên Trang Thông tin điện tử Viện Nhà nước và Pháp luật, ngày 15/01/2026.
135. VNPT (2024), *Trung tâm dữ liệu - "Xương sống" của nền kinh tế số*, tại trang https://vnpt.com.vn/tin-tuc/trung-tam-du-lieu-xuong-song-cua-nen-kinh-te-so.html, [truy cập ngày 20/2/2025]
136. Trịnh Quang Vũ, Dương Tuấn Kiệt, Nguyễn Thị Thanh Hằng, Trần Thảo Mi, Lương Thị Ngân (2021), "Đề xuất xây dựng hệ sinh thái AI cho Việt Nam: Một mô hình sơ khởi nhưng mang tính quyết định thành bại", *Tạp chí Khoa học và Công nghệ Việt Nam* số 9 năm 2021, tr.13-17.
137. Phùng Thị Yến, Phạm Bùi Vi Thảo (2026), "Chuyển đổi số trong lập pháp và thi hành pháp luật: Trí tuệ nhân tạo và dữ liệu mở - Thách thức và cơ hội", Tạp chí điện tử *Pháp lý*, tại trang https://phaply.net.vn/chuyen-doi-so-trong-lap-phap-va-thi-hanh-phap-luat-tri-tue-nhan-tao-va-du-lieu-mo-thach-thuc-va-co-hoi-a260712.html, [truy cập ngày 20/6/2026]

*** Tài liệu nước ngoài**

138. AB-602 Depiction of individual using digital or electronic technology: sexually explicit material: cause of action, USA, 2019, tại trang https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602, [truy cập ngày 10/2/2026]
139. AB-730 Elections: deceptive audio or visual media, USA, 2019, tại trang https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730, [truy cập ngày 10/2/2026]
140. ACE Electoral Knowledge Network, "Legal Framework", tại trang https://aceproject.org/ace-en/topics/lf/default/mobile_browsing, [truy cập ngày 30/4/2026]
141. Administrative Conference of the United States (2021), *Notice-and-Comment Rulemaking*, tại trang https://www.acus.gov/sites/default/files/documents/IIB014-Rulemaking.pdf, [truy cập ngày 5/5/2025]
142. Alan Dafoe (2018), *AI Governance: A Research Agenda*, Governance of AI Program, Future of Humanity Institute, University of Oxford.
143. Alan Watson (1974), *Legal Transplants: An Approach to Comparative Law*, Scottish Academic Press.
144. Alina Köchling & Marius Claus Wehner (2020), "Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development", *Business Research* Vol 13, p.795-848
145. Amir Rabet Ghannadi (2026), "Artificial Intelligence and International Law: Challenges and Opportunities", *Legal Studies in Digital Age*, 5(1), p.1-15.
146. Andreas Theodorou, Virginia Dignum (2020), "Towards Ethical and Socio-legal Governance in AI", *Nature Machine Intelligence*, Vol 2, p.1-3
147. Andrew D. Selbst, et al. (2019), "Fairness and Abstraction in Sociotechnical Systems", in *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*)*, p.59-68.
148. Ankur Mallick, Kevin Hsieh, Behnaz Arzani, Gauri Joshi (2022), *Matchmaker: Data Drift Mitigation in Machine Learning for Large-Scale Systems*, tại trang https://www.microsoft.com/en-us/research/wp-content/uploads/2022/01/MLSYS2022.pdf, [truy cập ngày 10/3/2026]
149. Annette Elisabeth Töller (2011), *Voluntary Approaches to Regulation-Patterns, Causes and Effects*, In book: The Handbook of the Politics of Regulation, Chapter: 36, p.499-510, David Levi-Faur (ed.), Edward Elgar.

150. Annie Badman, Matthew Kosinski, *What is metadata?*, tại trang https://www.ibm.com/think/topics/metadata, [truy cập ngày 3/3/2026]
151. Anu Bradford (2020), *The Brussels Effect: How the European Union Rules the World*, Oxford University Press.
152. Archie Drake, Perry Keller, Irene Pietropaoli, Anuj Puri, Spyros Maniatis, Joe Tomlinson và các cộng sự (2022), "Legal contestation of artificial intelligence-related decision-making in the United Kingdom: reflections for policy", *International Review of Law, Computers & Technology*, Volume 36 Issue 2, p.251-285.
153. ASEAN (2023), *Digital Economy Framework Agreement (DEFA): ASEAN to leap forward its digital economy and unlock US$2 Tn by 2030*, tại trang https://asean.org/asean-defa-study-projects-digital-economy-leap-to-us2tn-by-2030/, [truy cập ngày 5/4/2024]
154. ASEAN (2024), *ASEAN Guide on AI Governance and Ethics*, tại trang https://asean.org/book/asean-guide-on-ai-governance-and-ethics/, [truy cập ngày 5/4/2024]
155. AWS, *What is hyperparameter tuning?*, tại trang https://aws.amazon.com/what-is/hyperparameter-tuning/, [truy cập ngày 10/3/2026]
156. B.J. Copeland (2026), *Artificial Intelligence*, Encyclopedia *Britannica*, tại trang https://www.britannica.com/technology/artificial-intelligence, [truy cập 25/6/2026]
157. Bart Custers, Edward Fosch-Villaronga (2022), "*Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice"*, *Information Technology and Law Series*, vol. 35, p.570-576
158. Bergur Thormundsson (2024), *Distribution of overall AI companies worldwide 2023, by country*, tại trang https://www.statista.com/statistics/1452378/ai-distribution-by-country-worldwide/, [truy cập ngày 30/1/2026]
159. Benjamin Mueller (2021), *How Much Will the Artificial Intelligence Act Cost Europe?*, Center for Data Innovation, tại trang https://itif.org/publications/2021/07/26/how-much-will-artificial-intelligence-act-cost-europe/, truy cập ngày 12/10/2026
160. Bernard Marr & Matt Ward (2019), *Ứng dụng trí tuệ nhân tạo để dẫn đầu*, NXB Thế giới.
161. C2PA, *Advancing digital content transparency and authenticity*, tại trang https://c2pa.org, [truy cập ngày 15/3/2026]
162. CAC (2022), *Internet Information Service Algorithm Recommendation Management Regulations*, tại trang https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm, [truy cập ngày 20/2/2025]

163. CAC (2022), *Regulations on the Administration of Deep Synthesis of Internet Information Services*, tại trang https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm, [truy cập ngày 20/2/2025]
164. CAC (2025), *Measures for Labelling of AI-Generated Synthetic Content*, tại trang https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm, [truy cập ngày 3/3/2026]
165. CAC (2025), *The AI Safety Governance Framework 2.0*, tại trang https://www.cac.gov.cn/2025-09/15/c_1759653448369123.htm, [truy cập ngày 3/3/2026]
166. California Legislative Information (2024), *SB-942 California AI Transparency Act*, tại trang https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942, [truy cập ngày 3/3/2026]
167. Carl Folke, Thomas Hahn, Per Olsson & Jon Norberg (2005), "*Adaptive Governance of Social-Ecological Systems", Annual Review of Environment and Resources*, 30, p.441-473.
168. Carlo Casonato, Giulia Olivato (2024), "AI Regulation in Europe: Exploring the Artificial Intelligence Act", in book Digital Environments and Human Relations, p.87-111
169. Cath Corinne (2018), *Governing artificial intelligence: ethical, legal and technical opportunities and challenges*, Philosophical Transactions of the Royal Society A.376, tháng 10/2018, tại trang https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0080 [truy cập ngày 17/12/2023]
170. CGTN (2019), *Developing Responsible AI: Governance Principles for New Generation AI issued*, tại trang https://news.cgtn.com/news/7749444d774d4464776c6d636a4e6e62684a4856/index.html?from=timeline&isappinstalled=0, [truy cập ngày 14/6/2024]
171. Claudio Novelli, Giorgio Bongiovanni & Giovanni Sartor (2022), "A conceptual framework for legal personality and its application to AI*"*, *Jurisprudence*, Volume 13 Issue 2, p.194-219, tại trang https://www.tandfonline.com/doi/full/10.1080/20403313.2021.2010936 [truy cập ngày 17/12/2023]
172. Colin Scott (2004), "Regulation in the Age of Governance: The Rise of the Post-Regulatory State", Chapters, in: Jacint Jordana & David Levi-Faur (ed.), The Politics of Regulation, chapter 7, Edward Elgar Publishing.
173. Colorado General Assembly (2021), *Restrict Insurers' Use of External Consumer Data*, tại trang https://leg.colorado.gov/bills/sb21-169, [truy cập ngày 3/3/2026]

174. Court of Justice of the European Union (2015), Judgment of the Court (Second Chamber) of 15 January 2015. Ryanair Ltd v PR Aviation BV., tại trang https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:62014CJ0030, [truy cập ngày 10/2/2026]
175. Cranfield University (2006), *Regulatory tools: Legal & Policy Framework*, Bedfordshire MK43 0AL, UK, tại trang https://assets.publishing.service.gov.uk/media/57a08c44ed915d3cfd00129e/R8320-Sum19.pdf, [truy cập ngày 1/1/2024]
176. Chao Yang, Cui Huang (2022), *Quantitative mapping of the evolution of AI policy distribution, targets and focuses over three decades in China*, Technological Forecasting and Social Change, Volume 174, tại trang https://www.sciencedirect.com/science/article/pii/S0040162521006211 [truy cập ngày 17/12/2023]
177. Charles F. Sabel & Jonathan Zeitlin (2012), *Experimentalist Governance*, in book THE OXFORD HANDBOOK OF GOVERNANCE, DAVID LEVI-FAUR (ED.), OXFORD UNIVERSITY PRESS.
178. Damian Chalmers, Gareth Davies, Giorgio Monti, Veerle Heyvaert (2024), *European Union Law: Text and Materials*, Cambridge University Press, Cambridge
179. Daniel J. Gervais (2020), "The Machine as Author", *Iowa Law Review*, vol. 105, p.2053-2106.
180. Danielle K. Citron (2008), *Technological Due Process*, 85 Washington University Law Review 1249.
181. Daron Acemoglu & Pascual Restrepo (2017), "Robots and Jobs: Evidence from US Labor Markets", NBER Working Paper 23285; Daron Acemoglu & Pascual Restrepo (2018), "Artificial Intelligence, Automation and Work", NBER Working Paper 24196.
182. Data Science Council of America (2025), *Data Drift: What It Is, Why It Matters, and How to Tackle It*, tại trang https://www.dasca.org/world-of-data-science/article/data-drift-what-it-is-why-it-matters-and-how-to-tackle-it, [truy cập ngày 10/3/2026]
183. Dave Bergmann, *What is fine-tuning?*, tại trang https://www.ibm.com/think/topics/fine-tuning, truy cập ngày 30/12/2025]
184. David Collingridge (1980), *The Social Control of Technology*, Frances Pinter, London.
185. Deirdre K. Mulligan, Kenneth A. Bamberger (2018), "Saving Governance-By-Design", *California Law Review*, Vol.106, p.697-784

186. Deloitte (2025), *AI for Business: APAC Trends in AI Platform Adoption*, tại trang https://www.deloitte.com/southeast-asia/en/services/consulting/analysis/ai-business-apac-trends-platform-adoption.html, [truy cập ngày 15/6/2026]
187. Deloitte Insights (2018), *State of AI in the Enterprise, 2nd Edition*, Deloitte Development LLC, 2018., tại trang https://www2.deloitte.com/content/dam/insights/us/articles/4780_State-of-AI-in-the-enterprise/DI_State-of-AI-in-the-enterprise-2nd-ed.pdf, [truy cập ngày 15/12/2022]
188. Dennis F. Thompson (1980), "The moral responsibility of public officials: The problem of many hands", *American Political Science Review*, 74(4), p.905-916.
189. Doerte Peters (2010), *Policies and Legal Framework*, Further Resources: SSWM Background, tại trang https://sswm.info/sswm-university-course/module-3-ecological-sanitation-and-natural-systems-wastewater-treatment-2/policies-and-legal-framework, [truy cập ngày 1/1/2024]
190. Dongjun Hwang, Sungwon Woo, Tom Gao, Raymond Luo, Sunghwan Baek (2024), *Invisible Watermarks: Attacks and Robustness*, tại trang arXiv:2412.12511 [cs.CV], [truy cập ngày 15/3/2026]
191. Enoch Chi Ngai Lim, Chi Eung Danforn Lim (2025), "Polycentric AI Governance: A Multi-Stakeholder Approach to Distributed Responsibility and Ethical Technology Management", *International Journal of Advanced AI Applications* 1(4), p.77-97.
192. Eran Kahana (2023), *AI Life Cycle Core Principles*, Stanford Law School, tại trang https://law.stanford.edu/2023/03/17/ai-life-cycle-core-principles/, [truy cập ngày 12/8/2026].
193. European Center for Constitutional and Human Rights, tại trang https://www.ecchr.eu/en/glossary/hard-law-soft-law/, [truy cập ngày 15/3/2026]
194. European Commission (2022), *Proposal for a Directive on adapting non contractual civil liability rules to artificial intelligence*, tại trang https://commission.europa.eu/topics/business-and-industry/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en, [truy cập ngày 15/3/2026]
195. European Commission (2024), *AI Act*, tại trang https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai, [truy cập ngày 20/2/2025]
196. European Commission (2024), *European AI Office*, tại trang https://digital-strategy.ec.europa.eu/en/policies/ai-office, [truy cập ngày 20/2/2025]
197. European Commission (2025), *The General-Purpose AI Code of Practice*, tại

trang https://digital--strategy-ec-europa-eu.translate.goog/en/policies/contents-code-gpai?_x_tr_sl=en&_x_tr_tl=vi&_x_tr_hl=vi&_x_tr_pto=tc, [truy cập ngày 30/12/2025]

198. European Commission, *CE marking*, tại trang https://single-market-economy.ec.europa.eu/single-market/goods/ce-marking_en, [truy cập ngày 15/3/2026]
199. European Commission, Expert Group on Liability and New Technologies - New Technologies Formation (2019), *Liability for Artificial Intelligence and Other Emerging Digital Technologies*, Luxembourg: Publications Office of the European Union.
200. European Commission, *What is the Commission's proposal to simplify the AI Act implementation?*, tại trang https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai#1720699867912-2, [truy cập ngày 15/3/2026]
201. European Law Institute (2022), *Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration*, Vienna
202. European Parliament (2017), *Report with Recommendations to the Commission on Civil Law Rules on Robotics* (2015/2103(INL)), Committee on Legal Affairs, Rapporteur: Mady Delvaux, A8-0005/2017
203. European Parliament (2018), *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Technical Aspects*, tại trang https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf, [truy cập ngày 20/2/2025]
204. European Parliament (2019), *Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC*, tại trang https://eur-lex.europa.eu/eli/dir/2019/790/oj, [truy cập ngày 20/2/2025]
205. European Parliament (2024), *Artificial Intelligence Act: MEPs adopt landmark law*, tại trang https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law, [truy cập ngày 14/6/2024]
206. European Parliament and Council of the European Union (2024), *Regulation (EU) 2024/1689 of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*.
207. Everett M. Rogers (2003), *Diffusion of Innovations* (5th ed.), Simon and Schuster, New York.

208. Fabian Hinder, Valerie Vaquet, Barbara Hammer (2024), "One or two things we know about concept drift-a survey on monitoring in evolving environments. Part A: detecting concept drift", *Front. Artif. Intell.*, Sec. Machine Learning and Artificial Intelligence Volume 7 - 2024, tại trang https://doi.org/10.3389/frai.2024.1330257, [truy cập ngày 10/3/2026]
209. Finale Doshi-Velez & Been Kim (2017), "Towards A Rigorous Science of Interpretable Machine Learning", *arXiv*, tại trang https://arxiv.org/abs/1702.08608, [truy cập ngày 1/6/2026]
210. Finale Doshi-Velez, Mason Kortz (2017), *Accountability of AI Under the Law: The Role of Explanation*, Berkman Klein Center for Internet & Society working paper, tại trang https://dash.harvard.edu/entities/publication/73120379-2f8b-6bd4-e053-0100007fdf3b, [truy cập ngày 22/5/2026]
211. Frank Pasquale (2015), *The Black Box Society*, Havard University Press, Cambridge.
212. Garth Thornton (1996), *Legislative Drafting,* Butterworths, Reed Elsevier UK.
213. Gary E Marchant (2011), *Addressing the Pacing Problem*, The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight, Springer Netherlands, p.199-295
214. George H. Stankey, Roger N. Clark, Bernard T. Bormann (2005), *Adaptive management of natural resources: theory, concepts, and management institutions*. General Technical Report PNW-GTR-654, US Department of Agriculture, Forest Service, Pacific Northwest Research Station, tại trang https://research.fs.usda.gov/treesearch/20657, [truy cập ngày 15/5/2026]
215. Global Initiative against Transnational Organized Crime (2024), *Criminal exploitation of deepfakes in South East Asia*, tại trang https://globalinitiative.net/analysis/deepfakes-ai-cyber-scam-south-east-asia-organized-crime/, [truy cập ngày 1/8/2026].
216. Grand View Research (2022), *Artificial Intelligence Market Size, Share &Trend Analysis Report By Solution, By Technology (Deep Learning, Machine Learning, Natural Language Processing, Machine Vision), By End Use, By Region, And Segment Forecast, 2022 - 2030*, San Francisco, CA 94105, United States, tại trang https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market, [truy cập ngày 20/11/2025]
217. Grand View Research (2026), *U.S. Artificial Intelligence Market Size & Outlook,* tại trang https://www.grandviewresearch.com/horizon/outlook/artificial-intelligence-market/united-states, [truy cập ngày 30/1/2026]

218. Giandomenico Majone (1989), *Evidence, Argument, and Persuasion in the Policy Process*, Yale University Press, New Haven
219. Gillian K. Hadfield, Jack Clark (2026), "Regulatory Markets: The Future of AI Governance", 65 Jurimetrics J., p.195-240
220. Gunther Teubner (1998)), "Legal Irritants: Good Faith in British Law or How Unifying Law Ends Up in New Divergences", *Modern Law Review* Vol 61, p11-32
221. H.L.A. Hart (1961), *The Concept of Law*, Clarendon Press, Oxford.
222. Helen Xanthaki (2014), *Drafting Legislation: Art and Technology of Rules for Regulation*, Hart Publishing, Oxford.
223. High-Level Expert Group on Artificial Intelligence (2018), *A definition of AI: Main capabilities and scientific disciplines* (European Commission B-1049 Brussels, 2018), p.1
224. Huw Roberts, et al. (2021), "The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation", *AI & Society*, 36(1), p.59-77.
225. Huw Roberts, et al. (2023), "Artificial Intelligence Regulation in the United Kingdom: A Path to Good Governance and Global Leadership?", *Internet Policy Review*, vol. 12, no. 2, p.1-31
226. IAPP (2025), *Global AI Governance Law and Policy: China*, tại trang https://iapp.org/resources/article/global-ai-governance-china, [truy cập ngày 3/3/2026]
227. IBM (2023), *What is artificial intelligence (AI)*?, tại trang https://www.ibm.com/topics/artificial-intelligence>, [truy cập ngày 5/6/2023]
228. IMDA (2023), *Singapore launches AI Verify Foundation to shape the future of international AI standards through collaboration*, tại trang https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2023/singapore-launches-ai-verify-foundation, [truy cập ngày 15/3/2026]
229. IMDA (2024), *Singapore proposes framework to foster trusted Generative AI development*, tại trang https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2024/public-consult-model-ai-governance-framework-genai
230. IMDA, *Artificial Intelligence in Singapore*, tại trang https://www.imda.gov.sg/about-imda/emerging-technologies-and-research/artificial-intelligence, [truy cập ngày 22/6/2026]
231. Infocomm Media Development Authority and AI Verify Foundation (2024), *Model AI Governance Framework for Generative AI*, Singapore.
232. Iris H.-Y. Chiu, Ernest W.K. Lim (2021), "Managing corporations' risk in adopting artificial intelligence: A corporate responsibility paradigm", *Washington University Global Studies Law Review*, Vol.20, p.347-389

233. ISO (2017), *ISO/IEC JTC 1/SC 42 Artificial Intelligence*, tại trang https://www.iso.org/committee/6794475.html, [truy cập ngày 3/3/2026]
234. ISO (2026), *Standards*, tại trang https://www.iso.org/standards.html, [truy cập ngày 28/8/2026]
235. Jacques Bughin, et al. (2018), *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy*, McKinsey Global Institute.
236. James Moor (2006), "The Dartmouth Summer Research Project on Artificial Intelligence Conference: The next fifty years", *AI Magazine* Volume 27 Number 4, p.87.
237. Jan-Erik Schirmer (2019), *Teilrechtsfähigkeit for Autonomous Systems: Introducing a 'Halfway' Status*, tại trang https://blogs.law.ox.ac.uk/business-law-blog/blog/2019/08/teilrechtsfahigkeit-autonomous-systems-introducing-halfway-status, [truy cập ngày 15/4/2026]
238. Janneke de Snaijer (2024), *Trusting Robots: Limiting Due Diligence Obligations in Robot-Assisted Surgery under Swiss Criminal Law*, in book Human-Robot Interaction in Law and Its Narratives, Cambridge University Press, Cambridge.
239. Jeff Heaton (2015), *Artificial Intelligence for Humans*, CreateSpace Independent Publishing Platform
240. Jeremy Waldron (2010), *The Rule of Law and the Importance of Procedure*, NYU School of Law, Public Law. Research Paper No. 10-73
241. Jim Holdsworth, Ivan Belcic, Cole Stryker, *What is model drift?*, tại trang https://www.ibm.com/think/topics/model-drift, [truy cập ngày 10/3/2026]
242. John McCarthy (2007), *What Is Artificial Intelligence?*, Stanford University, tại trang http://jmc.stanford.edu/articles/whatisai.html, [truy cập ngày 9/5/2026]
243. Jonas Schuett (2022), "Defining the scope of AI regulations", *Law, Innovation and Technology*, Volume 15, Issue 1, p.60-82, tại trang https://www.tandfonline.com/doi/abs/10.1080/17579961.2023.2184135 [truy cập ngày 17/12/2023]
244. Jonathan Briskman (2025), *State of AI Apps Report 2025: Why Apps Across Verticals are Becoming AI-Powered*, tại trang https://sensortower.com/blog/state-of-ai-apps-report-2025, [truy cập ngày 30/12/2025]
245. Joon Sung Park, et al. (2023), "Generative Agents: Interactive Simulacra of Human Behavior", UIST '23: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, p.1-22

246. Julia Black (2001), "Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a 'Post-Regulatory' World", *Current Legal Problems 54(1)*, p.103-146
247. Julia Black (2003), "Enrolling Actors in Regulatory Systems: Examples from UK Financial Services Regulation", *Public Law*, p63-91
248. Julia Black (2005), "*The Emergence of Risk-Based Regulation and the New Public Risk Management in the United Kingdom"*, *Public Law*, p.512-549.
249. Julia Black (2008), "Constructing and contesting legitimacy and accountability in polycentric regulatory regimes", *Regulation & Governance* 2, p.137-164.
250. Kai - Fu Lee (2020), *Các Siêu Cường AI - Trung Quốc, Thung Lũng Silicon Và Trật Tự Thế Giới Mới*, NXB Trẻ, TP. Hồ Chí Minh.
251. Karen Yeung, Martin Lodge (ed.) (2019), *Algorithmic Regulation*, Oxford University Press, Oxford
252. Katerina Yordanova, Natalie Bertels (2024), "Regulating AI: Challenges and the Way Forward Through Regulatory Sandboxes", *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, Law, Governance and Technology Series, Volume 58, Springer, p.441-456.
253. Keepnet (2026), *Deepfake Statistics & Trends 2026: Growth, Risks, and Future Insights*, tại trang https://keepnetlabs.com/blog/deepfake-statistics-and-trends, [truy cập ngày 14/3/2026]
254. Kenneth W. Abbott, et al. (2000), "The Concept of Legalization", *International Organization* 54, 3, p.401-419.
255. Kok-Lim Alvin Yau, HeeJeong Jasmine Lee, Yung-Wey Chong, Mee Hong Ling, Aqeel Raza Syed, Celimuge Wu, and Hock Guan Goh (2021), "Augmented Intelligence: Surveys of Literature and Expert Opinion to Understand Relations Between Human Intelligence and Artificial Intelligence", *IEEE Access* (Volume 9), 136744 - 136761, tại trang https://ieeexplore.ieee.org/abstract/document/9548047 [truy cập ngày 4/1/2024]
256. Kyohei Yamamoto, and Yasunori Tabei (2025), "Japanese AI Regulation and Competition Law", The Law & Technology & Economics of AI (ed. Adrian Kuenzler, Thibault Schrepel & Volker Stocker), *Network Law Review*, Summer 2025.
257. Lawrence B. Solum (1992), "Legal Personhood for Artificial Intelligences", *North Carolina Law Review*, vol. 70, no. 4, p.1231-1287.
258. Louis G.Tornatzky, Mitchell Fleischer, Alok K. Chakrabarti (1990), *The Processes of Technological Innovation*, Lexington Books, Maryland.

259. Louis Kaplow (1995), "A Model of the Optimal Complexity of Legal Rules", The *Journal of Law, Economics, and Organization*, Volume 11, Issue 1, p.150-163
260. Luc J. Wintgens (2012), *Legisprudence: Practical Reason in Legislation*, Ashgate Publishing Limited, London
261. Luciano Floridi, et al. (2018), "An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", *Minds and Machines* 28 (2018), p.689-707
262. Mahabir Mohapatra (2025), *Retrain or Refresh? Updating Machine Learning Models the Right Way*, tại trang https://mhmohapatra.medium.com/retrain-or-refresh-updating-machine-learning-models-the-right-way-820098a9a3b2, [truy cập ngày 10/3/2026]
263. Mar Negreiro (2025), *Children and deepfakes*. European Parliamentary Research Service, tại trang https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775855/EPRS_BRI(2025)775855_EN.pdf, [truy cập ngày 3/3/2026]
264. Marco Tulio Ribeiro, Sameer Singh & Carlos Guestrin (2016), "Why Should I Trust You?: Explaining the Predictions of Any Classifier", KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p.1135 - 1144
265. Mariarosaria Taddeo, Tom McCutcheon, Luciano Floridi (2019), "Trusting Artificial Intelligence in Cybersecurity is a Double-Edged Sword", *Nature Machine Intelligence* 1(12), p.557-560
266. Marly van Assen, Scott J. Lee, Carlo N. De Cecco (2020), "Artificial intelligence from A to Z: From neural network to legal framework", *European Journal of Radiology* Volume 129, August 2020, tại trang https://www.sciencedirect.com/science/article/abs/pii/S0720048X20302722, [truy cập ngày 3/1/2024]
267. Marta Bo (2024), Are Programmers in or out of Control? The Individual Criminal Responsibility of Programmers of Autonomous Weapons and Self-Driving Cars, in book Human-Robot Interaction in Law and Its Narratives, Cambridge University Press, Cambridge.
268. Marton Varju, Kitti Mezei (ed.) (2025), *The Challenges of Artificial Intelligence for Law in Europe*, Data Science, Machine Intelligence, and Law, Springer, Switzerland.

269. MAS (2025), *MAS/TCRS/2025/06: Cyber Risks Associated with Deepfakes*, tại trang https://www.mas.gov.sg/regulation/circulars/cyber-risks-associated-with-deepfakes, [truy cập ngày 22/6/2026]
270. Matthew U. Scherer (2016), "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies", *Harvard Journal of Law & Technology* 29(2): p.353-400.
271. Max Tegmark (2019), *LIFE 3.0 - Loài người trong Kỷ nguyên Trí tuệ nhân tạo*, NXB Thế giới, Hà Nội
272. *Merriam-Webster Dictionary*, *Artificial Intelligence*, tại trang http://merriam-webster.com/dictionary/artificial%20intelligence, [truy cập ngày 25/6/2026]
273. Michael Veale, Frederik Zuiderveen Borgesius (2021), "Demystifying the Draft EU Artificial Intelligence Act", *Computer Law Review International*, 22(4), p.97-112.
274. Milo Foster (2025), *The AI Workshop: The Complete Beginner's Guide to AI: Your A-Z Guide to Mastering Artificial Intelligence for Life, Work, and Business-No Coding Required*, Funtacular Books, Illinois, United States
275. Ministry of Home Affairs (2025), *Second Reading of the Criminal Law (Miscellaneous Amendments) Bill - Opening Speech by Ms Sim Ann, Senior Minister of State, Ministry of Foreign Affairs and Ministry of Home Affairs*, tại trang https://www.mha.gov.sg/media-room/newsroom/second-reading-of-the-criminal-law-miscellaneous-amendments-bill-opening-speech/, [truy cập ngày 22/6/2026]
276. Ministry of Science and Technology (2019), *Governance Principles for New Generation Artificial Intelligence - Developing RAI*, tại trang https://www.gov.cn/zhengce/zhengceku/2020-08/09/5533454/files/bf4f158874434ad096636ba297e3fab3.pdf, [truy cập ngày 14/6/2024]
277. MSIT (2025), *Basic Act on the Development of Artificial Intelligence and the Establishment of a Foundation for Trustworthiness*, tại trang https://www.msit.go.kr/eng/bbs/view.do?bbsSeqNo=42&mId=4&mPid=2&nttSeqNo=1071&pageIndex=&sCode=eng&searchOpt=ALL&searchTxt=, [truy cập ngày 5/5/2026]
278. Nadia Naffi (2025), *Deepfakes and the crisis of knowing*, tại trang https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing, [truy cập ngày 14/3/2026]
279. Naftali Tishby & Ravid Schwartz-Ziv (2017), "Opening the Black Box of Deep Neural Networks via Information", *arXiv*, tại trang https://arxiv.org/abs/1703.00810, [truy cập ngày 1/6/2026]

280. National Conference of State Legislatures (2025), *Artificial Intelligence 2025 Legislation*, tại trang https://www.ncsl.org/technology-and-communication/artificial-intelligence-2025-legislation, [truy cập ngày 3/3/2026]
281. National Information Security Standardization Technical Committee (2021), *Information security technology - Security specification and assessment methods for machine learning algorithms*, tại trang https://www.tc260.org.cn/file/2021-08-04/6b530404-858b-4c9d-8d89-a83239ec5712.pdf, [truy cập ngày 14/6/2024]
282. National Information Security Standardization Technical Committee (2021), *Cybersecurity Standards Practice Guide - Artificial Intelligence Ethical Safety Risk Prevention Guidelines*, tại trang https://www.tc260.org.cn/file/zn10.pdf, [truy cập ngày 14/6/2024]
283. National Information Technology Standardisation (2023), *AI - Code of practice for data labeling of machine learning*, tại trang https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=2BE1BE0093F37778CFB2071DD490EECD, [truy cập ngày 14/6/2024]
284. National Institute of Standards and Technology (2023), *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, tại trang https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf, [truy cập ngày 30/12/2025]
285. National Institute of Standards and Technology (2026), *AI Agent Standards Initiative*, tại trang https://www.nist.gov/caisi/ai-agent-standards-initiative, [truy cập ngày 15/3/2026]
286. Nathalie A. Smuha (2021), "From a 'Race to AI' to a 'Race to AI Regulation': Regulatory Competition for Artificial Intelligence", *Law, Innovation and Technology*, 13(1), p.1-28
287. Nathalie A. Smuha (ed.) (2025), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, Cambridge University Press, Cambridge.
288. Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, Tobi Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, Sukrut Oak (2025), "The AI Index 2025 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025.
289. Nils J. Nilsson (1982), *Principles of Artificial Intelligence*, Springer-Verlag Heidelberg New York
290. NRGI Reader (2015), *Legal Framework: Navigating the Web of Laws and Contracts Governing Extractive Industries*, tại trang https://resourcego

vernance.org/sites/default/files/nrgi_Legal-Framework.pdf, [truy cập ngày 1/1/2024]

291. OECD (2012), *Recommendation of the Council on Regulatory Policy and Governance*, OECD Publishing, Paris.
292. OECD (2019), *Recommendation of the Council on Artificial Intelligence*, tại trang https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449, [truy cập ngày 20/2/2025]
293. OECD (2021), *OECD Regulation Policy Outlook 2021*, OECD Publishing, Paris, tại trang https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/10/oecd-regulatory-policy-outlook-2021_c5274577/38b0fdb1-en.pdf, [truy cập ngày 5/5/2026]
294. OECD (2021), *OECD Work on Implementing Trustworthy AI*, tại trang https://wp.oecd.ai/app/uploads/2021/10/OECD-Work-on-Implementing-Trustworthy-AI.pdf, [truy cập ngày 20/2/2025]
295. OECD (2022), *OECD Framework for classification of AI systems*, tại trang https://www.oecd.org/content/dam/oecd/en/publications/reports/2022/02/oecd-framework-for-the-classification-of-ai-systems_336a8b57/cb6d9eca-en.pdf, [truy cập ngày 20/2/2025]
296. OECD (2023), *Regulatory sandboxes in artificial intelligenc*e, OECD Digital Economy Papers No.356, tại trang https://www.oecd-ilibrary.org/docserver/8f80a0e6-en.pdf?expires=1718443853&id=id&accname=guest&checksum=35EF1E9A915D9E2C7ED423932DAA44AE, [truy cập ngày 14/6/2024]
297. OECD (2024), *AI principles*, tại trang https://www.oecd.org/en/topics/ai-principles.html, [truy cập ngày 5/5/2025]
298. OECD (2024), *Explanatory memorandum on the updated OECD definition of an AI system*, tại trang https://www.oecd.org/en/publications/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_623da898-en.html, [truy cập ngày 9/5/2026]
299. OECD (2024), *Global Partnership on Artificial Intelligence*, tại trang https://www.oecd.org/en/about/programmes/global-partnership-on-artificial-intelligence.html, [truy cập ngày 30/12/2025]
300. OECD (2025), *Towards a Common Reporting Framework for AI Incidents*, tại trang https://www.oecd.org/en/publications/towards-a-common-reporting-framework-for-ai-incidents_f326d4ac-en.html, [truy cập ngày 30/12/2025]
301. OECD (2026), *The agentic AI landscape and its conceptual foundations*, tại trang

https://www.oecd.org/en/publications/the-agentic-ai-landscape-and-its-conceptual-foundations_396cf758-en.html, [truy cập ngày 15/3/2026]

302. OECD.AI (2026), *Top countries in AI publications in time*, data from OpenAlex, tại trang https://oecd.ai/en/data?selectedArea=ai-research&selectedVisualization=16719&visualizationFiltersHash=eyJlbnRpdGllcyI6eyJjb3VudHJ5IjpbIlZOTSJdfSwiZGF0ZSI6WyIyMDEwIiwiMjAyNSJdfQ%3D%3D, [truy cập ngày 13/3/2026]
303. Olivia J. Erdélyi, Judy Goldsmith (2018), "Regulating Artificial Intelligence: Proposal for a Global Solution", in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18)*, p. 95-101.
304. Organisation for Economic Cooperation and Development (OECD), *Regulatory Policy Outlook 2021* (Paris: OECD Publishing, 2021), tại trang https://doi.org/10.1787/38b0fdb1-en. [truy cập ngày 26/2/2026]
305. Ortwin Renn (2008), *Risk Governance: Coping with Uncertainty in a Complex World*, Earthscan, London.
306. Oxford Insights (2025), *Government AI Readiness Index 2025*, England, WR13 6PL, United Kingdom, tại trang https://oxfordinsights.com/ai-readiness/government-ai-readiness-index-2025/, [truy cập ngày 30/1/2026]
307. Patrick Caughill (2017), *This AI has officially been granted residence*, World Economic Forum, tại trang https://www.weforum.org/stories/emerging-technologies/this-ai-has-officially-been-granted-residence-877b423c-191c-4d39-b6c9-305d27d8b0c2, [truy cập ngày 12/10/2026]
308. Pei Wang (2019), "On defining artificial intelligence", *Journal of Artificial General Intelligence* 10(2) 1-37, tại trang https://sciendo.com/downloadpdf/journals/jagi/10/2/article-p1.pdf [truy cập ngày 3/1/2024].
309. Peter Sihon, Matthijs M. Maas, Luke Kemp (2020), "Should Artificial Intelligence Governance be Centralised? Design Lessons from History", AIES '20, February 7-8, 2020, New York
310. Pindrop (2024), *How Singapore is Taking Action Against Deepfakes*, tại trang https://www.pindrop.com/article/singapore-taking-action-against-deepfakes/, [truy cập ngày 22/6/2026]
311. Public Relations Office, Cabinet Office, Government of Japan (2025), *Act on Promotion of Research and Development, and Utilization of Artificial Intelligence-related Technology Now in Full Effect*, tại trang https://www.gov-online.go.jp/hlj/en/november_2025/november_2025-08.html, [truy cập ngày 30/4/2026]
312. PwC China (2023), *Regulatory and legislation: China's Interim Measures for the Management of Generative Artificial Intelligence Services officially implemented*,

tại trang https://www.pwccn.com/en/tmt/interim-measures-for-generative-ai-services-implemented-aug2023.pdf?ref=keymakr.com, [truy cập ngày 20/2/2025]

313. R. Edward Freeman (1997), *Strategic Management: A Stakeholder Approach* (Pitman 1984); Ronald K. Mitchell, Bradley R. Agle và Donna J. Wood, "Toward a Theory of Stakeholder Identification and Salience", 22(4) Academy of Management Review 853.
314. Ranjan Sapkota et al. (2025), "AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges", *Information Fusion 126 (2025) 103599*.
315. Rathenau Institute (2017), *Human Rights in the Robot Age*. Strasbourg: Parliamentary Assembly of the Council of Europe, tại trang https://www.rathenau.nl/en/digitalisering/human-rights-robot-age, [truy cập ngày 1/1/2024]
316. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (*General Data Protection Regulation*), tại trang https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng, [truy cập ngày 6/9/2024]
317. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (*Artificial Intelligence Act*), tại trang https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689#cpt_I, [truy cập ngày 6/9/2024]
318. Rex Martinez (2019), "Artificial intelligence: Distinguishing between types and definitions", *Nevada Law Journal*, Vol.19:3, 1015-1041, tại trang https://scholars.law.unlv.edu/cgi/viewcontent.cgi?article=1799&context=nlj [truy cập ngày 5/1/2024]
319. Robert Alexy (2002), *A Theory of Constitutional Rights*, Julian Rivers tr, Oxford University Press.
320. Robert Baldwin & Julia Black (2010), "*Really Responsive Risk-Based Regulation", Law & Policy*, 32(2), p.181-213.
321. Robert Baldwin, Martin Cave, Martin Lodge (2012), *Understanding Regulation: Theory, Strategy, and Practice*, Oxford University Press, Oxford
322. Robert K. Yin (2017), Case Study Research and Applications: Design and Methods, SAGE Publications
323. Roger Brownsword, Eloise Scotford, Karen Yeung (ed.) (2017), *The Oxford Handbook of Law, Regulation and Technology*, Oxford University Press, Oxford.

324. Roger Brownsword, Morag Goodwin (2012), *Law and the Technologies of the Twenty-First Century: Text and Materials*, Cambridge University Press, Cambridge
325. Ryan Abbott (2020), *The Reasonable Robot: Artificial Intelligence and the Law*, Cambridge University Press, Cambridge.
326. Ryan Calo (2017), "Artificial Intelligence Policy: A Primer and Roadmap", *UC Davis Law Review* 51: p.399-435.
327. Sabine Gless, Helena Whalen-Bridge (ed.) (2024), *Human-Robot Interaction in Law and Its Narratives*, Cambridge University Press, Cambridge.
328. SAE (2021), SAE Levels of Driving Automation, tại trang https://www.sae.org/news/blog/sae-levels-driving-automation-clarity-refinements, [truy cập ngày 1/1/2024]
329. Sandra Wachter, Brent Mittelstadt, and Chris Russell (2018), "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR", *Harvard Journal of Law & Technology*, 31(2), p.841-887
330. Sebastian Raschka (2025), *Build a Large Language Model (From Scratch)*, Manning Publications Co. Shelter Island NY 11964.
331. Sheila Jasanoff (ed.) (2004), *States of Knowledge: The Co-production of Science and Social Order*, Routledge, London.
332. Sherry R. Arnstein (1969), "A Ladder of Citizen Participation", 35(4) Journal of the American Institute of Planners 216.
333. Shunyu Yao, et al. (2023), "ReAct: Synergizing Reasoning and Acting in Language Models", *arXiv*, tại trang https://arxiv.org/abs/2210.03629, [truy cập ngày 15/6/2026]
334. Sitki Selim Dolanay (2023), Chapter "Artificial Intelligence, Smart Robots, Types of Artificial Intelligence and a New Economic Order", In book: *An Overview on Business, Management and Economic* Research Vol.2 (pp.138-160), tại trang https://www.researchgate.net/publication/374104150_Artificial_Intelligence_Smart_Robots_Types_of_Artificial_Intelligence_and_a_New_Economic_Order [truy cập ngày 4/1/2024]
335. standards.ieee.org
336. Stanford Institute for Human-Centered AI (2025), *AI Index Report 2025*, tại trang https://hai.stanford.edu/ai-index/2025-ai-index-report, [truy cập ngày 15/6/2026]
337. State Council (2017), *Next Generation Artificial Intelligence Development Plan*, tại trang https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm, [truy cập ngày 14/6/2024]

338. Stuart J. Russell, Peter Norvig (2010), *Artificial Intelligence: A Modern Approach*, 3rd Edition, Prentice Hall.
339. Stuart Russell (2019), *Human Compatible: Artificial Intelligence and the Problem of Control*, Viking, New York.
340. Sun, Y., Sheng, D., Zhou, Z. et al (2024), "AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content", *Humanit Soc Sci Commun* 11, 1278 (2024). https://doi.org/10.1057/s41599-024-03811-x.
341. Texas Legislature Online (2025), *Responsible Artificial Intelligence Governance Act, tại trang* https://capitol.texas.gov/tlodocs/89R/analysis/html/HB00149S.htm, *[truy cập ngày 3/3/2026]*
342. Tom Krantz, Alexandra Jonker, *What is ground truth?*, tại tranghttps://www.ibm.com/think/topics/ground-truth, [truy cập ngày 10/3/2026]
343. Tshilidzi Marwala, Letlhokwa George Mpedi (2024), *Artificial Intelligence and the Law*, Palgrave Macmillan Singapore, Singapore.
344. The 116th United States Congress (2020), *AI in Government Act*, tại trang https://www.congress.gov/bill/116th-congress/house-bill/2575, [truy cập ngày 15/3/2026]
345. The 117th United States Congress (2022), H.R.4346 - *CHIPS and Science Act, tại trang* https://www.congress.gov/bill/117th-congress/house-bill/4346, *[truy cập ngày 3/3/2026]*
346. The 119th United States Congress (2025), *S.146 - TAKE IT DOWN Act*, tại trang https://www.congress.gov/bill/119th-congress/senate-bill/146, [truy cập ngày 3/3/2026]
347. The European Parliament (2017), *Civil Law Rules on Robotics*, P8_TA(2017)0051
348. The Government of Japan (2024), *The Hiroshima AI Process: Leading the Global Challenge to Shape Inclusive Governance for Generative AI*, tại trang https://www-japan-go-jp.translate.goog/kizuna/2024/02/hiroshima_ai_process.html?_x_tr_sl=en&_x_tr_tl=vi&_x_tr_hl=vi&_x_tr_pto=tc, [truy cập ngày 15/3/2026]
349. The White House (2025), *America's AI Action Plan, tại trang* https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf*, [truy cập ngày 15/3/2026]*
350. The White House (2025), *Ensuring a National Policy Framework for Artificial Intelligence, tại trang* https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/*, [truy cập ngày 3/3/2026]*

351. The White House (2025), *Removing Barriers to American Leadership in Artificial Intelligence, tại trang* https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/*, [truy cập ngày 15/3//2026]*
352. Themistoklis Tzimas (2021), *Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective*, Springer, Switzerland.
353. Thomas Weigend (2024), Forms of Robot Liability: Criminal Robots and Corporate Criminal Responsibility, in book Human-Robot Interaction in Law and Its Narratives, Cambridge University Press, Cambridge.
354. UN General Assembly (2024), *Enhancing International Cooperation on Capacity-Building of Artificial Intelligence*, tại trang https://digitallibrary.un.org/record/4054005?v=pdf, [truy cập ngày 30/12/2025]
355. UN General Assembly (2024), *Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development*, tại trang https://digitallibrary.un.org/record/4043244?v=pdf, [truy cập ngày 30/12/2025]
356. UNESCO (2021), *Recommendations on the Ethics of Artificial Intelligence*, tại trang https://www.unesco.org/en/artificial-intelligence/recommendation-ethics, [truy cập ngày 3/6/2024]
357. UNESCO (2023), *Readiness assessment methodology: a tool of the Recommendation on the Ethics of Artificial Intelligence*, tại trang https://www.unesco.org/ethics-ai/en/ram, [truy cập ngày 30/12/2025]
358. United Nations (2024), *Global Digital Compact*, tại trang https://www.un.org/digital-emerging-technologies/global-digital-compact, [truy cập ngày 30/12/2025]
359. United Nations (2024), *Governing AI for Humanity*, tại trang https://www.un-ilibrary.org/content/books/9789211067873, [truy cập ngày 30/12/2025]
360. United Nations Development Group (2003), *The Human Rights Based Approach to Development Cooperation: Towards a Common Understanding Among UN Agencies*, tại trang https://unsdg.un.org/resources/human-rights-based-approach-development-cooperation-towards-common-understanding-among-un, [truy cập ngày 30/4/2026]
361. United Nations Statistics Division (2025), SDG Indicator Metadata, tại trang https://unstats.un.org/sdgs/metadata/files/Metadata-05-0a-02.pdf, [truy cập ngày 30/4/2026]
362. United States Court of Appeals for the ninth circuit (2022*), hiQ Labs, Inc. v. LinkedIn Corporation,* tại trang

https://cdn.ca9.uscourts.gov/datastore/opinions/2022/04/18/17-16783.pdf, [truy cập ngày 10/2/2026]

363. Urs Gasser, Virgilio Almeida (2017), "A Layered Model for AI Governance", *IEEE Internet Computing*, 21(6), p.58-62.
364. Vinodkumar Prabhakaran, et al. (2022), *A Human Rights-Based Approach to Responsible AI*, arXiv, https://doi.org/10.48550/arXiv.2210.02667
365. Viswanath Venkatesh, et al. (2003), "User Acceptance of Information Technology: Toward a Unified View" *MIS Quarterly*, 27(3), p.425-478.
366. World Bank, "Establishing the PPP Framework", PPP Legal Resource Center, tại trang https://ppp.worldbank.org/PPP_Online_Reference_Guide/Establishing_PPP_Framework, [truy cập ngày 30/4/2026]
367. World Bank, "Legal Framework", Identification for Development - ID4D Practitioner's Guide, tại trang https://id4d.worldbank.org/guide/legal-framework, [truy cập ngày 30/4/2026]
368. Wesley Newcomb Hohfeld (1913), "Some Fundamental Legal Conceptions as Applied in Judicial Reasoning", 23 Yale Law Journal 16.
369. World Economic Forum (2017), *A robot has just been granted citizenship of Saudi Arabia*, tại trang https://www.weforum.org/stories/2017/10/a-robot-has-just-been-granted-citizenship-of-saudi-arabia/, [truy cập ngày 15/4/2026]

PHỤ LỤC

PHỤ LỤC 1

BẢNG TỔNG HỢP SO SÁNH KINH NGHIỆM QUỐC TẾ VỀ XÂY DỰNG KHUNG PHÁP LÝ AI

Chiều phân tích	Nội dung	LHQ/ UNESCO	OECD	EU	Hoa Kỳ	Trung Quốc	Nhật Bản	Hàn Quốc	Singapore
Mô hình khung pháp lý	**Tên mô hình**	Quản trị chuẩn mực phi ràng buộc	Nguyên tắc và công cụ đo lường	Tập trung toàn diện ex ante hard law	Phân tán đa tầng tự nguyện	Tích lũy theo công nghệ cụ thể	Agile governance luật mềm	Nhà nước phát triển cân bằng	Facilitative - công cụ tự đánh giá
	Văn bản trung tâm	Khuyến nghị Đạo đức AI (2021); Nghị quyết A/RES/78/311 (2024)	OECD AI Principles (2024, cập nhật từ 2019)	EU AI Act Quy định 2024/1689/EU	NIST AI RMF 1.0 (2023); EO 14179 (2025)	Biện pháp QLTM Dịch vụ AI Tạo sinh (8/2023); Quy định Deep Synthesis (1/2023)	Hướng dẫn AI cho Doanh nghiệp METI (2024)	AI Basic Act (2024, hiệu lực 1/2026)	AI Governance Framework (2020); AI Verify
	Tầng VBQPPL	Không có (soft law thuần túy)	Không có (principles-based)	Đầy đủ - 4 tầng rủi ro; chuỗi provider-deployer	Không có luật AI liên bang; quy định ngành từng phần	Quy định theo công nghệ cụ thể; chưa có luật tổng thể	Không có luật AI cứng; hướng dẫn ngành	Luật khung + quy định ngành	Không có luật AI; AI Governance Framework
	Tầng tiêu chuẩn kỹ thuật	Khuyến nghị kỹ thuật không chính thức	Phân loại hệ thống AI (2022)	Harmonised standards (CEN/CENELEC) tạo suy đoán tuân thủ	NIST AI RMF + Profiles theo lĩnh vực; affirmative defense TRAIGA	GB/T độc lập, không tương thích ISO/IEC	Viện dẫn ISO/IEC không ngày trong hướng dẫn	Tiêu chuẩn quốc gia + tiêu chuẩn ngành	AI Verify phản ánh sang ISO/IEC + OECD; hướng MRA
	Tầng luật mềm	Trọng tâm - 11 nguyên tắc + RAM	Trọng tâm - Principles + Observatory	Hỗ trợ - hướng dẫn EU AI Office; Bộ quy tắc GPAI	Trọng tâm - NIST AI RMF tự nguyện	Hạn chế - quy định cứng là chủ đạo	Trọng tâm - hướng dẫn ngành comply-or-explain	Hỗ trợ - hướng dẫn ủy ban AI	Trọng tâm - AI Governance Framework + AI Verify

Chiều phân tích	Nội dung	LHQ/ UNESCO	OECD	EU	Hoa Kỳ	Trung Quốc	Nhật Bản	Hàn Quốc	Singapore
Mô hình khung pháp lý	**Tầng án lệ/lẽ công bằng**	Nguyên tắc quốc tế tham chiếu	Không trực tiếp	CJEU đang phát triển	Tòa án tiểu bang tích cực; tranh chấp AI ngày càng nhiều	Beijing Internet Court: án lệ AI về nhân cách quyền (2021, 2024)	Tòa án dân sự/thương mại	Tòa án Hàn Quốc đang hình thành	Tòa án Singapore theo common law
	Tầng thiết chế	ITES; Working Groups UNESCO; OECD AI Policy Observatory	AI Policy Observatory toàn cầu	EU AI Office + cơ quan AI quốc gia + EU Database	NIST AI Safety Institute; AISIC; FTC; FDA; SEC theo ngành	CAC + 6 bộ ngành phối hợp; đăng ký thuật toán	AI Strategy Council; Study Groups theo bộ	National AI Commission (xuyên ngành)	IMDA + MAS + nhóm đồng thiết kế
	Cơ chế thích ứng	Báo cáo 4 năm + hướng dẫn cập nhật	Observatory liên tục; cập nhật Principles định kỳ	Rà soát bắt buộc 3 năm (Đ.112); cập nhật phụ lục rút gọn	Rà soát NIST 2028; cập nhật sắc lệnh theo nhiệm kỳ	Cập nhật quy định theo công nghệ mới xuất hiện	Cập nhật hướng dẫn 12-18 tháng không qua quy trình pháp lý	Ủy ban AI rà soát hàng năm; cập nhật danh mục rút gọn	Cập nhật AI Verify 3 lần/năm; hướng MRA ASEAN
	Nhà nước	UNESCO Secretariat + 193 quốc gia	OECD Secretariat + 47 thành viên	EU Commission + Nghị viện + Hội đồng; AI HLEG 52 chuyên gia	NIST (tiêu chuẩn kỹ thuật); OMB (khu vực công); Quốc hội (hạn chế)	CAC + 6 bộ ngành phối hợp ký ban hành	METI + MIC + Bộ Y tế + Hội đồng AI Quốc gia	MSIT + Tiểu ban QH đặc biệt về AI	IMDA + MAS + MOH
	Cơ quan lập pháp	Phê chuẩn khuyến nghị (thủ tục đồng thuận)	Phê duyệt Principles ở cấp Bộ trưởng	Nghị viện bổ sung điều khoản GPAI quan trọng; đồng quyết định	Quốc hội: Ủy ban lưỡng đảng AI (12/2024); hạn chế lập pháp trực tiếp	Không có vai trò cơ quan lập pháp độc lập	Không có can thiệp lập pháp; phê chuẩn ngân sách	Tiểu ban QH đặc biệt tham gia thiết kế; đồng thuận đa đảng	Không có vai trò lập pháp AI riêng

Chiều phân tích	Nội dung	LHQ/ UNESCO	OECD	EU	Hoa Kỳ	Trung Quốc	Nhật Bản	Hàn Quốc	Singapore
Chủ thể xây dựng	**Khu vực tư nhân**	Xã hội dân sự toàn cầu tham vấn; không có đồng thiết kế thực chất	Business at OECD; TUAC người lao động; xã hội dân sự	Tham vấn rộng 1.300+ phản hồi; AlgorithmWatch, Access Now vận động	AISIC (200+ tổ chức); cam kết tự nguyện BigTech; ảnh hưởng lobby lớn	Doanh nghiệp lớn (Alibaba, Baidu, ByteDance) tham vấn ngắn	Hội đồng tư vấn đa tầng; Study Groups; đồng thuận ngành	Startup tham vấn đặc biệt về miễn trừ tỷ lệ; dẫn đến thay đổi thiết kế	60+ doanh nghiệp đồng thiết kế AI Verify; nhóm thường trực cập nhật
	Học thuật	Nghiên cứu nền UNESCO; peer review Principles	Nghiên cứu kinh tế số; phân tích chính sách	AI HLEG; các đại học châu Âu nghiên cứu nền	Brookings, Stanford HAI, MIT; NIST hợp tác học thuật	Học viện Khoa học Xã hội TQ soạn mô hình luật; TC260 tiêu chuẩn	RIKEN, NII tư vấn kỹ thuật	STEPI nghiên cứu chi phí tuân thủ; viện công lập	Học thuật Singapore + NTU tham gia AI Verify
	Tổ chức quốc tế	Phối hợp đa phương; định hình chuẩn mực	Phân tích so sánh; hỗ trợ kỹ thuật	"Brussels Effect" xuất khẩu chuẩn mực	NIST: ảnh hưởng qua tiêu chuẩn kỹ thuật toàn cầu	Đóng góp cho ISO/IEC SC 42; TC260 song song	Tham gia OECD; dẫn dắt G7 về AI (2023)	Tham gia OECD; đóng góp AI Basic Act như mô hình	Chia sẻ AI Verify với ASEAN; đàm phán MRA
	Điểm mạnh chủ thể	Tính chính đáng toàn cầu; đại diện địa lý đầy đủ	Chất lượng phân tích; uy tín học thuật	Đa chủ thể có cấu trúc; trách nhiệm giải trình rõ	Uy tín kỹ thuật NIST; tính độc lập khỏi chính trị	Điều phối đa cơ quan nhanh; CAC quyết đoán	Đồng thuận ngành cao; Study Groups chuyên sâu	Tiểu ban QH kỹ thuật; bảo vệ startup thực chất	Đồng thiết kế thực sự; nhóm thường trực cập nhật
	Điểm yếu chủ thể	Thiếu chuyên gia kỹ thuật AI; đồng thuận loãng nội dung	Đại diện nước đang phát triển hạn chế	Quy trình 6 năm; lobby BigTech ảnh hưởng lớn	Quốc hội bế tắc; xáo trộn chính sách theo nhiệm kỳ	Thiếu tiếng nói SME; không có thiết chế độc lập	Quy trình đồng thuận chậm; ít minh bạch	Còn mới, chưa kiểm chứng hiệu quả dài hạn	Phụ thuộc văn hóa tuân thủ cao; khó áp dụng ở nơi khác

Chiều phân tích	Nội dung	LHQ/ UNESCO	OECD	EU	Hoa Kỳ	Trung Quốc	Nhật Bản	Hàn Quốc	Singapore
Quy trình xây dựng	**Giai đoạn chẩn đoán**	Nghiên cứu nền ủy thác + tham vấn khu vực; 2019-2020	AIGO 50 chuyên gia 18 tháng; phân tích dữ liệu AI quốc gia	AI HLEG 2018-2020; Sách trắng tham vấn 2020	NIST RFI; nhiều hội thảo; phân tích không gian vấn đề	Nội bộ cơ quan + ý kiến doanh nghiệp lớn ngắn	Study Groups thường xuyên; khảo sát thực tiễn hàng năm	STEPI nghiên cứu chi phí; phân tích so sánh quốc tế	Không tương ứng; thay bằng kiểm thử thực tiễn
	Giai đoạn thiết kế kiến trúc	Cấu trúc 4 giá trị → 11 nguyên tắc → 10 chính sách	Cấu trúc 5 nguyên tắc → 5 khuyến nghị chính sách	Phụ lục kỹ thuật → định nghĩa chức năng; kiến trúc 4 tầng rủi ro	Cấu trúc Govern-Map-Measure-Manage; voluntary-rights-preserving	Không có kiến trúc tổng thể; tích lũy theo công nghệ	Không có kiến trúc tổng thể; cập nhật liên tục	Cân bằng rủi ro-hệ sinh thái-niềm tin trong một luật	IMDA xác định nguyên tắc; doanh nghiệp điền nội dung
	Giai đoạn tạo lập các thành tố	4 vòng tham vấn khu vực; đồng thuận 193 quốc gia	Thương lượng 47 thành viên + đối tác	Soạn thảo 2021; đàm phán trilogue 2022-2023; ban hành 2024	NIST 2 vòng dự thảo công khai; RFI; workshops	Dự thảo 1 tháng; lấy ý kiến 30 ngày; ban hành	Hội đồng nghiên cứu → hướng dẫn; nemawashi đồng thuận trước	Lập pháp 3 năm; điều trần kỹ thuật; sửa đổi sau tham vấn startup	Kiểm thử AI Verify 18 tháng; đồng thiết kế nhóm thường trực
	Giai đoạn vận hành/ thích ứng	Báo cáo quốc gia 4 năm; working groups chuyên đề	Observatory liên tục; cập nhật Principles theo công nghệ	Rà soát 3 năm (Đ.112); Ủy ban cập nhật danh mục; AI Office giám sát	NIST rà soát 2028; sắc lệnh cập nhật theo nhiệm kỳ	Cập nhật theo công nghệ mới; hướng đến luật tổng thể	Cập nhật hướng dẫn 12-18 tháng; không qua quy trình pháp lý	Ủy ban AI rà soát hàng năm; cập nhật danh mục rút gọn	Cập nhật AI Verify 3 lần/năm; đàm phán MRA ASEAN

Chiều phân tích	Nội dung	LHQ/ UNESCO	OECD	EU	Hoa Kỳ	Trung Quốc	Nhật Bản	Hàn Quốc	Singapore
	Tính minh bạch quy trình	Cao - biên bản họp công khai; tham vấn khu vực	Cao - tài liệu làm việc công bố; báo cáo phản hồi	Rất cao - toàn bộ tài liệu công khai; báo cáo phản hồi 1.300+	Cao - NIST dự thảo công khai; workshops; comments	Trung bình - lấy ý kiến ngắn; không báo cáo phản hồi chi tiết	Thấp - đồng thuận qua kênh phi chính thức (nemawashi)	Cao - điều trần QH công khai; báo cáo phản hồi tham vấn	Cao - kiểm thử mở; mã nguồn mở AI Verify
	Thời gian từ khởi động đến ban hành	2 năm (2019-2021)	18 tháng (mỗi chu kỳ cập nhật)	6 năm (2018-2024)	2 năm/AI RMF 1.0; cập nhật liên tục	3 tháng/văn bản; nhanh nhất	12-18 tháng/cập nhật hướng dẫn	3 năm (2022-2024)	18 tháng/AI Verify v1.0; cập nhật 6 tháng
Kỹ thuật xây dựng	**RIA**	RAM tự đánh giá 5 chiều có hướng dẫn; thiếu kiểm tra chéo	Phân tích so sánh quốc gia; Observatory đo lường	RIA 400+ trang; 3 lựa chọn chính sách; phân tích chi phí theo quy mô	RFI + hội thảo → dự thảo → rà soát 2028; OMB bắt buộc cơ quan liên bang	Tham vấn 30 ngày; không có RIA đầy đủ; ưu tiên tốc độ	Khảo sát thực tiễn hàng năm 800+ doanh nghiệp; RIA liên tục sau ban hành	STI nghiên cứu chi phí tuân thủ doanh nghiệp trước soạn thảo	Field RIA: kiểm thử thực tiễn 60+ doanh nghiệp 18 tháng trước ban hành
	Định nghĩa chức năng	Chức năng rộng; OECD 2024 lần đầu tách biệt "AI system" vs "agentic AI"	Chức năng theo tác động; định nghĩa agentic AI theo chuỗi hành động	Chuyển từ Phụ lục liệt kê (2021) → định nghĩa chức năng (2024); bài học từ thất bại	AI RMF: chức năng rộng; NIST 2026 khởi động tiêu chuẩn cho AI agents tự chủ	Liệt kê tên công nghệ → vấn đề khi xuất hiện kiến trúc mới	Theo tác động ("ảnh hưởng đến con người không?"); không cần kiến trúc kỹ thuật	Chức năng + thẩm quyền diễn giải chính thức của Ủy ban AI	Tích hợp định nghĩa vào bộ câu hỏi AI Verify; áp dụng nhất quán

Chiều phân tích	Nội dung	LHQ/ UNESCO	OECD	EU	Hoa Kỳ	Trung Quốc	Nhật Bản	Hàn Quốc	Singapore
Kỹ thuật xây dựng	**Ủy quyền linh hoạt**	Báo cáo định kỳ 4 năm + hướng dẫn thực hiện cập nhật	Observatory cập nhật liên tục; không có ủy quyền cứng	3 cấp: Quy định gốc → Văn bản ủy quyền Ủy ban → Hướng dẫn AI Office; sunset 3 năm	Sắc lệnh hành pháp linh hoạt nhưng bất ổn theo nhiệm kỳ; không có sunset	Ủy quyền theo cơ quan chủ trì ngành; thiếu điều phối tổng thể	Không qua quy trình pháp lý; Study Groups cập nhật 12-18 tháng	Ủy ban AI diễn giải định nghĩa (30 ngày); cập nhật danh mục rút gọn	Ủy quyền cho nhóm đồng thiết kế thường trực; IMDA phê duyệt 30 ngày
	Phân định chuỗi giá trị	Không áp dụng	Không áp dụng trực tiếp	Provider-deployer đầy đủ nhất; quy tắc "vượt phạm vi kiểm định trở thành provider"; Đ.25, 28	Phân định theo ngành chuyên biệt; không có mô hình chuỗi giá trị thống nhất	Nhà cung cấp dịch vụ chịu toàn bộ; không phân định provider-deployer	Phân định theo ngành, không theo chuỗi giá trị	Provider-deployer + chứng nhận nền tảng giảm gánh nặng triển khai cho SME	Tự xác định vị trí qua bộ câu hỏi AI Verify; hệ thống hiển thị nghĩa vụ tương ứng
	Miễn trừ tỷ lệ	Differentiated implementation theo năng lực quốc gia	Không áp dụng trực tiếp	Thiếu - chỉ điều chỉnh nhỏ cho SME; chi phí tuân thủ cao gây loại trừ	Ngưỡng cao ($500M doanh thu SB 53 California); TRAIGA safe harbor NIST RMF	Theo phạm vi "cho công chúng"; không có ngưỡng quy mô doanh nghiệp	Comply-or-explain hoàn toàn linh hoạt; không có ngưỡng cứng	Ba tầng ngưỡng cụ thể (siêu nhỏ → nhỏ → lớn) sau khi startup phản hồi	Tự đánh giá AI Verify cho mọi quy mô; spot-check rủi ro; không ngưỡng cứng
	Luật hóa tiêu chuẩn	Viện dẫn tiêu chuẩn như "thực hành tốt" không tạo suy đoán	Phân loại hệ thống AI như công cụ kỹ thuật tham chiếu	Harmonised standards tạo "presumption of conformity"	Affirmative defense TRAIGA Texas (mạnh hơn suy đoán	GB/T độc lập; TC260 phối hợp tiêu chuẩn song song với quy	Viện dẫn ISO/IEC không ngày trong hướng dẫn; "suy	Tiêu chuẩn ngành + chứng nhận nền tảng tạo gánh nặng	AI Verify phản ánh sang OECD + ISO/IEC; đàm phán MRA

Chiều phân tích	Nội dung	LHQ/ UNESCO	OECD	EU	Hoa Kỳ	Trung Quốc	Nhật Bản	Hàn Quốc	Singapore
				chính thức (Đ.40); viện dẫn không ngày	tuân thủ); không có ở liên bang	định; không tương thích ISO/IEC	đoán mềm" qua comply-or-explain	giảm cho deployer	với EU, Hàn, Úc
Điểm mạnh nổi bật		Tính chính đáng toàn cầu; bao trùm địa lý; RAM thực tiễn	Uy tín phân tích; Observatory so sánh quốc gia; cập nhật kịp thời	Kiến trúc toàn diện nhất; chuỗi trách nhiệm rõ; Brussels Effect; sunset clause	AI RMF thực tiễn nhất thế giới; affirmative defense kỹ thuật; AISIC đổi mới	Tốc độ ban hành nhanh nhất; đăng ký thuật toán; gắn nhãn deepfake sớm nhất; phối hợp tiêu chuẩn song song	Linh hoạt cập nhật cao nhất; chi phí thể chế thấp; comply-or-explain thực tiễn	Cân bằng rủi ro-hệ sinh thái tốt nhất; miễn trừ tỷ lệ đầy đủ; Ủy ban AI xuyên ngành	Đồng thiết kế thực sự; AI Verify sáng tạo nhất; hướng MRA; cập nhật nhanh nhất
Điểm yếu cơ cấu		Không ràng buộc; loãng nội dung vì đồng thuận 193 nước; báo cáo 4 năm quá chậm	Thiếu đại diện nước đang phát triển; không có cơ chế thực thi	Chi phí tuân thủ rất cao; 6 năm xây dựng; ít bảo vệ SME; phức tạp hóa quá mức	Bất ổn chính sách theo nhiệm kỳ (Biden→Trump); thiếu bảo vệ quyền công dân; phân tán tiểu bang tốn kém	Định nghĩa công nghệ cụ thể lỗi thời; GB/T không tương thích quốc tế; thiếu miễn trừ tỷ lệ	Thiếu ràng buộc; phụ thuộc văn hóa tuân thủ cao; thiếu minh bạch quy trình	Còn quá mới; chưa có dữ liệu hiệu quả dài hạn; có thể xáo trộn nếu thay đổi chính phủ	Phụ thuộc văn hóa tuân thủ cao; khó nhân rộng cho nước khác; thiếu ràng buộc rủi ro cao
Mức phù hợp với Việt Nam		★★★☆☆ Chuẩn mực tham chiếu tối thiểu	★★★★☆ Phương pháp luận và công cụ đo lường	★★★☆☆ Học kiến trúc, không sao chép toàn bộ	★★☆☆☆ Học công cụ kỹ thuật; không học mô hình tổng thể	★★☆☆☆ Thận trọng có chọn lọc	★★★★☆ Cao - luật + luật mềm ngành + sandbox	★★★★★ Rất cao - bối cảnh tương đồng nhất	★★★★☆ Cao - công cụ tự đánh giá + đồng thiết kế

PHỤ LỤC 2
BỘ CÔNG CỤ TỰ ĐÁNH GIÁ VN-AI VERIFY

GIỚI THIỆU KHÁI QUÁT VN-AI VERIFY

* Mức độ kế thừa AI Verify của Singapore

Việc đề xuất VN-AI Verify "học hỏi có chọn lọc" từ AI Verify Singapore cần được cụ thể hóa thành bốn cấp độ kế thừa, mỗi cấp độ có hàm ý pháp lý và kỹ thuật khác nhau.

Cấp độ 1, Kế thừa về triết lý thiết kế (kế thừa hoàn toàn): nguyên tắc "công cụ tự đánh giá do khu vực tư nhân đồng thiết kế với nhà nước giữ vai trò phê duyệt" được kế thừa nguyên vẹn từ Singapore.

Cấp độ 2, Kế thừa về cấu trúc kỹ thuật (kế thừa có điều chỉnh cấu trúc): cấu trúc ba module (phân loại rủi ro - kiểm tra tuân thủ - đánh giá thiên lệch) không sao chép cấu trúc gốc của AI Verify Singapore (vốn tổ chức theo 11 nguyên tắc quản trị AI theo khung Model AI Governance Framework) mà là cấu trúc được thiết kế riêng cho Việt Nam, phản ánh trực tiếp sang điều khoản cụ thể của Luật AI 2025 và Nghị định 142 - một cấu trúc mà bản thân AI Verify Singapore không có (vì Singapore không có luật AI chuyên biệt). Đây là điểm khác biệt cấu trúc quan trọng nhất cần được nêu rõ: AI Verify Singapore là công cụ kiểm thử kỹ thuật đối chiếu với khung nguyên tắc mềm; VN-AI Verify là công cụ tự đánh giá tuân thủ đối chiếu với nghĩa vụ pháp lý cứng. Hai công cụ có chức năng pháp lý khác nhau về bản chất dù cùng hình thức "bảng câu hỏi tự đánh giá".

Cấp độ 3, Kế thừa về công cụ kỹ thuật nền (kế thừa có chọn lọc từng phần): AI Verify Toolkit của Singapore, từ năm 2023, đã được công bố dưới dạng mã nguồn mở trên GitHub (theo AI Verify Foundation). Đây là tài sản kỹ thuật có thể được Việt Nam tiếp cận và điều chỉnh hợp pháp theo giấy phép mã nguồn mở tương ứng (cần được nhóm phát triển VN-AI Verify xác minh cụ thể loại giấy phép và điều kiện sử dụng khi triển khai), đặc biệt các thư viện kiểm thử kỹ thuật về thiên lệch thuật toán (tương ứng Module C của VN-AI Verify) là bộ phận có khả năng kế thừa trực tiếp cao nhất về mặt kỹ thuật thuần túy (thuật toán đo lường thiên lệch không phụ thuộc vào hệ thống pháp luật quốc gia), trong khi Module A và Module B (gắn với quy định pháp luật Việt Nam) không thể kế thừa trực tiếp mà phải xây dựng mới hoàn toàn.

Cấp độ 4, Kế thừa về quy trình phát triển (kế thừa có điều chỉnh về mức độ ràng buộc): quy trình kiểm thử thực địa 18 tháng với 60+ doanh nghiệp tình nguyện của Singapore được kế thừa về mặt phương pháp (đồng thiết kế, kiểm thử lặp lại trước khi ra mắt chính thức) nhưng được điều chỉnh bổ sung yếu tố ràng buộc pháp lý (suy đoán tuân thủ, affirmative defense) mà AI Verify Singapore thuần túy tự nguyện, không gắn với luật cứng không có.

Như vậy, VN-AI Verify kế thừa Singapore ở mức triết lý (100%) và phương pháp phát triển (kế thừa có điều chỉnh), kế thừa có chọn lọc ở mức công cụ kỹ thuật (chỉ với bộ phận không gắn quy định pháp luật cụ thể, như module kiểm thử thiên lệch), không kế thừa mà xây dựng mới hoàn toàn ở mức cấu trúc pháp lý (ánh xạ điều khoản, cơ chế suy đoán tuân thủ).

*** Chủ thể sở hữu VN-AI Verify**

Về quyền sở hữu trí tuệ đối với công cụ, cần phân định rõ hai lớp quyền sở hữu khác nhau, tương ứng với cấu trúc phát triển đồng thiết kế công-tư.

Quyền sở hữu đối với khung pháp lý (Module A, B) là bộ phận do Bộ Khoa học và Công nghệ chủ trì xây dựng nội dung (phản ánh câu hỏi sang điều khoản pháp luật là hành vi diễn giải và áp dụng pháp luật, không thể giao cho chủ thể tư nhân sở hữu độc quyền vì sẽ tạo ra vấn đề về tính chính đáng công quyền). Quyền sở hữu đối với nội dung này thuộc về Nhà nước, do Bộ Khoa học và Công nghệ đại diện quản lý, tương tự quyền sở hữu nhà nước đối với văn bản hướng dẫn pháp luật khác.

Quyền sở hữu đối với mã nguồn kỹ thuật của công cụ (nền tảng phần mềm, thuật toán kiểm thử kỹ thuật) được đề xuất công bố theo giấy phép mã nguồn mở (như đã xác lập tại lộ trình phát triển, "phiên bản 1.0 mã nguồn mở, miễn phí"). Theo mô hình này, không có chủ thể tư nhân hay nhà nước nào giữ độc quyền sở hữu trí tuệ đối với mã nguồn, mà mã nguồn thuộc về cộng đồng sử dụng theo điều kiện giấy phép đã chọn (cần được Bộ Khoa học và Công nghệ quyết định cụ thể loại giấy phép mã nguồn mở áp dụng tại thời điểm công bố chính thức).

Về chủ thể được thừa nhận là "chủ sở hữu chính thức" trên phương diện hành chính, Bộ Khoa học và Công nghệ là chủ thể chính thức đứng tên ban hành và công bố VN-AI Verify tương tự vai trò của IMDA đối với AI Verify tại Singapore.

*** Chủ thể vận hành VN-AI Verify**

Cần phân biệt ba lớp vận hành khác nhau, tránh nhầm lẫn "vận hành công cụ" với "vận hành hạ tầng kỹ thuật" và "vận hành nội dung".

Lớp vận hành hạ tầng kỹ thuật: đơn vị được Bộ Khoa học và Công nghệ giao nhiệm vụ hoặc ký hợp đồng dịch vụ công nghệ thông tin để duy trì máy chủ, giao diện web/ứng dụng của công cụ tự đánh giá. Đây là chức năng kỹ thuật thuần túy, có thể giao cho đơn vị sự nghiệp công lập thuộc Bộ hoặc đơn vị cung cấp dịch vụ công nghệ thông tin theo cơ chế đấu thầu, không đòi hỏi thẩm quyền pháp lý đặc biệt.

Lớp vận hành nội dung và cập nhật: Cục Công nghệ số và AI của Bộ Khoa học và Công nghệ chịu trách nhiệm cập nhật nội dung phản ánh pháp lý khi Luật AI 2025, Nghị định 142 hoặc Danh mục rủi ro cao được sửa đổi theo đúng chu kỳ cập nhật 90 ngày kể từ ngày văn bản sửa đổi có hiệu lực đã đề xuất tại phần nguyên tắc thiết kế bộ câu hỏi.

Lớp tích hợp kết quả vào hệ thống quản lý nhà nước: kết quả VN-AI Verify sau khi doanh nghiệp hoàn thành được tích hợp và lưu trữ trong AMIS. Đây là chức năng kết nối giữa hai hệ thống, không phải chức năng "vận hành" riêng của VN-AI Verify mà là điểm giao thoa giữa hai công cụ.

*** Chủ thể chịu trách nhiệm**

Trách nhiệm về độ chính xác và cập nhật của nội dung phản ánh pháp lý thuộc về Bộ Khoa học và Công nghệ, với tư cách chủ thể phê duyệt chính thức nội dung phản ánh tương đương. Nếu một câu hỏi trong VN-AI Verify dẫn sai hoặc lỗi thời so với quy định pháp luật hiện hành, trách nhiệm hành chính về việc cập nhật chậm trễ thuộc về đơn vị đầu mối được giao nhiệm vụ, không thuộc về doanh nghiệp sử dụng công cụ.

Trách nhiệm về độ tin cậy của kết quả tự đánh giá do doanh nghiệp cung cấp thuộc về chính doanh nghiệp thực hiện tự đánh giá, Bộ Khoa học và Công nghệ không xác nhận

hay bảo đảm tính trung thực của thông tin doanh nghiệp tự khai báo khi thực hiện VN-AI Verify. Đây là căn cứ trực tiếp cho việc thiết kế điều khoản suy đoán tuân thủ đã đề xuất phải có khả năng bác bỏ chính vì trách nhiệm về độ trung thực của nội dung khai báo thuộc về doanh nghiệp, không phải Nhà nước, nên suy đoán tuân thủ phải là suy đoán có thể bác bỏ, không phải suy đoán tuyệt đối.

Trách nhiệm về hạ tầng kỹ thuật và bảo mật vận hành công cụ thuộc về đơn vị vận hành hạ tầng kỹ thuật theo hợp đồng dịch vụ hoặc nhiệm vụ được giao bao gồm trách nhiệm bảo đảm an toàn thông tin đối với dữ liệu doanh nghiệp nhập vào công cụ trong quá trình tự đánh giá.

NỘI DUNG VN-AI VERIFY

I. Module A: Phân loại rủi ro hệ thống AI

Cơ sở pháp lý: Điều 9 Luật AI 2025 (Luật số 134/2025/QH15) và Điều 6 Nghị định 142/2026/NĐ-CP. Module vận hành theo cơ chế phân nhánh, câu trả lời "Có" tại câu A1 đến A6 dẫn đến phân loại rủi ro cao ngay lập tức mà không cần trả lời tiếp.

Nhóm A.0 - Xác định phạm vi áp dụng

A.0.1. Hệ thống của tổ chức/doanh nghiệp bạn có phải là "hệ thống AI" theo Điều 2 Luật AI 2025 không, tức là hệ thống máy tính sử dụng mô hình học máy, mạng nơ-ron nhân tạo, xử lý ngôn ngữ tự nhiên, thị giác máy tính hoặc phương pháp suy luận tự động tạo ra đầu ra (dự đoán, khuyến nghị, quyết định, nội dung) ảnh hưởng đến môi trường thực tế hoặc ảo?

(Nếu Không → Hệ thống không thuộc phạm vi điều chỉnh của Luật AI 2025. Kết thúc Module A)

A.0.2. Hệ thống có thuộc một trong các trường hợp được miễn trừ không: hệ thống AI phục vụ hoạt động nội bộ không tiếp xúc với bên ngoài tổ chức; hệ thống AI trong giai đoạn nghiên cứu và phát triển chưa triển khai thương mại; hoặc hệ thống AI thuộc danh mục bí mật nhà nước, quốc phòng, an ninh theo quy định của Chính phủ?

(Nếu Có → Hệ thống thuộc chế độ miễn trừ. Ghi nhận loại miễn trừ và kết thúc Module A)

A.0.3. Tổ chức/doanh nghiệp của bạn đóng vai trò nào đối với hệ thống AI này? (Chọn tất cả các vai trò áp dụng)

Lựa chọn: Nhà phát triển mô hình AI nền tảng / Nhà cung cấp hệ thống AI (đưa hệ thống ra thị trường) / Bên triển khai (tích hợp và vận hành hệ thống của nhà cung cấp khác vào sản phẩm/dịch vụ của mình) / Người dùng cuối tổ chức (sử dụng hệ thống cho mục đích nội bộ có tác động đến cá nhân bên ngoài).

Ghi chú: Một tổ chức có thể có nhiều vai trò đồng thời. Nghĩa vụ khác nhau theo từng vai trò sẽ được hiển thị tại Module B.

Nhóm A.1 - Kiểm tra điều kiện cấm tuyệt đối (Rủi ro không thể chấp nhận)

Câu trả lời "Có" cho bất kỳ câu nào từ A.1.1 đến A.1.6 → Hệ thống thuộc danh mục bị cấm theo Điều 9.1 Luật AI 2025. Tổ chức phải dừng triển khai và không tiếp tục Module A.

A.1.1. Hệ thống có sử dụng kỹ thuật dưới ngưỡng nhận thức tác động đến hành vi người dùng mà họ không nhận thức được, theo cách gây hại cho họ hoặc người khác không?

A.1.2. Hệ thống có khai thác các điểm yếu đặc biệt của một nhóm người cụ thể (người cao tuổi suy giảm nhận thức, trẻ em, người có rối loạn tâm thần) để tác động đến hành vi của họ theo cách gây hại không?

A.1.3. Hệ thống có thực hiện chấm điểm xã hội đối với cá nhân dựa trên hành vi hoặc đặc điểm cá nhân và sử dụng điểm số đó để tước đoạt quyền lợi không liên quan đến bối cảnh thu thập dữ liệu không?

A.1.4. Hệ thống có sử dụng nhận dạng cảm xúc trong môi trường công sở hoặc giáo dục để đánh giá, phân loại hoặc ra quyết định ảnh hưởng đến quyền lợi của người lao động hoặc học sinh không?

A.1.5. Hệ thống có thực hiện phân loại cá nhân theo đặc điểm sinh trắc học suy luận được (chủng tộc, quan điểm chính trị, tôn giáo, khuynh hướng tình dục) nhằm mục đích phân biệt đối xử không?

A.1.6. Hệ thống có phân tích ảnh khuôn mặt thu thập từ không gian công cộng một cách ngẫu nhiên để xây dựng cơ sở dữ liệu nhận diện mà không có căn cứ pháp lý cụ thể không?

Nhóm A.2 - Kiểm tra điều kiện rủi ro cao

Câu trả lời "Có" cho bất kỳ câu nào từ A.2.1 đến A.2.12 → Hệ thống được phân loại Rủi ro cao. Chuyển sang Module B với chế độ nghĩa vụ đầy đủ.

A.2.1. Hạ tầng quan trọng Hệ thống AI có được sử dụng để vận hành, điều khiển hoặc ra quyết định ảnh hưởng trực tiếp đến hoạt động của hệ thống điện, nước, giao thông, mạng viễn thông hoặc cơ sở hạ tầng tài chính quốc gia không?

A.2.2. Giáo dục và đào tạo Hệ thống AI có được sử dụng để chấm điểm bài thi, đánh giá năng lực học sinh/sinh viên, hoặc ra quyết định ảnh hưởng đến khả năng tiếp cận cơ hội giáo dục hoặc đào tạo nghề của cá nhân không?

A.2.3. Tuyển dụng và quản lý nhân sự Hệ thống AI có được sử dụng để sàng lọc hồ sơ ứng viên, xếp hạng ứng viên, đưa ra quyết định tuyển dụng, phân công công việc, đánh giá hiệu suất có ảnh hưởng đến lương thưởng hoặc chấm dứt hợp đồng lao động không?

A.2.4. Dịch vụ thiết yếu Hệ thống AI có tự động đánh giá mức tín dụng, xét duyệt khoản vay, phê duyệt hợp đồng bảo hiểm, hoặc quyết định khả năng tiếp cận dịch vụ tài chính thiết yếu của cá nhân không?

A.2.5. Y tế Hệ thống AI có được sử dụng để chẩn đoán bệnh, hỗ trợ quyết định lâm sàng có ảnh hưởng trực tiếp đến phác đồ điều trị, hoặc vận hành thiết bị y tế tiếp xúc với bệnh nhân không?

A.2.6. Thực thi pháp luật và tư pháp Hệ thống AI có được sử dụng bởi cơ quan công an, viện kiểm sát hoặc tòa án để đánh giá mức độ nguy hiểm của cá nhân, hỗ trợ điều tra, hoặc cung cấp thông tin ảnh hưởng đến quyết định tố tụng không?

A.2.7. Dịch vụ công và quản lý hành chính Hệ thống AI có tự động ra quyết định về quyền lợi phúc lợi xã hội, trợ cấp, cấp phép, hoặc các quyền hành chính ảnh hưởng đến cá nhân không?

A.2.8. Quản lý biên giới và di trú Hệ thống AI có được sử dụng để đánh giá rủi ro trong kiểm soát biên giới, xét duyệt thị thực hoặc quyết định có liên quan đến di trú không?

A.2.9. Ảnh hưởng đến nhóm dân số lớn Hệ thống AI có tạo ra quyết định tự động hoặc bán tự động tác động đến quyền lợi pháp lý hoặc điều kiện sống của hơn 50.000 người thường xuyên tại Việt Nam không?

A.2.10. Mức độ tự chủ cao Hệ thống có hoạt động theo mô hình AI agent - tự chủ lập kế hoạch và thực thi chuỗi hành động không cần phê duyệt từng bước của con người - trong môi trường có tác động đến người thật hoặc tài sản thật không?

A.2.11. Sinh trắc học Hệ thống có xử lý dữ liệu sinh trắc học (khuôn mặt, giọng nói, vân tay, mống mắt, dáng đi) để xác minh hoặc nhận diện danh tính trong bối cảnh có hậu quả pháp lý hoặc kinh tế với người được nhận diện không?

A.2.12. Tích hợp vào sản phẩm rủi ro cao Hệ thống AI có được tích hợp là thành phần không thể tách rời của một sản phẩm hoặc dịch vụ đã được pháp luật chuyên ngành xếp vào loại rủi ro cao (thiết bị y tế loại II, III; phương tiện tự lái; hệ thống kiểm soát công nghiệp) không?

Nhóm A.3 - Đánh giá rủi ro trung bình

Được hỏi khi tất cả câu A.2 đều trả lời "Không". Câu trả lời "Có" cho bất kỳ câu nào từ A.3.1 đến A.3.5 → Rủi ro trung bình. Câu trả lời "Không" cho tất cả → Rủi ro thấp.

A.3.1. Hệ thống AI có tiếp xúc trực tiếp với người dùng cuối theo cách người dùng có thể nhầm tưởng là đang tương tác với con người không (chatbot, trợ lý ảo, deepfake có chủ ý giải trí)?

A.3.2. Hệ thống AI có tạo ra nội dung văn bản, hình ảnh, âm thanh hoặc video tổng hợp được phân phối công khai và có khả năng ảnh hưởng đến nhận thức của người dùng về thực tế không?

A.3.3. Hệ thống AI có đưa ra khuyến nghị cá nhân hóa (nội dung, sản phẩm, dịch vụ) có khả năng ảnh hưởng đến quyết định tiêu dùng hoặc quan điểm của người dùng theo quy mô lớn không?

A.3.4. Hệ thống AI có thu thập và xử lý dữ liệu hành vi của người dùng theo thời gian thực để điều chỉnh trải nghiệm theo cách có thể tác động đến hành vi không?

A.3.5. Hệ thống AI có liên quan đến xử lý dữ liệu cá nhân nhạy cảm theo Luật Bảo vệ dữ liệu cá nhân 2025 (sức khoẻ, tài chính, vị trí, niềm tin tôn giáo, quan điểm chính trị) dù không thuộc các danh mục rủi ro cao ở Nhóm A.2 không?

Kết quả Module A

Sau khi hoàn thành Module A, hệ thống xuất ra: (i) kết quả phân loại rủi ro kèm căn cứ điều khoản; (ii) danh sách nghĩa vụ áp dụng theo vai trò và mức rủi ro; và (iii) mã định danh phân loại để tải lên Cổng thông tin điện tử một cửa về AI theo Điều 6 Nghị định 142.

II. Module B: Kiểm tra tuân thủ nghĩa vụ

Module B được kích hoạt theo kết quả Module A. Doanh nghiệp Rủi ro thấp trả lời B.1 và B.2. Doanh nghiệp Rủi ro trung bình trả lời B.1, B.2, B.3 và B.4. Doanh nghiệp Rủi ro cao trả lời toàn bộ B.1 đến B.7. Với mỗi câu hỏi, doanh nghiệp chọn một trong bốn mức: Đã thực hiện đầy đủ / Đang thực hiện một phần / Chưa thực hiện / Không áp dụng (kèm lý do).

B.1 - Nghĩa vụ minh bạch và thông báo (áp dụng mọi mức rủi ro)

Cơ sở pháp lý: Điều 16 Nghị định 142.

B.1.1. Người dùng cuối có được thông báo rõ ràng rằng họ đang tương tác với hoặc bị ảnh hưởng bởi một hệ thống AI trước khi tương tác bắt đầu không, theo hình thức mà người dùng thực sự đọc và hiểu được (không phải chỉ trong điều khoản sử dụng)?

B.1.2. Thông báo đó có được thực hiện bằng tiếng Việt, ngôn ngữ rõ ràng, không sử dụng thuật ngữ kỹ thuật không cần thiết, và phù hợp với đặc điểm của nhóm người dùng mục tiêu không?

B.1.3. Đối với hệ thống AI tương tác (chatbot, trợ lý ảo), hệ thống có tự động tiết lộ bản chất AI của mình ngay khi người dùng hỏi hoặc khi có dấu hiệu nhầm lẫn không?

B.1.4. Người dùng có được cung cấp thông tin liên hệ để yêu cầu giải thích về cách hệ thống AI đưa ra quyết định hoặc khuyến nghị ảnh hưởng đến họ không?

B.1.5. Chính sách về việc tổ chức sử dụng AI như thế nào trong sản phẩm/dịch vụ có được công bố trên website hoặc tài liệu chính thức không?

B.2 - Nghĩa vụ về nội dung do AI tạo sinh và gắn nhãn watermarking (áp dụng mọi mức rủi ro có tạo nội dung)

Cơ sở pháp lý: Điều 18 Nghị định 142.

B.2.1. Toàn bộ nội dung văn bản, hình ảnh, âm thanh hoặc video do hệ thống AI tạo sinh và phân phối đến người dùng bên ngoài có được đánh dấu bằng biện pháp kỹ thuật máy đọc được không?

B.2.2. Watermark đó có được nhúng vào nội dung theo cách bền vững, khó xóa trong điều kiện sử dụng thông thường (chia sẻ mạng xã hội, chuyển đổi định dạng cơ bản) không?

B.2.3. Watermark có chứa ít nhất các thông tin: nhận diện nội dung do AI tạo sinh; tên hoặc mã định danh của tổ chức triển khai; và thời gian tạo ra nội dung không?

B.2.4. Hệ thống có áp dụng nhãn hiển thị để người dùng bình thường nhận biết nội dung do AI tạo ra, đặc biệt với hình ảnh và video tổng hợp không?

B.2.5. Tổ chức có lưu trữ nhật ký nội dung AI được tạo ra trong thời hạn pháp luật quy định, có thể cung cấp cho cơ quan có thẩm quyền khi được yêu cầu không?

B.3 - Nghĩa vụ hồ sơ kỹ thuật và tài liệu hóa (rủi ro trung bình và cao)

Cơ sở pháp lý: Điều 14 Nghị định 142.

B.3.1. Hệ thống AI có hồ sơ kỹ thuật mô tả mục đích sử dụng, phạm vi áp dụng, đối tượng người dùng và nhóm đối tượng bị ảnh hưởng không?

B.3.2. Hồ sơ có mô tả kiến trúc tổng thể và nguyên lý hoạt động của hệ thống ở mức đủ để cơ quan có thẩm quyền hiểu được cách hệ thống đưa ra đầu ra không (không nhất thiết phải tiết lộ mã nguồn)?

B.3.3. Hồ sơ có ghi rõ nguồn gốc, quy mô, đặc điểm nhân khẩu học và phương pháp làm sạch của bộ dữ liệu huấn luyện không?

B.3.4. Hồ sơ có ghi rõ các rủi ro đã được nhận diện, các trường hợp sử dụng không được phép và điều kiện sử dụng an toàn không?

B.3.5. Hồ sơ kỹ thuật có được cập nhật khi hệ thống trải qua thay đổi đáng kể về mô hình, dữ liệu huấn luyện hoặc mục đích sử dụng không?

B.3.6. Tổ chức có duy trì nhật ký quyết định AI ghi lại đầu vào, đầu ra và tham số hệ thống trong thời gian đủ dài để phục vụ giải quyết tranh chấp không?

B.4 - Nghĩa vụ báo cáo sự cố (rủi ro trung bình và cao)

Cơ sở pháp lý: Điều 22 Nghị định 142.

B.4.1. Tổ chức có quy trình nội bộ xác định thế nào là "sự cố AI" cần báo cáo không, với phân loại mức độ nghiêm trọng rõ ràng?

B.4.2. Quy trình có đảm bảo tổ chức phát hiện sự cố nghiêm trọng có thể thông báo cho Bộ KH&CN trong vòng 72 giờ không?

B.4.3. Tổ chức có khả năng gửi báo cáo đầy đủ về sự cố trong vòng 72 giờ, bao gồm mô tả sự cố, đánh giá sơ bộ nguyên nhân, số người bị ảnh hưởng và biện pháp khắc phục tức thời không?

B.4.4. Có đơn vị hoặc cá nhân được phân công rõ ràng chịu trách nhiệm tiếp nhận thông tin về sự cố AI, đánh giá mức độ và kích hoạt quy trình báo cáo không?

B.4.5. Sau mỗi sự cố AI đáng kể, tổ chức có thực hiện phân tích nguyên nhân gốc rễ và điều chỉnh hệ thống trong thời hạn 30 ngày không?

B.4.6. Tổ chức có kênh tiếp nhận phản ánh về sự cố AI từ người dùng cuối và nhân viên nội bộ, đảm bảo phản ánh đó được xử lý và không bị phủ nhận không?

B.5 - Nghĩa vụ đánh giá tác động và quản lý rủi ro (rủi ro cao)

Cơ sở pháp lý: Điều 12-13 Nghị định 142 và Điều 9 Luật AI 2025.

B.5.1. Tổ chức có thực hiện đánh giá tác động hệ thống AI trước khi triển khai hệ thống rủi ro cao lần đầu không?

B.5.2. Đánh giá tác động có xem xét tác động phân phối theo giới tính, dân tộc, địa lý và nhóm tuổi tại Việt Nam không?

B.5.3. Tổ chức có hệ thống quản lý rủi ro AI đang hoạt động, không phải chỉ là tài liệu, bao gồm các kiểm soát kỹ thuật và quy trình đang được thực thi thực tế không?

B.5.4. Đánh giá tác động có được cập nhật khi hệ thống trải qua thay đổi đáng kể hoặc khi phát sinh bằng chứng về rủi ro mới từ thực tiễn vận hành không?

B.5.5. Kết quả đánh giá tác động có được người ra quyết định cấp cao (CEO, HĐQT) xem xét và phê duyệt chính thức không, chứ không chỉ là trách nhiệm của nhóm kỹ thuật?

B.6 - Nghĩa vụ kiểm soát của con người và khả năng can thiệp (rủi ro cao)

Cơ sở pháp lý: Điều 10 Luật AI 2025.

B.6.1. Người vận hành hệ thống AI rủi ro cao có khả năng giám sát, hiểu và đánh giá đầu ra của hệ thống không?

B.6.2. Có biện pháp kỹ thuật cho phép dừng, tạm dừng hoặc hủy bỏ quyết định tự động của hệ thống AI trong thời gian thực khi phát hiện bất thường không?

B.6.3. Với hệ thống AI ra quyết định ảnh hưởng đến quyền lợi cá nhân, có con người được phân công xem xét và phê duyệt trước khi quyết định AI có hiệu lực không, hoặc ít nhất có quy trình xem xét lại khi có yêu cầu không?

B.6.4. Người vận hành AI có được đào tạo đủ để nhận diện khi hệ thống AI đưa ra đầu ra bất thường, không nhất quán hoặc có dấu hiệu sai lệch không?

B.6.5. Với AI agent tự chủ, có giới hạn hành động được thiết lập kỹ thuật để hệ thống không thể thực hiện một số loại hành động nhất định mà không có sự phê duyệt của con người không?

B.7 - Nghĩa vụ đăng ký và thông báo với cơ quan quản lý (rủi ro cao)

Cơ sở pháp lý: Điều 5, 6 Nghị định 142.

B.7.1. Tổ chức đã hoàn thành thủ tục thông báo kết quả phân loại rủi ro hệ thống AI lên Cổng thông tin điện tử một cửa về AI chưa?

B.7.2. Kết quả đánh giá sự phù hợp của hệ thống AI rủi ro cao đã được thực hiện bởi tổ chức đánh giá được công nhận và kết quả đã được nộp lên Cổng chưa?

B.7.3. Tổ chức có lịch trình cụ thể để thực hiện đánh giá sự phù hợp định kỳ và cập nhật kết quả khi hệ thống AI thay đổi đáng kể không?

B.7.4. Hệ thống AI có được cấp mã định danh trên Cổng thông tin điện tử một cửa về AI và mã đó có được gắn vào tài liệu kỹ thuật, hồ sơ sự cố và nhật ký vận hành không?

III. Module C: Đánh giá rủi ro thiên lệch thuật toán

Module C bắt buộc với hệ thống AI rủi ro cao trong các lĩnh vực tín dụng, tuyển dụng, dịch vụ công và y tế. Tùy chọn cho hệ thống rủi ro trung bình.

C.1 - Đánh giá thiên lệch trong dữ liệu huấn luyện

C.1.1. Bộ dữ liệu huấn luyện có đại diện đầy đủ cho 54 dân tộc tại Việt Nam theo tỷ lệ gần với phân bố dân số thực tế không, hay phần lớn dữ liệu phản ánh người Kinh ở đô thị?

C.1.2. Tỷ lệ nam/nữ trong dữ liệu huấn luyện là bao nhiêu? Nếu lệch hơn 60/40 theo một giới tính, tổ chức có thực hiện biện pháp kỹ thuật bù đắp không?

C.1.3. Dữ liệu huấn luyện có đại diện đủ cho các vùng địa lý khác nhau của Việt Nam (Bắc, Trung, Nam; đô thị, nông thôn, miền núi) không? Nếu không, mô hình có được kiểm thử về hiệu năng khác nhau theo vùng không?

C.1.4. Dữ liệu huấn luyện có chứa dữ liệu lịch sử phản ánh các quyết định phân biệt đối xử trong quá khứ không (ví dụ: dữ liệu tuyển dụng từ khi phụ nữ bị hạn chế trong một số vai trò; dữ liệu tín dụng từ khi nhóm dân tộc thiểu số bị từ chối một cách có hệ thống)? Nếu có, biện pháp xử lý là gì?

C.1.5. Tổ chức có thực hiện kiểm thử thống kê so sánh hiệu năng của mô hình theo các nhóm được bảo vệ (giới tính, dân tộc, tuổi, khuyết tật) chứ không chỉ đo hiệu năng tổng thể không?

C.2 - Kiểm thử thiên lệch trong đầu ra mô hình

C.2.1 Kiểm thử ngang bằng nhân khẩu học Tỷ lệ kết quả tích cực (phê duyệt tín dụng, nhận việc làm, phê duyệt yêu cầu dịch vụ) của mô hình có khác nhau đáng kể giữa các nhóm giới tính, dân tộc hoặc khu vực địa lý khi các đặc điểm liên quan đến kết quả được kiểm soát không? Mức chênh lệch được phép tối đa là bao nhiêu?

C.2.2 Kiểm thử cân bằng cơ hội Với nhóm người thực sự đủ điều kiện (nợ tốt, năng lực đạt yêu cầu), tỷ lệ được mô hình phân loại đúng có nhất quán giữa các nhóm dân số khác nhau không?

C.2.3 Kiểm thử thiên lệch ngôn ngữ Đối với mô hình NLP/LLM xử lý tiếng Việt: hệ thống có được kiểm thử với văn bản có giọng điệu địa phương (Bắc, Trung, Nam), tiếng lóng giới trẻ, cách diễn đạt của người cao tuổi không? Hiệu năng có nhất quán không?

C.2.4 Kiểm thử thiên lệch theo tên và địa danh Tổ chức có kiểm thử xem mô hình có đưa ra kết quả khác nhau cho hồ sơ giống hệt nhau chỉ khác ở tên (tên Kinh so với tên dân tộc thiểu số) hoặc địa chỉ (đô thị so với nông thôn/miền núi) không?

C.2.5 Kiểm thử thiên lệch theo độ tuổi Mô hình có hiệu năng nhất quán với người dùng thuộc các nhóm tuổi khác nhau không, đặc biệt với người cao tuổi sử dụng cách diễn đạt hoặc hành vi tương tác khác với nhóm trẻ tuổi?

C.3 - Cơ chế giám sát thiên lệch liên tục

C.3.1. Tổ chức có giám sát định kỳ (ít nhất hàng quý) sự thay đổi trong phân phối đầu ra của mô hình theo các nhóm nhân khẩu học không, tức là phát hiện khi thiên lệch xuất hiện hoặc tăng lên trong quá trình vận hành?

C.3.2. Khi phát hiện mức độ thiên lệch vượt ngưỡng được tổ chức xác định trước, có quy trình tự động kích hoạt rà soát mô hình không, và quy trình đó có thời hạn xử lý rõ ràng không?

C.3.3. Tổ chức có cơ chế tiếp nhận phản hồi về thiên lệch từ người dùng cuối, đặc biệt từ các nhóm bị thiệt thòi, và phản hồi đó có được đưa vào quy trình cải thiện mô hình không?

C.3.4. Kết quả kiểm thử thiên lệch có được lưu trữ và so sánh theo thời gian để phát hiện xu hướng không, chứ không chỉ kiểm thử một lần trước triển khai?

C.4 - Giải trình về kết quả thiên lệch

C.4.1. Tổ chức có tài liệu giải thích kết quả kiểm thử thiên lệch không, bao gồm chỉ số nào được đo, kết quả là gì, ngưỡng chấp nhận được dựa trên căn cứ nào và biện pháp giảm thiểu nào đã được thực hiện?

C.4.2. Khi mô hình thể hiện thiên lệch vượt ngưỡng chấp nhận nhưng tổ chức quyết định tiếp tục vận hành, có tài liệu ghi rõ lý do và biện pháp giảm thiểu thay thế không?

C.4.3. Tổ chức có thể cung cấp kết quả kiểm thử thiên lệch cho cơ quan quản lý hoặc bên thứ ba độc lập được ủy quyền khi được yêu cầu không?

Kết quả Module C

Module C tạo ra *Báo cáo đánh giá thiên lệch* gồm: điểm số thiên lệch theo từng nhóm nhân khẩu học; so sánh với ngưỡng chấp nhận do tổ chức tự xác định; các lĩnh vực cần cải thiện ưu tiên; và khuyến nghị kiểm thử bổ sung. Báo cáo này là thành phần bắt buộc của Hồ sơ kỹ thuật theo B.3 và là đầu vào để AMIS theo dõi xu hướng thiên lệch theo ngành.

IV. Nguyên tắc thiết kế câu hỏi VN-AI VERIFY

Ba nguyên tắc quyết định chất lượng của toàn bộ bộ câu hỏi cần được ghi rõ trong tài liệu phát triển để duy trì nhất quán qua các phiên bản cập nhật.

Nguyên tắc thứ nhất là *phản ánh hai chiều bắt buộc*: mỗi câu hỏi phải có điều khoản pháp luật tương ứng được ghi rõ, và mỗi điều khoản pháp luật tạo ra nghĩa vụ kỹ thuật phải có ít nhất một câu hỏi đại diện trong Module B.

Nguyên tắc thứ hai là *khả năng kiểm chứng*: mỗi câu hỏi phải có thể trả lời dựa trên bằng chứng có thể xuất trình (tài liệu, nhật ký, biên bản kiểm thử). Câu hỏi không thể kiểm chứng không nên đưa vào bộ công cụ.

Nguyên tắc thứ ba là *cập nhật có kiểm soát*: khi Luật AI 2025 hoặc Nghị định 142 được sửa đổi, nhóm phát triển VN-AI Verify phải cập nhật bộ câu hỏi trong vòng 90 ngày kể từ ngày văn bản sửa đổi có hiệu lực, với nhật ký thay đổi công khai ghi rõ câu hỏi nào được thêm, sửa hay xóa theo điều khoản nào.

PHỤ LỤC 3
BỘ CÔNG CỤ KỸ THUẬT LẬP PHÁP AI

Công cụ 1: Test chức năng cho định nghĩa AI

Mỗi định nghĩa AI trong văn bản pháp lý phải vượt qua ba câu hỏi kiểm tra tuần tự trước khi được giữ nguyên.

Câu hỏi 1 - Test thay thế công nghệ: Nếu thay thế công nghệ này bằng công nghệ khác có cùng chức năng nhưng kiến trúc kỹ thuật hoàn toàn khác, định nghĩa này có còn bao phủ không?. Nếu không, định nghĩa đang mô tả kiến trúc kỹ thuật, không phải chức năng, do đó cần viết lại.

Câu hỏi 2 - Test ổn định thời gian: Nếu 5 năm sau, công nghệ AI phát triển theo hướng [kịch bản cụ thể], định nghĩa này còn phù hợp không?. Nếu không, cần bổ sung cơ chế diễn giải linh hoạt hoặc điều khoản ủy quyền cập nhật.

Câu hỏi 3 - Test ranh giới: Định nghĩa này có bao phủ tác nhân AI (tự chủ chuỗi hành động), hệ thống AI đa tác nhân (nhiều agents tương tác), và AI tích hợp phần cứng (embedded AI) không?. Nếu không rõ, cần bổ sung điều khoản làm rõ hoặc điều khoản diễn giải.

Ba câu hỏi này được trình bày dưới dạng bảng kiểm tra (checklist) kèm ví dụ minh họa cho từng loại AI phổ biến tại Việt Nam. Kết quả kiểm tra là phần bắt buộc của hồ sơ soạn thảo.

Công cụ 2: Bảng phân định nghĩa vụ chuỗi giá trị

Mẫu bảng chuẩn hóa để cán bộ soạn thảo phân định nghĩa vụ theo chuỗi giá trị AI, gồm sáu cột: chủ thể (nhà phát triển mô hình / nhà cung cấp hệ thống / nhà triển khai / người dùng); hành vi điều chỉnh; nghĩa vụ chính; nghĩa vụ bổ sung khi chủ thể trước không thực hiện; căn cứ pháp lý tham chiếu; và ngưỡng kích hoạt nghĩa vụ.

Bảng phải được điền đầy đủ cho mọi điều khoản tạo ra nghĩa vụ trong văn bản AI trước khi điều khoản được đưa vào dự thảo chính thức. Điều này buộc cán bộ soạn thảo phải tư duy về *ai* phải làm gì thay vì chỉ xác định *điều gì* phải được thực hiện.

Mẫu bảng bao gồm ba tình huống mặc định cần kiểm tra cho mọi điều khoản: tình huống nhà cung cấp mô hình nền tảng nước ngoài không tuân thủ; tình huống nhà triển khai sửa đổi hệ thống vượt phạm vi kiểm định; và tình huống tác nhân AI hành động tự chủ gây thiệt hại.

Công cụ 3: Mẫu báo cáo AI-RIA ba chiều

Phần bổ sung 1 - Phân tích kịch bản công nghệ: Bảng ba cột đánh giá tác động của từng yêu cầu pháp lý dự kiến trong ba kịch bản (chu kỳ đổi mới AI chậm 4-5 năm/vừa 2-3 năm/nhanh 12-18 tháng), xác định điều khoản nào lỗi thời sớm nhất trong từng kịch bản và cơ chế ủy quyền nào cần thiết kế từ đầu.

Phần bổ sung 2 - Chi phí tuân thủ theo quy mô: Bảng ước lượng chi phí tuân thủ (ban đầu, vận hành hàng năm, thích ứng khi cập nhật) theo ba tầng: doanh nghiệp lớn (trên

500 nhân viên), SME (11-499 nhân viên) và micro/startup (dưới 10 nhân viên hoặc dưới 3 năm tuổi). Số liệu phải có nguồn xác định: khảo sát doanh nghiệp, tham chiếu chi phí từ thị trường tương đồng (Hàn Quốc, Singapore) hoặc ước lượng từ chuyên gia độc lập.

Phần bổ sung 3 - Đánh giá năng lực thể chế: Bảng xác định cho mỗi nghĩa vụ kiểm định hoặc đánh giá sự phù hợp: số tổ chức có năng lực thực hiện hiện tại, thời gian xây dựng năng lực bổ sung và cơ chế chuyển tiếp trong thời gian năng lực chưa đủ.

Công cụ 4: Quy trình kiểm tra tỷ lệ cho điều khoản hạn chế quyền

Mọi điều khoản AI hạn chế quyền hiến định (nhận diện khuôn mặt, phân tích hành vi, ra quyết định tự động ảnh hưởng đến quyền lợi) phải trải qua quy trình kiểm tra tỷ lệ ba bậc trước khi đưa vào văn bản.

Bậc 1 - Mục tiêu hợp pháp: Điều khoản phục vụ mục tiêu nào trong số các mục tiêu được liệt kê tại Khoản 2 Điều 14 Hiến pháp 2013? Mục tiêu đó có thực sự đặt ra trong bối cảnh AI không, hay chỉ là biện hộ hình thức?

Bậc 2 - Tính phù hợp: Có bằng chứng thực nghiệm nào cho thấy điều khoản này thực sự góp phần đạt được mục tiêu đã xác định không?

Bậc 3 - Tính cần thiết tối thiểu: Có biện pháp nào ít xâm phạm quyền hơn mà đạt được cùng mục tiêu không? Nếu có, phải giải thích tại sao biện pháp ít xâm phạm hơn không được chọn. Nếu không, ghi rõ lý do kỹ thuật và pháp lý.

Kết quả kiểm tra ba bậc được tài liệu hóa trong hồ sơ soạn thảo và công bố kèm dự thảo văn bản tạo ra trách nhiệm giải trình lập pháp thực chất.

Công cụ 5: Template điều khoản thích ứng chuẩn

Thư viện điều khoản mẫu cho năm loại điều khoản thích ứng phổ biến nhất trong văn bản AI.

Mẫu 1 - Điều khoản rà soát bắt buộc: "Bộ Khoa học và Công nghệ thực hiện đánh giá toàn diện [tên văn bản] sau mỗi [X năm] kể từ ngày có hiệu lực và công bố báo cáo đánh giá kèm đề xuất sửa đổi nếu phát sinh bất cập cơ cấu. Báo cáo phải dựa trên dữ liệu định lượng từ AMIS về tuân thủ, sự cố và tác động kinh tế-xã hội".

Mẫu 2 - Điều khoản viện dẫn không ngày: "[Yêu cầu kỹ thuật] được xác định theo [tên TCVN] phiên bản hiện hành tại thời điểm đánh giá".

Mẫu 3 - Điều khoản suy đoán tuân thủ: "[Chủ thể] tuân thủ [tên TCVN] được chỉ định theo [điều khoản ủy quyền] được suy đoán đáp ứng yêu cầu [điều khoản cụ thể] của [tên văn bản]. Suy đoán có thể bị bác bỏ khi [điều kiện bác bỏ]".

Mẫu 4 - Điều khoản ủy quyền diễn giải: "Khi xuất hiện hệ thống AI hoặc ứng dụng AI mới có đặc điểm kỹ thuật chưa được xác định rõ theo [tên văn bản], [cơ quan được ủy quyền] ban hành diễn giải chính thức trong thời hạn [X ngày] kể từ ngày nhận đề nghị. Diễn giải có giá trị áp dụng đến khi có văn bản sửa đổi chính thức".

Mẫu 5 - Điều khoản tuân thủ hoặc giải thích: "[Bộ chuyên ngành] ban hành hướng dẫn áp dụng [tên văn bản] trong lĩnh vực [X]. Tổ chức, doanh nghiệp không tuân thủ

hướng dẫn phải cung cấp bằng văn bản cho [cơ quan thẩm quyền] giải thích cách thức đạt được mục tiêu bảo vệ tương đương trong thời hạn [Y ngày] khi được yêu cầu".

Mẫu 6 - Điều khoản nghĩa vụ hậu chấm dứt vận hành và kế thừa nghĩa vụ: "Khi chấm dứt vận hành hệ thống AI [đối tượng], [chủ thể vận hành] có nghĩa vụ thông báo trước cho người dùng trong thời hạn tối thiểu [X ngày]; xử lý dữ liệu người dùng theo lựa chọn của người dùng; lưu trữ hồ sơ kỹ thuật và nhật ký vận hành trong thời hạn [Y năm] kể từ ngày chấm dứt; và cập nhật trạng thái trên Cơ sở dữ liệu quốc gia về AI trong thời hạn [Z ngày]. Trường hợp [chủ thể vận hành] chấm dứt tồn tại trước khi hệ thống được chính thức ngừng vận hành, nghĩa vụ nêu trên được chuyển giao cho [chủ thể kế thừa theo quy định pháp luật về giải thể, phá sản]".

Công cụ 6: Hướng dẫn tham vấn kỹ thuật chuyên sâu

Quy trình tổ chức tham vấn kỹ thuật chuyên sâu cho văn bản AI quan trọng, gồm ba phần.

Phần 1 - Danh sách câu hỏi kỹ thuật bắt buộc: Bộ câu hỏi chuẩn cần đặt cho kỹ sư AI trong phiên tham vấn kỹ thuật như "Yêu cầu kỹ thuật X có thể thực hiện bằng công nghệ hiện tại không?", "Chi phí triển khai ước tính là bao nhiêu cho doanh nghiệp có 100 nhân viên?", "Có giải pháp kỹ thuật thay thế nào đạt cùng mục tiêu bảo vệ với chi phí thấp hơn không?". Danh sách câu hỏi chuẩn hóa tránh tình trạng phiên tham vấn thiên về vận động chính sách thay vì cung cấp thông tin kỹ thuật.

Phần 2 - Mẫu biên bản tham vấn kỹ thuật: Cấu trúc biên bản ghi rõ câu hỏi đặt ra, câu trả lời kỹ thuật nhận được, đánh giá của cán bộ soạn thảo về hàm ý pháp lý và điều chỉnh dự thảo (nếu có). Biên bản này là tài liệu hồ sơ công khai.

Phần 3 - Mẫu Báo cáo phản hồi ý kiến tham vấn: Cấu trúc công khai sau tham vấn, phân loại ý kiến thành bốn nhóm: tiếp nhận và phản ánh vào dự thảo; tiếp nhận nhưng xử lý theo cách khác (giải thích); không tiếp nhận (giải thích lý do); và cần nghiên cứu thêm (thời hạn cụ thể).

PHỤ LỤC 4
HỆ THỐNG THÔNG TIN QUẢN LÝ AI (AMIS)

1. Giới thiệu

AMIS (nền tảng kỹ thuật) là một hệ thống thông tin bao gồm cơ sở dữ liệu, phần mềm phân tích, giao diện truy vấn và cơ chế kết nối dữ liệu, không có tư cách pháp nhân, không có thẩm quyền pháp lý, không tự mình chịu trách nhiệm pháp lý về bất kỳ hành vi nào.

Cơ quan quản lý và vận hành AMIS là chủ thể pháp lý có tư cách quyết định, có thẩm quyền được pháp luật trao, và là chủ thể chịu trách nhiệm pháp lý về việc thiết lập, vận hành, khai thác và bảo vệ dữ liệu trong AMIS. Đề xuất là Cục Công nghệ số và AI thuộc Bộ Khoa học và Công nghệ.

Cơ sở pháp lý của AMIS: Nghị định 142/2026/NĐ-CP đã thiết lập Cơ sở dữ liệu quốc gia về AI và Cổng thông tin điện tử một cửa về là căn cứ pháp lý nền tảng đã có sẵn cho hạ tầng dữ liệu mà AMIS được xây dựng dựa trên và mở rộng thêm chức năng phân tích. Tuy nhiên, cần lưu ý Nghị định 142 thiết lập Cơ sở dữ liệu quốc gia về AI như một kho lưu trữ hành chính (tiếp nhận thông báo, cấp mã định danh), trong khi AMIS là tầng phân tích và phản hồi chính sách đặt phía sau kho lưu trữ đó. Do đó, cần bổ sung căn cứ pháp lý riêng cho chức năng phân tích và phản hồi chính sách của AMIS, không thể suy diễn hoàn toàn từ căn cứ pháp lý đã có cho Cơ sở dữ liệu quốc gia.

Đề xuất bổ sung căn cứ pháp lý: Nghị định 142 sửa đổi cần bổ sung điều khoản: *"Bộ Khoa học và Công nghệ xây dựng và vận hành Hệ thống thông tin quản lý AI (AMIS) trên cơ sở Cơ sở dữ liệu quốc gia về AI, có chức năng tiếp nhận, tổng hợp, phân tích dữ liệu về tuân thủ, sự cố và kết quả tự đánh giá của hệ thống AI, phục vụ công tác quản lý nhà nước và đề xuất hoàn thiện chính sách, pháp luật về AI"*. Đây là căn cứ pháp lý bổ sung cần thiết để chuyển hóa từ "kho lưu trữ hành chính" (đã có) sang "công cụ phân tích chính sách chủ động" (đang đề xuất).

Cơ sở pháp lý về bảo vệ dữ liệu trong AMIS: Luật Bảo vệ dữ liệu cá nhân 2025 là căn cứ pháp lý áp dụng bắt buộc đối với mọi dữ liệu cá nhân được xử lý trong AMIS (ví dụ dữ liệu về nạn nhân sự cố AI, dữ liệu người dùng bị ảnh hưởng). AMIS với tư cách một hệ thống xử lý dữ liệu của cơ quan nhà nước phải tuân thủ đầy đủ Luật Bảo vệ dữ liệu cá nhân 2025 như bất kỳ hệ thống thông tin nào khác của Nhà nước.

2. Kiến trúc kỹ thuật và pháp lý của AMIS

Module 1 Thu thập và chuẩn hóa dữ liệu: AMIS tiếp nhận bốn luồng dữ liệu từ Cổng thông tin một cửa theo thời gian thực: *(i)* thông báo sự cố AI theo Điều 19 Nghị định 142 (72 giờ-5 ngày); *(ii)* kết quả tự phân loại rủi ro của nhà cung cấp; *(iii)* kết quả VN-AI Verify; và *(iv)* báo cáo định kỳ về tình trạng vận hành hệ thống AI rủi ro cao.

Chuẩn hóa dữ liệu sự cố theo phân loại bốn cấp học từ EASA (cơ quan hàng không EU): sự cố an toàn trực tiếp; sự cố ảnh hưởng đến quyền người dùng; sự cố kỹ thuật không ảnh hưởng trực tiếp; và gần xảy ra sự cố (*near-miss*). Phân loại bốn cấp này quan trọng vì cho phép phân tích xu hướng theo loại sự cố, không phải tất cả sự cố đều có hàm ý chính sách như nhau.

Module 2 Phân tích xu hướng và cảnh báo chính sách: AMIS thực hiện phân tích định kỳ theo chu kỳ quý, tạo ra ba sản phẩm thông tin có giá trị pháp lý khác nhau.

Báo cáo xu hướng quý: Thống kê sự cố theo loại, ngành, quy mô doanh nghiệp và điều khoản pháp luật liên quan, ẩn danh hóa thông tin doanh nghiệp nhưng giữ nguyên thông tin ngành và loại sự cố. Báo cáo này được công bố công khai và là đầu vào cho Báo cáo học hỏi chính sách hàng năm.

Cảnh báo xu hướng: Khi dữ liệu cho thấy loại sự cố nhất định tăng vượt ngưỡng thống kê xác định trong hai quý liên tiếp, AMIS tự động phát cảnh báo lên đầu mối AI của Bộ KH&CN trong 48 giờ. Cơ chế tự động này là điều kiện để cơ chế phản ứng nhanh qua nghị quyết Chính phủ 30 ngày hoạt động hiệu quả.

Báo cáo học hỏi chính sách năm: Tổng hợp bằng chứng từ ba nguồn dữ liệu sự cố AMIS, kết quả VN-AI Verify và phản hồi từ doanh nghiệp qua Cổng thông tin thành đề xuất cụ thể cho Bộ KH&CN về điều khoản nào trong Nghị định 142 cần sửa đổi và theo hướng nào. Báo cáo này là đầu vào bắt buộc cho mỗi kỳ sửa đổi Nghị định, thể chế hóa vòng phản hồi từ thực tiễn đến lập pháp.

Module 3 Đăng ký và tra cứu: AMIS vận hành Sổ đăng ký hệ thống AI rủi ro cao với danh sách công khai gồm: tên hệ thống, nhà cung cấp, phân loại rủi ro, kết quả kiểm định và tình trạng tuân thủ. Đây là cơ chế minh bạch quan trọng với hai tác dụng: *(i)* người dùng có thể tra cứu hệ thống AI đang tác động đến mình; và *(ii)* cơ quan thực thi có thể đối chiếu báo cáo sự cố với hồ sơ đăng ký để phát hiện doanh nghiệp không báo cáo.

3. Chủ thể có thẩm quyền đối với AMIS

Thẩm quyền đối với AMIS được phân bổ theo ba cấp độ chủ thể có tính chất khác nhau.

Chủ thể có thẩm quyền quyết định chính sách và thiết lập hệ thống: Bộ Khoa học và Công nghệ, với tư cách cơ quan đầu mối theo Chương VII Nghị định 142, có thẩm quyền quyết định về cấu trúc, phạm vi dữ liệu, và mục đích sử dụng của AMIS.

Chủ thể có thẩm quyền truy cập và khai thác dữ liệu: bao gồm nhiều cơ quan có thẩm quyền khác nhau theo phạm vi truy cập được phân quyền, không phải chỉ Bộ Khoa học và Công nghệ độc quyền truy cập.

Chủ thể có thẩm quyền giám sát việc vận hành AMIS: là chủ thể giám sát *chính hoạt động của cơ quan quản lý AMIS*, không phải chủ thể vận hành hệ thống.

4. Chế độ bảo vệ dữ liệu cá nhân trong AMIS

Nguyên tắc áp dụng: mọi dữ liệu cá nhân được xử lý trong AMIS (thông tin về nạn nhân sự cố AI, thông tin người dùng hệ thống AI bị ảnh hưởng khi được đề cập trong báo cáo sự cố) phải tuân thủ đầy đủ các nguyên tắc của Luật Bảo vệ dữ liệu cá nhân 2025, đặc biệt nguyên tắc giới hạn mục đích (dữ liệu cá nhân trong báo cáo sự cố chỉ được sử dụng

cho mục đích phân tích chính sách và xử lý sự cố, không được sử dụng cho mục đích khác); nguyên tắc tối thiểu hóa dữ liệu (chỉ thu thập dữ liệu cá nhân cần thiết, không thu thập toàn bộ nội dung chi tiết vượt quá nhu cầu phân tích); nguyên tắc ẩn danh hóa khi công bố (Báo cáo xu hướng quý và Báo cáo học hỏi chính sách năm phải được ẩn danh hóa hoàn toàn thông tin định danh cá nhân và doanh nghiệp trước khi công bố công khai).

Đề xuất cơ chế cụ thể: AMIS cần áp dụng kiến trúc phân tách hai lớp dữ liệu, lớp dữ liệu chi tiết có định danh (chỉ truy cập được bởi nhóm cán bộ được phân quyền cụ thể phục vụ xử lý sự cố trực tiếp) và lớp dữ liệu tổng hợp đã ẩn danh hóa (phục vụ phân tích xu hướng và công bố báo cáo).

5. Chế độ bảo vệ bí mật kinh doanh

Vấn đề pháp lý cốt lõi là AMIS tiếp nhận dữ liệu từ doanh nghiệp (báo cáo sự cố, kết quả VN-AI Verify, thông tin đăng ký hệ thống AI rủi ro cao) có khả năng chứa đựng bí mật kinh doanh như thông tin về kiến trúc hệ thống, phương pháp huấn luyện mô hình, dữ liệu khách hàng, chiến lược kinh doanh liên quan đến sản phẩm AI. Cần có cơ chế bảo vệ tương xứng để tránh việc nghĩa vụ báo cáo minh bạch với Nhà nước trở thành kênh rò rỉ bí mật kinh doanh ra bên ngoài hoặc tới đối thủ cạnh tranh.

Đề xuất cơ chế cụ thể: *(i)* nghĩa vụ bảo mật của cán bộ có quyền truy cập AMIS cần được quy định trong Nghị định 142 sửa đổi, với chế tài kỷ luật và trách nhiệm bồi thường thiệt hại nếu để lộ thông tin bí mật kinh doanh của doanh nghiệp mà không có căn cứ pháp lý; *(ii)* cơ chế đánh dấu và phân loại thông tin bí mật kinh doanh khi doanh nghiệp nộp báo cáo cho phép doanh nghiệp chủ động đánh dấu phần thông tin có tính chất bí mật kinh doanh trong hồ sơ nộp cho AMIS; *(iii)* nguyên tắc chỉ công bố thông tin tổng hợp, ẩn danh hóa ở cấp độ ngành (không phải cấp độ từng doanh nghiệp) trong các báo cáo xu hướng công khai bảo đảm dữ liệu phục vụ mục tiêu phân tích chính sách công cộng nhưng không tạo ra thiệt hại cạnh tranh cho doanh nghiệp đã báo cáo trung thực.

6. Phạm vi truy cập AMIS

Đề xuất phân quyền truy cập theo bốn cấp độ, tương ứng với nguyên tắc phân tách dữ liệu đã nêu:

Cấp độ 1 - Truy cập đầy đủ (cán bộ nghiệp vụ trực tiếp): cán bộ thuộc Cục Công nghệ số và AI của Bộ Khoa học và Công nghệ được phân công trực tiếp xử lý sự cố và phân tích chính sách truy cập cả lớp dữ liệu chi tiết có định danh, trong phạm vi công việc được giao và theo nguyên tắc cần biết.

Cấp độ 2 - Truy cập theo phạm vi thẩm quyền chuyên ngành (cơ quan phối hợp): các cơ quan có thẩm quyền quản lý theo Chương VII Nghị định 142 (NHNN đối với dữ liệu sự cố AI trong lĩnh vực ngân hàng, Bộ Y tế đối với lĩnh vực y tế, cơ quan điều tra đối với dữ liệu phục vụ điều tra hình sự) truy cập dữ liệu chi tiết nhưng giới hạn trong phạm vi lĩnh vực hoặc vụ việc thuộc thẩm quyền, không truy cập toàn bộ hệ thống.

Cấp độ 3 - Truy cập dữ liệu tổng hợp đã ẩn danh hóa (cơ quan giám sát, nghiên cứu): cơ quan giám sát độc lập, cơ quan thanh tra và các tổ chức nghiên cứu được cấp

phép truy cập lớp dữ liệu tổng hợp đã ẩn danh hóa phục vụ mục đích giám sát hoặc nghiên cứu chính sách, không truy cập dữ liệu chi tiết có định danh.

Cấp độ 4 - Tiếp cận công khai (công chúng, doanh nghiệp): Báo cáo xu hướng quý và Báo cáo học hỏi chính sách năm được công bố công khai trên Cổng thông tin điện tử một cửa về AI; Sổ đăng ký hệ thống AI rủi ro cao (tên hệ thống, nhà cung cấp, phân loại rủi ro, tình trạng tuân thủ) được công khai để người dùng tra cứu, không bao gồm dữ liệu sự cố chi tiết hay thông tin bí mật kinh doanh.

7. Thời hạn lưu trữ dữ liệu trong AMIS

Đề xuất phân biệt thời hạn theo loại dữ liệu, không áp dụng một thời hạn chung cho toàn bộ hệ thống.

Dữ liệu sự cố AI: lưu trữ tối thiểu theo thời hiệu khởi kiện dân sự (hiện hành theo Bộ luật Dân sự 2015) kể từ thời điểm sự cố được ghi nhận, để bảo đảm dữ liệu vẫn khả dụng phục vụ giải quyết tranh chấp có thể phát sinh sau đó. Đề xuất cụ thể tối thiểu 3 năm, có thể kéo dài nếu sự cố đang trong quá trình tố tụng.

Dữ liệu đăng ký hệ thống AI rủi ro cao và kết quả VN-AI Verify: lưu trữ trong suốt thời gian hệ thống AI còn vận hành, và tiếp tục lưu trữ thêm một khoảng thời gian sau khi hệ thống chấm dứt vận hành (đề xuất tối thiểu 3 năm kể từ ngày chấm dứt, thống nhất với đề xuất về nghĩa vụ lưu trữ hồ sơ hậu chấm dứt vận hành đã trình bày tại phần giải pháp về giai đoạn thứ tám của vòng đời AI) bảo đảm dữ liệu vẫn khả dụng cho việc xử lý tranh chấp có thể phát sinh sau khi hệ thống đã ngừng hoạt động.

Dữ liệu tổng hợp đã ẩn danh hóa phục vụ phân tích xu hướng dài hạn: không áp dụng thời hạn lưu trữ cố định, được lưu trữ dài hạn phục vụ phân tích xu hướng chính sách qua nhiều năm, vì dữ liệu đã ẩn danh hóa không còn phát sinh rủi ro về quyền riêng tư cá nhân tương ứng với thời hạn lưu trữ kéo dài.

8. Cơ chế giám sát việc vận hành AMIS

Cơ chế giám sát AMIS cần bảo đảm ba chiều đồng thời, không chỉ dựa vào giám sát nội bộ của chính cơ quan vận hành.

Giám sát hành chính độc lập: đề xuất cơ chế thanh tra định kỳ (ít nhất hàng năm) do Thanh tra Bộ Khoa học và Công nghệ hoặc cơ quan thanh tra chuyên ngành thực hiện đối với việc tuân thủ quy định về bảo vệ dữ liệu cá nhân và bí mật kinh doanh trong vận hành AMIS độc lập tương đối với đơn vị trực tiếp vận hành hệ thống.

Giám sát tư pháp: khi phát sinh tranh chấp liên quan đến việc AMIS xử lý dữ liệu sai quy định (ví dụ để lộ bí mật kinh doanh, xử lý dữ liệu cá nhân trái mục đích), tòa án có thẩm quyền xem xét theo thủ tục tố tụng hành chính hoặc dân sự tương ứng, không có quy định miễn trừ trách nhiệm tư pháp cho cơ quan vận hành AMIS.

Giám sát xã hội: đề xuất cơ chế báo cáo minh bạch định kỳ về hoạt động của AMIS (không phải nội dung dữ liệu cụ thể, mà là thống kê về số lượng truy cập theo cấp độ, số lượng yêu cầu từ cơ quan phối hợp, kết quả thanh tra) được công bố công khai hàng năm, cho phép tổ chức xã hội dân sự và báo chí giám sát cách thức hệ thống vận hành mà không cần truy cập dữ liệu nhạy cảm.

9. Trách nhiệm khi hệ thống phát cảnh báo sai

Đây là vấn đề pháp lý quan trọng và tinh tế nhất trong toàn bộ thiết kế AMIS, đòi hỏi phân biệt rõ các loại "cảnh báo sai" khác nhau vì hệ quả trách nhiệm khác nhau.

Loại thứ nhất, Cảnh báo dương tính giả: hệ thống AMIS phân tích dữ liệu và phát cảnh báo xu hướng sự cố đối với một loại hệ thống AI hoặc một doanh nghiệp cụ thể, nhưng thực tế không có vấn đề đáng kể (do sai sót trong thuật toán phân tích, hoặc do dữ liệu đầu vào không đầy đủ dẫn đến kết luận sai). Về trách nhiệm, cảnh báo tự động từ hệ thống AMIS không tự động tạo ra hậu quả pháp lý bất lợi trực tiếp cho doanh nghiệp. Cảnh báo chỉ có chức năng kích hoạt quy trình xem xét nội bộ. Do đó, nếu cảnh báo sai dẫn đến việc cán bộ tiến hành kiểm tra hành chính đối với doanh nghiệp mà kết quả kiểm tra không phát hiện vi phạm, doanh nghiệp không phải chịu bất kỳ hậu quả bất lợi nào từ chính cảnh báo sai đó (không bị coi là có "tiền sử vi phạm"), nhưng cũng không có quyền yêu cầu bồi thường thiệt hại từ Nhà nước cho việc bị kiểm tra, trừ khi việc kiểm tra được tiến hành trái thẩm quyền hoặc trái thủ tục theo quy định của Luật Thanh tra và pháp luật xử lý vi phạm hành chính hiện hành.

Loại thứ hai, Cảnh báo âm tính giả: hệ thống AMIS không phát hiện xu hướng sự cố đang thực sự tồn tại (do thuật toán phân tích chưa đủ tinh vi, hoặc do dữ liệu báo cáo từ doanh nghiệp không đầy đủ/không trung thực), dẫn đến việc cơ quan quản lý không can thiệp kịp thời và thiệt hại tiếp tục xảy ra cho người dùng. Về trách nhiệm, cần phân biệt hai nguyên nhân khác nhau. Nếu nguyên nhân do doanh nghiệp báo cáo không trung thực (vi phạm nghĩa vụ báo cáo sự cố theo Điều 22 Nghị định 142), trách nhiệm chính thuộc về doanh nghiệp báo cáo sai, không phải hệ thống AMIS hay cơ quan vận hành. Nếu nguyên nhân do thuật toán phân tích của AMIS chưa đủ khả năng phát hiện xu hướng dù dữ liệu đầu vào đã đầy đủ và trung thực, đây là hạn chế kỹ thuật của công cụ quản lý nhà nước. Về nguyên tắc chung của trách nhiệm nhà nước, cơ quan vận hành AMIS không tự động chịu trách nhiệm bồi thường cho người dùng bị thiệt hại chỉ vì công cụ phân tích chưa hoàn hảo (tương tự nguyên tắc cơ quan nhà nước không chịu trách nhiệm tuyệt đối vì mọi công cụ giám sát đều có giới hạn kỹ thuật), trừ khi có thể chứng minh có yếu tố lỗi cụ thể của cán bộ hoặc cơ quan (như đã biết về hạn chế kỹ thuật nghiêm trọng nhưng không có biện pháp khắc phục hợp lý trong thời gian dài) theo quy định về trách nhiệm bồi thường nhà nước hiện hành.

PHỤ LỤC 5
BIÊN BẢN TỔNG HỢP KẾT QUẢ PHỎNG VẤN CHUYÊN GIA

Đề tài nghiên cứu: ***Xây dựng khung pháp lý cho phát triển, ứng dụng trí tuệ nhân tạo ở Việt Nam***
Nghiên cứu sinh: Nguyễn Quỳnh Trang
Cơ sở đào tạo: Học viện Chính trị quốc gia Hồ Chí Minh
Ngành: Lý luận và Lịch sử Nhà nước và Pháp luật
Tổng số cuộc phỏng vấn: 15 cuộc phỏng vấn sâu bán cấu trúc

A. CĂN CỨ PHÁP LÝ VÀ ĐẠO ĐỨC NGHIÊN CỨU

Toàn bộ hoạt động phỏng vấn chuyên gia trong khuôn khổ đề tài nghiên cứu này được thực hiện theo các nguyên tắc đạo đức nghiên cứu sau:

Về sự đồng thuận: mỗi chuyên gia tham gia đã được cung cấp đầy đủ thông tin mô tả mục đích nghiên cứu, cách thức sử dụng dữ liệu và cơ chế bảo mật thông tin trước khi cuộc phỏng vấn bắt đầu.

Về bảo mật danh tính: toàn bộ thông tin nhận dạng cá nhân bao gồm họ tên, tên tổ chức, chức danh cụ thể và bất kỳ chi tiết nào có thể dùng để suy ra danh tính được mã hóa và thay thế bằng ký hiệu chuyên gia theo hệ thống mã hóa mô tả tại Mục B dưới đây. Danh sách tương ứng giữa ký hiệu và danh tính thật được lưu trữ tách biệt, trong tệp được mã hóa, chỉ nghiên cứu sinh có quyền truy cập.

B. HỆ THỐNG MÃ HÓA CHUYÊN GIA

Ký hiệu chuyên gia được cấu trúc theo định dạng **CG-[Nhóm]-[Số thứ tự],** trong đó **CG** = Chuyên gia; **Nhóm** = Lĩnh vực chuyên môn chính bao gồm 4 nhóm (**CS** = Chính sách và lập pháp; **DN** = Doanh nghiệp công nghệ / thực hành pháp luật tư nhân; **HT** = Học thuật - pháp lý, luật học; **NC** = Nghiên cứu - viện nghiên cứu, tư vấn độc lập); **Số thứ tự** = 01 đến 03 hoặc 01 đến 04 tuỳ số lượng chuyên gia trong nhóm.

C. MÔ TẢ MẪU PHỎNG VẤN

C.1. Tiêu chí và lý do lựa chọn chuyên gia

C.1.1. Tiêu chí lựa chọn

Việc lựa chọn chuyên gia áp dụng phương pháp chọn mẫu có chủ đích theo tiêu chí, không áp dụng chọn mẫu ngẫu nhiên vì mục tiêu nghiên cứu đòi hỏi thông tin chuyên sâu từ người có kinh nghiệm trực tiếp, không phải thông tin đại diện cho một tổng thể. Bốn tiêu chí lựa chọn sau đây được áp dụng đồng thời.

Thứ nhất, mức độ liên quan trực tiếp đến hoạt động xây dựng hoặc thực thi khung pháp lý AI: ưu tiên người có tham gia trực tiếp (soạn thảo, thẩm định, thẩm tra, tư vấn chính sách) hoặc chịu tác động trực tiếp và có kinh nghiệm thực tiễn đáng kể (doanh nghiệp AI, tư vấn pháp lý chuyên trách công nghệ) hơn người chỉ có kiến thức gián tiếp qua nghiên cứu tài liệu.

Thứ hai, đa dạng vị trí trong hệ sinh thái xây dựng pháp luật: bảo đảm chuyên gia được lựa chọn bao phủ đủ bốn nhóm chủ thể có lợi ích và góc nhìn khác biệt về khung

pháp lý AI (cơ quan chính sách và lập pháp, doanh nghiệp và thực hành pháp luật tư nhân, học thuật pháp lý và viện nghiên cứu/tư vấn độc lập) nhằm tránh thiên lệch do chỉ thu thập ý kiến từ một phía có lợi ích đồng nhất.

Thứ ba, mức độ chuyên sâu về công nghệ-pháp lý: trong mỗi nhóm, ưu tiên người có kinh nghiệm kết hợp cả hiểu biết pháp lý lẫn hiểu biết kỹ thuật AI ở mức tối thiểu vì đây chính là năng lực mà luận án xác định là thiếu hụt cơ cấu quan trọng nhất, nên bản thân người được phỏng vấn cần có đủ nền tảng để đánh giá vấn đề này một cách có căn cứ.

Thứ tư, khả năng tiếp cận thông tin đặc thù: một số chuyên gia được lựa chọn không chỉ vì vị trí công tác mà vì có quyền tiếp cận thông tin cụ thể cần thiết cho luận án.

C.1.2. Lý do lựa chọn

Quy mô mẫu 15 chuyên gia được xác định theo nguyên tắc bão hòa thông tin, áp dụng cho nghiên cứu phỏng vấn giới tinh hoa/chuyên gia. Đây là loại hình phỏng vấn định tính có đặc điểm khác biệt với phỏng vấn đối tượng đại chúng do số lượng người có đủ điều kiện tiếp cận thông tin chuyên sâu về một lĩnh vực hẹp (xây dựng khung pháp lý AI Việt Nam giai đoạn 2017-2026) vốn dĩ hạn chế. Kinh nghiệm nghiên cứu định tính cho thấy bão hòa chủ đề trong phỏng vấn giới tinh hoa thường đạt được trong khoảng 12-20 phỏng vấn khi mẫu được phân tầng theo nhóm chủ thể có sự khác biệt về vai trò. Quyết định dừng ở 15 chuyên gia được đưa ra sau khi quan sát thấy các phỏng vấn ở cuối mỗi nhóm không còn tạo ra chủ đề mới có ý nghĩa mà chủ yếu củng cố hoặc bổ sung chi tiết cho các chủ đề đã xuất hiện từ các phỏng vấn trước đó trong cùng nhóm.

C.1.3. Cơ cấu theo nhóm chuyên môn

Nhóm	**Mô tả chung**	**Số lượng**
CG-CS	Chuyên gia chính sách pháp luật với kinh nghiệm tham gia xây dựng văn bản quy phạm pháp luật liên quan đến công nghệ thông tin, chuyển đổi số hoặc trực tiếp tham gia quá trình soạn thảo, góp ý Luật AI 2025	4
CG-DN	Chuyên gia pháp lý doanh nghiệp công nghệ hoặc đại diện có kinh nghiệm thực tiễn về tuân thủ pháp luật AI trong môi trường kinh doanh	4
CG-HT	Giảng viên tại các cơ sở giáo dục đại học với chuyên môn về luật dân sự, luật quốc tế, sở hữu trí tuệ, quyền con người trong kỷ nguyên số	4
CG-NC	Nhà nghiên cứu độc lập, chuyên gia viện nghiên cứu chính sách với kinh nghiệm nghiên cứu về AI, kinh tế số hoặc quản trị công nghệ	3
Tổng		**15**

Cơ cấu 4-4-4-3 giữa bốn nhóm phản ánh chủ đích của nghiên cứu, không phải tỷ lệ đại diện của bốn nhóm trong thực tế: nhóm Chính sách - Lập pháp được phân bổ đồng đều (4 người) với nhóm Doanh nghiệp và Học thuật vì ba nhóm có thông tin trực tiếp nhất về quá trình soạn thảo - đối tượng trung tâm của luận án; nhóm Nghiên cứu có ít hơn 1 người để bảo đảm mỗi góc nhìn đều được đại diện đầy đủ mà không làm phình to quy mô mẫu vượt quá khả năng phân tích chuyên sâu bằng phương pháp định tính.

Các đặc điểm chung của mẫu

Về phân bổ giới tính: 13 chuyên gia nam và 02 chuyên gia nữ. Về số năm kinh nghiệm: từ 06 đến 23 năm trong lĩnh vực chuyên môn chính, trung bình là 14 năm. Về phạm vi kinh nghiệm địa lý: tất cả chuyên gia đều có kinh nghiệm làm việc với bối cảnh pháp lý Việt Nam; trong đó 02 chuyên gia có kinh nghiệm quốc tế trực tiếp về quản trị AI tại khu vực ASEAN, châu Á-Thái Bình Dương hoặc cấp độ toàn cầu.

Hạn chế về cơ cấu mẫu này được thiết kế để tối đa hóa *chiều sâu và đa dạng góc nhìn*, không phải để bảo đảm *tính đại diện thống kê* cho bất kỳ tổng thể nào (toàn bộ cán bộ hoạch định chính sách AI, toàn bộ doanh nghiệp AI Việt Nam, toàn bộ học giả pháp lý). Kết quả phỏng vấn vì vậy không thể được diễn giải theo nghĩa "X% chuyên gia cho rằng..." như một tỷ lệ suy rộng cho tổng thể, mà chỉ có thể được diễn giải theo nghĩa "trong số 15 chuyên gia được phỏng vấn, có N người nêu quan điểm...".

C.2. Cách thức tiếp cận chuyên gia

Việc tiếp cận chuyên gia được thực hiện theo phương thức kết hợp hai kênh. *Kênh thứ nhất* là tiếp cận trực tiếp chuyên gia (áp dụng cho phần lớn chuyên gia) kèm thư mời có nêu rõ mục đích nghiên cứu, phạm vi câu hỏi dự kiến, cam kết bảo mật danh tính và quyền từ chối trả lời bất kỳ câu hỏi nào. *Kênh thứ hai* là phương pháp chọn mẫu có kiểm soát, sau mỗi phỏng vấn, người nghiên cứu đề nghị chuyên gia giới thiệu người khác có kinh nghiệm phù hợp tiêu chí, đồng thời chủ động kiểm tra để tránh hiện tượng toàn bộ mẫu tự giới thiệu trong một mạng lưới quan hệ hẹp có cùng quan điểm bằng cách chủ động bổ sung liên hệ độc lập không qua giới thiệu cho ít nhất một nửa số chuyên gia trong mỗi nhóm.

D. PHƯƠNG PHÁP VÀ QUY TRÌNH THU THẬP DỮ LIỆU

D.1. Thời gian và hình thức phỏng vấn

Phỏng vấn được thực hiện trong khoảng thời gian từ tháng 2 đến tháng 5/2026 là giai đoạn được lựa chọn có chủ đích vì đây là thời điểm sau khi Luật AI 2025 có hiệu lực (01/3/2026) và giai đoạn đầu Nghị định 142 có hiệu lực (01/5/2026), cho phép thu thập đồng thời cả đánh giá về quá trình soạn thảo Luật lẫn kỳ vọng/lo ngại về giai đoạn thực thi sắp tới.

Hình thức phỏng vấn kết hợp trực tiếp (đối với chuyên gia tại Hà Nội và TP. Hồ Chí Minh có thể sắp xếp gặp mặt) và trực tuyến qua nền tảng gọi video (đối với chuyên gia ở xa hoặc theo yêu cầu thuận tiện của người được phỏng vấn), với thời lượng mỗi phỏng vấn từ 60 đến 90 phút. Toàn bộ phỏng vấn sử dụng năm chủ đề định hướng lớn, phân tầng từ tổng quát đến cụ thể, trong đó các câu hỏi được điều chỉnh linh hoạt theo nhóm chuyên gia và theo diễn biến thực tế của cuộc trò chuyện. Đặc trưng của phỏng vấn sâu bán cấu trúc là cho phép người phỏng vấn đặt câu hỏi tiếp nối khi phát hiện hướng trả lời có giá trị thông tin cao, do vậy không có bộ câu hỏi dùng chung cho tất cả chuyên gia.

Hạn chế ở sự khác biệt về hình thức (trực tiếp so với trực tuyến) có khả năng ảnh hưởng đến chất lượng và độ cởi mở của thông tin thu thập được. Nghiên cứu phương pháp luận định tính ghi nhận phỏng vấn trực tuyến đôi khi hạn chế khả năng quan sát phi ngôn ngữ và có thể làm giảm mức độ chi tiết của câu trả lời tự phát. Hạn chế này được kiểm soát một phần bằng cách bảo đảm mọi phỏng vấn bất kể hình thức đều tuân theo cùng chủ đề định hướng và cùng quy trình ghi âm, phiên mã như nhau.

D.2. Chủ đề định hướng

Các câu hỏi được tổ chức theo năm chủ đề lớn, phân loại từ tổng quát đến cụ thể:

Chủ đề 1 - Đánh giá bức tranh toàn cảnh AI: Quan điểm cá nhân của chuyên gia về AI trong thời đại số, thực trạng phát triển, ứng dụng AI ở Việt Nam, so với 5 năm trước và tầm nhìn 5-10 năm tới.

Chủ đề 2 - Đánh giá tổng thể khung pháp lý hiện hành: Nhận định chung của chuyên gia về mức độ đầy đủ, đồng bộ và tính khả thi của khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam tính đến thời điểm phỏng vấn; so sánh với kinh nghiệm quốc tế mà chuyên gia có tham chiếu.

Chủ đề 3 - Đánh giá việc xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam: Đánh giá chuyên sâu căn cứ vào chuyên môn và kinh nghiệm cụ thể của từng chuyên gia.

Chủ đề 4 - Khoảng trống và thách thức: Xác định các khoảng trống và thách thức thực tiễn quan trọng, nguyên nhân gốc rễ, hậu quả nếu không được giải quyết.

Chủ đề 5 - Đề xuất và định hướng hoàn thiện: Khuyến nghị cụ thể về nội dung cần hoàn thiện trong xây dựng khung pháp lý.

D.3. Quy trình phân tích dữ liệu

Dữ liệu phỏng vấn được phân tích theo phương pháp phân tích chủ đề kết hợp với phân tích nội dung định tính, theo quy trình gồm sáu bước: làm quen với dữ liệu, tạo mã ban đầu, tìm kiếm chủ đề, xem xét chủ đề, định nghĩa và đặt tên chủ đề và trình bày kết quả.

D.3.1. Xử lý ý kiến khác biệt và phân kỳ

Nguyên tắc xử lý ý kiến khác biệt được luận án tuân thủ nhất quán là: *ý kiến thiểu số hoặc phân kỳ không bị loại bỏ khỏi báo cáo kết quả để tạo ra bức tranh đồng thuận giả tạo* mà được trình bày như một phần cấu thành của kết quả nghiên cứu. Ba loại phân kỳ được xử lý theo cách thức khác nhau tùy bản chất.

Phân kỳ giữa các nhóm chuyên gia (ví dụ nhóm Chính sách - Lập pháp ưu tiên tốc độ ban hành văn bản trong khi nhóm Học thuật ưu tiên chiều sâu chẩn đoán bằng chứng) được trình bày như bằng chứng về sự khác biệt góc nhìn hợp lý xuất phát từ vị trí và trách nhiệm khác nhau của từng nhóm, không phải như một bên đúng và một bên sai và luận án cho rằng tính đa chiều thực sự của vấn đề cần được cân nhắc đồng thời cả hai góc nhìn khi thiết kế giải pháp.

Phân kỳ trong cùng một nhóm chuyên gia được xử lý bằng cách trình bày cả hai lập luận đầy đủ kèm căn cứ và người nghiên cứu tránh việc ngầm ủng hộ quan điểm này hơn quan điểm kia chỉ vì nó phù hợp hơn với giả thuyết ban đầu của luận án.

Ý kiến đơn lẻ không được xác nhận bởi chuyên gia khác được xử lý thận trọng nhất, được ghi nhận trong Biên bản tổng hợp nhưng không được nâng lên thành "phát hiện chung" trong phần phân tích chính. Khi được sử dụng làm bằng chứng minh họa ở Chương 3 hoặc Chương 4 luôn gắn kèm nguồn cụ thể (ví dụ "chuyên gia CG-DN-03 mô tả...") thay vì được khái quát hóa thành "các chuyên gia cho rằng...".

D.3.2. Kiểm chứng thành viên

Sau khi hoàn thành phiên mã và trước khi đưa trích dẫn cụ thể vào Biên bản tổng hợp và vào nội dung luận án, người nghiên cứu gửi lại cho từng chuyên gia được phỏng vấn bản tóm tắt nội dung trả lời của họ (không phải toàn bộ bản phiên mã, để tôn trọng thời

gian của chuyên gia) kèm các trích dẫn trực tiếp dự kiến sử dụng, đề nghị xác nhận tính chính xác hoặc đề nghị điều chỉnh nếu có sai lệch về ý nghĩa so với điều họ muốn truyền đạt. Quy trình kiểm chứng thành viên này phục vụ hai mục đích: bảo đảm độ chính xác của việc diễn giải (tránh người nghiên cứu hiểu sai hoặc trích dẫn ngoài ngữ cảnh) và tăng cường tính chính đáng đạo đức của nghiên cứu (chuyên gia có quyền kiểm soát cách phát ngôn của mình được trình bày công khai).

D.3.3. Kiểm soát thiên lệch của người nghiên cứu

Do nghiên cứu định tính không tách rời hoàn toàn vai trò của người nghiên cứu khỏi quá trình tạo ra và diễn giải dữ liệu, luận án áp dụng ba biện pháp kiểm soát thiên lệch tương ứng với thực hành phản tư trong nghiên cứu định tính.

Biện pháp thứ nhất, ghi chép nhật ký phản tư: trong suốt quá trình phỏng vấn và phân tích, người nghiên cứu ghi chép nhật ký phản tư định kỳ, ghi nhận các giả định cá nhân, phản ứng cảm xúc và những điểm có khả năng ảnh hưởng đến cách đặt câu hỏi hoặc cách diễn giải.

Biện pháp thứ hai, thiết kế câu hỏi tránh dẫn dắt: bộ câu hỏi định hướng được thiết kế theo nguyên tắc câu hỏi mở, tránh đưa giả định của người nghiên cứu vào câu hỏi.

Biện pháp thứ ba, chủ động tìm kiếm bằng chứng phản bác: sau khi các chủ đề chính đã hình thành từ đa số phỏng vấn, người nghiên cứu chủ động rà soát lại toàn bộ dữ liệu để tìm kiếm các trường hợp ngoại lệ hoặc mâu thuẫn với chủ đề đã hình thành.

E. TỔNG HỢP NỘI DUNG PHỎNG VẤN THEO CHỦ ĐỀ

E.1. Chủ đề 1 - Đánh giá bức tranh toàn cảnh AI

Nhận định hội tụ

Tất cả chuyên gia đều đánh giá AI là công nghệ tiên tiến sẽ làm thay đổi thế giới với tốc độ phát triển vượt bậc, nhanh hơn rất nhiều so với chúng ta tưởng tượng. Phần lớn chuyên gia (12/15) coi AI là công cụ, do đó vẫn cần sự điều phối, quản lý, giám sát của con người.

Về thực trạng phát triển, ứng dụng AI, chuyên gia CG-DN-01 nhận định các công nghệ công ty công nghệ lớn ở Việt Nam đều có năng lực phát triển AI, gần như tất cả mọi ngành nghề và lĩnh vực ở Việt Nam hiện nay đều đã áp dụng AI, phổ biến nhất dưới dạng chatbot. Đã có một số thay đổi nhất định về cách sử dụng công nghệ, thay vì trước đây là Google thì bây giờ là ChatGPT. Dẫn đầu nghiêng về lĩnh vực công nghệ vì công ty cung cấp phần mềm thường ứng dụng hoặc tích hợp AI vào trong sản phẩm luôn. Thứ hai, về giao thông rất phổ biến phạt nguội và phát hiện những hành vi vi phạm thông qua các camera AI gắn ở trên đường phố. Những lĩnh vực khác sẽ ứng dụng AI vào công việc hằng ngày, ví dụ trong ngành giáo dục, chống đạo văn là một vấn đề quan ngại, phải nâng cao nhận thức cho học sinh làm sao để ứng dụng AI một cách công bằng. Tiếp theo về nhân sự, việc sử dụng AI để screening CV rất phổ biến. Công nghệ và nguồn lực không phải rào cản mà chính xác hơn phải là việc tuân thủ đạo đức, cách thức sử dụng AI, không lạm dụng AI, không khiến AI thay thế hoàn toàn con người. So với 5 năm trước về mặt công nghệ đã phát triển với tốc độ chóng mặt: ChatGPT nổi lên vào khoảng cuối năm 2023. Trước đó gần như mọi người đều chưa biết đến AI nhưng ngay khi ChatGPT ra đời đã thay đổi góc

nhìn của mọi người và đã có rất nhiều model khác được ra đời với sự tham gia của rất nhiều tập đoàn công nghệ lớn. Còn về mặt pháp lý, 5 năm trước Việt Nam chỉ mới có nghị định bảo vệ dữ liệu năm 2023 trong khi GDPR của châu Âu thì đã có từ năm 2018. So sánh 5 năm trước với thời điểm hiện tại là một khoảng cách rất lớn về mọi mặt tại thị trường Việt Nam. Còn tầm nhìn từ 5 đến 10 năm sau, nhà nước Việt Nam vẫn đang định hướng sẽ đẩy mạnh và ưu tiên phát triển công nghệ AI là một trong những lĩnh vực trọng tâm. Ví dụ như việc xây dựng các trung tâm về dữ liệu, các trung tâm hoặc lab về AI được đặt ở khu công nghệ cao. Về pháp lý đi theo việc ban hành luật về AI thì các văn bản dưới luật, nghị định, thông tư hướng dẫn hoặc thông tư của các bộ ngành sẽ dần dần được ban hành và thực thi. Về việc AI quá phổ biến có thể thay thế một số ngành nghề và AI trở thành một kỹ năng bắt buộc bên cạnh tin học văn phòng.

Chuyên gia CG-CS-02 đánh giá doanh nghiệp là tầng lớp năng động, thực tế nhất nên thường ứng dụng công nghệ đầu tiên. Trong thực tiễn kinh doanh, AI là một áp lực cạnh tranh chứ không chỉ là kiểu khẩu hiệu tuyên truyền nữa. Đặc biệt là doanh nghiệp nhỏ và vừa có mức độ thích ứng rất cao so với các nước. Thứ hai, Việt Nam có những lợi thế để ứng dụng AI. Một là tỷ lệ người có smartphone cao, chất lượng Internet, hạ tầng viễn thông của Việt Nam rất tốt. Chính vì thế việc ứng dụng rất nhanh và dân số Việt Nam rất là trẻ. Rào cản là nhiều người dùng AI chưa biết cách bảo mật như thế nào khi up tài liệu lên ứng dụng AI cũng là một rủi ro, nguy cơ. Có những hiệu ứng, tác động chưa nhìn thấy được và sẽ có những hệ quả về sau cho nên cách tiếp cận khôn ngoan là bộ máy chính quyền phải đi smart hơn so với xã hội.

Chuyên gia CG-NC-02 cho rằng AI đang đóng vai trò quan trọng trong việc tăng hiệu quả công việc, nhưng việc sử dụng còn tự phát do chưa có khung pháp lý rõ ràng (Luật AI mới có hiệu lực từ 1/3/2026). Có sự phân hóa người dùng thành 2 nhóm: (1) Nhóm tích cực: Những người có năng lực chuyên môn giỏi và ham học hỏi trở nên xuất chúng hơn nhờ AI; (2) Nhóm tiêu cực: Đại đa số có xu hướng trở nên lười nhác tư duy, thậm chí dẫn đến tình trạng "rỗng não", "thối não" vì lạm dụng AI mà không có định hướng. Dẫn chứng vòng luẩn quẩn trong giáo dục: tình trạng giáo viên dùng AI ra đề, học sinh dùng AI làm bài và giáo viên lại dùng AI chấm bài, thể hiện sự lạm dụng và thiếu tư duy. Triết lý của thời đại mới: Luôn hướng tới việc "Làm việc cùng AI, không bị lệ thuộc AI".

Tính bảo mật là rào cản lớn nhất đối với việc phát triển và ứng dụng AI, đặc biệt đối với các ứng dụng yêu cầu an ninh cao như phòng họp không giấy tờ. Việc phải đưa dữ liệu lên Cloud để xử lý nhưng lại thiếu hạ tầng viễn thông và giải pháp bảo mật đủ mạnh là một vấn đề nan giải. Tiếp theo là đạo đức dữ liệu và quyền riêng tư, đây là mảng hổng rất lớn trong cộng đồng AI Việt Nam, chia làm hai thái cực: (1) Nhiều người sử dụng AI vô tội vạ mà không quan tâm đến bảo mật, coi như "nhà không có khóa cửa". (2) Ngược lại, có nhóm người sợ hãi quá mức, không dám sử dụng AI dù không hiểu rõ về bảo mật. Nguyên tắc đơn giản: "Hãy tưởng tượng bạn đưa thông tin đó lên Facebook ở chế độ public. Nếu bạn thấy ok thì hãy đưa lên AI, nếu không thì không nên".

Nhận định phân kỳ

Chuyên gia nhóm doanh nghiệp cho rằng AI tại thời điểm phỏng vấn đã bắt đầu hoạt động độc lập, AI không chỉ là công cụ mà là cộng sự. Ví dụ dẫn chứng là Claude Cowork và sản phẩm doanh nghiệp đang cung cấp theo hướng dùng AI thay thế con người trong tất cả các phòng ban, lãnh đạo chỉ cần kiểm tra 1-2 lần/ngày nếu có nhu cầu. Trong tương lai gần, AI được dự đoán sẽ tự động hoàn toàn, không cần sự giám sát của con người. Thậm chí trong nhiều trường hợp, AI làm việc còn chính xác hơn con người do không phải chịu sự chi phối về mặt cảm xúc.

Chuyên gia nhóm học thuật nhìn nhận vấn đề chủ thể không nhất thiết phải là một chủ thể sinh học nhất định. AI không phải không thể là chủ thể, câu chuyện đáng bàn hơn là vì sao chúng ta cần cấp chủ thể? Bản chất cấp tư cách chủ thể là cấp khả năng tham gia quan hệ độc lập, một sản nghiệp riêng làm cơ sở kinh tế để ra quyết định, tổ chức đời sống độc lập, không phụ thuộc các thành viên khác, chịu trách nhiệm toàn bộ về tài sản đó và kéo theo hệ quả về tố tụng là một chủ thể độc lập có thể đi kiện và bị kiện. Nếu hiểu theo nghĩa đó thì AI bình thường không được nhìn nhận là một chủ thể, nhưng nếu gắn lên một phần cứng nữa ví dụ robot thông minh thì có nên thừa nhận tư cách chủ thể không?

Trích dẫn đại diện theo nhóm quan điểm

Chuyên gia CG-NC-02 khẳng định AI không thể và không nên hoạt động hoàn toàn độc lập mà không cần sự giám sát của con người. Rủi ro từ hệ thống tự động: Các hệ thống tự động (agentic AI) vẫn đang gây tranh cãi và tiềm ẩn nhiều rủi ro về bảo mật, mất kiểm soát. Đặc biệt tại Việt Nam, sự can thiệp và giám sát của con người là vô cùng cần thiết để đảm bảo hiệu quả và an toàn.

Chuyên gia CG-DN-01 cho rằng AI có nhiều ứng dụng và nhiều dạng mô hình. Physical AI có robot hoặc các ứng dụng thông minh, ví dụ như smart home. Còn có những dạng khác như là computer vision về hình ảnh, nổi tiếng nhất hiện nay là những mô hình ngôn ngữ lớn. Tùy thuộc vào dạng AI nào, ứng dụng như thế nào, tại thời điểm nào sẽ có cách nhìn nhận khác nhau. AI đang hỗ trợ cho các công việc của con người, đa phần mang tính chất khuyến nghị. Đôi khi nó sẽ làm lu mờ việc ai sẽ phải quyết định phần nào, ai làm việc gì. Nhưng AI không phải là chủ thể độc lập.

E.2. Chủ đề 2 - Đánh giá tổng thể khung pháp lý hiện hành

Đa số chuyên gia (12/15) đồng thuận rằng "khung pháp lý AI" có nội hàm rộng hơn "hệ thống văn bản QPPL về AI". Tuy nhiên, có sự khác biệt đáng kể giữa các nhóm về mức độ nhận thức thực tiễn về các tầng phi văn bản.

Chuyên gia nhóm Chính sách - lập pháp CS-03 xác nhận nhận thức học thuật nhưng chỉ ra khoảng cách với thực tiễn: *"Trong nghiên cứu chiến lược, chúng tôi hiểu rằng khung pháp lý bao gồm cả tiêu chuẩn kỹ thuật, luật mềm và án lệ. Nhưng trong thực tiễn thẩm định văn bản, công cụ và thẩm quyền của chúng tôi dừng lại ở tầng văn bản QPPL. Đây không phải vấn đề nhận thức mà là vấn đề phân công chức năng."*

Đại diện nhóm Doanh nghiệp, chuyên gia CG-DN-01 mô tả trải nghiệm thực tế từ góc độ doanh nghiệp AI: *"Chúng tôi phải tuân thủ cùng lúc Luật AI 2025, Nghị định 142,*

quy định của NHNN, Luật BVDLCN và một loạt hướng dẫn khác. Bộ phận pháp chế của chúng tôi mất ba tuần chỉ để xác định hệ thống nghĩa vụ áp dụng cho sản phẩm AI của mình". Còn chuyên gia DN-03 nhấn mạnh vấn đề của startup: *"Chúng tôi là startup không thể thuê luật sư chuyên về AI do không đủ tài chính. Đây là khó khăn rất lớn trong quá trình tuân thủ các quy định về AI"*.

Chuyên gia nhóm Học thuật CG-HT-04 cung cấp phân tích từ góc độ lý luận, 'khung pháp lý' là khái niệm rộng hơn 'khung pháp luật'. Pháp luật Việt Nam đang dần thừa nhận điều này qua quy định về án lệ và lẽ công bằng (Điều 6 BLDS 2015). Vấn đề là các văn bản AI hiện nay chưa được thiết kế có ý thức về mối quan hệ giữa các nguồn pháp luật khác nhau.

Về VBQPPL chuyên biệt và liên quan, tất cả 15 chuyên gia tham gia phỏng vấn đều đánh giá việc ban hành Luật AI 2025 là bước tiến pháp lý có tính bước ngoặt, đặc biệt trong bối cảnh hệ thống pháp luật trước đó thiếu trục kết nối xuyên ngành. Chuyên gia CG-CS-01 đặt vấn đề theo cách sau: Luật AI 2025 quan trọng không chỉ là một khung pháp lý đơn thuần mà còn là một "tuyên ngôn tầm nhìn" chiến lược, giải quyết bài toán khó nhất trong quản trị công nghệ mới: Tìm ra điểm cân bằng tinh tế giữa việc mở đường cho đổi mới sáng tạo và việc thiết lập các lằn ranh đỏ để kiểm soát rủi ro.

Thuận lợi cho Việt Nam là một trong những quốc gia tiên phong ban hành luật AI, thể hiện sự nhạy bén của chính phủ. Nước ta cũng có lợi thế kế thừa và tham khảo luật AI từ các quốc gia khác nhờ đội ngũ nhà khoa học Việt Nam đang làm việc trên khắp thế giới. Chuyên gia CG-CS-02 cho rằng Việt Nam ban hành Luật AI khẳng định đây là lĩnh vực quan trọng, thể hiện quá trình chuyển đổi hiện nay phát triển lên kinh tế tri thức, tăng trưởng dựa trên khoa học công nghệ, lấy khoa học công nghệ làm động lực, làm trọng tâm. Vì luật AI gửi thông điệp như vậy cho nên sắp tới Việt Nam sẽ dành nguồn lực về hạ tầng, về nhân lực, về đào tạo nhân lực về chính sách cho lĩnh vực này phát triển. Lợi thế là định hình, định danh được và bộ máy quản lý cũng hình dung được trách nhiệm của mình, tránh tình trạng không ai chịu trách nhiệm. Hệ quả là sẽ có thể có những kìm hãm cho sự phát triển của AI nếu quản lý quá chặt, quá thận trọng. Đây cũng là một quá trình đánh đổi vừa lợi thế mà vừa không có lợi thế.

Chuyên gia CG-DN-01 nhận định các yếu tố bên ngoài liên quan đến mức độ phổ biến của AI cần thiết phải có một khung pháp lý hoặc một cơ chế nào đó để quản lý việc phát triển, sử dụng, nghiên cứu AI trong thực tế tại Việt Nam. Vì luật AI đã trở nên phổ biến ở các khu vực khác trên thế giới, Việt Nam cũng có một khung pháp lý đáng tin cậy sẽ trở thành một trong những lợi thế cho Việt Nam để tự quảng bá là một môi trường đầu tư tốt và thu hút thêm các doanh nghiệp nước ngoài, thu hút thêm vốn đầu tư về công nghệ. Yếu tố bên trong do Việt Nam đã có định hướng rất rõ ràng về việc ưu tiên các chính sách phát triển công nghệ, xây dựng hạ tầng phát triển công nghệ và thu hút vốn đầu tư nước ngoài, khuyến khích các doanh nghiệp vừa và nhỏ, khởi nghiệp trong lĩnh vực công nghệ, đặc biệt liên quan đến AI được hưởng những chính sách ưu đãi.

Theo chuyên gia CG-DN-04, thuận lợi là quyết tâm cao, có sự đồng thuận từ cấp cao nhất của Đảng, Nhà nước, Chính phủ trong việc ban hành luật này.

Chuyên gia CG-HT-01 đánh giá chúng ta đang đi theo xu hướng của thế giới là đưa ra khung quy tắc chung đảm bảo an toàn, an ninh trong việc phát triển AI. Trước khi có luật AI, việc ứng dụng và phát triển AI không phải là không được điều chỉnh mà vẫn có thể được điều chỉnh bởi các quy định của pháp luật thông thường. Trong luật AI cũng có quy định khi ứng dụng AI vào các lĩnh vực cụ thể như an toàn thông tin mạng, bảo vệ dữ liệu cá nhân, sở hữu trí tuệ vẫn phải tuân thủ pháp luật của các lĩnh vực cụ thể đó. Cho nên luật AI hiện nay giống như là một biện pháp bổ sung cho các luật hiện hành đang điều chỉnh tất cả các vấn đề cụ thể nói chung.

Bên cạnh đó, thách thức đặt ra theo chuyên gia CG-DN-01, khó khăn bên trong là bối cảnh năng lực của Việt Nam hiện tại để có thể áp dụng một đạo luật về công nghệ chuyên sâu với rất nhiều yêu cầu kỹ thuật thì mọi người chưa sẵn sàng, các doanh nghiệp chưa sẵn sàng và chưa có thói quen về điều đó. Những nước đã ban hành luật AI bao gồm khu vực châu Âu nói chung và Hàn Quốc có một lợi thế là họ đã có khung pháp lý về dữ liệu rất sớm giúp cho doanh nghiệp quen với việc báo cáo, đánh giá tác động, tuân thủ… nói chung là cẩn trọng hơn khi sử dụng, áp dụng những công nghệ mới vào trong thực tiễn của doanh nghiệp hoặc trong đời sống của mọi người. Giống như GDPR trở thành một trong những thông lệ quốc tế quá phổ biến và được sử dụng rất nhiều nên các nước khác cũng phải dần dần siết chặt lại về vấn đề bảo vệ dữ liệu cá nhân và phải thực hiện đánh giá tác động, ký hợp đồng data processing để đảm bảo việc xử lý các dữ liệu này không ảnh hưởng đến nhân quyền. Khi ban hành luật AI dựa trên đó thì trong luật AI của họ có những chi tiết ví dụ như nếu hệ thống của bạn đã đáp ứng các tiêu chuẩn của châu Âu về ngành cụ thể trong lĩnh vực y tế, xe hơi, linh kiện điện tử thì coi như cũng đã đáp ứng một phần hoặc là thậm chí là đầy đủ các yêu cầu về kỹ thuật của đạo luật AI. Nhưng ở Việt Nam chưa hình thành một nhận thức chung hoặc thực hành chung, chỉ sau năm 2023 mọi người mới dần biết đến khái niệm đánh giá tác động dữ liệu khi chuyển dữ liệu ra nước ngoài, đó là một trong những tài liệu đầu tiên mà người Việt Nam phải chuẩn bị khi xử lý các loại dữ liệu xuyên biên giới. Nhưng luật AI Việt Nam đang áp khá nặng các yêu cầu kỹ thuật ví dụ như việc phải ghi lại nhật ký hoạt động của hệ thống, đánh giá rủi ro, ghi lại cơ chế quản lý rủi ro hệ thống, đánh giá được tác động và chứng nhận được sự phù hợp của hệ thống AI ở Việt Nam. Đây đều là những thuật ngữ mới đối với doanh nghiệp Việt Nam, đặc biệt là doanh nghiệp Việt Nam chưa từng làm việc với các thị trường quốc tế sẽ trở thành khó khăn.

Chuyên gia CG-DN-01 cho rằng thực tế quá ít thông tin để tham khảo vì chỉ có vài nước đã ban hành luật như châu Âu, Nhật, Hàn Quốc. Mỹ không có luật chung mà có một số đạo luật của bang và cũng đã bỏ đi khá nhiều, chủ yếu là các quy tắc. Trung Quốc cũng chưa ban hành thành luật, tức là những nước phát triển nhất về AI trên thế giới chưa đưa thành luật. Đó là một khó khăn cho đơn vị xây dựng luật AI ở Việt Nam hiện nay. Hơn nữa AI thay đổi rất nhanh nên có thể có nhiều thứ xây dựng xong đã bị lạc hậu.

Cũng có sự đồng thuận cao về khoảng cách giữa văn bản pháp lý và năng lực thực thi. Khi đã có luật rồi vấn đề là luật đi vào cuộc sống đến đâu, cần thời gian để kiểm chứng. Chuyên gia CG-NC-02 nhận định: Việc luật pháp quy định tiêu chuẩn năng lực AI cho từng nhóm chức danh (nhà phát triển, người triển khai, người ra quyết định) là bài toán "quá khó" và "ngược". Thiếu chuyên gia đủ năng lực: Không ai đủ năng lực để xây dựng khung năng lực này khi thực tiễn chưa diễn ra.

Về ba yếu tố chưa được xây dựng đầy đủ, kết quả tổng hợp từ 15 chuyên gia cho thấy:

Yếu tố được đề cập nhiều nhất (12/15): Tầng tiêu chuẩn kỹ thuật. Chuyên gia CS-03, CS-04, DN-01, DN-02, DN-03, DN-04, HT-01, HT-02, NC-03 và ba chuyên gia khác cùng đề cập.

Yếu tố được đề cập nhiều thứ hai (11/15): Cơ chế điều phối liên ngành có thực quyền.

Yếu tố được đề cập nhiều thứ ba (10/15): Cơ chế thích ứng và học hỏi từ thực tiễn (ví dụ điều khoản rà soát bắt buộc, hệ thống phân tích dữ liệu sự cố).

E.3. Chủ đề 3 - Đánh giá việc xây dựng khung pháp lý cho phát triển, ứng dụng AI ở Việt Nam

E.3.1. Chủ thể xây dựng

Phần lớn chuyên gia (11/15) đánh giá các cơ quan nhà nước có thẩm quyền đã rất nỗ lực cho ra đời Luật AI và Nghị định 142.

Tuy nhiên, vẫn còn những thách thức đáng kể. Chuyên gia CG-DN-02 nhận định năng lực kỹ thuật để giám sát AI còn yếu. Cán bộ quản lý thiếu nền tảng để đánh giá hệ thống AI rủi ro cao một cách thực chất. Hệ quả với chính sách là Luật AI nếu thiết kế theo tư duy 'quản lý doanh nghiệp có tổ chức' sẽ bỏ qua hoàn toàn nhóm đang thực sự tạo ra chuyển động thị trường. Cần cơ chế tiếp cận và hướng dẫn tuân thủ đơn giản, chi phí thấp, phù hợp với doanh nghiệp siêu nhỏ và người dùng phi chính thức. Cần bổ sung góc nhìn tiếp cận (accessibility) và trao quyền (empowerment) vào khung quản trị AI, không chỉ góc nhìn kiểm soát rủi ro.

Chuyên gia CG-CS-01 cho rằng nếu các văn bản dưới luật không quy định rõ ràng về thủ tục, thời gian và cơ quan chịu trách nhiệm thẩm định, doanh nghiệp sẽ rơi vào tình trạng "chờ duyệt" thủ tục hành chính, làm lỡ nhịp đổi mới sáng tạo và làm mất lợi thế cạnh tranh của chu kỳ phát triển sản phẩm công nghệ ngắn. Trong xây dựng Luật AI 2025, tầm ảnh hưởng thực chất còn đến từ một số doanh nghiệp công nghệ lớn có quan hệ tốt với Bộ được mời tham vấn nhiều hơn các startup nhỏ. Điều đó tạo ra nguy cơ văn bản phản ánh lợi ích của doanh nghiệp lớn hơn doanh nghiệp nhỏ.

CG-DN-02 đồng tình nêu vấn đề đại diện: *"Trong các hội thảo tham vấn về Luật AI 2025 mà tôi được mời, phần lớn tham gia là doanh nghiệp lớn, hiệp hội ngành và chuyên gia học thuật. Startup AI, freelance developer và doanh nghiệp số hóa vừa áp dụng AI vào quy trình không có kênh tham vấn nhưng họ là nhóm chịu tác động lớn nhất từ gánh nặng tuân thủ".*

Về dữ liệu huấn luyện AI, đây là vấn đề được cả bốn nhóm chuyên gia đề cập như một khoảng trống đáng lo ngại. Chuyên gia CG-NC-01 xác định rào cản lớn nhất cho

doanh nghiệp là dữ liệu, đặc biệt dữ liệu đầu vào cho việc phát triển ứng dụng AI. Dữ liệu này vừa thiếu vừa tản mát rời rạc. Dữ liệu sạch để phát triển AI cũng đang còn ở mức hạn chế.

Chuyên gia CG-DN-04 cho rằng vấn đề bảo vệ dữ liệu còn quan trọng hơn cả AI, vì AI là công nghệ dựa trên rất nhiều về chất lượng dữ liệu. Hiện nay có nhiều cơ quan nhà nước đang bắt đầu tiến trình xây dựng các bộ dữ liệu cho Việt Nam. Nó là một trong những yếu tố quan trọng sẽ quyết định việc thành công của AI ở Việt Nam hay không.

Về mặt khoa học dữ liệu, chuyên gia CG-NC-02 đồng ý rằng ai kiểm soát càng nhiều dữ liệu thì càng có quyền lực. Quan điểm cá nhân ủng hộ việc nhà nước quản lý càng nhiều dữ liệu càng tốt nếu điều đó mang lại lợi ích cho người dân, khác với quan điểm phương Tây coi đây là sự can thiệp thô bạo.

E.3.2. Những vấn đề pháp lý liên quan đến AI

Về phân loại rủi ro, chuyên gia nhóm chính sách và học thuật nhìn chung đánh giá tích cực về hướng tiếp cận dựa trên rủi ro trong Luật AI 2025, nhưng đặt nghi vấn về tính khả thi của hệ thống phân loại trong Dự thảo Quyết định TTg do tiêu chí định lượng chưa thực sự hợp lý.

Theo chuyên gia CG-CS-02, khó khăn nhất là chúng ta quản lý hình thức rủi ro nhưng phân loại rủi ro lấy tiêu chí rủi ro là số lượng người dùng thì có những phần mềm, chương trình AI có số lượng người dùng rất lớn mặc dù thân thiện và người ta đã sử dụng quen rồi bỗng nhiên một ngày bị đưa vào dạng rủi ro cao. Khi đưa vào rủi ro cao thì quản lý chặt hơn, quản lý chặt hơn thì chỉ thấy thủ tục nhiều hơn. Gốc rễ vẫn là làm sao tinh thần của luật và các văn bản pháp luật phải theo hướng thúc đẩy sự phát triển. Việt Nam nên đi sau, học hỏi và khai thác kinh nghiệm của các quốc gia đi trước.

Chuyên gia CG-DN-03 và CG-NC-02 nhận định quản lý dựa trên rủi ro là đúng, nhưng tiêu chí phân loại mức độ rủi ro không phải là yếu tố quan trọng nhất vì công nghệ thay đổi rất nhanh, ngày hôm nay một hệ thống AI thuộc mức độ rủi ro trung bình nhưng ngày mai có thể xếp vào rủi ro cao. Điều quan trọng là doanh nghiệp tự đánh giá, chuẩn bị đầy đủ về mặt kỹ thuật, pháp lý và có phương án dự phòng cho mình.

Chuyên gia CG-NC-01 chỉ ra Dự thảo Quyết định đang áp nghĩa vụ trách nhiệm đối với các giải pháp AI gần giống như rủi ro cao, nên sửa đổi đối với các AI rủi ro trung bình để xác định tiêu chí rõ ràng giữa cao với trung bình. Cũng có ý kiến cho rằng không nên xếp các hệ thống AI theo lĩnh vực mà phải theo cách ứng dụng và tính chất của công cụ kết hợp với nhau. Ví dụ xếp phân loại rủi ro theo số lượng người dùng là không phù hợp bởi vì nhà cung cấp không thể biết sau này có bao nhiêu người dùng. Người dùng cũng có nhiều đối tượng. Ví dụ chỉ toàn người dùng trẻ tuổi, số lượng nhiều thì cũng không phản ánh hết được.

Về trách nhiệm pháp lý và giải trình, chuyên gia CG-DN-01 cho rằng trách nhiệm pháp lý về bản chất cũng là luật dân sự hoặc luật hình sự. Lỗi của ai thì sẽ do người đó chịu trách nhiệm và bồi thường thiệt hại và lỗi đó có thật sự vi phạm theo luật hay không. Khi AI xảy ra lỗi nghĩa là trường hợp vi phạm diễn ra ở nhiều giai đoạn khác nhau của AI

từ quá trình thu thập dữ liệu, sử dụng dữ liệu để training phát triển vì một mục đích nào đó hoặc khi đưa vào sử dụng thì người sử dụng mới là người vi phạm chứ không phải là bên cung cấp mô hình. Ở các hệ thống pháp luật khác nhau, tài phán khác nhau thì cách phán quyết và cân nhắc lỗi nằm ở bên nào sẽ khác nhau nên có thể tham khảo các nước khác nhau và các án lệ trên thế giới.

Chuyên gia CG-HT-01 chia sẻ trên thực tế đối với những chủ thể ứng dụng, phát triển AI là một nguồn nguy hiểm cao độ để đảm bảo nên mua bảo hiểm giảm thiểu thiệt hại. Có nghĩa là các chủ thể sẽ cùng nhau chia sẻ bớt thiệt hại.

Chuyên gia CG-HT-02 nhìn dưới góc độ dân sự về bồi thường thiệt hại phải dần dần giảm bớt quy định. Đã có Bộ Luật Dân sự quy định khung về trách nhiệm bồi thường thiệt hại ngoài hợp đồng, nghị quyết 02 hướng dẫn bộ luật dân sự, một số chế định đặc thù về trách nhiệm sản phẩm, điều cần làm là trên cơ sở nền tảng đó có thể gia cố phát triển cho AI như thế nào.

Về bảo vệ quyền con người và SHTT, chuyên gia CG-DN-01 đánh giá cơ bản Việt Nam khá đầy đủ. Quan trọng là áp dụng thực tế và hướng dẫn thực hành như thế nào. Vấn đề đang thiếu là phần liên quan đến sự an toàn, nghĩa là phải đảm bảo AI không được gây hại, gây hại ở góc độ nào? Ví dụ chỉ gây hại ở góc độ tính mạng con người, sức khỏe hay gây hại ở góc độ xã hội, AI cướp mất việc làm của con người… Ví dụ án lệ về SHTT ở Trung Quốc đã từng xử thắng cho một người sử dụng AI để tạo ra một hình ảnh thì họ cho là quyền IP thuộc về người sử dụng AI, nhưng ở Czech việc này là vi phạm. Còn ở Hoa Kỳ đang có rất nhiều án kiện về việc vi phạm luật bản quyền khi sử dụng những tác phẩm sáng tạo vào việc training AI.

Chuyên gia CG-HT-01 nhận xét trong giai đoạn đầu để phục vụ phát triển AI, việc cho phép sử dụng văn bản và dữ liệu đã được công bố hợp pháp để phục vụ mục đích nghiên cứu khoa học, thử nghiệm hoặc huấn luyện AI không vi phạm quyền sở hữu trí tuệ. Trừ trường hợp người công bố yêu cầu và đăng ký rằng ai muốn sử dụng phải xin phép. Còn nếu như họ không đặt ra yếu tố đó thì do thỏa thuận hợp đồng giữa người sáng lập, người tạo ra sản phẩm với người công bố. Nếu được trong tương lai pháp luật có thể bổ sung vấn đề có trả phí hay không và trong trường hợp nào dùng có trả phí.

Về ứng dụng AI trong khu vực công và quản trị nhà nước, chuyên gia CG-DN-04 cho rằng các ứng dụng AI trong nhà nước là một xu hướng không thể đảo ngược. Nhà nước đang tập trung rất nhiều vào xây dựng các trung tâm dữ liệu, các ứng dụng như hỏi đáp, tự động hóa công việc để phục vụ công chức.

Bên cạnh đó, chuyên gia CG-NC-01 lo ngại việc nhà nước sử dụng các ứng dụng cài đặt trong điện thoại của người dân để giám sát và chấm điểm công dân. Đây là một vấn đề rất đáng phải bàn luận nhưng hiện nay hầu như không được đề cập đến ở các diễn đàn công khai, các diễn đàn lớn. Kể cả các hội thảo, hội nghị của các cơ quan nhà nước, các hội thảo của chuyên gia, các trường, hầu như không ai nói. Theo quan điểm cá nhân chuyên gia, không nên theo dõi diện rộng đối với công dân. Như công ty đằng sau Claude của Mỹ nhất quyết không chịu nhượng bộ chính phủ Mỹ trong việc cho phép sử dụng AI

để theo dõi công dân diện rộng. Nếu theo dự thảo danh mục AI có rủi ro cao thì những ứng dụng của chính phủ ở mức rủi ro cao nhất nhưng dường như nằm ngoài phạm vi bị giám sát hoặc lên án. Trong khi các doanh nghiệp và tổ chức tư nhân nếu sử dụng như thế thì không được. Tinh thần vẫn phải áp dụng trách nhiệm đối với tất cả các chủ thể như nhau. Hy vọng thực hiện trên thực tế cũng vậy.

Chuyên gia CG-CS-02 cho rằng không thể dùng cách tiếp cận truyền thống, những trải nghiệm truyền thống đối với AI. Phân loại, định danh, quản lý là các lĩnh vực phải quan tâm đến AI. Những gì ảnh hưởng đến an ninh quốc gia, ảnh hưởng nghiêm trọng đến người tiêu dùng, đến thuần phong, mỹ tục mới cần nhà nước phải có phản ứng. Còn lại là khuyến khích phát triển một cách lành mạnh chứ không giới hạn trong năng lực quản lý. Đặc biệt phải khuyến khích sử dụng AI, khai thác những lợi thế của AI không chỉ trong lĩnh vực doanh nghiệp cung cấp những sản phẩm dịch vụ cho người dùng mà trong quản lý nhà nước, nâng cao năng lực nhà nước trong quá trình chuyển đổi số, cải cách hành chính ở Việt Nam. Ví dụ như trong thuế, sử dụng AI để xác định những người trốn thuế. Sử dụng AI để có thể phân loại rủi ro đối với hàng hoá xuất nhập khẩu, thì có thể giảm rất nhiều biên chế và nâng cao hiệu quả. Chủ yếu AI trong giám sát giao thông, an ninh trật tự, an toàn xã hội. Việt Nam là một nước nhiều thiên tai, tại sao không sử dụng AI trong dự báo thời tiết để tăng mức độ chính xác, cực kỳ tốt cho nông nghiệp, nông dân. Việt Nam cần phải thành lập những trung tâm nghiên cứu ứng dụng AI để khai thác AI tốt nhất trong giáo dục, văn hóa, thể thao. Nhà nước nên đầu tư vào ứng dụng AI, nhà nước phải đi trước để xem có thể khai thác ứng dụng AI như thế nào.

E.3.3. Quy trình xây dựng

Về thể chế giám sát và cơ quan quản lý AI, 13/15 chuyên gia ủng hộ thành lập cơ quan AI chuyên trách có thẩm quyền liên ngành; hai chuyên gia (CG-CS-01 và CG-DN-03) cho rằng điều kiện tiên quyết là xây dựng năng lực kỹ thuật trước khi mở rộng bộ máy hành chính.

Chuyên gia CG-CS-02 đề xuất thành lập các trung tâm gồm những chuyên gia giỏi nhất đứng đầu, nhà nước bảo trợ nghiên cứu tất cả những lĩnh vực pháp luật thế giới quy định thế nào, chính xác ra sao để thảo luận quá trình hình thành, những điều cần phải lưu ý trong nước. Khi đó, những trung tâm này sẽ là đầu mối về tri thức, thông tin và khuyến nghị quyết định chính sách của Việt Nam chứ không phải dựa trên một xu thế nhất thời. Nếu về mặt khoa học công nghệ phải làm chủ AI rất khó đoán thì về chính sách, về mặt tác động, nghiên cứu thì cần phải nghiên cứu sâu.

Về kiểm định và chứng nhận, tất cả chuyên gia nhóm doanh nghiệp đều nhấn mạnh sự thiếu hụt tổ chức kiểm định AI được công nhận trong nước. Chuyên gia CG-NC-02 chia sẻ rằng: Việt Nam đang thiếu năng lực trong hệ sinh thái kiểm định AI. Thị trường sẽ tự đào thải và công nhận các chứng chỉ, đơn vị uy tín khi AI phát triển rộng rãi. Chuyên gia CG-CS-02 khuyến khích cả tư nhân đầu tư sẽ phù hợp, tức là những cơ sở được nhà nước xác nhận có đủ năng lực kiểm định, đánh giá. Cách này là một cách tiếp cận tốt, chứ không phải nhà nước tự đầu tư, cử người đi làm mà nhà nước thúc đẩy các doanh nghiệp khác để

nhà nước khảo sát, xác nhận, giám sát. Các phòng lab tư nhân đầu tư máy móc để thu phí chứ không nhất thiết là nhà nước làm.

Về thúc đẩy AI có trách nhiệm và xây dựng nguồn lực, chuyên gia CG-NC-01 tập trung kiểm soát rủi ro song hành hỗ trợ doanh nghiệp, nhất là các doanh nghiệp nhỏ và vừa. Cụ thể giảm gánh nặng tuân thủ của doanh nghiệp khi phải tự đánh giá, tuân thủ hay tự đánh giá xếp loại, theo tiêu chuẩn trong Nghị định quy định Bộ KHCN thiết kế một công cụ điện tử để giúp doanh nghiệp đánh giá và tuân thủ. Vấn đề là công cụ đó phải chính xác, hiệu quả. Có một lĩnh vực là lao động việc làm cần chú ý để sửa đổi cho bắt kịp với tốc độ phát triển của AI. Nếu không người lao động, đặc biệt là gig workers trên thực tế bị thiệt thòi, ví dụ chạy xe ôm Grab bị thuật toán đẩy vào thế yếu, gặp những bất lợi rất lớn. Ngoài ra cần nâng cao nhận thức và năng lực của các chủ thể để tuân thủ quy định của luật AI và các nghị định hướng dẫn.

Về hợp tác và hội nhập quốc tế, đa số chuyên gia khi được hỏi đều đồng ý không cần phải điều chỉnh quy định nội tại cho phù hợp với chuẩn mực quốc tế vì mỗi quốc gia có sự tự chủ nhất định về việc thiết kế và thành lập hệ thống pháp lý riêng của mình.

Chuyên gia CG-HT-01 nhấn mạnh một trong những vấn đề cần quan tâm nhất trong lĩnh vực công nghệ số đó chính là hợp tác để đảm bảo an ninh mạng giữa các quốc gia trên thế giới. Để đảm bảo được thì yêu cầu về kỹ thuật chủ yếu là hợp tác trong siết chặt an ninh về mặt kỹ thuật là cách hiệu quả nhất. Ví dụ như trường hợp lừa đảo ở Campuchia trong thời gian vừa qua hoặc là việc sử dụng AI bởi các cái tội phạm ở bên ngoài lãnh thổ để gọi điện cho những người ở Việt Nam, có rất nhiều người là nạn nhân của những cuộc gọi điện do AI giả giọng nói, giả khuôn mặt. Nếu các quốc gia trên thế giới hợp tác với nhau trong mảng đảm bảo an ninh mạng, mỗi quốc gia sẽ tự điều chỉnh trong hệ thống pháp luật của quốc gia mình phù hợp với sự hợp tác đó.

Chuyên gia CG-DN-01 cho rằng đầu tiên vẫn phải tuân thủ những cam kết hoặc chuẩn mực mà Việt Nam đã và đang cam kết trong quá khứ bao gồm các hiệp định thương mại tự do, bảo đảm Việt Nam là thị trường công bằng, tin cậy, sẵn sàng. Đừng khiến cho hệ thống pháp lý Việt Nam trở thành một trong những rào cản khiến cho việc gia nhập thị trường khó khăn hay là thách thức đối với các doanh nghiệp nước ngoài. Còn việc hoàn thiện khung pháp lý như thế nào để phù hợp với chuẩn mực quốc tế cũng tương tự như vậy vì đặc thù của Internet và AI là xuyên biên giới, xóa mờ ranh giới giữa các quốc gia với nhau nên việc tuân thủ pháp lý Việt Nam cần cân nhắc đến việc thực hành ở các khu vực khác.

Theo chuyên gia CG-CS-02, Việt Nam không nên đi tiên phong mà áp dụng những đạo luật, quy định, thảo luận của các nước đi trước để phù hợp với mình. Không nên cực đoan, không nên đi theo mô hình nào chặt chẽ quá. Còn về tham gia các hiệp định thì nên cân nhắc, vì sau này có thể có những phức tạp phe nhóm. Chắc chắn công nghệ sẽ trở thành vũ khí giống như thương mại trong xung đột, Việt Nam cần tính bài toán dài hạn. Mục tiêu của Việt Nam là phát triển AI theo định hướng phát triển kinh tế, hòa bình, cùng có lợi đồng thời cố gắng để làm chủ công nghệ.

E.3.4. Kỹ thuật xây dựng

Tổng hợp đánh giá theo mức độ vận dụng (Tốt / Đạt một phần / Chưa đạt):

Kỹ thuật	Tốt	Đạt một phần	Chưa đạt	Nhận xét tổng hợp
1. AI-RIA	1/15	6/15	8/15	Có báo cáo tác động nhưng thiếu ba chiều đặc thù dành riêng cho AI
2. Định nghĩa chức năng	11/15	3/15	1/15	Được vận dụng tốt nhất
3. Ủy quyền linh hoạt	5/15	8/15	2/15	Đúng hướng
4. Phân định chuỗi giá trị	2/15	7/15	6/15	Có khung nhưng thiếu chi tiết vận hành; khoảng trống xuyên biên giới
5. Miễn trừ tỷ lệ	1/15	5/15	9/15	Khoảng trống lớn nhất ảnh hưởng đến SME
6. Luật hóa tiêu chuẩn	0/15	4/15	11/15	Ý định đúng nhưng thiếu TCVN và cơ chế suy đoán tuân thủ

Về kỹ thuật định nghĩa chức năng và AI agents, chuyên gia CG-CS-03 xác nhận từ phía soạn thảo: *"Khi soạn thảo, AI agents chưa bùng nổ mạnh như bây giờ. Định nghĩa 'hệ thống AI' trong Luật AI 2025 về nguyên tắc có thể bao phủ AI agents, nhưng không rõ ràng về các yêu cầu đặc thù như giám sát có ý nghĩa của con người, cơ chế dừng khẩn cấp hay ranh giới hành động".*

Chuyên gia CG-DN-03 nhận xét từ góc độ startup AI: *"Chúng tôi có AI agents xử lý các tác vụ tự động. Để hạn chế tối đa việc AI agent gây thiệt hại cho khách hàng thì chúng tôi chịu trách nhiệm thiết kế từ khâu kỹ thuật ban đầu và giám sát chặt chẽ toàn bộ quy trình".*

Về kỹ thuật luật hoá tiêu chuẩn, Chuyên gia CG-CS-01 cho rằng để đối phó với deepfake tinh vi, mã độc AI hay AI tự chủ, chúng ta không cần thiết phải liên tục sửa Luật, mà cần nhanh chóng ban hành các Nghị định, Tiêu chuẩn/Quy chuẩn kỹ thuật (TCVN/QCVN) và duy trì một cơ chế quản lý linh hoạt, cho phép cập nhật liên tục các rủi ro mới vào danh mục kiểm soát của Bộ Khoa học và Công nghệ.

E.4. Chủ đề 4 - Khoảng trống và thách thức

Phân tích chủ đề 4 xác định được năm khoảng trống pháp lý được đề cập với tần suất cao và mức độ đồng thuận đáng kể:

Khoảng trống 1 - Chế tài chuyên biệt: Mười hai trong số mười lăm chuyên gia xác định sự thiếu vắng chế tài hành chính và dân sự đặc thù cho AI là khoảng trống nghiêm trọng nhất, tạo ra sự không cân xứng giữa nghĩa vụ và hậu quả không tuân thủ.

Khoảng trống 2 - Hạ tầng kiểm định: Toàn bộ chuyên gia nhóm doanh nghiệp và phần lớn nhóm chính sách đồng thuận rằng thiếu hạ tầng kiểm định AI trong nước là yếu tố cản trở trực tiếp khả năng thực thi Luật AI 2025.

Khoảng trống 3 - Dữ liệu huấn luyện và bản quyền: Chín chuyên gia coi đây là vấn đề chưa được giải quyết đầy đủ, với đặc thù là không thể giải quyết đơn lẻ mà cần phối hợp giữa Luật AI, Luật Sở hữu trí tuệ và các điều ước quốc tế.

Khoảng trống 4 - AI trong khu vực công: Tám chuyên gia nhấn mạnh thiếu khung pháp lý đặc thù điều chỉnh việc cơ quan nhà nước sử dụng AI trong ra quyết định hành chính ảnh hưởng đến quyền và nghĩa vụ của công dân.

Khoảng trống 5 - Cơ chế thích ứng động: Bảy chuyên gia đặt câu hỏi về khả năng của hệ thống pháp luật hiện tại trong việc cập nhật kịp thời khi có đột phá công nghệ, ví dụ điển hình là mô hình ngôn ngữ lớn và hệ thống AI tác nhân tự chủ (*agentic AI*) vẫn chưa có quy định riêng.

E.5. Chủ đề 5 - Đề xuất và định hướng hoàn thiện

Đề xuất được đa số chuyên gia ủng hộ *(trên 10/15 chuyên gia, tức hơn 66%)*

Thứ nhất, ưu tiên hoàn thiện hệ thống chế tài trong Nghị định hướng dẫn Luật AI, bảo đảm tương xứng với mức chế tài của EU AI Act theo tỷ lệ phù hợp với điều kiện Việt Nam.

Thứ hai, đầu tư khẩn cấp vào xây dựng hạ tầng kiểm định AI trong nước, bao gồm đào tạo đội ngũ kiểm định viên và công nhận ít nhất một tổ chức kiểm định AI quốc gia.

Thứ ba, xây dựng cơ chế MRA (*Mutual Recognition Agreement*) về kiểm định AI với ASEAN và các đối tác chiến lược, tận dụng lợi thế là một trong số ít quốc gia đã có luật AI thống nhất.

Thứ tư, ban hành hướng dẫn riêng về AI trong khu vực công, bổ sung vào khoảng trống mà Luật AI 2025 và Nghị định hướng dẫn chưa xử lý đầy đủ.

Thứ năm, đề xuất có một khoảng thời gian để thử áp dụng luật này trong những lĩnh vực đặc thù nhất định để xem xét mức độ áp dụng thi hành được đến đâu, có khó khăn hay thách thức gì trong thực tiễn, sau đó mới áp dụng hoàn toàn cho tất cả các ngành nghề. Phải có lộ trình hợp lý cho việc áp dụng và thi hành.

Đề xuất được thiểu số chuyên gia đề xuất nhưng có giá trị học thuật cao

Chuyên gia CG-DN-02 cho rằng ban hành luật chỉ là bước một. Nếu không đầu tư đồng thời vào năng lực thực thi và giáo dục hệ sinh thái, luật sẽ tồn tại trên giấy.

Theo chuyên gia CG-DN-04, sau khi có luật AI, nên tập trung vào khuyến khích các doanh nghiệp đầu tư và phát triển AI ở Việt Nam. Một là giúp cho việc xây dựng hạ tầng AI ở Việt Nam rẻ hơn, tốt hơn, nhanh hơn. Hai là làm sao để có các tập dữ liệu của quốc gia tốt để có thể phát triển được AI chủ quyền ở Việt Nam. Thứ ba là đầu tư về con người, nguồn lực của các doanh nghiệp để họ có thể tuyển dụng được các nhân sự AI tốt không chỉ ở Việt Nam mà cả trên thế giới về Việt Nam.

Chuyên gia CG-NC-02 đề xuất hợp tác với Hiệp hội các nhà khoa học dữ liệu của Mỹ để tạo ra các cuộc thi chứng chỉ về năng lực khoa học dữ liệu/AI. Ưu tiên bổ sung nội dung về đào tạo nhận thức bài bản về AI cho toàn bộ người dân như một nền tảng quan trọng trước và song song với việc ban hành luật. Bởi lẽ, đa số người dân không biết luật và việc hiểu về AI là tiền đề để tuân thủ pháp luật.

Chuyên gia CG-DN-01 cho rằng cách làm luật của Việt Nam nên chậm một chút và cần nhiều chuyên gia, nhiều thành phần xã hội tham gia hơn. Xen kẽ với đó là phải có

nghiên cứu khảo sát thực tế về mức độ ảnh hưởng của luật vào thị trường và việc phát triển AI ở Việt Nam thì sẽ hợp lý hơn là ngay lập tức ban hành. Thà các quy phạm của chúng ta giải quyết đúng cốt lõi vấn đề thì hơn hẳn so với việc chạy theo những lần xây dựng luật nhanh chóng, thiếu cơ chế phản ứng, chính sách, thiếu lắng nghe ý kiến của thị trường.

Theo chuyên gia CG-CS-02, cần theo dõi thực thi sát sao. Bộ KHCN, Bộ Công an, Bộ GD&ĐT có thể tổ chức các cuộc tổng kết đánh giá nhanh trong quá trình thực hiện luật hay thậm chí có thể tổ chức gặp mặt đối thoại doanh nghiệp trong ngành nghề đó.

Chuyên gia CG-NC-01 kiến nghị phải tập trung làm rõ hơn áp đặt các nghĩa vụ trách nhiệm cho khối khu vực công trong sử dụng AI vì cơ quan nhà nước lâu nay chưa có AI đã có quyền lực rất lớn so với các chủ thể khác. Bây giờ có thêm AI, bao nhiêu quyền lực ở AI nữa. Nếu không quy định thật chặt đối với cơ quan nhà nước, dễ xảy ra nguy cơ lạm dụng. Ví dụ như không được dùng AI để theo dõi diện rộng công dân chẳng hạn. Nhưng hiện nay không rõ trong luật AI.

Chuyên gia CG-HT-02 cho rằng sau khi đã có luật khung phải tiếp cận từ góc độ bottom-up. Trên cơ sở luật AI phải cho phép các lĩnh vực có quy chuẩn riêng để phát triển, phải linh hoạt, chứ không nên cứng nhắc sẽ hình thành các bộ khuyến nghị về đạo đức của từng lĩnh vực. Chúng ta phải thấy được tiềm năng điều chỉnh của hệ thống luật mềm soft law bởi vì nó phản ánh được nhu cầu thực tiễn của đời sống, là một cơ chế trong đó tiếng nói của cộng đồng được coi trọng. Một luận điểm rất cơ bản là về bản chất luật tư theo chiều ngang, mối quan hệ giữa các chủ thể có địa vị bình đẳng, tự độc lập với nhau, khác hoàn toàn so với luật công là mối quan hệ theo chiều dọc. Do vậy nó luôn hàm chứa tính đa dạng về nguồn luật, về quy phạm và có tính mở, nói cách khác là luật tư tự bản thân luôn có tính đàn hồi, linh hoạt, cân đối, điều chỉnh để hấp thụ được những hiện tượng phát sinh mới ở trong đời sống số.

Chuyên gia CG-CS-01 nhận thấy để Luật AI thực sự đi vào đời sống, chúng ta phải lấp đầy các khoảng trống trên bằng những Nghị định hướng dẫn thi hành có tính thực tiễn cao, linh hoạt. Đặc biệt, cần sớm có Bộ quy tắc đạo đức, tiêu chuẩn phân loại rủi ro chi tiết và hành lang pháp lý bảo vệ cơ chế "sử dụng hợp lý" dữ liệu. Có như vậy, doanh nghiệp Việt mới đủ tự tin vươn lên làm chủ công nghệ lõi. Cần quy định rõ quy trình đánh giá sự phù hợp phải minh bạch, quy định thời gian xử lý tối đa (ví dụ 15-30 ngày) và xác định rõ cơ quan chịu trách nhiệm. Ưu tiên cho phép doanh nghiệp tự đánh giá dựa trên bộ công cụ và hồ sơ mẫu do Nhà nước cung cấp miễn phí, cơ quan quản lý chỉ thực hiện hậu kiểm đối với các hệ thống không thuộc lĩnh vực quá nhạy cảm.

F. KIỂM CHỨNG DỮ LIỆU VÀ ĐỘ TIN CẬY

F.1. Tam giác hóa

Dữ liệu phỏng vấn được đối chiếu với hai nguồn độc lập: phân tích văn bản pháp lý trực tiếp và tổng quan tài liệu học thuật quốc tế. Sự hội tụ giữa ba nguồn (phỏng vấn, phân tích pháp lý và tài liệu học thuật) tăng cường độ tin cậy của các kết luận rút ra.

F.2. Giới hạn của phương pháp

Nghiên cứu này thừa nhận một số giới hạn cần được làm rõ. Về tính đại diện, mẫu phỏng vấn 15 chuyên gia không thể đại diện cho toàn bộ cộng đồng chuyên gia AI và pháp lý Việt Nam; kết quả phản ánh quan điểm của nhóm chuyên gia được lựa chọn theo tiêu chí mục đích, không phải mẫu ngẫu nhiên. Về thời điểm, quan điểm của chuyên gia phản ánh bối cảnh tại thời điểm phỏng vấn và có thể thay đổi theo diễn biến của hệ thống pháp luật. Về tính chủ quan, nghiên cứu sinh thừa nhận vai trò của mình trong quá trình diễn giải dữ liệu, mặc dù đã cố gắng tối đa hóa tính khách quan thông qua kiểm chứng thành viên và tham chiếu đa nguồn.

G. KẾT LUẬN TỔNG HỢP VÀ Ý NGHĨA ĐỐI VỚI LUẬN ÁN

Kết quả phỏng vấn chuyên gia cung cấp ba đóng góp chính cho luận án. Thứ nhất, xác nhận từ góc độ thực tiễn các phát hiện phân tích pháp lý được trình bày trong các chương lý luận, tăng cường tính thuyết phục của các luận điểm. Thứ hai, bổ sung chiều cạnh thực tiễn và kinh nghiệm nghề nghiệp mà phân tích văn bản pháp lý thuần túy không thể cung cấp. Thứ ba, định hướng cho các đề xuất giải pháp dựa trên tính khả thi thực tiễn, không chỉ dựa trên tính hoàn hảo lý thuyết.

Đặc biệt, sự phân kỳ quan điểm giữa các nhóm chuyên gia thay vì được coi là hạn chế của nghiên cứu, phản ánh tính chất đa chiều và còn tranh luận của chủ đề, do đó được trình bày trong luận án như bằng chứng về tính phức tạp không thể được giải quyết bằng các câu trả lời đơn giản.