

# CHÍNH SÁCH VỀ TRÍ TUỆ NHÂN TẠO (AI) Ở VIỆT NAM: *CÂN BẰNG GIỮA THÚC ĐẨY PHÁT TRIỂN VÀ KIỂM SOÁT RỦI RO*



## **Nhóm nghiên cứu**

**Nguyễn Quang Đồng**

**Nguyễn Lan Phương**

**Trần Thị Tuyết**

**Nguyễn Đức Lam**

**(Viện Nghiên cứu Chính sách và Phát triển Truyền thông)**

# MỤC LỤC

Nhóm nghiên cứu .....	1
Danh mục bảng .....	4
Danh mục hình .....	4
Danh mục từ viết tắt .....	5
Lời cảm ơn .....	6
Tóm tắt báo cáo .....	7
<b>I. GIỚI THIỆU TỔNG QUAN .....</b>	<b>16</b>
1. Bối cảnh nghiên cứu .....	16
2. Mục đích, phạm vi nghiên cứu .....	16
3. Cách tiếp cận; phương pháp .....	17
4. Giá trị và hạn chế của nghiên cứu .....	17
<b>II. CƠ HỘI VÀ RỦI RO, THÁCH THỨC CỦA AI .....</b>	<b>18</b>
<b>2.1. CƠ HỘI, LỢI ÍCH DO AI MANG LẠI .....</b>	<b>18</b>
2.1.1. Tạo các giá trị gia tăng .....	18
2.1.2. Tăng năng suất lao động, tạo việc làm .....	19
2.1.3. Thay đổi mô hình kinh doanh .....	19
2.1.4. Góp phần thay đổi các ngành kinh tế .....	20
2.1.5. Thúc đẩy sự phát triển, hạnh phúc của con người, quyền con người .....	21
<b>2.2. RỦI RO, TÁC ĐỘNG TIÊU CỰC CỦA AI ĐỐI VỚI CON NGƯỜI, XÃ HỘI .....</b>	<b>21</b>
2.2.1. Nhận diện rủi ro của AI .....	21
2.2.2. Tác động của AI đối với lao động, việc làm .....	23
2.2.3. Tác động của AI đối với các giá trị cốt lõi của con người, xã hội .....	26
2.2.4. Tác động của AI đối với các quyền con người .....	27
2.2.5. Thách thức trong kiểm soát rủi ro của AI .....	28
<b>III. CHÍNH SÁCH VỀ TRÍ TUỆ NHÂN TẠO TRÊN THẾ GIỚI .....</b>	<b>30</b>
<b>3.1. CÁC CÔNG CỤ CHÍNH SÁCH ĐIỀU CHỈNH AI TRÊN THẾ GIỚI .....</b>	<b>30</b>
3.1.1. Các bộ nguyên tắc về AI .....	30
3.1.2. Khuôn khổ chính sách về AI .....	32
3.1.3. Khuôn khổ pháp luật liên quan đến AI .....	32
3.1.4. Hướng dẫn; tiêu chuẩn kỹ thuật .....	36
3.1.5. Quy tắc ứng xử tự nguyện .....	36
<b>3.2. PHÁT TRIỂN HỆ SINH THÁI AI .....</b>	<b>37</b>
3.2.1. Hạ tầng và dữ liệu .....	37
3.2.2. Nhân lực .....	38
3.2.3. Giáo dục, đào tạo .....	38
3.2.4. Nghiên cứu .....	39
3.2.5. Phát triển thị trường AI; đầu tư, nguồn tài chính .....	39

<b>3.3. CHÍNH SÁCH KIỂM SOÁT RỦI RO CỦA AI: MỘT SỐ CÁCH TIẾP CẬN .....</b>	<b>41</b>
3.3.1. Khung quản trị rủi ro toàn diện .....	41
3.3.2. Kiểm soát dựa trên mức độ rủi ro .....	42
3.3.3. Kiểm soát rủi ro theo vòng đời hệ thống AI .....	43
3.3.4. Dựa trên các chuẩn mực quốc tế về các giá trị chung của loài người .....	44
3.3.5. Kiểm soát rủi ro ở khu vực công; thúc đẩy đổi mới sáng tạo trong khu vực tư .....	45
3.3.6. Coi trọng an ninh quốc gia trong khi phát triển AI .....	46
<b>IV. HỆ SINH THÁI VÀ CHÍNH SÁCH VỀ TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM .....</b>	<b>47</b>
<b>4.1. HIỆN TRẠNG HỆ SINH THÁI AI Ở VIỆT NAM .....</b>	<b>47</b>
4.1.1. Đầu tư vào thị trường AI ở Việt Nam: sức nóng và thực tế .....	47
4.1.2. Hạ tầng, dữ liệu dành cho AI .....	50
4.1.3. Nhân lực AI .....	53
4.1.4. Giáo dục, đào tạo, nghiên cứu khoa học về AI .....	55
4.1.5. Nguồn tài chính và cơ chế tài chính cho AI .....	57
<b>4.2. CHÍNH SÁCH, PHÁP LUẬT VỀ AI Ở VIỆT NAM .....</b>	<b>58</b>
4.2.1. Chính sách, pháp luật hiện có về AI .....	58
4.2.2. Khoảng cách giữa chính sách và nhu cầu thực tế .....	59
<b>V. KHUYẾN NGHỊ CHÍNH SÁCH VỀ AI Ở VIỆT NAM .....</b>	<b>62</b>
<b>5.1. CÁCH TIẾP CẬN CHUNG .....</b>	<b>62</b>
5.1.1. Cân bằng giữa kiểm soát rủi ro với phát triển AI .....	62
5.1.2. Kết hợp các công cụ chính sách khác nhau .....	63
5.1.3. Tiếp cận tổng thể, thu hút các bên liên quan .....	64
<b>5.2. CHÍNH SÁCH HỖ TRỢ, THÚC ĐẨY PHÁT TRIỂN AI .....</b>	<b>65</b>
5.2.1. Đầu tư, lựa chọn ưu tiên phân khúc thị trường AI .....	65
5.2.2. Chính sách phát triển hạ tầng, dữ liệu cho AI .....	67
5.2.3. Chính sách việc làm, phát triển nguồn nhân lực AI .....	68
5.2.4. Chính sách giáo dục, nghiên cứu AI .....	71
5.2.5. Khuyến nghị về cơ chế tài chính .....	72
<b>5.3. CHÍNH SÁCH KIỂM SOÁT AI .....</b>	<b>73</b>
5.3.1. Đánh giá, điều chỉnh rủi ro theo vòng đời của hệ thống AI .....	73
5.3.2. Đánh giá, điều chỉnh theo mức độ rủi ro của AI .....	73
5.3.3. Hướng đến AI có trách nhiệm và đáng tin cậy .....	73
5.3.4. Giải pháp lập pháp .....	74
5.3.5. Tích hợp quyền con người trong kiểm soát rủi ro của AI .....	75
<b>PHỤ LỤC: RỦI RO VI PHẠM QUYỀN CON NGƯỜI THEO VÒNG ĐỜI CỦA HỆ THỐNG AI .....</b>	<b>78</b>

## Danh mục bảng

<b>Bảng 1:</b> Rủi ro, tác hại theo vòng đời AI .....	23
<b>Bảng 2:</b> Các nguyên tắc phổ biến trên thế giới về AI có trách nhiệm .....	31
<b>Bảng 3:</b> So sánh tình hình xây dựng pháp luật về AI ở Hoa Kỳ và EU .....	33
<b>Bảng 4:</b> Ưu, nhược điểm của từng cách tiếp cận .....	35
<b>Bảng 5:</b> Đầu tư vào AI trên toàn cầu từ 2017 – 2024 .....	40
<b>Bảng 6:</b> Kiểm soát rủi ro theo vòng đời AI .....	44
<b>Bảng 7:</b> Ví dụ rủi ro tiềm ẩn .....	78

## Danh mục hình

<b>Hình 1:</b> Cơ hội, lợi ích do AI mang lại .....	18
<b>Hình 2:</b> Rủi ro AI thay thế việc làm chân tay .....	25
<b>Hình 3:</b> Rủi ro AI thay thế việc làm trí óc .....	25
<b>Hình 4:</b> Công cụ chính sách điều chỉnh AI .....	30
<b>Hình 5:</b> Các mức độ rủi ro AI .....	43
<b>Hình 6:</b> Hệ sinh thái AI .....	47
<b>Hình 7:</b> Mức độ áp dụng AI .....	49
<b>Hình 8:</b> Đầu tư của doanh nghiệp vào AI ở một số nước Đông Nam Á .....	50
<b>Hình 9:</b> Nhu cầu nhân lực CNTT .....	53
<b>Hình 10:</b> Tình hình đào tạo đại học ngành CNTT tại thời điểm 2023 .....	55
<b>Hình 11:</b> Cộng sinh giữa con người với AI trong việc làm .....	69

## Danh mục từ viết tắt

<b>AI</b>	Trí tuệ nhân tạo
<b>LHQ</b>	Liên Hợp Quốc
<b>EC</b>	Cộng đồng châu Âu
<b>OECD</b>	Tổ chức Hợp tác và Phát triển Kinh tế
<b>UDHR</b>	Tuyên Ngôn Quốc Tế Nhân Quyền
<b>ICCPR</b>	Công ước Quốc tế về các Quyền Dân sự và Chính trị
<b>ISO</b>	Tổ chức tiêu chuẩn hoá quốc tế
<b>IEC</b>	Ủy ban Kỹ thuật Điện Quốc Tế
<b>GPU</b>	Bộ xử lý đồ họa
<b>SME</b>	Doanh nghiệp vừa và nhỏ
<b>R&amp;D</b>	Nghiên cứu và Phát triển
<b>LLM</b>	Mô hình ngôn ngữ lớn
<b>UAE</b>	Các Tiểu Vương Quốc Ả Rập Thống Nhất
<b>RMF</b>	Khung quản lý rủi ro
<b>NIST</b>	Viện Tiêu chuẩn và Công nghệ Quốc Gia
<b>API</b>	Giao diện lập trình ứng dụng
<b>ICESCR</b>	Công ước Quốc tế về các Quyền Kinh tế, Xã hội và Văn hóa
<b>STEM</b>	Khoa học, Công nghệ, Kỹ thuật và Toán học
<b>NQ/TW</b>	Nghị quyết Trung ương
<b>KHCN</b>	Khoa học Công nghệ
<b>SHTT</b>	Sở hữu Trí tuệ
<b>KCN</b>	Khu công nghiệp
<b>UI</b>	Giao diện người dùng
<b>UX</b>	Trải nghiệm người dùng
<b>QĐ</b>	Quyết định

# Lời cảm ơn

Báo cáo nghiên cứu **“Chính sách về Trí tuệ nhân tạo (AI) ở Việt Nam: Cân bằng giữa thúc đẩy phát triển và kiểm soát rủi ro”** do Viện Nghiên cứu Chính sách và Phát triển Truyền thông (IPS) thực hiện trong năm 2024. Nhóm nghiên cứu gồm có **Nguyễn Quang Đồng, Nguyễn Lan Phương, Trần Thị Tuyết, Nguyễn Đức Lam**.

Chúng tôi trân trọng cảm ơn cán bộ, công chức của các cơ quan Nhà nước ở trung ương và địa phương, các chuyên gia, chủ doanh nghiệp đã nhận lời trao đổi quan điểm, ý kiến về các vấn đề liên quan đến chính sách phát triển AI ở Việt Nam.

Viện Nghiên cứu Chính sách và Phát triển Truyền thông trân trọng cảm ơn Hội Truyền thông số Việt Nam luôn đồng hành, ủng hộ thực hiện báo cáo này và những hoạt động nghiên cứu khác của Viện.

Cuối cùng, xin trân trọng cảm ơn Ủy ban Nhân quyền Úc, Đại sứ quán Úc tại Việt Nam đã hỗ trợ thực hiện nghiên cứu này.

# Tóm tắt báo cáo

Trong năm 2024, Viện Nghiên cứu Chính sách và Phát triển Truyền thông (IPS) đã thực hiện nghiên cứu và hoàn thành báo cáo “Chính sách về trí tuệ nhân tạo ở Việt Nam: Cân bằng giữa thúc đẩy phát triển và quản trị rủi ro”. Báo cáo khẳng định lại nhận định chung trên thế giới là trí tuệ nhân tạo (AI) có thể mang lại rất nhiều cơ hội, lợi ích cho nền kinh tế, sự phát triển của đất nước, doanh nghiệp, cá nhân con người, bảo vệ và thúc đẩy quyền con người; đồng thời tiềm ẩn nhiều rủi ro, tác động tiêu cực đối với con người, xã hội. Báo cáo đánh giá tổng thể các điều kiện sẵn sàng cho sự phát triển hệ sinh thái AI gồm hạ tầng, dữ liệu, con người, đầu tư, tài chính; các công cụ chính sách có thể áp dụng để điều chỉnh AI, cách tiếp cận trong kiểm soát rủi ro của AI trên thế giới và đối chiếu với thực trạng ở Việt Nam. Sau đây là tóm tắt các nhận định chính về hiện trạng chính sách, thực tế phát triển AI ở Việt Nam; các khuyến nghị chính sách thúc đẩy phát triển và kiểm soát rủi ro của AI, trong đó có bảo đảm quyền con người ở Việt Nam.

## 1. Hệ sinh thái AI ở Việt Nam

**Thị trường AI ở Việt Nam:** Nhu cầu ngày càng tăng đối với AI đã tạo ra một số kết quả tích cực đối với thị trường AI ở Việt Nam. Theo một nghiên cứu, quy mô thị trường ước tính năm 2024 là 753,40 triệu đô la Mỹ; 3,4 tỷ đô la Mỹ vào năm 2030 với tốc độ tăng trưởng hàng năm (CAGR) 2024-2030 là 28,63%. AI đã được ứng dụng trong các tổ chức, doanh nghiệp ở các ngành như ngân hàng, thương mại điện tử, bán lẻ, sản xuất, y tế, giáo dục v.v Theo PitchBook Data Inc., Việt Nam đứng thứ hai ở Đông Nam Á sau Singapore về số lượng AI startup năm 2024 với 765 công ty nhờ chi phí hoạt động thấp hơn, nguồn nhân lực đa dạng cho phép các công ty khởi nghiệp tạo ra sản phẩm chất lượng cao, trong khi vẫn giữ được chi phí chấp nhận được.

Tuy nhiên, vẫn còn những điểm hạn chế trong thị trường AI ở Việt Nam. Theo Chỉ số sẵn sàng AI năm 2024 của CISCO, chỉ có 22% các tổ chức tại Việt Nam đã hoàn toàn sẵn sàng triển khai và tận dụng các công nghệ AI. Các doanh nghiệp gia đình Việt Nam mới đang ở giai đoạn đầu áp dụng AI; với phần lớn (64%) vẫn chưa bắt đầu tìm hiểu về AI. Doanh nghiệp Nhà nước hoặc doanh nghiệp có vốn Nhà nước thường dè chừng trong việc thử nghiệm AI hơn so với doanh nghiệp tư nhân; đời chủ trương hoặc quy trình đấu thầu gây tốn kém về thời gian; lo ngại về trách nhiệm.

Các dự án đầu tư vào AI của khu vực tư ở Việt Nam hầu hết có nguồn vốn huy động được còn nhỏ. Đơn cử như 765 AI startup đã nhắc đến ở trên chỉ gọi vốn được 47,3 triệu đô la trong 9 tháng đầu năm 2024, trung bình mỗi startup chỉ nhận được gần 62.000 đô la. Các doanh nghiệp Việt Nam đầu tư AI với quy mô nhỏ, nhưng rải rác ở nhiều lĩnh vực. Điều này có thể là dấu hiệu cho thấy không có định hướng nào từ nhà nước trong việc phát triển các ngành công nghiệp AI quan trọng để có được khả năng cạnh tranh tập trung.

Phân tích thực trạng trên đây cho thấy, mục tiêu của Chiến lược quốc gia đến năm 2025 có “05 thương hiệu AI uy tín” dường như khó đạt được, khi mà chưa có thương hiệu AI nào được xác định cho đến thời điểm cuối năm 2024. Do đó, mục tiêu “10 thương hiệu AI” năm 2030 cũng có thể không đạt được nếu không có sự điều chỉnh chính sách.

**Hạ tầng, dữ liệu dành cho AI:** Tương tự như trên thế giới, hầu hết doanh nghiệp Việt Nam không đủ nguồn lực để xây dựng các mô hình LLM cho riêng mình, hoặc không có cloud riêng. Doanh nghiệp Việt Nam, kể cả các doanh nghiệp lớn chủ yếu vẫn đang phát triển, ứng dụng AI trên cơ sở hạ tầng tính toán, hạ tầng dữ liệu của các công ty nước ngoài.

Cloud của các doanh nghiệp trong nước chất lượng không đủ để chạy hệ thống AI. Hơn nữa, hầu hết các doanh nghiệp cloud trong nước cung cấp dịch vụ cho thuê trung tâm dữ liệu (không phù hợp với doanh nghiệp vừa và nhỏ), mà chưa cung cấp dịch vụ theo thời gian thực; hỗ trợ khách hàng vừa và nhỏ chưa tốt. Bên cạnh đó, Việt Nam không tham gia phát triển các phần mềm nguồn mở trên thế giới và không tạo ảnh hưởng đến sự phát triển của các phần mềm nguồn mở.

Dữ liệu ở Việt Nam có thể là nhiều, nhưng dữ liệu đã xử lý, có cấu trúc vẫn đang thiếu; dữ liệu còn nằm rải rác và không được sử dụng nhiều để phân tích, đưa ra quyết định. Dữ liệu ở khu vực công



ở Việt Nam nhiều hơn so với dữ liệu của doanh nghiệp, nhưng mức độ, phạm vi dữ liệu mở của Chính phủ, dữ liệu mở quốc gia còn hạn chế. Doanh nghiệp Việt Nam chưa đầu tư xứng đáng vào xây dựng hệ thống dữ liệu của mình; cản trở khả năng khai thác giá trị kinh tế từ dữ liệu và gây khó khăn trong việc ứng dụng AI vào hoạt động quản trị, kinh doanh.

**Nhân lực AI:** Năng lực kỹ sư CNTT nói chung của Việt Nam trên phương diện nhất định được đánh giá khá cao, cụ thể như: theo khảo sát SkillValue, năng lực lập trình viên Việt Nam nằm trong TOP10 thế giới; theo UNESCO Statistics, Việt Nam nằm trong số 10 quốc gia đào tạo nhiều kỹ sư phần mềm nhất; hạng 6 trong số các quốc gia hàng đầu về gia công phần mềm. Tuy nhiên, hàng năm Việt Nam vẫn thiếu hụt từ 150.000 đến 200.000 lập trình viên, kỹ sư có chuyên môn cao và các kỹ năng mềm, ngoại ngữ.

Tại thời điểm 2024, khoảng 2000 chuyên gia AI làm việc trong nước, chủ yếu sống tại Hà Nội, Thành phố Hồ Chí Minh và Đà Nẵng, nơi các công ty liên quan đến AI đặt trụ sở chính. Chính phủ đang đặt ra mục tiêu đào tạo ít nhất 5.000 kỹ sư có chuyên môn AI vào năm 2030. Mặc dù vậy, con số này còn có khoảng cách xa so với nhu cầu nhân sự AI ở Việt Nam. Trong đó, các nhà khoa học về AI (nghiên cứu ra các mô hình AI, thiết kế, chỉnh sửa mô hình AI) còn thiếu nhiều ở Việt Nam. Còn kỹ sư phần mềm AI có thể tìm được, nhưng thiếu người có kinh nghiệm và năng lực tốt. Đồng thời, các ngành không phải là CNTT đang thiếu nhân sự có kiến thức, kỹ năng biết cách ứng dụng AI vào quản trị doanh nghiệp, kinh doanh, cung cấp dịch vụ. Ở cấp độ doanh nghiệp, nhân sự hiện tại của nhiều doanh nghiệp Việt Nam chưa được đào tạo bài bản về AI (kiến thức, kỹ năng về AI của họ có thường do yêu thích công nghệ, hoặc đáp ứng do làm trong môi trường cần AI).

**Giáo dục, đào tạo; nghiên cứu khoa học về AI:** Nói chung, giáo dục đào tạo các ngành CNTT ở Việt Nam có xu hướng tăng lên cả về số lượng sinh viên và chương trình đào tạo. Về giáo dục - đào tạo AI nói riêng, hiện tại Việt Nam đã có những trường đại học thuộc khối CNTT đã có các chương trình nhằm đáp ứng sự phát triển của công nghệ AI như đào tạo khoa học máy tính, khoa học dữ liệu, trí tuệ nhân tạo. Bên cạnh đó, một số trường đại học khác không thuộc khối CNTT đã đưa AI vào chương trình đào tạo của mình, ví dụ kết hợp khoa học máy tính, khoa học dữ liệu và kinh doanh.

Chỉ một số ít trường đại học có các hoạt động nghiên cứu trong lĩnh vực AI, chủ yếu là các trường đại học kỹ thuật, công nghệ; nghiên cứu AI chưa thu hút được sự chú ý của các trường đại học kinh tế và kinh doanh. Trung tâm Nghiên cứu và đào tạo AI và mạch tích hợp (IC) tại Đà Nẵng được hy vọng góp phần thu hút các nhà nghiên cứu, nhà đổi mới và nhà đầu tư chất lượng cao về AI. Đại học FPT là đơn vị đầu tiên sử dụng siêu máy tính để nghiên cứu AI tại Việt Nam, giúp FPT đẩy nhanh quá trình nghiên cứu tập trung vào xử lý hình ảnh và ngôn ngữ trên nền tảng FPT.AI và phát triển sản phẩm.

Tuy nhiên, các doanh nghiệp, trường đại học có nguồn lực như FPT để đầu tư mạnh vào R&D không nhiều ở Việt Nam. Trong khi đó, còn thiếu sự hỗ trợ của Chính phủ về hạ tầng tính toán, dữ liệu cho đào tạo, cho R&D đối với AI. Hơn nữa, quy trình thủ tục phê duyệt đề tài nghiên cứu của nhà nước hiện nay gây tốn kém về thời gian, khiến các phòng thí nghiệm kỹ thuật ở trường đại học không bắt kịp với tốc độ phát triển của công nghệ trên thế giới.

**Nguồn tài chính và cơ chế tài chính cho AI:** Trong khi Nhà nước sẽ là bên “mua” lớn nhất các sản phẩm, dịch vụ dựa trên AI, nguồn tài chính, cơ chế tài chính cho AI hiện nay từ ngân sách nhà nước còn gặp nhiều vướng mắc, đặc biệt là định giá các công việc, sản phẩm để lập dự toán, quyết toán các dự án liên quan đến AI trong khu vực Nhà nước.

Về đầu tư của tư nhân vào AI, theo khảo sát của GenAI Fund, trong các nguồn tài chính của 250 công ty khởi nghiệp GenAI tại ASEAN, 50% là do tự lực hoặc được tài trợ bởi các nhà đầu tư thiên thần (angel-funded), 41% đã đảm bảo được nguồn tài trợ tiền hạt giống (pre-seed funding) hoặc hạt giống (seed funding). Với khoảng 70 trong số 250 công ty này có trụ sở ở Việt Nam, những con số đó cho thấy những hàm ý chính sách để thúc đẩy, phát triển các nguồn tài chính cho doanh nghiệp khởi nghiệp AI ở Việt Nam.

## 2. Chính sách, pháp luật về AI ở Việt Nam

Mặc dù một số văn bản đã ban hành hoặc đang xây dựng, nhưng khuôn khổ chính sách, pháp luật về AI của Việt Nam vẫn còn hạn chế. Chiến lược quốc gia về AI của Việt Nam ban hành năm 2021 đã lạc hậu so với sự phát triển quá nhanh của AI như GenAI; còn chung chung, chưa rõ trọng tâm, trọng điểm; chưa có các chính sách, quy định chi tiết hơn để tạo ra lộ trình rõ ràng phía trước. Chiến lược cũng không nhắc đến vai trò của các bên tham gia khác như doanh nghiệp vừa và nhỏ, các công ty khởi nghiệp, các tổ chức xã hội, những người hưởng lợi từ AI. Khuôn khổ chính sách, pháp luật chưa chú trọng làm rõ khái niệm, nội dung cụ thể về đạo đức AI ở tầm quốc gia.

Tại thời điểm 2024, khuôn khổ chính sách, pháp luật hiện hành vẫn còn những hạn chế hoặc chưa điều chỉnh đối với một số vấn đề liên quan đến thúc đẩy phát triển AI và kiểm soát rủi ro của AI. Cụ thể như:

- Các văn bản quy phạm pháp luật hiện hành chưa có quy định rõ ràng, cụ thể đủ mạnh để tạo cơ chế hỗ trợ, ưu đãi, cơ chế tài chính khuyến khích, thúc đẩy sự phát triển của AI, cả trong khu vực Nhà nước và tư nhân.
- Chưa có các quy định cụ thể về cấp độ rủi ro, các biện pháp tương ứng cần phải thực hiện để kiểm soát các hệ thống AI; trách nhiệm các bên trong kiểm soát rủi ro từ AI.
- Hệ thống pháp luật vẫn thiếu các quy định rõ ràng về địa vị pháp lý của các thực thể AI trong các mối quan hệ xã hội. Pháp luật dân sự quy định, chủ thể pháp lý phải là cá nhân hoặc thể nhân, chứ không phải máy móc hoặc chương trình máy tính. Tuy nhiên, việc xác định trách nhiệm pháp lý trong một số trường hợp xảy ra lỗi, vi phạm khi triển khai, ứng dụng các hệ thống AI trong y tế, giáo dục, nông nghiệp v.v trở nên khó khăn hơn vì tính chất “hộp đen” trong quá trình AI đưa thông tin đầu ra cho con người.
- Liên quan đến vấn đề sở hữu trí tuệ, câu hỏi về trách nhiệm pháp lý đối với hành vi vi phạm SHTT trong bối cảnh nội dung do AI tạo ra vẫn là một vấn đề gây tranh cãi, đặc biệt là đối với các nền tảng cho phép người dùng tải lên tài liệu. Luật SHTT hiện hành tại Việt Nam không đề cập cụ thể đến vấn đề nội dung do AI tạo ra xâm phạm quyền SHTT.
- Pháp luật Việt Nam còn thiếu những quy định cụ thể điều chỉnh việc kiểm soát rủi ro đối với quyền riêng tư, sử dụng dữ liệu cá nhân, bảo mật trong phát triển, triển khai, ứng dụng AI, ví dụ như giám sát, theo dõi ở các địa điểm công cộng; công nghệ nhận dạng khuôn mặt hoặc công nghệ sinh trắc học khác.
- Dự định sử dụng các công cụ AI trong “chủ động ngăn chặn tội phạm” đòi hỏi phải thảo luận thêm về những tác động của công nghệ giám sát, xử lý thuật toán và chất lượng của các tập dữ liệu cho mục đích dự đoán tội phạm. Bất chấp tiềm năng về hiệu quả của chúng, các báo cáo về sự thiên vị, phân biệt đối xử dẫn đến sai sót trong tư pháp hình sự ở Hoa Kỳ, Vương quốc Anh là những bài học cảnh báo cho Việt Nam.
- Luật An ninh mạng năm 2018, Nghị định 147 yêu cầu chủ sở hữu nền tảng phải chủ động ngăn chặn, phát hiện và xóa nội dung có hại theo yêu cầu của các cơ quan có thẩm quyền. Tuy nhiên, không rõ các nền tảng tạo nội dung với sự hỗ trợ của AI có thể bị cáo buộc vi phạm hay không, vì việc tạo nội dung dựa trên thuật toán chủ yếu là tự động.
- Pháp luật lao động, việc làm chưa có quy định về một số vấn đề mới nổi liên quan đến sử dụng AI tại nơi làm việc như: tuyển dụng lao động; theo dõi hành vi của nhân viên nhằm đo lường và đảm bảo năng suất; giảm thiểu rủi ro mất việc làm do công nghệ AI.
- Cơ chế khiếu nại, khiếu kiện, giải quyết tranh chấp nói chung, cũng như khiếu kiện về thiệt hại, vi phạm do việc triển khai, sử dụng AI gây ra là một khoảng trống lớn ở Việt Nam. Tòa án các cấp ở Việt Nam chưa xem xét một vụ kiện nào liên quan đến AI; và dường như chưa sẵn sàng đối mặt với lĩnh vực mới mẻ, phức tạp này.

### 3. Khuyến nghị chính sách cho Việt Nam

Để tận dụng lợi ích của AI đồng thời giảm thiểu các rủi ro đối với xã hội, con người, Việt Nam nên xem xét các nhóm giải pháp sau:

#### 3.1. Cách tiếp cận chung

**Cân bằng giữa kiểm soát rủi ro với phát triển AI:** Các quy định, biện pháp kiểm soát, giảm thiểu rủi ro xã hội và bảo vệ quyền con người nhưng không làm cản trở hay đình trệ các tiến bộ và cải cách sáng tạo; cân bằng được yêu cầu vừa tạo ra môi trường và độ linh hoạt cho sự phát triển AI, đồng thời kiểm soát, giảm thiểu các mối nguy cơ, rủi ro với xã hội. Điều đáng chú ý là ở nhiều nước, cách tiếp cận thúc đẩy, hỗ trợ phát triển AI, đổi mới, sáng tạo không có nghĩa là tránh né ban hành các quy định để kiểm soát rủi ro. Tuy nhiên, việc ban hành các quy định mới trước hết cần xem xét khuôn khổ pháp luật hiện tại đã đủ khả năng để xử lý các vấn đề liên quan đến AI hay chưa; nếu không đủ mới cân nhắc ban hành quy định mới. Đối với nhiều vấn đề liên quan đến AI, có thể lựa chọn cách tiếp cận “luật mềm” như các nguyên tắc đạo đức, bộ quy tắc ứng xử, hướng dẫn kỹ thuật. Pháp luật có thể cần can thiệp, nhưng hiện chưa phải thời điểm quy định chặt chẽ, để dành không gian phát triển AI.

**Kết hợp các công cụ chính sách khác nhau:** Như kinh nghiệm các nước cho thấy, để điều chỉnh các vấn đề liên quan đến AI, chỉ riêng pháp luật là không đủ, mà cần áp dụng các công cụ chính sách khác nhau: Các nguyên tắc về AI có trách nhiệm, đáng tin cậy; khuôn khổ chính sách, pháp luật (gồm cả các luật hiện hành và quy định mới về AI); cơ chế thử nghiệm có kiểm soát (sandbox); các tiêu chuẩn trong từng ngành; hướng dẫn, bộ quy tắc ứng xử.

Cụ thể, cần sửa đổi, bổ sung Chiến lược quốc gia về AI cho phù hợp với những diễn biến mới trong quá trình phát triển của công nghệ AI; đề ra tầm nhìn, mục tiêu lộ trình phát triển AI; chú trọng đồng thời cả thúc đẩy, hỗ trợ phát triển AI cùng với quản trị rủi ro của AI; bảo đảm AI có trách nhiệm, đáng tin cậy, tuân theo các giá trị của con người, quyền con người trong suốt vòng đời của hệ thống AI ưu tiên phát triển AI theo lĩnh vực hẹp trên cơ sở ứng dụng các mô hình AI đã được phát triển trên thế giới. Chiến lược cũng cần giao trách nhiệm cụ thể, rõ ràng, hợp lý hơn cho các bộ, ngành; vai trò của các bên liên quan trong phát triển AI ở Việt Nam.

Trong xây dựng khuôn khổ pháp luật về AI, như kinh nghiệm trên thế giới cho thấy, có thể áp dụng hai cách tiếp cận chính: theo chiều rộng, tức là một khung chung, có thể là một đạo luật riêng bao quát về AI; và theo chiều hẹp, tức là các văn bản pháp luật khác nhau điều chỉnh các nội dung cụ thể liên quan đến AI. Cần xây dựng cơ chế thử nghiệm có kiểm soát (sandbox) tạo ra không gian cho các doanh nghiệp có thể thử nghiệm các sản phẩm, dịch vụ đổi mới chưa được điều chỉnh bởi khuôn khổ hiện hành; được miễn trừ các quy định pháp lý hoặc quy trình tuân thủ có liên quan; với thời hạn thử nghiệm nhất định, từ 06 – 24 tháng. Các thông tin kỹ thuật, thị trường và dữ liệu được thu thập trong quá trình thử nghiệm giúp các cơ quan quản lý đánh giá khung pháp lý hiện hành có phù hợp hay cần được điều chỉnh.

Cần tiếp tục xây dựng các Bộ tiêu chuẩn, quy chuẩn kỹ thuật liên quan đến AI có tính chất tự nguyện, theo các tiêu chuẩn quốc tế của ISO như các tiêu chuẩn về quản trị AI trong doanh nghiệp; về vòng đời của hệ thống AI; về đạo đức AI; các nội dung cần chú ý về xã hội, môi trường liên quan đến AI. Bên cạnh đó, tổ chức, doanh nghiệp, cá nhân có thể theo cơ chế tuân thủ tự nguyện các bộ quy tắc ứng xử, hoặc các bộ tiêu chuẩn, quy chuẩn công nghệ. Có thể xây dựng bộ quy tắc ứng xử AI trong từng ngành nghề có sử dụng AI hướng dẫn cụ thể, đáp ứng theo đặc thù của ngành nghề.

**Tiếp cận tổng thể, thu hút các bên liên quan:** Xét đến bản chất xuyên suốt của AI, cần tránh các cách tiếp cận phân mảnh trong điều chỉnh AI. Để phối hợp các cơ quan khác nhau, có thể cân nhắc phương án thiết lập cơ chế liên ngành có tính đại diện, hiệu quả và cân bằng các quan điểm, góc nhìn khác nhau để chính sách phù hợp với lợi ích rộng lớn hơn của công chúng. Các bên liên quan trong xã hội đa dạng cần được tham gia từ đầu trong vòng đời của hệ thống AI; không chỉ các chuyên gia kỹ thuật, doanh nghiệp, mà còn cả các chuyên gia pháp lý, các chuyên gia về quyền con người, đại diện từ cộng đồng, các tổ chức xã hội, và các nhóm khác có thể bị ảnh hưởng bởi hệ thống AI. Việc ban hành chính sách điều chỉnh AI cần hướng tới công bằng xã hội,

bảo đảm quyền, lợi ích, có sự tham gia của các nhóm người dễ tổn thương như người khuyết tật, dân tộc thiểu số, người nghèo, phụ nữ, trẻ em.

### **Một số nguyên tắc trong xây dựng khuôn khổ chính sách về AI:**

**Thúc đẩy phát triển AI có đạo đức:** Cần cân nhắc, thảo luận kỹ lưỡng những vấn đề đạo đức trong phát triển và triển khai các hệ thống AI; trong đó đảm bảo rằng các thuật toán AI được thiết kế giảm thiểu thiên kiến, thúc đẩy sự công bằng.

**Minh bạch và trách nhiệm giải trình:** Công chúng cần được thông tin về cách thức AI được sử dụng; về các thuật toán được sử dụng trong các quy trình ra quyết định; giúp người dân hiểu và đánh giá các quyết định do hệ thống AI đưa ra có ảnh hưởng đến quyền lợi của họ. Cần tạo các cơ chế để các nhà phát triển và người sử dụng AI chịu trách nhiệm về các vi phạm quyền con người. Điều này bao gồm việc thành lập các cơ quan quản lý để giám sát các ứng dụng AI và đảm bảo tuân thủ các tiêu chuẩn về quyền con người.

**Giám sát sử dụng AI:** Cần xem xét thành lập thiết chế giám sát, theo dõi việc triển khai AI và đảm bảo tuân thủ các hướng dẫn đạo đức và yêu cầu pháp lý. Chẳng hạn, ở khá nhiều nước có các Ủy ban đạo đức AI xem xét việc sử dụng AI trong các dịch vụ công, đưa ra các khuyến nghị để đảm bảo rằng các hệ thống AI không vi phạm quyền lợi của cá nhân. Có thể lồng ghép nội dung về AI vào các cơ chế khiếu nại, khiếu kiện hiện hành ở Việt Nam, đảm bảo rằng có các quy trình rõ ràng để các cá nhân có thể thách thức các quyết định do AI đưa ra, đặc biệt là trong các lĩnh vực như thực thi pháp luật, dịch vụ xã hội, tuyển dụng, sử dụng lao động.

### **3.2. Chính sách hỗ trợ, thúc đẩy phát triển hệ sinh thái AI**

**Phát triển thị trường AI:** Đối với Việt Nam, nên ưu tiên theo hướng phát triển các giải pháp, ứng dụng, công cụ AI trong từng lĩnh vực hẹp để ứng dụng vào kinh doanh, công việc ở các ngành thì phù hợp hơn, vì ít tốn kém hơn; dữ liệu cấu trúc có sẵn hơn (phần lớn dữ liệu của cơ quan nhà nước là dữ liệu có cấu trúc); con người hiểu rõ dữ liệu và thuật toán thực hiện, dễ phân biệt đúng - sai và dễ kiểm soát AI hơn. Việt Nam có thể có thể xây dựng các giải pháp AI của mình (thay vì chỉ làm gia công phần mềm cho bên ngoài) bằng cách thúc đẩy hỗ trợ các công ty khởi nghiệp hơn, tận dụng lợi thế hiện tại của nhân lực, đồng thời nâng cao năng lực cho lực lượng lao động theo kịp GenAI.

Với nguồn lực khan hiếm và hạn chế của Việt Nam, việc lựa chọn ưu tiên tập trung một số lĩnh vực để nghiên cứu, ứng dụng, tạo điều kiện thuận lợi cho đầu tư vào AI có vẻ phù hợp, thu hút và thúc đẩy đầu tư từ các khu vực công và tư nhân, từ các doanh nghiệp nước ngoài. Các lĩnh vực như vậy có thể là khử cacbon và các công nghệ sạch để hỗ trợ mục tiêu phát thải ròng bằng không; các ứng dụng trong lĩnh vực nông nghiệp và thực phẩm để đảm bảo an ninh lương thực; các ứng dụng trong chăm sóc sức khỏe.

Đầu tư vào nâng cao năng lực của SMEs là rất quan trọng, với các biện pháp, cơ chế hỗ trợ SMEs tiếp cận nguồn lực, hưởng lợi ích từ việc khai thác và triển khai AI. Cụ thể như chính sách cải thiện khả năng tiếp cận vốn, bao gồm thông qua các chương trình cho vay và tài trợ lãi suất thấp được thiết kế để hỗ trợ quá trình chuyển đổi bắt kịp với AI; tư vấn miễn phí cho doanh nghiệp áp dụng AI an toàn, hiệu quả; tập huấn các kỹ năng chuyên sâu về AI. Hợp tác của chính phủ với các công ty công nghệ lớn có nguồn lực mạnh về tài chính, công nghệ, con người là một giải pháp chính sách được áp dụng ở khá nhiều nước, ví dụ các nước Đông Nam Á với Google.

### **Chính sách phát triển hạ tầng, dữ liệu cho AI**

Chính sách quốc gia cần coi hạ tầng như siêu máy tính, điện toán đám mây, năng lực tính toán là nền tảng không thể thiếu để nghiên cứu, phát triển, ứng dụng AI. Cần áp dụng chính sách “Điện toán đám mây công cộng trên hết”, ưu tiên cơ sở hạ tầng và dịch vụ CNTT dựa trên đám mây công cộng và các phương án sử dụng đám mây khác. Vì đám mây công cộng sử dụng cơ sở hạ tầng dùng chung và tập hợp tài nguyên trong các trung tâm dữ liệu phân tán, nên nó có xu hướng mang lại hiệu quả hoạt động cao hơn nhiều trong các trường hợp sử dụng so với đám mây riêng. Việc ưu tiên các mô hình đám mây công cộng sẽ cho phép các doanh nghiệp mở rộng quy mô tài nguyên

điện toán và lưu trữ của mình trong khi vẫn đảm bảo tính bảo mật, là điều rất cần thiết cho khối lượng công việc AI, không cần phải đầu tư nhiều vốn.

Việc tối ưu hóa hơn nữa dữ liệu công khai để đào tạo các mô hình AI sẽ thúc đẩy đổi mới có ý nghĩa, phù hợp với bối cảnh Việt Nam. Cần thực hiện kết hướng tới tối ưu hóa các bộ dữ liệu mở của Nhà nước, tạo điều kiện tiếp cận dữ liệu tiếng Việt có giá trị và chất lượng cao để sử dụng đào tạo các mô hình AI. Đồng thời, để tăng cơ hội, điều kiện tiếp cận dữ liệu – một nguồn lực cho nghiên cứu, phát triển, ứng dụng AI, rất cần đến chính sách về dữ liệu mở, luồng dữ liệu xuyên biên giới đáng tin cậy.

**Chính sách phát triển nguồn nhân lực AI:** Để giải quyết vấn đề nhân lực trong bối cảnh phát triển nhanh chóng của AI, câu hỏi mấu chốt ở đây là: AI làm được gì và không làm được gì. Khi đã xác định rõ câu trả lời, có thể định hướng chung về phát triển nhân lực, việc làm kể cả từ cấp độ cá nhân người lao động, tổ chức/doanh nghiệp, và chính sách quốc gia. Đối với một số công việc, cần chấp nhận thực tế AI có thể thay thế hoàn toàn con người. Ở các phần còn lại, việc con người cộng sinh với AI ở các mức độ khác nhau là cần thiết để phát triển.

AI thay đổi những yêu cầu về mặt kỹ năng, cho nên nhân sự cần thay đổi và đáp ứng được nhu cầu của thị trường lao động bằng cách tham gia đào tạo và cập nhật kiến thức. Người sử dụng lao động phải có vai trò trong việc hỗ trợ nâng cao và đào tạo lại kỹ năng của họ trong bối cảnh công nghệ phát triển. Các doanh nghiệp, tổ chức sẽ cần phải thiết lập các quan hệ đối tác đào tạo AI liên ngành mới để đảm bảo rằng người lao động trong mọi ngành đều sẵn sàng khai thác AI.

Cần sự can thiệp và hỗ trợ của chính phủ trong việc xây dựng các chính sách công để tăng cường giáo dục, đào tạo, hỗ trợ người lao động. Chính phủ có thể hợp tác với khu vực tư nhân, các trường đại học để thiết kế các chương trình đào tạo phù hợp với nhu cầu của từng ngành; khuyến khích các doanh nghiệp phát triển các chương trình liên ngành, phản ánh toàn bộ các kỹ năng cần thiết. Việc tổng hợp thông tin về các rủi ro và cơ hội mà AI mang lại cho các nhóm người lao động cụ thể sẽ cho phép các nhà chính sách có các biện pháp hỗ trợ xác thực khác nhau, nắm bắt lợi ích của AI (năng suất tăng và tăng trưởng kinh tế) mà không làm gia tăng bất bình đẳng. Cần đặc biệt hỗ trợ người lao động có trình độ thấp và trung bình, có mức lương thấp, cộng đồng ở nông thôn, miền núi phải đối mặt với những trở ngại đáng kể khi tái gia nhập thị trường việc làm.

Bản thân các cơ quan Nhà nước cũng cần nhiều chuyên môn về AI hơn để khai thác hiệu quả AI. Chính phủ nên xây dựng và mở rộng quy mô các chương trình nâng cao năng lực AI “nội bộ” cho các công chức; tích hợp các nội dung liên quan đến công nghệ số như sử dụng AI vào các chương trình đào tạo, bồi dưỡng cán bộ, công chức hiện nay ở Việt Nam.

**Chính sách giáo dục:** Việt Nam cần một hệ thống giáo dục và đào tạo chuẩn bị cho người lao động có thể học hỏi sử dụng AI trong suốt sự nghiệp, coi AI là một thành phần cốt lõi của hệ thống giáo dục từ tiểu học và phát triển chuyên môn về sau. Cần cập nhật khuôn khổ chương trình giảng dạy, tăng cường giáo dục STEM với trọng tâm vào kiến thức về AI; tích hợp các khái niệm về AI vào chương trình giảng dạy của trường ngay từ khi còn nhỏ; mở rộng các tài nguyên học tập có thể truy cập; phát triển các chương trình cung cấp cho người lao động kinh nghiệm thực tế trong việc áp dụng AI; các sáng kiến đào tạo AI cho các doanh nghiệp vừa và nhỏ và các vùng nông thôn; đầu tư mạnh vào giáo dục STEM từ khi còn nhỏ, bồi dưỡng một thế hệ những người có tư duy về AI, giải quyết vấn đề.

Các kiến thức, kỹ năng AI ở cấp đại học cần được trang bị hướng tới hai nhóm nhân lực chính: nhóm phát triển mô hình AI (nhân sự trong lĩnh vực công nghệ, với năng lực kỹ thuật đặc thù) và nhóm ứng dụng mô hình AI (nhân sự trong các lĩnh vực kinh doanh, marketing, chăm sóc khách hàng v.v biết cách sử dụng hiệu quả các công cụ công nghệ mới). Ngay cả những chuyên gia trong lĩnh vực công nghệ và dữ liệu cũng cần liên tục trau dồi kiến thức về cách các mô hình này hoạt động để tự tin ứng dụng chúng trong công việc.

Sẽ có những vị trí việc làm mới mà doanh nghiệp cần các trường đại học đào tạo để tuyển dụng, chẳng hạn như chuyên gia về ngôn ngữ học, kiểm soát chất lượng AI, chỉnh lý AI (AI editors), kỹ sư (prompt engineers). Mức độ mà AI có thể ảnh hưởng đến từng nhóm công việc của tổ chức, doanh nghiệp với mức độ khác nhau - hoàn toàn tự động hóa, ảnh hưởng nhiều, một phần, hoặc

không bị ảnh hưởng. Như vậy, chương trình đào tạo nhân lực AI cho các tổ chức, doanh nghiệp cũng cần xuất phát từ đặc điểm này.

**Nghiên cứu về AI:** Học hỏi kinh nghiệm các nước, cần có các chính sách hỗ trợ nghiên cứu và phát triển AI thông qua nguồn tài trợ có mục tiêu, thành lập các trung tâm nghiên cứu lớn, hợp tác giữa các trường đại học và các ngành kinh tế. Một phần nguồn lực cả từ Nhà nước và khu vực tư vẫn nên dành cho nghiên cứu nền tảng về AI để nắm được công nghệ cốt lõi về AI. Để tham khảo, tại thời điểm 2024, Chiến lược quốc gia trên dưới 20 nước trong số hơn 80 nước đã đề ra mục tiêu, giải pháp phát triển các mô hình AI nền tảng. Mặt khác, các mô hình nguồn mở đã nhanh chóng bắt kịp về khả năng trong vài năm qua, mở ra cơ hội cạnh tranh cho các nước như Việt Nam. Bên cạnh đó, không phải tất cả các nghiên cứu AI đều yêu cầu cơ sở vật chất giống nhau. Do đó, chiến lược cơ sở hạ tầng AI cần hướng tới các lĩnh vực trọng tâm, giúp hỗ trợ các công nghệ AI cụ thể có mục tiêu hơn, quản lý các nguồn lực một cách hiệu quả.

Chính phủ Việt Nam cần tạo cơ chế ở cấp Bộ để liên kết giữa tất cả các trường đại học tham gia vào các hoạt động nghiên cứu AI rộng rãi hơn. Cần cung cấp nhiều kinh phí hơn cho các trường đại học và viện để tiến hành nghiên cứu AI và tạo điều kiện trao đổi kiến thức và chuyên gia giữa các tổ chức trong nước và nước ngoài. Cả Nhà nước và khu vực tư đều có thể hỗ trợ các nhà nghiên cứu thông qua các chương trình như khuôn khổ chuyển giao công nghệ, học bổng và hỗ trợ trực tiếp cho nghiên cứu. Cần khuyến khích các hoạt động, chương trình của khu vực tư, doanh nghiệp, tổ chức, cá nhân phổ biến kiến thức, kỹ năng AI.

### 3.3. Chính sách kiểm soát AI

Có thể học hỏi có chọn lọc các chính sách kiểm soát rủi ro từ việc phát triển, sử dụng AI trên thế giới, như Khung quản lý rủi ro AI (AI RMF) do Viện Tiêu chuẩn và Công nghệ Quốc gia (NIST) Hoa Kỳ công bố đầu năm 2023; được cập nhật các nội dung về AI tạo sinh năm 2024; Đạo luật AI của EU, Các nguyên tắc về AI của OECD.

**Quản trị rủi ro của AI một cách cân bằng:** Trên phương diện quản trị rủi ro của AI, có thể đạt mục tiêu cân bằng với thúc đẩy phát triển AI bằng các cách thức sau:

- Đánh giá, điều chỉnh AI theo các mức độ rủi ro để có phương án phù hợp với mức độ rủi ro đã được xác định dựa trên các dấu hiệu, nguyên tắc; mỗi cấp độ rủi ro có những biện pháp kiểm soát phù hợp. Cách tiếp cận này đặt trọng tâm vào các biện pháp tiền kiểm (ex ante), ngăn ngừa không để rủi ro lan rộng gây ra những hậu quả nặng nề, hơn là trông chờ vào các biện pháp áp đặt trách nhiệm thị trường về sau, hoặc khiếu kiện dân sự ở tòa án kéo dài, tốn kém.
- Các công cụ chính sách cũng cần được áp dụng tương thích với mức độ rủi ro: đối với rủi ro mức độ cao nhất, những quy định pháp luật khắt khe, có tính bắt buộc thực hiện là cần thiết; đối với rủi ro ở mức độ tiếp theo, các yêu cầu pháp lý nên “nhẹ nhàng” hơn; đối với các rủi ro thấp, chỉ cần áp dụng hướng dẫn, tiêu chuẩn kỹ thuật, bộ quy tắc ứng xử.
- Sự cân bằng với thúc đẩy, phát triển AI đòi hỏi kiểm soát rủi ro theo phạm vi, mức độ nắm giữ quyền lực của các chủ thể tham gia vào vòng đời hệ thống AI; theo đó, chủ thể nào nắm giữ quyền lực nhiều hơn thì phải chịu sự kiểm soát chặt chẽ hơn và ngược lại. Trong đó, cần kiểm soát chặt rủi ro từ AI ở khu vực công.
- Cơ hội và rủi ro của AI đối với xã hội, con người, quyền con người có thể được đánh giá, điều chỉnh, ứng phó theo từng công đoạn của vòng đời hệ thống, đòi hỏi biện pháp, đối sách phù hợp. Trong vòng đời của hệ thống AI, cần xác định rõ ràng, cụ thể trách nhiệm của từng chủ thể: Nhà phát triển AI, nhà triển khai triển khai/ứng dụng, người dùng cuối.
- Trong quá trình xây dựng, thực thi chính sách về AI, cần cân bằng giữa các giá trị đối lập, chẳng hạn như quyền riêng tư và tính minh bạch. Cách tiếp cận phù hợp là tùy theo ngữ cảnh, trong đó các bên liên quan ưu tiên các giá trị dựa trên ứng dụng cụ thể của hệ thống AI; thích ứng các loại hệ thống AI với bối cảnh khác nhau.

## Giải pháp lập pháp:

Trong xây dựng pháp luật về AI, cách tiếp cận hiệu quả nhất là kết hợp vừa ban hành khung pháp luật chung về AI, vừa có những quy định cho các công cụ AI cụ thể. Khung pháp luật chung về AI cần có các nội dung lớn sau đây: đưa ra khái niệm rõ ràng về AI, hệ thống AI, mô hình AI để làm điểm xuất phát cho các nội dung điều chỉnh khác về công nghệ này; xác định các nguyên tắc cốt lõi trong quản trị AI gồm: lấy con người, các giá trị của con người làm căn cứ cho toàn bộ vòng đời của hệ thống AI, minh bạch, trách nhiệm giải trình, đáng tin cậy; xác định các mức độ rủi ro và các biện pháp thích ứng với mức độ rủi ro; xác định trách nhiệm cụ thể của các chủ thể liên quan trong vòng đời của hệ thống AI gồm bên phát triển, bên triển khai/ứng dụng, người dùng AI. Trong đó, cần quy định trách nhiệm của các bộ, ngành về việc xây dựng, thực hiện các bộ tiêu chuẩn kỹ thuật, bộ quy tắc ứng xử về đạo đức AI có tính chất tự nguyện.

Đặc biệt, khung pháp luật chung cần quy định rõ ràng, chặt chẽ đối với các trường hợp sau đây:

- Các hệ thống AI phải bị hạn chế trong những trường hợp chúng gây ra mối đe dọa đáng kể và sắp xảy ra đối với tính mạng hoặc sự an toàn của cá nhân hoặc nhóm người.
- Theo Điều 16.1 Hiến pháp 2013, cần cấm các hệ thống AI gây ra phân biệt đối xử, đảm bảo quyền bình đẳng cho tất cả mọi người. Điều này bao gồm cả phân biệt đối xử trực tiếp (ví dụ: dựa trên độ tuổi, giới tính, tôn giáo) và phân biệt đối xử gián tiếp (ví dụ: thông qua điểm số xã hội dẫn đến đối xử bất lợi). Khung này cần yêu cầu tính minh bạch về thuật toán trong quy trình ra quyết định của AI để ngăn ngừa sự thiên vị và kết quả phân biệt đối xử; ví dụ thuật toán tuyển dụng, bắt buộc kiểm toán để ngăn ngừa các kết quả thiên vị trong các quyết định tuyển dụng.
- Bảo vệ quyền riêng tư và dữ liệu cá nhân: Các hệ thống AI không được xâm phạm đời tư, bí mật cá nhân hoặc bí mật gia đình của cá nhân mà không có sự đồng ý rõ ràng. Việc thu thập, lưu trữ, sử dụng và phổ biến dữ liệu cá nhân, bao gồm dữ liệu sinh trắc học và hình ảnh cá nhân, phải tuân theo các quy định và giám sát chặt chẽ. Cần xây dựng các hướng dẫn rõ ràng, thủ tục đơn giản và minh bạch để có được sự đồng ý của cá nhân khi sử dụng dữ liệu của họ trong các hệ thống AI; với các điều khoản về việc rút lại sự đồng ý bất cứ lúc nào.
- Bảo vệ danh dự, phẩm giá và uy tín: Các hệ thống AI phải bị cấm thực hiện các hành động xâm phạm danh dự, phẩm giá và uy tín của cá nhân, phù hợp với các nguyên tắc văn hóa và pháp lý của Việt Nam. Khung này cần giải quyết các tác động về mặt đạo đức của công nghệ AI, ngăn chặn việc sử dụng sai mục đích có thể gây tổn hại đến danh tiếng của cá nhân hoặc dẫn đến kỳ thị xã hội.
- Việc sử dụng AI trong điều tra tội phạm phải được quản lý chặt chẽ, chỉ cho phép truy cập dữ liệu cá nhân theo sự cho phép của tòa án, với các điều khoản dành các tình huống khẩn cấp.
- Bảo vệ các nhóm dễ bị tổn thương: Cần có các biện pháp bảo vệ đặc biệt dành cho các nhóm yếu thế, chẳng hạn như người già, người khuyết tật hoặc người có hoàn cảnh kinh tế khó khăn. Các hệ thống AI khai thác những điểm yếu này, có khả năng gây ra tác hại về mặt tâm lý hoặc hành vi, cần được quản lý chặt chẽ. Điều này bao gồm đảm bảo rằng các công nghệ AI không thao túng hoặc ép buộc những cá nhân dễ bị tổn thương hành động trái với lợi ích của họ.

Đồng thời, cần ban hành các văn bản quy phạm pháp luật, có thể là nghị định, thông tư quy định cụ thể hơn về các vấn đề kỹ thuật, chuyên sâu liên quan đến AI, chẳng hạn như điều chỉnh thuật toán để đảm bảo sự minh bạch, công bằng, tránh phân biệt đối xử; sử dụng công nghệ nhận diện khuôn mặt; sử dụng AI trong một số lĩnh vực tiềm ẩn nhiều rủi ro như tuyển dụng, sử dụng, đánh giá lao động v.v

### **Tích hợp các giá trị của con người trong kiểm soát rủi ro của AI**

Điều chỉnh AI dựa trên các chuẩn mực quốc tế về các giá trị chung của loài người như dân chủ, pháp quyền, quyền con người là một cách tiếp cận được thống nhất rộng khắp trên thế giới. Điểm


hình là Nghị quyết của Đại Hội đồng Liên Hợp quốc nhấn mạnh: "quyền con người và các quyền tự do cơ bản phải được tôn trọng, bảo vệ và thúc đẩy trong suốt vòng đời của các hệ thống trí tuệ nhân tạo". Công ước khung của EU về AI với quyền con người, dân chủ, pháp quyền đặt ra mục tiêu và yêu cầu các quốc gia thành viên phải đảm bảo mọi hoạt động trong vòng đời của các hệ thống AI hoàn toàn phù hợp với quyền con người, dân chủ, pháp quyền.

Việt Nam đã ký kết, tham gia hầu hết các điều ước quốc tế có nội dung liên quan đến quyền con người; Hiến pháp năm 2013 dành một chương riêng (Chương II) về quyền con người, quyền công dân; nhiều đạo luật có các nội dung cụ thể hóa quy định của Hiến pháp về quyền con người. Dựa trên các căn cứ pháp lý này, cần phải thực hiện các biện pháp nhằm đảm bảo rằng các hoạt động trong vòng đời các hệ thống AI nhất quán với các nghĩa vụ bảo vệ quyền con người được quy định trong các văn kiện luật quốc tế có hiệu lực và pháp luật quốc gia; không đánh đổi một bước lùi dù nhỏ nhất về quyền con người cho bất cứ một phát triển nào của công nghệ.

Việc triển khai, ứng dụng AI nếu có mục đích hoặc hệ quả giới hạn quyền con người cũng cần phải đều cần phải tuân thủ các điều kiện sau đây: (1) Việc giới hạn quyền phải được quy định rõ ràng trong luật để người dân có thể hiểu rõ và tự điều chỉnh hành vi của họ; có những biện pháp thích hợp để phòng ngừa việc lạm dụng quy định giới hạn quyền, hoặc tùy tiện đặt ra các giới hạn mới; (2) Biện pháp hạn chế phải hợp lý, cần thiết, và tương xứng; (3) Chỉ được đặt ra giới hạn quyền nhằm mục đích thúc đẩy phúc lợi chung của cộng đồng hoặc để bảo vệ an ninh quốc gia, an toàn của cộng đồng, trật tự công cộng, sức khỏe cộng đồng, đạo đức xã hội.

Cần chuyển đổi các giá trị cốt lõi của con người, quyền con người thành các đặc điểm kỹ thuật và tính năng kỹ thuật trong hệ thống AI như: thiết kế các thuật toán ưu tiên công bằng, phát triển giao diện người dùng tôn trọng quyền riêng tư, hoặc tạo ra các cơ chế cho tính minh bạch và trách nhiệm. Ví dụ, nếu công bằng là một giá trị quan trọng, quá trình thiết kế sẽ liên quan đến việc tạo ra các thuật toán được kiểm tra để loại trừ thiên kiến và đảm bảo rằng kết quả của hệ thống AI đảm bảo tính công bằng.





# I.

# GIỚI THIỆU TỔNG QUAN

## 1. BỐI CẢNH NGHIÊN CỨU

Trong kỷ nguyên số hóa, công nghệ trí tuệ nhân tạo (AI) đang trở thành động lực quan trọng thúc đẩy sự phát triển kinh tế - xã hội trên toàn cầu. Với khả năng xử lý dữ liệu khổng lồ và học hỏi không ngừng, AI đã và đang mở ra những tiềm năng mới trong mọi lĩnh vực, từ y tế, giáo dục, giao thông cho đến nông nghiệp, công nghiệp và quản lý đô thị. Từ việc tối ưu hóa sản xuất, cải thiện năng suất lao động đến việc hỗ trợ ra quyết định chính xác và hiệu quả hơn, AI mang đến cơ hội to lớn để thúc đẩy sự phát triển toàn diện. Cụ thể như theo dự báo của Goldman Sachs, AI có thể thúc đẩy tăng trưởng năng suất toàn cầu thêm 1,5% mỗi năm trong vòng 10 năm và tăng GDP toàn cầu thêm khoảng 7 nghìn tỷ USD nếu được triển khai hiệu quả<sup>1</sup>.

Tuy nhiên, đi liền với những lợi ích vượt trội, AI cũng đặt ra không ít thách thức, đặc biệt là những rủi ro liên quan đến quyền con người. Chẳng hạn, việc triển khai AI trên diện rộng có thể làm gia tăng nguy cơ xâm phạm quyền riêng tư, khi khối lượng dữ liệu cá nhân khổng lồ được thu thập và phân tích một cách tự động. Đồng thời, các thuật toán AI, nếu không được giám sát và điều chỉnh hợp lý, có thể vô tình duy trì hoặc khuếch đại các định kiến xã hội, dẫn đến tình trạng bất bình đẳng, kỳ thị, phân biệt đối xử trong nhiều khía cạnh của đời sống. Người được mệnh danh là “cha đẻ” của AI, Giáo sư Geoffrey Hinton, chủ nhân giải Turing 2018 và Nobel Vật lý 2024, đã quyết định từ bỏ công việc tại Google để có thể công khai cảnh báo về các rủi ro của công nghệ này đem lại. Không phải ngẫu nhiên mà theo một khảo sát ở 32 nước, những người được hỏi vừa hào hứng vừa lo lắng về AI: 53% cho biết họ hào hứng với các sản phẩm và dịch vụ có sử dụng AI, so với 50% trả lời, AI khiến họ lo lắng<sup>2</sup>.

Lợi ích và rủi ro mà AI mang lại đòi hỏi các Chính phủ phải có chính sách kiểm soát rủi ro của AI song hành với phát triển của bản thân công nghệ AI và thị trường (ứng dụng) công nghệ. Một trong những thách thức lớn nhất hiện nay là đảm bảo sự cân bằng giữa phát triển AI và bảo vệ các giá trị cốt lõi mà nhân loại đã mất hàng nghìn năm để có được và gìn giữ như tự do, bình đẳng, công bằng. Các quốc gia cần xây dựng khung chính sách phù hợp để thúc đẩy tiến bộ đồng thời quản trị rủi ro của AI.

Trong bối cảnh đó, Việt Nam cần có những bước đi mang tính chiến lược, bước đầu là đánh giá chính xác mức độ sẵn sàng cho phát triển và quản trị AI. Điều này bao gồm việc xem xét các khía cạnh như hạ tầng kỹ thuật, dữ liệu, nguồn nhân lực chất lượng cao, giáo dục, nghiên cứu, định hướng chính sách và quy định pháp luật liên quan đến AI.

Trước thực tế này, Viện Nghiên cứu Chính sách và Phát triển Truyền thông (IPS) đã thực hiện nghiên cứu “Chính sách về trí tuệ nhân tạo (AI) ở Việt Nam: Cân bằng giữa thúc đẩy phát triển và quản trị rủi ro”. Báo cáo của IPS tìm hiểu tính chất kép của AI – vừa mang lại cơ hội, vừa tiềm ẩn rủi ro – và đưa ra các khuyến nghị chính sách thúc đẩy phát triển và kiểm soát rủi ro của AI, trong đó có bảo đảm quyền con người ở Việt Nam.

## 2. MỤC ĐÍCH, PHẠM VI NGHIÊN CỨU

Nghiên cứu này hướng tới các mục đích sau:

<sup>1</sup> <https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent>

<sup>2</sup> The Ipsos AI Monitor 2024: Changing attitudes and feelings about AI and the future it will bring | Ipsos

- Đánh giá hiện trạng phát triển công nghệ AI tại Việt Nam trên các phương diện: cơ sở hạ tầng kỹ thuật, dữ liệu, nguồn nhân lực, tài chính.
- Đánh giá rủi ro của AI đối với xã hội, con người trong bối cảnh Việt Nam và khuôn khổ pháp luật hiện hành để quản lý rủi ro.
- Đề xuất các khuyến nghị chính sách nhằm thúc đẩy phát triển công nghệ AI đi đôi với quản trị rủi ro, đảm bảo phù hợp với thực tiễn và định hướng phát triển của Việt Nam.

Phạm vi nghiên cứu là công nghệ AI, các chính sách và văn bản quy phạm pháp luật, điều kiện cơ sở hạ tầng công nghệ, dữ liệu, nguồn nhân lực và các yếu tố liên quan đến phát triển và quản trị AI tại Việt Nam có tham khảo kinh nghiệm quốc tế về vấn đề này.

### 3. CÁCH TIẾP CẬN, PHƯƠNG PHÁP

Nghiên cứu này dựa trên cách tiếp cận của một số nghiên cứu và khung đánh giá liên quan đến mức độ sẵn sàng trong phát triển và quản trị rủi ro của AI trên toàn cầu như: Báo cáo xếp hạng mức độ sẵn sàng triển khai AI của các quốc gia (Government AI Readiness Index) do Oxford Insights và IDRC thực hiện; đánh giá sẵn sàng về AI (AI Readiness Assessment) do Chương trình Phát triển Liên Hợp Quốc (UNDP); Chỉ số AI toàn cầu (Global AI Index) do Tortoise Media thực hiện. Các khung này đều đánh giá hiện trạng phát triển của AI ở một quốc gia theo các trụ cột chính với các tiêu chí cụ thể như: Khung chính sách, pháp luật; cơ sở hạ tầng và dữ liệu; thị trường các công ty AI; đầu tư nghiên cứu và phát triển; giáo dục; chất lượng nhân lực.

Các khung đánh giá nói trên cung cấp cái nhìn toàn cảnh về trình độ phát triển AI và khuôn khổ chính sách dành cho công nghệ này trên toàn cầu, giúp từng quốc gia như Việt Nam tham khảo về năng lực công nghệ và quản trị AI của mình. Tuy nhiên, do ở phạm vi toàn cầu, nên các báo cáo trên chưa thể đem đến bức tranh chi tiết, sâu sắc về mỗi quốc gia, nghiên cứu tập trung nhiều khía cạnh định lượng nên thiếu đi các lý giải, phân tích định tính gắn với bối cảnh chính trị, văn hóa, xã hội của từng quốc gia.

Chính vì vậy, nghiên cứu của IPS tham khảo cách tiếp cận, các trụ cột cần xem xét từ các khung nói trên, đối chiếu với thực tiễn, bối cảnh Việt Nam để điều chỉnh cho phù hợp, đưa ra những khuyến nghị chính sách vừa dựa trên bằng chứng, vừa có cơ sở lý thuyết bao quát.

Nghiên cứu được thực hiện thông qua sự kết hợp của các phương pháp sau đây:

- **Nghiên cứu tài liệu (desk study):** Tổng hợp và phân tích các nghiên cứu, báo cáo, văn bản chính sách, pháp luật trong nước và quốc tế về phát triển và quản trị AI.
- **Phòng vấn và tọa đàm:** Tiến hành phỏng vấn với các chuyên gia, nhà hoạch định chính sách, đại diện doanh nghiệp hoạt động trong lĩnh vực AI (tổng cộng gần 50 người). Đồng thời, các buổi tọa đàm chuyên sâu được tổ chức để thu thập ý kiến đa chiều từ các bên liên quan.
- **Trường hợp điển hình:** Từ các tài liệu nghiên cứu, phỏng vấn sâu, phân tích các trường hợp điển hình để minh chứng cho các nhận định, khuyến nghị về chính sách AI ở Việt Nam.
- **Phân tích định tính:** Sử dụng dữ liệu và thông tin thu thập được từ các khảo sát định tính có liên quan để phân tích, đưa ra đánh giá về mức độ sẵn sàng về AI ở Việt Nam.

### 4. GIÁ TRỊ VÀ HẠN CHẾ CỦA NGHIÊN CỨU

Tại Việt Nam hiện nay, các nghiên cứu chính sách về phát triển, quản trị AI còn ít, mang tính chất riêng lẻ, tập trung vào các khía cạnh cụ thể của AI. Nghiên cứu này có cách tiếp cận tổng thể, nhằm đánh giá các khía cạnh khác nhau về mức độ sẵn sàng của quốc gia cho sự phát triển AI, cũng như nhận diện và quản trị các rủi ro liên quan đến công nghệ này ở cấp độ chính sách.

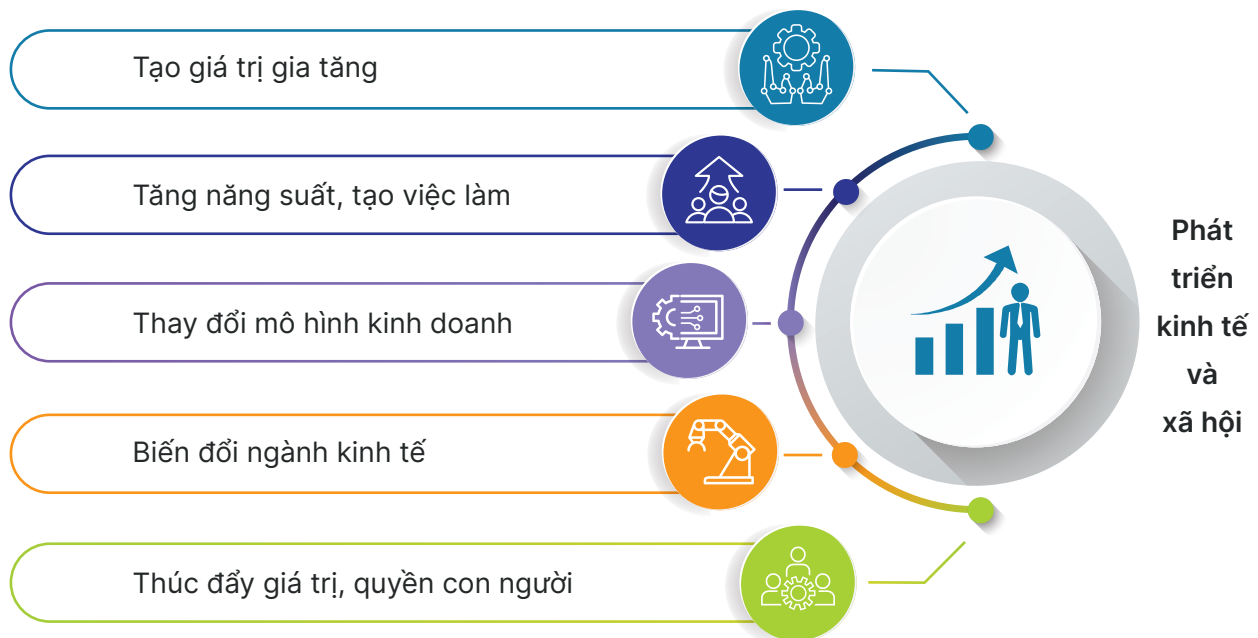
Tuy nhiên, do thời gian thực hiện, cũng như nguồn lực có hạn, tốc độ phát triển nhanh chóng của AI trên thực tế, báo cáo có thể chưa bao quát đầy đủ tất cả các khía cạnh của chính sách phát triển, kiểm soát AI, cũng như các vấn đề mới phát sinh về AI.

Mặc dù vậy, nghiên cứu này được kỳ vọng sẽ là tài liệu tham khảo chính sách quan trọng, góp phần thúc đẩy đối thoại chính sách và xây dựng các giải pháp phù hợp cho phát triển công nghệ AI an toàn, đáng tin cậy, có trách nhiệm tại Việt Nam.

# II. CƠ HỘI VÀ RỦI RO, THÁCH THỨC CỦA AI

## 2.1. CƠ HỘI, LỢI ÍCH DO AI MANG LẠI

AI có thể mang lại rất nhiều cơ hội, lợi ích cho nền kinh tế, sự phát triển của đất nước, doanh nghiệp, cá nhân con người, bảo vệ và thúc đẩy quyền con người trên nhiều phương diện như hình dưới đây thể hiện<sup>3</sup>.



Hình 1. Cơ hội, lợi ích do AI mang lại

### 2.1.1. Tạo các giá trị gia tăng

AI mang lại lợi ích thuyết phục nhất là các giá trị gia tăng (additional values) được tạo ra trong quá trình sản xuất và cung cấp dịch vụ. Các giá trị gia tăng nổi bật mà AI tạo ra như tăng năng suất và hiệu quả (tự động hóa các công việc lặp đi lặp lại, tối ưu hóa quy trình, v.v), cải thiện chất lượng sản phẩm và dịch vụ (cá nhân hóa, kiểm soát chất lượng, v.v), đổi mới sáng tạo (tăng tốc độ đổi mới sáng tạo dựa trên phân tích nguồn dữ liệu lớn hiện có), mở ra những khả năng mới (hỗ trợ người khuyết tật, thiết kế chương trình giáo dục cá nhân hóa, v.v)<sup>4</sup>. Chẳng hạn, tự động hóa (Automation) có thể được coi là yếu tố giúp tăng cường sức lao động cho con người, giải phóng nguồn nhân lực có giá trị cho công việc mang tính chiến lược và sáng tạo hơn. Điều này chắc chắn sẽ cải thiện hiệu quả và năng suất trong các ngành từ sản xuất, hậu cần, đến chăm sóc sức khỏe, tài chính.

<sup>3</sup> Xem thêm: OECD, *Assessing potential future AI risks, benefits and policy imperatives*, OECD AI Paper No.27, 2024.

<sup>4</sup> Australia's Chief Scientist, *'Rapid response Information Report: Generative AI'* Australian Government (June 1, 2023).

### 2.1.2. Tăng năng suất lao động, tạo việc làm

AI có tác động to lớn đến năng suất lao động và tạo nhiều việc làm mới trong nhiều lĩnh vực. Một khảo sát của PwC cho thấy, các ngành đặc biệt tiếp xúc với AI đang có mức tăng trưởng năng suất lao động cao hơn gần 5 lần, tạo động lực chính thúc đẩy tăng trưởng kinh tế và nâng cao mức sống<sup>5</sup>. Khảo sát CEO toàn cầu năm 2024 của PwC xác nhận, 84% CEO các công ty đã áp dụng AI tin AI sẽ giúp tăng hiệu quả làm việc của nhân viên, có nghĩa là làm những việc cũ nhanh hơn và tìm ra những cách mới, sử dụng AI để tạo ra giá trị. 70% CEO cho biết AI sẽ thay đổi đáng kể cách công ty của họ tạo ra, cung cấp và nắm bắt giá trị trong ba năm tới; 46% cho biết AI sẽ tăng lợi nhuận và 41% cho biết AI sẽ tăng doanh thu. Bên cạnh đó, 61% các nhà đầu tư tin rằng việc áp dụng AI nhanh chóng là rất quan trọng để tạo ra giá trị.

Theo một khảo sát toàn cầu khác, trong đó có Việt Nam, 60% người lao động ở Việt Nam được hỏi cho rằng AI giúp họ tăng năng suất/hiệu quả làm việc (so với tỷ lệ ở Châu Á Thái Bình Dương là 41%) và 58% coi đó là cơ hội để học các kỹ năng mới (so với Châu Á Thái Bình Dương là 34%). Các ngành như Công nghệ, Truyền thông và Viễn thông, cũng như Dịch vụ tài chính, có tiềm năng lớn nhất để cải thiện năng suất thông qua AI<sup>6</sup>. Đại diện các doanh nghiệp ở Việt Nam được khảo sát trong nghiên cứu của IPS cũng cho biết, việc ứng dụng AI đã giúp nhân viên của họ tăng năng suất lao động lên nhiều lần, tiết kiệm được nhiều thời gian, công sức. Tăng năng suất không chỉ có nghĩa là làm những việc cũ nhanh hơn, mà còn có nghĩa là tìm ra những cách mới, được hỗ trợ bởi AI để tạo ra giá trị.

Theo các ấn phẩm, báo cáo khác nhau trên thế giới, AI có thể tạo những việc làm mới. Chẳng hạn, theo Báo cáo Tương lai việc làm năm 2023 của Diễn đàn Kinh tế Thế giới, trong 5 năm tới, dự kiến sẽ có 69 triệu việc làm được tạo ra<sup>7</sup>. Tương tự, trong khuôn khổ nghiên cứu này, các chuyên gia và doanh nhân được hỏi đều cho rằng, AI có thể tạo ra các cơ hội việc làm mới trong lĩnh vực công nghệ và các ngành khác của Việt Nam. Khi nhu cầu về chuyên môn AI tăng lên, nhu cầu ngày càng tăng đối với các chuyên gia được đào tạo về khoa học dữ liệu, học máy, phát triển phần mềm, cũng như các công việc đòi hỏi cách làm việc với AI. Điều này sẽ mở ra con đường sự nghiệp mới cho những người trẻ tuổi, đồng thời thúc đẩy tăng trưởng kinh tế và đổi mới sáng tạo.

### 2.1.3. Thay đổi mô hình kinh doanh

Một nghiên cứu cho thấy 91% các doanh nghiệp nhỏ ở Hoa Kỳ sử dụng AI đã thành công trong việc tăng doanh thu, tiếp cận và thu hút khách hàng, tăng năng suất<sup>8</sup>. Còn theo khảo sát của PwC, 82% doanh nghiệp gia đình ở Việt Nam bày tỏ sự quan tâm tích cực trong việc tìm hiểu về GenAI; 67% chia sẻ sự đồng thuận tích cực về tiềm năng của GenAI như là một công cụ mạnh mẽ để chuyển đổi kinh doanh. Bên cạnh đó, 67% cho rằng AI sẽ mang lại cơ hội lớn để các doanh nghiệp gia đình tiên phong trong việc sử dụng công nghệ một cách có trách nhiệm. Về mặt cá nhân, gần hai phần ba (58%) tin rằng việc ủng hộ các sáng kiến AI sẽ giúp họ tạo dựng tên tuổi và hỗ trợ vai trò lãnh đạo trong doanh nghiệp của mình. Họ coi GenAI là chất xúc tác cho chuyển đổi kinh doanh, với các lợi ích được dự đoán như tăng năng suất của nhân viên (39%), tiết kiệm chi phí (39%) và nâng cao trải nghiệm của khách hàng (36%). Hơn một nửa trong số họ tin rằng GenAI sẽ đóng vai trò quan trọng trong việc nâng cao năng lực kỹ thuật số cho lực lượng lao động của họ, vượt qua mức trung bình toàn cầu là 41%. 73% dự đoán rằng AI sẽ giúp doanh nghiệp nâng cao tính cạnh tranh trên thị trường trong ba năm tới, mang lại nhiều áp lực cho doanh nghiệp để dẫn đầu trong bối cảnh bị chi phối mạnh mẽ bởi AI<sup>9</sup>. Các chủ doanh nghiệp, chuyên gia được hỏi trong nghiên cứu của IPS cũng chung nhận định, doanh nghiệp Việt Nam đang ngày càng chịu sức hút từ tiềm năng chuyển đổi của AI; xem AI không chỉ đơn thuần là một sự đổi mới công nghệ mà còn là chất xúc tác để tái định hình hoạt động kinh doanh, chiến lược và trải nghiệm của khách hàng, góp phần định vị doanh nghiệp để đạt được thành công trong thời đại kỹ thuật số.

<sup>5</sup> PwC, 2024 AI Jobs Barometer, <https://www.pwc.com/gx/en/issues/artificial-intelligence/ai-jobs-barometer.html>

<sup>6</sup> <https://www.pwc.com/vn/en/publications/vietnam-publications/workforce-hopes-fears-vietnam-2023.html>

<sup>7</sup> WEF, Future of Jobs Report 2023.

<sup>8</sup> Constant Contact, Small Business Now: An AI Awakening, 2023.

<sup>9</sup> <https://www.pwc.com/vn/en/publications/vietnam-publications/nextgen-survey-2024.html>

#### 2.1.4. Góp phần thay đổi các ngành kinh tế

AI có thể tạo ra tác động sâu sắc đến các ngành cụ thể ở Việt Nam; từ việc cải thiện hiệu quả sản xuất, đến chăm sóc sức khỏe, nông nghiệp, giao thông vận tải, marketing, logistics, chứng khoán, ngân hàng, truyền thông<sup>10</sup>. Qua đó, AI có tiềm năng mang lại những thay đổi đáng kể cho nền kinh tế, giúp Việt Nam đẩy nhanh con đường phát triển của mình.

Chẳng hạn, đại diện của một số ngân hàng, công ty chứng khoán khi trả lời phỏng vấn sâu thuộc nghiên cứu này cho biết, họ đã sử dụng các công cụ AI để chăm sóc khách hàng; tổng hợp, phân tích dữ liệu về thị trường, khách hàng. VPBank đã ứng dụng trí tuệ nhân tạo vào các lĩnh vực như giao dịch ngoại tệ, tín dụng cá nhân và ngân hàng số, phát triển hệ thống giám sát giao dịch dựa trên trí tuệ nhân tạo để phát hiện gian lận và rủi ro. Vietcombank đã hợp tác với Công ty FPT Smart Cloud để phát triển nền tảng chatbot chăm sóc khách hàng VCB Digibot<sup>11</sup>.

Ngành giao thông vận tải cũng ứng dụng AI rất nhiều vào hoạt động của mình. Ví dụ, Hệ thống giao thông thông minh (ITS Solution) là công nghệ được sử dụng để giải quyết các vấn đề về giao thông đường bộ như giám sát an ninh công cộng, cảnh báo ùn tắc giao thông, phát hiện vi phạm giao thông, hệ thống đếm lưu lượng phương tiện. Hệ thống này đã được sử dụng ở nhiều nước phát triển trên thế giới và bước đầu được áp dụng tại một số thành phố lớn tại Việt Nam như Hà Nội, Thành phố Hồ Chí Minh, Bình Dương, Quảng Ninh.

Ngành y tế tại Việt Nam cũng đang từng bước ứng dụng AI vào nhiều hoạt động như ra quyết định lâm sàng, hồ sơ sức khỏe điện tử, chẩn đoán. Các công nghệ AI, chẳng hạn như khai thác văn bản và học máy, có thể được sử dụng để phân tích dữ liệu y tế và cải thiện quy trình ra quyết định. Mặt khác, robot có thể hỗ trợ nhiều nhiệm vụ chăm sóc sức khỏe khác nhau, bao gồm chăm sóc người cao tuổi, phục hồi chức năng và can thiệp phẫu thuật. Việc triển khai AI trong các trường hợp cấp cứu và chấn thương vẫn đang trong giai đoạn đầu, nhưng ngày càng có nhiều sự quan tâm và công nhận về những lợi ích tiềm năng của nó<sup>12</sup>. DrAidTM do Công ty VinBrain sản xuất, hỗ trợ chẩn đoán các rối loạn tim và phổi, có thể đánh giá các chấn thương xương từ X-quang. Một nỗ lực khác để giảm bớt gánh nặng cho nhân viên y tế là sử dụng các cuộc gọi robot, sử dụng AI để tự động gọi bệnh nhân, kiểm tra sức khỏe và cập nhật thông tin y tế của họ.

AI có thể được ứng dụng nhiều trong ngành nông nghiệp Việt Nam, một trong những quốc gia sản xuất nông nghiệp lớn nhất thế giới<sup>13</sup>. Việc sử dụng công nghệ như AI có thể tối ưu hóa năng suất cây trồng, vật nuôi thông qua giám sát, phân tích và quản lý nông nghiệp chính xác, thông minh. Công nghệ AI có thể giúp nông dân tối ưu hóa năng suất, mùa vụ, sử dụng phân bón; thu thập dữ liệu về tình trạng đất, thời tiết, sự phát triển của cây trồng, lượng nước sử dụng và các yếu tố khác để đưa ra quyết định về việc gieo hạt, bón phân, kiểm soát dịch hại và các biện pháp canh tác nông nghiệp khác. AI còn hỗ trợ trong quản lý nhà nước trong ngành nông nghiệp như quy hoạch, thiết kế đất nông nghiệp; cảnh báo, giám sát bằng không ảnh.

AI đã và đang tác động đến lĩnh vực sản xuất ở Việt Nam như điện tử, dệt may và giày dép, giúp các ngành này trở nên hiệu quả hơn và cải thiện chất lượng. Các robot được hỗ trợ AI hoạt động 24/7, cải thiện năng suất và giảm nhu cầu lao động của con người. Ví dụ, chúng được lập trình để xác định và giải quyết vấn đề nhanh chóng, giảm thời gian chết và giảm hoặc loại bỏ lỗi. Với việc sử dụng robot hỗ trợ AI, các nhà sản xuất Việt Nam có thể tự động hóa các nhiệm vụ lặp đi lặp lại và nguy hiểm, giảm nguy cơ tai nạn và giải phóng người lao động để tập trung vào các nhiệm vụ có giá trị cao hơn. Ngoài ra, việc sử dụng AI trong quản lý chuỗi cung ứng và hậu cần giúp tối ưu hóa quy trình, giảm lãng phí và tăng hiệu quả.

<sup>10</sup> Các cuộc phỏng vấn sâu của IPS với đại diện của một số doanh nghiệp Việt Nam, 2023-2024.

<sup>11</sup> IPS, Tọa đàm về phát triển AI ở Việt Nam, Hà Nội, 11/2023.

<sup>12</sup> Doan Thu T N, Nguyen Q K, Taylor-Robinson A W (September 11, 2023) Healthcare in Vietnam: Harnessing Artificial Intelligence and Robotics to Improve Patient Care Outcomes. Cureus 15(9): e45006. doi:10.7759/cureus.45006

<sup>13</sup> Bộ NNPTNT, tọa đàm về ứng dụng AI trong ngành nông nghiệp, Hà Nội, 18/10/2024.

### 2.1.5. Thúc đẩy sự phát triển, hạnh phúc của con người, quyền con người

Nghị quyết của LHQ lưu ý rằng, việc sử dụng trí tuệ nhân tạo có thể góp phần vào việc bảo vệ quyền con người, thúc đẩy sự phát triển, hạnh phúc của cá nhân.

**Tiếp cận thông tin và dịch vụ:** AI có thể tạo điều kiện tiếp cận thông tin và dịch vụ tốt hơn, đặc biệt là cho các cộng đồng bị thiệt thòi. Ví dụ, các ứng dụng AI có thể dịch theo thời gian thực, giúp cải thiện giao tiếp và tiếp cận thông tin, vượt rào cản ngôn ngữ đối với người dân tộc thiểu số, người nhập cư. Nhất là ở Việt Nam có đến hơn 50 cộng đồng dân tộc thiểu số, các ứng dụng AI như vậy có thể giúp ích rất nhiều cho đồng bào thụ hưởng các quyền con người.

**Sức khỏe và phúc lợi:** Các ứng dụng AI trong chăm sóc sức khỏe có thể cải thiện đáng kể kết quả cho bệnh nhân. Các thuật toán AI có thể phân tích lượng lớn dữ liệu y tế để phát hiện bệnh sớm, cá nhân hóa kế hoạch điều trị và dự đoán xu hướng sức khỏe, góp phần vào quyền được chăm sóc sức khỏe.

**Giáo dục:** AI có thể phổ cập hóa giáo dục bằng cách cung cấp trải nghiệm học tập và tài nguyên cá nhân hóa cho học sinh bất kể vị trí địa lý. Điều này có thể giúp thực hiện quyền được học tập bằng cách làm cho việc học tập trở nên dễ tiếp cận và hiệu quả hơn.

**Hỗ trợ pháp lý và tiếp cận công lý:** AI có thể hỗ trợ trong các quy trình pháp lý bằng cách cung cấp thông tin và trợ giúp pháp lý cho những người không có khả năng chi trả. Các nền tảng dựa trên AI có thể cung cấp hướng dẫn về quyền và thủ tục pháp lý, do đó cải thiện việc tiếp cận công lý, quyền trong tố tụng hình sự, tố tụng dân sự. Ở Việt Nam, công cụ trợ lý ảo dựa trên AI của ngành tòa án dành cho người dân AI dự kiến được ứng dụng trong vài năm tới có thể hỗ trợ người dân thực hiện các quyền tố tụng của mình tốt hơn.

**Hỗ trợ người khuyết tật:** Khá nhiều công cụ AI có thể hỗ trợ người khuyết tật trong cuộc sống, công việc, học tập, đặc biệt là với người khiếm thị. Điều này thúc đẩy quyền của người khuyết tật được hoà nhập với cộng đồng được ghi nhận trong Công ước về quyền của người khuyết tật năm 2007. Ở Việt Nam, AI có tiềm năng cải thiện khả năng tiếp cận các dịch vụ thiết yếu, đặc biệt là ở các vùng nông thôn, miền núi. Các công cụ chăm sóc sức khỏe và giáo dục dựa trên AI có thể giúp thu hẹp khoảng cách giữa các cộng đồng đô thị và nông thôn, đồng bằng và miền núi.

## 2.2. RỦI RO, TÁC ĐỘNG TIÊU CỰC CỦA AI ĐỐI VỚI CON NGƯỜI, XÃ HỘI

### 2.2.1. Nhận diện rủi ro của AI

Để có chính sách, pháp luật điều chỉnh phù hợp, cần nhận diện rủi ro do AI gây ra gồm các loại rủi ro; các cấp độ rủi ro; cũng như rủi ro theo vòng đời của hệ thống AI.

#### 2.2.1.1. Các loại rủi ro của AI

AI đi kèm với những rủi ro như: Hội chứng ảo giác (Hallucination) - là một hiện tượng xảy ra khi một mô hình AI, đặc biệt là các mô hình ngôn ngữ lớn (LLM) như ChatGPT, tạo ra các thông tin, câu trả lời hoặc nội dung hoàn toàn không có thật hoặc không có cơ sở trong dữ liệu huấn luyện của nó; xói mòn quyền riêng tư; định kiến/thiên lệch; thông tin đầu ra không chính xác; thao túng hành vi của con người; vi phạm quyền sở hữu trí tuệ; thao tác bất lợi với dữ liệu hoặc đầu vào dùng để đào tạo AI nhằm đánh lừa hoặc làm sai lệch một mô hình AI; các mối đe dọa về an ninh mạng, an toàn số; thông tin giả mạo (Deepfakes); Lộ lọt thông tin bí mật thương mại của doanh nghiệp, dữ liệu cá nhân, thông tin độc quyền, bí quyết, tài liệu có bản quyền<sup>14</sup>.

Ở Việt Nam, các loại rủi ro của AI đã được nhận diện phần nào trong nhận thức của công chúng. Chẳng hạn, theo một khảo sát đối với thể hệ mới các doanh nghiệp gia đình ở Việt Nam (Next-Gen), hơn một nửa số doanh nghiệp được hỏi (58%) bày tỏ lo ngại về nguy cơ gia tăng rủi ro an

<sup>14</sup> Xem thêm: OECD, *Assessing potential future AI risks, benefits and policy imperatives*, OECD AI Paper No.27, 2024; xem danh mục các tác hại có thể xảy ra khi phát triển, sử dụng AI, các điểm khác biệt của rủi ro do AI gây ra so với các rủi ro của phần mềm truyền thống trong Khung quản lý rủi ro của AI do Viện Tiêu chuẩn và công nghệ quốc gia (NIST) của Hoa Kỳ đưa ra.

ninh mạng liên quan đến việc áp dụng GenAI; 52% đồng ý rằng GenAI có thể làm tăng những rủi ro truyền tải thông tin sai lệch; 52% lo ngại về rủi ro trách nhiệm pháp lý, bị ảnh hưởng đến danh tiếng; 39% cho rằng, GenAI có thể dẫn đến thiên kiến về nhóm khách hàng hay nhân viên cụ thể<sup>15</sup>. Điều này phản ánh phần nào nhận thức về các lỗ hổng có thể phát sinh từ việc sử dụng công nghệ mới này, chẳng hạn như vi phạm dữ liệu cá nhân, vi phạm quyền riêng tư, an ninh mạng, thông tin sai lệch.

### 2.2.1.2. Các cấp độ rủi ro của AI

Việc phân loại cấp độ rủi ro có tác dụng làm căn cứ để có mức độ kiểm soát, điều chỉnh phù hợp với từng cấp độ. Theo Đạo luật về AI của EU, có các mức độ rủi ro của AI như sau:

**Rủi ro không thể chấp nhận được:** các ứng dụng AI được coi là vi phạm các giá trị của Liên minh châu Âu, bao gồm:

- Thao túng/tác động vào tiềm thức con người: ví dụ như thao túng hành vi để ảnh hưởng đến việc bỏ phiếu hoặc đầu tư;
- Khai thác lỗ hổng dẫn đến hành vi có hại: ví dụ như một đồ chơi dành cho trẻ em khuyến khích cờ bạc;
- Phân loại sinh trắc học người dựa trên các đặc điểm nhạy cảm: ví dụ như một công cụ cho rằng có thể dự đoán niềm tin tôn giáo từ sinh trắc học;
- Mục đích chấm điểm công dân nói chung: ví dụ như cung cấp dịch vụ an sinh xã hội dựa trên đánh giá hành vi tuân thủ xã hội;
- Xác định sinh trắc học từ xa trong thời gian thực (trong không gian công cộng): ngoại trừ các mục đích được xác định trước như tìm kiếm mục tiêu nạn nhân, tội phạm, phòng chống khủng bố, tìm kiếm mục tiêu tội phạm nghiêm trọng hoặc nghi phạm (ví dụ buôn người, khai thác tình dục, cướp có vũ trang, tội phạm môi trường);
- Đánh giá trạng thái cảm xúc của một người tại nơi làm việc hoặc trong giáo dục, ngoại trừ việc đảm bảo an toàn;
- Kiểm tra dự đoán chỉ dựa trên việc lập hồ sơ hoặc đánh giá tính cách hoặc đặc điểm; thu thập hình ảnh khuôn mặt.

**Rủi ro cao:** Hệ thống AI rủi ro cao gồm: Hệ thống sinh trắc học và dựa trên sinh trắc học (ví dụ như nhận dạng sinh trắc học); quản lý và vận hành cơ sở hạ tầng quan trọng (như giao thông đường bộ, cung cấp năng lượng hoặc cơ sở hạ tầng kỹ thuật số); các công cụ dùng trong giáo dục và đào tạo nghề (ví dụ các công cụ dùng để đánh giá); quản lý việc làm và người lao động (ví dụ tuyển dụng, đánh giá hiệu suất hoặc phân bổ nhiệm vụ); truy cập vào các dịch vụ và phúc lợi tư nhân và công cộng thiết yếu (ví dụ bảo hiểm, phúc lợi xã hội); thực thi pháp luật (ví dụ đánh giá bằng chứng); quản lý di cư, tị nạn và kiểm soát biên giới; quản lý tư pháp, bầu cử.

**Rủi ro hạn chế:** Ba hệ thống AI gồm Chatbot; hệ thống nhận diện cảm xúc và phân loại sinh trắc học; hệ thống tạo ra nội dung 'deepfake' hoặc tổng hợp;

**Rủi ro tối thiểu:** Gồm các ứng dụng như bộ lọc thư rác hoặc trò chơi điện tử được hỗ trợ bởi AI.

### 2.2.1.3. Các rủi ro theo vòng đời của hệ thống AI

Hiện nay trên thế giới, khái niệm vòng đời của hệ thống AI (AI system lifecycle) do OECD đưa ra được thừa nhận, sử dụng trên thế giới trong xây dựng và hướng dẫn chính sách về AI, cả ở các tổ chức quốc tế như LHQ, EU, EC, cũng như các nước như Hoa Kỳ, Canada, Úc<sup>16</sup>. Theo OECD, vòng đời của hệ thống AI có các giai đoạn sau: i) “thiết kế, dữ liệu và các mô hình”; là một quá trình gồm các bước sau tùy thuộc vào ngữ cảnh: lập kế hoạch và thiết kế, thu thập và xử lý dữ liệu, cũng như xây dựng mô hình; ii) “xác minh và xác thực”, trong đó có thử nghiệm, đánh giá; iii) “triển khai”, bao gồm đưa vào sử dụng rộng rãi; và iv) “vận hành và giám sát”. Vòng đời này có thể được cải biến ở các nước hoặc các tổ chức, ví dụ chia theo 3 giai đoạn thiết kế mô hình, triển khai, sử dụng; nhưng bản chất vòng đời vẫn được duy trì.

<sup>10</sup> <https://www.pwc.com/vn/en/publications/vietnam-publications/nextgen-survey-2024.html>

<sup>11</sup> AI Principles Overview - OECD.AI

Các rủi ro của AI có thể xảy ra ở bất kỳ giai đoạn nào trong vòng đời này, và cần nhận diện chúng để có hành động phù hợp. Bảng dưới đây cho thấy ví dụ về các rủi ro theo 3 giai đoạn của vòng đời AI: xây dựng/thiết kế mô hình; triển khai/ứng dụng mô hình; sử dụng hệ thống AI.

**Bảng 1. Rủi ro, tác hại theo vòng đời AI<sup>17</sup>**

Nguồn gốc của rủi ro, tác hại	Loại rủi ro, tác hại	Ví dụ
<b>Hệ thống, mô hình AI bị lỗi</b>	Hệ thống vận hành kém	Đưa thông tin tư vấn đầu tư không chính xác
	Hệ thống bị thiên kiến	Chấm điểm tín dụng định kiến giới
	Hệ thống dễ tổn thương	Cơ sở hạ tầng thiết yếu bị lỗi
	Các lỗi về an toàn	Dữ liệu huấn luyện AI bị lộ lọt
<b>Triển khai sai lầm hoặc lợi dụng mô hình</b>	Các hệ thống gây sai lầm hoặc thiếu công bằng	Các xu hướng “hộp đen” của AI; đầu ra thiên kiến của AI
	Tấn công mạng dựa trên AI	Email lừa đảo nhắm vào cá nhân
	Thông tin sai lệch diện rộng	Thao túng xã hội bằng deepfakes
	Biến các hệ thống AI thành vũ khí	Vũ khí sinh học dựa trên AI
<b>Sử dụng không phù hợp, không cần trọng, quá mức</b>	Ảnh hưởng đến quyền con người ở phạm vi rộng	Xâm phạm quyền riêng tư qua việc sử dụng quá mức nhận diện khuôn mặt
	Tác động ngoại vi về mặt xã hội	Thuật toán lọc thông tin gợi ý cho người dùng dẫn đến phân cực
	Tác động ngoại vi về mặt kinh tế	Thất nghiệp do AI thay thế việc làm
	Tác động ngoại vi về mặt môi trường	Tăng tiêu thụ điện, thải cacbon do sử dụng AI quá mức

## 2.2.2. Tác động của AI đối với lao động, việc làm

### 2.2.2.1. Thay đổi quan hệ lao động

Trong quan hệ lao động truyền thống, người lao động phải thực hiện công việc mà người sử dụng lao động đưa ra, ngược lại họ có quyền được hưởng các chế độ về lương, thưởng, bảo hiểm theo thỏa thuận trong hợp đồng; hai bên phụ thuộc lẫn nhau. Tuy nhiên, với sự áp dụng rộng rãi AI trong các doanh nghiệp hiện nay, tính chất này có thể dần bị lu mờ. Cụ thể, các nhà sử dụng lao động hoàn toàn có thể sử dụng AI để thay mình đưa ra quyết định liên quan đến người lao động, dẫn tới trường hợp sai sót, gây ra hậu quả pháp lý cho doanh nghiệp. Nhiều công ty đã sử dụng AI để tự động đánh giá và xếp hạng nhân viên dựa trên nhiều tiêu chí như kinh nghiệm, kỹ năng, thành tích nổi bật. Điều này đặt ra những lo ngại đáng kể về giới hạn của công nghệ AI trong việc dự đoán và đánh giá cảm xúc của con người<sup>18</sup>. Mặc dù người lao động cũng có thể chịu sự đánh giá chủ quan của nhà quản lý, nhưng trong trường hợp này, các công cụ AI có thể được sử dụng làm cơ sở để chấm dứt hoặc cắt giảm việc làm; làm dấy lên lo ngại rằng nếu các nhà sử dụng lao động ngày càng áp dụng AI vào đánh giá nhân viên thì các quyền lợi của người lao động có thể sẽ bị ảnh hưởng rất nhiều.

<sup>17</sup> Human Technology Study, *The State of AI Governance in Australia*, 2023.

<sup>18</sup> Nazanin Andalibi, 'Emotion-tracking AI on the job: Workers fear being watched – and misunderstood', *The Conversation* (7 March 2024); Steven Greenhouse, "'Constantly monitored': the pushback against AI surveillance at work", *The Guardian* (7 January 2024)



### 2.2.2.2. Thay thế việc làm

Như đã đề cập, AI giúp tạo ra nhiều việc làm mới. Mặt khác, AI cũng làm cho nhiều người mất việc. Thông thường, các nhóm được xem là “lao động chất lượng thấp” sẽ phải đối mặt với nguy cơ cao bị thay thế bởi các ứng dụng AI, đặc biệt là trong khoảng 20 năm tiếp theo. Chẳng hạn, một nghiên cứu của OECD ở sáu quốc gia EU đã phát hiện ra rằng những người lao động trẻ, nam giới và những người có trình độ học vấn trung cấp, làm các công việc có tính chất lặp đi lặp lại có nhiều khả năng bị thay thế bởi tự động hóa. Một nghiên cứu của OECD nhận thấy, rủi ro mất việc làm nhiều hơn xảy ra đối với những người có trình độ học vấn thấp, lao động nữ và lớn tuổi do họ ít được tiếp cận hơn với các cơ hội việc làm liên quan đến AI và các công cụ AI nâng cao năng suất tại nơi làm việc<sup>19</sup>. Tuy nhiên, với những bước phát triển gần đây của AI, nhiều công việc đòi hỏi trình độ học vấn cao như tổng hợp, phân tích dữ liệu, marketing, chăm sóc khách hàng, truyền thông, thậm chí công việc chuyên sâu của ngành CNTT như viết mã (coding), lập trình cũng bị AI thay thế nhiều.

Vào năm 2021, một nhóm tác giả đã đánh giá tác động của AI đến việc làm ở Việt Nam và Lào dựa trên đánh giá sự phù hợp cho học máy của hơn 2.000 hoạt động công việc chi tiết cấu thành nên các nghề nghiệp<sup>20</sup>. Đối với Việt Nam, hầu hết người trả lời có các kỹ năng phù hợp ở mức độ vừa phải với máy học, cho thấy những người này có nguy cơ vừa phải bị thay thế bởi máy móc kỹ thuật số. Tuy nhiên, một số lượng đáng kể cũng thể hiện các kỹ năng rất phù hợp với máy học<sup>21</sup>. Kết quả phân tích theo giới tính cho thấy, tương tự như các nghiên cứu khác, lao động nữ bị ảnh hưởng nhiều hơn một chút bởi các công nghệ AI thay thế lao động so với lao động nam. Về độ tuổi, rủi ro bị thay thế bởi các công nghệ học máy nhiều nhất ở những người trong độ tuổi từ 25 đến 35; những người lao động trong nhóm tuổi lớn nhất (55-65 tuổi) chỉ bị ảnh hưởng ở mức độ vừa phải bởi các công nghệ học máy.

Theo đánh giá nói trên, nhiều nghề nghiệp ở Việt Nam có các hoạt động phù hợp với học máy, tức là chịu tác động nhiều từ các công cụ học máy; đặc biệt là “công nhân may mặc và các ngành nghề liên quan”, “nhân viên bán hàng tại cửa hàng” và “người làm nông nghiệp, lâm nghiệp và thủy sản”. Trong số ba nghề nghiệp này, các hoạt động của nhân viên bán hàng tại cửa hàng phù hợp nhất với máy học nói chung; mặc dù vẫn cần có người lao động nghề này để thực hiện một số hoạt động mà AI khó thay thế. Ngược lại, một số ít nghề nghiệp, chủ yếu là trong lĩnh vực dịch vụ có các hoạt động công việc không phù hợp lắm với công nghệ học máy. Mặc dù những nghề nghiệp này có thể được coi là an toàn xét về tác động thay thế lao động của AI, nhưng không có nhiều cơ hội cho những người lao động làm trong những nghề này cải thiện năng suất của họ thông qua AI. Đó là những nghề như “dọn dẹp và giúp việc gia đình, khách sạn và văn phòng”, “phục vụ và pha chế”.

Theo ý kiến của những người được IPS khảo sát, AI không phải lúc nào cũng có thể thay thế được con người, đồng thời luôn phải đặt lên bàn cân giữa đầu tư công nghệ và giá nhân công. Theo một chuyên gia AI, một số công việc như nhân công giá rẻ và khả năng của con người không thay thế được, ví dụ nhân công kiểm tra mức độ hoàn thiện của bề mặt sản phẩm trong nhà máy. Còn đại diện của một doanh nghiệp cung cấp giải pháp phần mềm đánh giá, hiện tại AI chưa thể thay thế lập trình viên, vì ở Việt Nam còn quá nhiều thứ cần phải lập trình. Doanh nghiệp này có đội ngũ 50 người trong vòng 5 năm chỉ lập trình phần mềm cho trang web dauthau.info nhưng vẫn còn 200 tính năng mới chưa được lập trình.

Trong hai cuốn sách về AI của tác giả Kai - Fu Lee<sup>22</sup>, các yếu tố sau là căn cứ để đánh giá mức độ ảnh hưởng của AI đối với việc làm: Mức độ tương tác xã hội ít hay nhiều; ít đòi hỏi sự khéo léo và môi trường được sắp xếp, hay cần độ khéo léo cao, môi trường phi tổ chức (đối với công việc chân tay); tính chất công việc dựa trên sự tối ưu hóa hay là dựa trên sự sáng tạo và chiến lược (đối với công việc trí óc). Tác động của AI đến việc làm được chia thành 4 vùng: Vùng nguy hiểm, vùng an toàn, vùng lớp phủ con người, vùng thay đổi chậm. Những công việc rơi vào “Vùng nguy hiểm” có nguy cơ bị thay thế cao trong những năm tới; là những công việc ít tương tác xã hội, tối đa hóa các biến số định lượng có thể ghi lại trong dữ liệu, không đòi hỏi nhiều về kỹ năng, môi trường làm

<sup>19</sup> OECD, *Who will be the workers most affected by AI?: A closer look at the impact of AI on women, low-skilled workers and other groups*, 10/2024.

<sup>20</sup> Francesco Carbonero et al., *The Impact of Artificial Intelligence on Labor Markets in Developing Countries: A New Method with an Illustration for Lao PDR and Viet Nam*, IZA – Institute of Labor Economics Discussion Paper Series, 12/2021. <https://docs.iza.org/dp14944.pdf>

<sup>21</sup> Francesco Carbonero et al., *ltdđ*.

<sup>22</sup> Kai-Fu Lee, *Các siêu cường AI: Trung Quốc, thung lũng Silicon và trật tự thế giới mới* (sách dịch), NXB Trẻ, 2020; Chen Qiufan và Kai-Fu Lee, *AI 2041 - 10 viễn cảnh cho tương lai* (sách dịch), NXB Thế Giới, 2024.

việc được sắp xếp sẵn. Trái ngược với vùng nguy hiểm, ở “Vùng an toàn” có các ngành nghề chưa thể bị thay thế bởi tự động hóa; là những công việc mang tính xã hội cao, yêu cầu nhiều về mặt sáng tạo, môi trường làm việc phi tổ chức. Tác động của AI ở các vùng “Lớp phủ con người” và “Thay đổi chậm” ít rõ ràng hơn: mặc dù hiện tại không thể thay thế hoàn toàn, nhưng việc tổ chức lại các nhiệm vụ hoặc những tiến bộ của công nghệ có thể dẫn đến việc cắt giảm việc làm trên diện rộng thuộc các vùng này. Hai sơ đồ dưới đây thể hiện rủi ro bị AI thay thế đối với lao động chân tay và lao động trí óc ở bốn vùng nói trên.



Hình 2. Rủi ro AI thay thế việc làm chân tay



Hình 3. Rủi ro AI thay thế việc làm trí óc

### 2.2.3. Tác động của AI đối với các giá trị cốt lõi của con người, xã hội

AI có thể dẫn đến những tác hại về thể chất, tâm lý, thiệt hại về tài sản hoặc tổn thất kinh tế đối với một cá nhân; cũng như các tác động tiêu cực có thể xảy ra trên khắp các lĩnh vực của nền kinh tế; tác hại đối với các nhóm dân cư, làm tăng mức độ nghiêm trọng của tác động. Ví dụ, các nhóm dễ bị tổn thương như trẻ em, có thể phải đối mặt với nguy cơ chịu tác hại lớn hơn từ hệ thống AI. Việc tích hợp AI vào các lĩnh vực khác nhau có thể thách thức việc bảo vệ và thúc đẩy các giá trị cốt lõi của con người, xã hội: phẩm giá, tự do, bình đẳng và đoàn kết/gắn kết xã hội.

#### (a) Tác động của AI lên phẩm giá

**Mất danh tính:** Hệ thống AI như chấm điểm công dân có thể khiến con người trở nên vô danh khi chỉ được coi như các điểm dữ liệu, biến số thống kê, hoặc như máy móc. Các quy trình ra quyết định tự động thiếu minh bạch có thể khiến con người bị đối xử một cách phi cá nhân, bỏ qua sự phức tạp trong danh tính và trải nghiệm của họ. Quan điểm giản đơn này có thể làm mất đi giá trị của con người, xói mòn sự tôn trọng đối với giá trị nội tại của con người.

**Mất sự tương tác cá nhân:** Trong các lĩnh vực như y tế hoặc dịch vụ khách hàng, AI có thể thay thế sự tương tác giữa con người, dẫn đến cảm giác không được tôn trọng và chăm sóc đúng mức. Sự thay đổi này có thể khiến mọi người cảm thấy như hoàn cảnh riêng của họ bị bỏ qua.

#### (b) Tác động của AI lên tự do

**Thao túng và kiểm soát hành vi:** AI có thể được sử dụng để ảnh hưởng hoặc thao túng hành vi, đặc biệt thông qua quảng cáo có mục tiêu hoặc nội dung cá nhân hóa có thể hình thành ý kiến và quyết định của con người, xâm phạm quyền tự chủ của cá nhân.

**Giám sát, theo dõi:** Các hệ thống giám sát dựa trên AI có thể xâm phạm quyền tự do cá nhân bằng cách theo dõi và phân tích hành động của các cá nhân mà không có sự đồng ý của họ. Điều này có thể tạo ra hiệu ứng sợ hãi, khiến mọi người thay đổi hành vi vì sợ bị theo dõi.

#### (c) Tác động của AI lên bình đẳng

**Thiên kiến và phân biệt đối xử:** Các hệ thống AI được đào tạo trên dữ liệu thiên lệch có thể duy trì hoặc thậm chí làm trầm trọng thêm các bất bình đẳng hiện có. Ví dụ, các công nghệ nhận diện khuôn mặt đã được chứng minh là có tỷ lệ lỗi cao hơn đối với một số nhóm nhân khẩu học nhất định. Các thuật toán AI được sử dụng trong quy trình tuyển dụng hoặc tư pháp hình sự có thể củng cố những thiên kiến về chủng tộc, giới tính, kinh tế, xã hội có trong dữ liệu huấn luyện. Hoặc là thuật toán trong các nền tảng điều hành lái xe công nghệ có thể làm gia tăng sự bất bình đẳng giữa chủ của các nền tảng đó với lái xe công nghệ<sup>23</sup>.

**Hệ thống quyết định tự động:** AI ngày càng được sử dụng trong các quy trình ra quyết định trong các lĩnh vực như chấm điểm tín dụng, thực thi pháp luật, dịch vụ xã hội. Khi các hệ thống này đưa ra các quyết định ảnh hưởng đến cuộc sống của cá nhân, chẳng hạn như từ chối cho vay hoặc xác định nghi phạm, chúng có thể làm như vậy theo cách bất công hoặc phân biệt đối xử, đặc biệt nếu không được điều chỉnh hoặc kiểm soát đúng cách.

**Hạn chế tiếp cận lợi ích của AI:** AI có thể làm sâu sắc thêm khoảng cách số, với các cộng đồng bị thiệt thòi có ít cơ hội tiếp cận lợi ích của các công nghệ AI. AI có thể làm trầm trọng thêm các bất bình đẳng hiện có bằng cách mang lại lợi ích cho những người có thể tiếp cận công nghệ trong khi gạt ra ngoài lề những người yếu thế, không chỉ ở phạm vi quốc gia, mà cả toàn cầu<sup>24</sup>. Đồng thời, hiện đang tồn tại khoảng cách giữa các quốc gia với nhau trong phát triển AI và hưởng lợi từ AI. Như trong tháng 6/2024, dự thảo Nghị quyết của LHQ đã bày tỏ quan ngại rằng khoảng cách về AI và các khoảng cách số khác giữa và trong các quốc gia tiếp tục gia tăng; các nước đang phát triển phải đối mặt với những thách thức riêng trong việc theo kịp tốc độ phát triển nhanh chóng

<sup>23</sup> The Plight of Platform Workers Under Algorithmic Management in Southeast Asia - Carnegie Endowment for International Peace | Carnegie Endowment for International Peace

<sup>24</sup> Kai Fu Lee, AI Powers, 2018; Kai Fu Lee, Ten Perspectives for AI 2041, 2024; Advancing a More Global Agenda for Trustworthy Artificial Intelligence - Carnegie Endowment for International Peace | Carnegie Endowment for International Peace

của AI, cụ thể là trong thiết kế, phát triển, triển khai, sử dụng trí tuệ nhân tạo, nguy cơ bị tụt hậu thêm càng gia tăng<sup>25</sup>.

Chẳng hạn, sự phát triển của AI (và các công nghệ số khác) dẫn đến phân hóa trong thị trường lao động toàn cầu. Các quốc gia đang phát triển sẽ đối mặt với tình trạng dư thừa lao động và gia tăng thất nghiệp. Lao động giá rẻ không còn là lợi thế cạnh tranh của các thị trường mới nổi ở khu vực Mỹ La-tinh và châu Á. Bên cạnh đó, cạnh tranh về nhân lực chất lượng cao sẽ diễn ra mạnh mẽ trên bình diện thế giới, khu vực và quốc gia<sup>26</sup>. Có sự chênh lệch đáng kể giữa các quốc gia về cơ sở hạ tầng số và mức độ tham gia của lực lượng lao động. Theo Kai-Fu Lee, các quốc gia tiên tiến trong phát triển AI, như Mỹ và Trung Quốc, sẽ có lợi thế lớn trong việc tạo ra các công việc mới và nâng cao năng suất, trong khi những quốc gia không đầu tư đủ vào AI có nguy cơ bị bỏ lại phía sau trong cuộc đua này<sup>27</sup>. Có sự phân bố lao động bất đối xứng, nơi các quốc gia như Ấn Độ, Bangladesh, Philippines chiếm phần lớn lực lượng lao động số, trong khi các công ty có trụ sở tại Mỹ, Canada và Tây Âu là những nhà tuyển dụng chính; châu Phi tụt hậu trong việc tham gia việc làm số do hạ tầng hạn chế<sup>28</sup>.

#### **(d) Tác động của AI lên gắn kết xã hội**

**Phân mảnh xã hội:** AI, đặc biệt là trong các nền tảng truyền thông xã hội và thông tin, có thể dẫn đến phân mảnh xã hội bằng cách thúc đẩy nội dung củng cố các niềm tin hiện có và chia rẽ các cộng đồng.

**Bỏ qua các nhóm dễ bị tổn thương:** Việc thiết kế, triển khai, sử dụng các hệ thống AI có thể ưu tiên hiệu quả và lợi nhuận, bỏ qua nhu cầu của các nhóm dễ bị tổn thương, cả ở địa bàn nhỏ, quốc gia, thậm chí toàn cầu<sup>29</sup>. Ví dụ, các hệ thống tự động trong các dịch vụ xã hội có thể không nhạy cảm với sự phức tạp của từng trường hợp cụ thể, dẫn đến hỗ trợ không đầy đủ.

#### **2.2.4. Tác động của AI đối với các quyền con người**

Dự thảo Nghị quyết của LHQ vào tháng 6/2024 bày tỏ quan ngại rằng, trong quá trình hình thành, thiết kế, phát triển, triển khai, sử dụng, đánh giá và điều chỉnh công nghệ AI, có thể tiềm ẩn rủi ro củng cố sự phân biệt đối xử có hệ thống vì lý do chủng tộc, giới tính, ảnh hưởng đến việc hưởng thụ các quyền con người, bao gồm các quyền kinh tế, xã hội và văn hóa<sup>30</sup>. Các quyền con người cụ thể chịu tác động tiêu cực từ AI như quyền về đời tư, quyền biểu đạt, các quyền về giáo dục, y tế, quyền bình đẳng và không bị phân biệt đối xử, v.v. Nổi bật nhất và nghiêm trọng nhất là việc sử dụng AI một cách sai lệch có thể làm gia tăng sự phân biệt đối xử, vi phạm các quyền riêng tư, dữ liệu cá nhân, an toàn cá nhân.

**Vi phạm quyền riêng tư:** Các điều khoản về quyền riêng tư trong các văn kiện quốc tế về quyền con người mà công nghệ AI dễ vi phạm<sup>31</sup>: **Điều 12 UDHR:** không ai phải chịu sự can thiệp tùy ý vào cuộc sống riêng tư, gia đình, nơi ở hoặc thư tín hay bị xúc phạm danh dự hoặc uy tín cá nhân; **Điều 17 của ICCPR:** Bảo vệ chống lại sự can thiệp tùy tiện vào quyền riêng tư, gia đình, nhà ở hoặc thư tín; **Điều 8 của Hiến chương EU về các quyền cơ bản:** Bảo đảm quyền bảo vệ dữ liệu, yêu cầu xử lý dữ liệu cá nhân một cách công bằng dựa trên sự đồng ý hoặc cơ sở hợp pháp. Đáng chú ý là như Nghị quyết của Đại hội đồng LHQ lưu ý, việc vi phạm quyền riêng tư có thể dẫn đến vi phạm các quyền con người khác như quyền tự do biểu đạt, tự do hội họp một cách hòa bình<sup>32</sup>.

Các hành động vi phạm quyền riêng tư như: Việc triển khai AI trong các hệ thống giám sát, chẳng hạn như công nghệ nhận diện khuôn mặt, có thể theo dõi cá nhân mà không có sự đồng ý của họ;

<sup>25</sup> UN General Assembly, seventy - eight session (6/2024), Draft Resolution: Enhancing international cooperation on capacity-building of artificial intelligence.

<sup>26</sup> Lorraine Charles, Shuting Xia and Adam P. Coutts, Digitalization and Employment: A Review, International Labour Organization, 2022.

<sup>27</sup> Kai-Fu Lee, AI Powers: China, Silicon Valley, and New World Order, 2018.

<sup>28</sup> Lorraine Charles, Shuting Xia and Adam P. Coutts, Digitalization and Employment: A Review, International Labour Organization, 2022.

<sup>29</sup> Advancing a More Global Agenda for Trustworthy Artificial Intelligence - Carnegie Endowment for International Peace | Carnegie Endowment for International Peace

<sup>30</sup> UN General Assembly, seventy - eight session (6/2024), Draft Resolution: Enhancing international cooperation on capacity-building of artificial intelligence.

<sup>31</sup> D. Majumdar & H.K. Chattopadhyay, AI and Human Rights: From Business and Policy Perspectives.

<sup>32</sup> UN General Assembly, Seventy-eighth session, Resolution No. 78/213 Promotion and protection of human rights in the context of digital technologies.

các hệ thống AI thu thập dữ liệu của cá nhân mà không có sự đồng ý rõ ràng hoặc hiểu biết về cách dữ liệu của họ sẽ được sử dụng. Ở Việt Nam, các chủ thể cả Nhà nước và tư nhân đang tăng cường giám sát bằng các công nghệ AI. Điều này có thể cải thiện trật tự, an toàn xã hội, nhưng cũng đặt ra lo ngại về vi phạm quyền riêng tư và rủi ro lạm dụng quyền lực của các cơ quan, tổ chức.

**Tác động đối với quyền có việc làm (Điều 23 UDHR):** Như đã đề cập, mất việc làm, thay thế lao động do việc triển khai, ứng dụng AI ngày càng tăng được coi là mối quan ngại về rủi ro của AI đối với quyền có việc làm. Ở các nước đang phát triển như Việt Nam, với cơ cấu lao động hiện hành, tác động của AI đối với người lao động sẽ là thách thức lớn, rủi ro thất nghiệp và khoảng cách bất bình đẳng ngày càng gia tăng.

**Ảnh hưởng đến tự do biểu đạt (Điều 19 UDHR):** Các thuật toán AI được các nền tảng truyền thông xã hội hoặc cơ quan công quyền sử dụng có thể ảnh hưởng và kiểm soát thông tin, dẫn đến việc kiểm duyệt và thao túng ý kiến công chúng; có thể ảnh hưởng quyền tự do biểu đạt và tiếp cận thông tin.

**Hạn chế quyền tự do cá nhân:** AI được sử dụng để theo dõi con người trên diện rộng, bằng các phần mềm nhận diện ngày càng tinh vi như phần mềm nhận diện khuôn mặt, nhận diện dáng đi... nhằm phát hiện, kiểm soát các nhóm và cá nhân nhất định, hoặc chấm điểm công dân<sup>33</sup>. Thực tế này đặt công dân vào tình trạng luôn cảm thấy bị theo dõi, không còn không gian riêng tư, tự do; cản trở họ thực hiện các quyền hợp pháp như tự do hiệp hội, hội họp, nhất là các nhóm dễ tổn thương. Hoặc là việc sử dụng AI ngày càng phổ biến trong hệ thống tư pháp hình sự ở các nước làm tăng nguy cơ xâm phạm các quyền tự do cá nhân. Ví dụ việc sử dụng phần mềm đánh giá rủi ro về việc tái phạm được sử dụng ở Mỹ để đưa ra quyết định về việc tạm giam hay cho tại ngoại, thậm chí xác định hình phạt với một người.

Nguy cơ, rủi ro đối với quyền con người có thể xảy ra trong tất cả các khâu thuộc vòng đời của hệ thống AI, từ việc thiết kế, sử dụng dữ liệu huấn luyện AI, triển khai, ứng dụng, vận hành, giám sát, cho đến người dùng cuối (xem chi tiết hơn trong Phụ lục ở cuối báo cáo này).

### 2.2.5. Thách thức trong kiểm soát rủi ro của AI

**Khó xác định trách nhiệm:** Xác định ai chịu trách nhiệm đối với thông tin, quyết định do các hệ thống AI đưa ra là một thách thức, đặc biệt là những hệ thống liên quan đến quyết định tự động; không rõ trách nhiệm thuộc về các nhà phát triển, nhà cung cấp dữ liệu hay cơ quan, tổ chức ứng dụng AI. Các hệ thống AI, đặc biệt là những hệ thống sử dụng các mô hình phức tạp như học sâu (deep learning) có quá trình ra quyết định thiếu rõ ràng. Điều này có thể dẫn đến khó thực thi các biện pháp bảo vệ pháp lý, khắc phục đối với tác hại, vi phạm phát sinh từ AI.

**Sai lệch ngữ cảnh:** Có xu hướng áp dụng các hệ thống AI phổ quát, không điều chỉnh chúng cho các ngữ cảnh pháp lý, văn hóa hoặc xã hội cụ thể mà chúng được sử dụng; không hiểu sâu về bối cảnh cụ thể hoặc nhu cầu và giá trị cụ thể của các cộng đồng<sup>34</sup>. Một hệ thống AI, hay thuật toán AI được thiết kế ở các nước phát triển được đưa vào áp dụng ở một nước có các điều kiện khác về kinh tế - xã hội, địa lý, tự nhiên, thì có thể sẽ gặp trở ngại, thiếu hiệu quả. Chẳng hạn, ở Tanzania, hơn 95% đường bộ là đường đất được trải nhựa một phần. Chính phủ sử dụng các công cụ được tích hợp AI để giúp ưu tiên phân bổ ngân sách cho việc bảo trì đường bộ. Tuy nhiên, các thuật toán lại không đáp ứng được yêu cầu vì chúng được đào tạo từ dữ liệu về những con đường bằng phẳng ở Hoa Kỳ và Châu Âu<sup>35</sup>. Hoặc một hệ thống AI trong một bối cảnh văn hóa nhất định có thể không tôn trọng hoặc hỗ trợ các quyền được coi trọng trong một bối cảnh khác; có thể làm suy yếu quyền con người do không đáp ứng nhu cầu đa dạng của các nhóm dân cư khác nhau.

**Hạn chế trong nhận thức về rủi ro của AI:** Ở Việt Nam, cả các cơ quan hoạch định, xây dựng chính sách, pháp luật, cũng như công chúng dường như chưa nhận thức hết những rủi ro tiềm ẩn trong nhiều ứng dụng AI liên quan đến quyền riêng tư, sự thiên kiến, an ninh mạng. Cụ thể như, không

<sup>33</sup> Nguyễn Văn Quân, Bùi Phú Châu, Ứng dụng trí tuệ nhân tạo và tác động tới quyền con người: Góc nhìn từ mô hình đánh giá tín nhiệm xã hội ở Trung Quốc, trong sách: NXB, 2020.

<sup>34</sup> Advancing a More Global Agenda for Trustworthy Artificial Intelligence - Carnegie Endowment for International Peace | Carnegie Endowment for International Peace

<sup>35</sup> Advancing a More Global Agenda for Trustworthy Artificial Intelligence - Carnegie Endowment for International Peace | Carnegie Endowment for International Peace

những người để ý, các tuyến đường mà tài xế Grab sử dụng để di chuyển trên mạng lưới đường phố phức tạp ở Hà Nội hay TP Hồ Chí Minh, các sản phẩm giảm giá trên Shopee hoặc Tiki có thể do thuật toán gợi ý. Trao đổi với IPS, một số chuyên gia AI, cũng như chuyên gia pháp lý nhận định, việc lộ lọt thông tin cá nhân hiện nay khiến người dân chỉ thấy phiền hà do bị gọi điện quảng cáo, chứ chưa nhìn thấy thiệt hại lớn do lộ lọt thông tin cá nhân. Người dân sẵn sàng cài ứng dụng VNeID do thấy sự thuận tiện khi được tích hợp các thông tin cá nhân khác nhau vào cùng một ứng dụng, nhờ được hướng dẫn cài đặt tận nơi, trong khi chưa nhìn thấy hết nguy cơ về việc bị giám sát, xâm phạm đời tư.

Một số doanh nghiệp đã triển khai AI với tôn chỉ “con người là trên hết” (human in the loop), coi AI chỉ là một công cụ cho con người; không phát triển phần mềm AI cho phép AI ra quyết định thay cho con người. Tuy nhiên, nhiều doanh nghiệp ở Việt Nam và khách hàng chưa quan tâm đến quy tắc ứng xử trong ứng dụng, sử dụng AI một cách an toàn, có trách nhiệm. Cụ thể như trong khi doanh nghiệp nước ngoài thường yêu cầu rất khắt khe về dữ liệu cá nhân, doanh nghiệp Việt Nam thường không coi trọng vấn đề này.

Hoặc là, bất chấp những rủi ro về quyền riêng tư, tính bảo mật và việc sử dụng dữ liệu cá nhân khác, công nghệ nhận diện khuôn mặt hoặc công nghệ sinh trắc học khác không gây ra nhiều lo ngại cho công chúng hoặc chính phủ; ít ai đặt câu hỏi về tính minh bạch, khả năng giải thích, trách nhiệm giải trình về các công nghệ này. Trong các cuộc phỏng vấn sâu thuộc khuôn khổ các nghiên cứu về AI của IPS, không nhiều người cả trong các cơ quan Nhà nước và doanh nghiệp nhận thức được các rủi ro từ việc phát triển, sử dụng AI trong công việc của cơ quan, tổ chức, doanh nghiệp hoặc trong cuộc sống mỗi người.

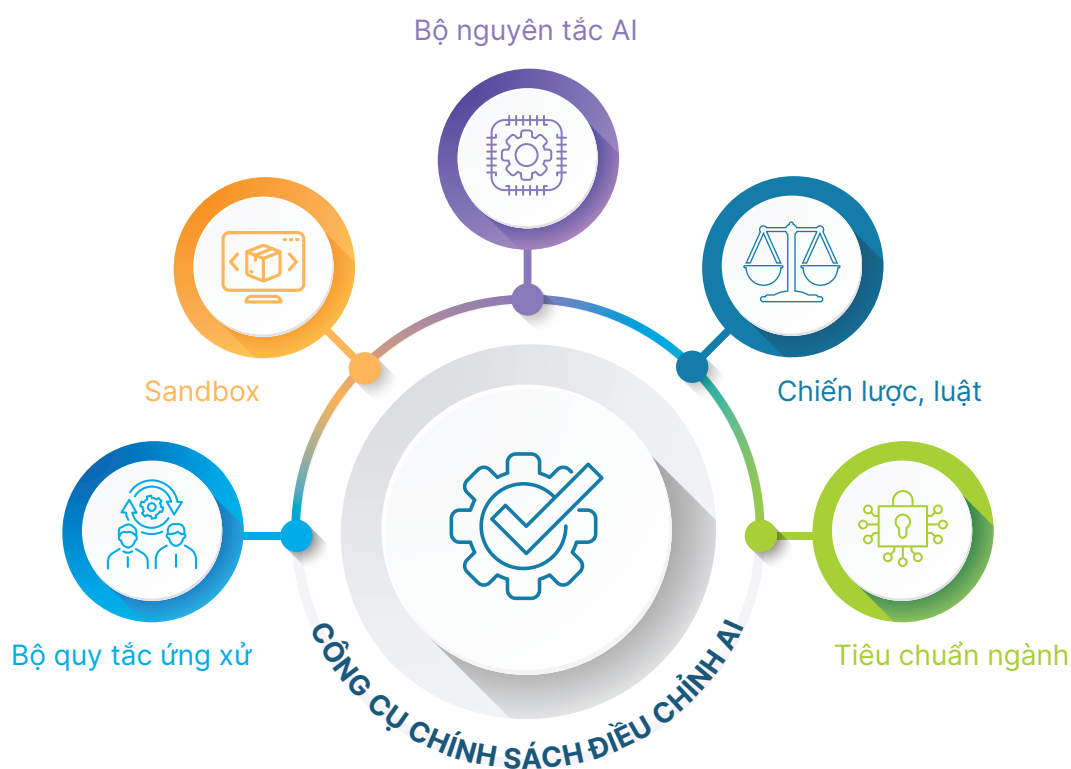
Đạo đức AI không chỉ vắng bóng trong các cuộc thảo luận về chính sách công, truyền thông, mà còn vắng bóng trong giáo dục đại học. Một số giảng viên khi được hỏi đã cho biết, đạo đức AI vẫn chưa được chú ý giảng dạy trong các trường đại học ở Việt Nam, cả đối với sinh viên học các ngành công nghệ và các ngành khác. Việc thiếu nhận thức sâu sắc là trở ngại đối với việc phát triển AI có trách nhiệm, đáng tin cậy, cũng như đối với việc xây dựng, ban hành chính sách về AI.

### III.

## CHÍNH SÁCH VỀ TRÍ TUỆ NHÂN TẠO TRÊN THẾ GIỚI

### 3.1. CÁC CÔNG CỤ CHÍNH SÁCH ĐIỀU CHỈNH AI TRÊN THẾ GIỚI

Trên thế giới có những công cụ chính sách khác nhau để điều chỉnh các vấn đề về AI: Các nguyên tắc đạo đức về AI có trách nhiệm, đáng tin cậy; khuôn khổ chính sách, pháp luật (gồm cả các luật hiện hành và quy định mới về AI); cơ chế thử nghiệm có kiểm soát (sandbox); các tiêu chuẩn trong từng ngành; hướng dẫn, bộ quy tắc ứng xử<sup>36</sup>.



Hình 4. Công cụ chính sách điều chỉnh AI

#### 3.1.1. Các bộ nguyên tắc về AI

Tại thời điểm cuối năm 2024, nhiều tổ chức quốc tế, cũng như nhiều nước đã ban hành các bộ nguyên tắc về AI có trách nhiệm, đáng tin cậy như UNESCO, OECD, ASEAN, Hoa Kỳ, Anh, Canada, Singapore, Úc.

<sup>36</sup> Có thể xem tổng hợp về các công cụ chính sách trên thế giới tại đây: [OECD's live repository of AI strategies & policies - OECD.AI](#), accessed on 27/10/2024; các sáng kiến chính sách cấp quốc gia tại đây: [Policies - OECD.AI](#)

**Bảng 2. Các nguyên tắc phổ biến trên thế giới về AI có trách nhiệm<sup>37</sup>**

Nguyên tắc	OECD	Hoa Kỳ	Canada	Singapore	Úc
An toàn, bảo mật	Hệ thống mạnh mẽ, an toàn, an ninh	Hệ thống an toàn, hiệu quả	Sự tích hợp của hệ thống; an ninh quốc gia		Đáng tin cậy và an toàn
Minh bạch	Minh bạch và giải thích được	Thông báo và giải thích	Minh bạch trong sử dụng và lợi ích	Ra quyết định minh bạch, giải thích được	Minh bạch và giải thích được
Giải thích được	Minh bạch và giải thích được	Thông báo và giải thích	Giải thích một cách tường minh	Quá trình ra quyết định giải thích được	Minh bạch và giải thích được
Công bằng và bình đẳng	Các giá trị vì con người và tính công bằng	Phòng chống phân biệt đối xử do thuật toán		Quá trình ra quyết định công bằng	Tính công bằng
Giải trình và giám sát	Trách nhiệm giải trình			Trách nhiệm giải trình	Trách nhiệm giải trình
Quyền từ chối		Các giải pháp thay thế và các cân nhắc của con người			Quyền thách thức quyết định của AI
Riêng tư		Tính riêng tư của dữ liệu	Bảo vệ thông tin cá nhân		Bảo vệ và bảo mật sự riêng tư

Trong số đó, nổi bật là Các nguyên tắc của OECD ra đời năm 2019, cập nhật năm 2024, nhằm mục đích đảm bảo sự hài hòa giữa các khung quản trị rủi ro AI của các quốc gia khác nhau; được nhiều tổ chức và các nước học tập, áp dụng. Các nguyên tắc về AI của OECD gồm có:

- **Tăng trưởng bao trùm, phát triển bền vững và hạnh phúc:** AI đáng tin cậy có khả năng đóng góp vào tăng trưởng và thịnh vượng chung cho tất cả - cá nhân, xã hội, hành tinh – và thúc đẩy các mục tiêu phát triển toàn cầu.
- **Quyền con người và các giá trị dân chủ, bao gồm công bằng và quyền riêng tư:** Các hệ thống AI nên được thiết kế theo cách tôn trọng pháp luật, quyền con người, các giá trị dân chủ và sự đa dạng, và nên bao gồm các biện pháp bảo đảm an toàn để có một xã hội công bằng.
- **Minh bạch và giải thích được:** Các hệ thống AI nên đảm bảo rằng mọi người biết là đang tương tác với chúng và có thể phản đối, khiếu nại kết quả do AI tạo ra.
- **Độ bền, bảo mật và an toàn:** Các hệ thống AI phải hoạt động một cách mạnh mẽ, bảo mật và an toàn trong suốt vòng đời của chúng; các rủi ro tiềm năng nên được đánh giá, quản lý liên tục.
- **Trách nhiệm giải trình:** Các tổ chức và cá nhân phát triển, triển khai hoặc vận hành các hệ thống AI cần chịu trách nhiệm về hoạt động đúng đắn của mô hình AI phù hợp với các nguyên tắc dựa trên giá trị (values-based principles) của OECD về AI.

Chính phủ Úc đã thông qua một bộ nguyên tắc có tính chất tự nguyện về đạo đức AI với các nội dung như sau<sup>38</sup>:

- **Lấy con người làm trung tâm:** Các hệ thống AI nên tôn trọng quyền con người, sự đa dạng và quyền tự chủ của cá nhân;

<sup>37</sup> Xem giới thiệu và tải Các nguyên tắc về AI của OECD tại đây: *OECD Legal Instruments, AI Principles Overview - OECD.AI*; báo cáo về thực tiễn các nước áp dụng Các nguyên tắc ở đây: *How countries are implementing the OECD Principles for Trustworthy AI - OECD.AI*

<sup>38</sup> Department of Industry, Science and Resources, Australia, *Supporting responsible AI: discussion paper*, 1 June 2023.



- **Công bằng:** Các hệ thống AI nên có tính bao trùm và dễ tiếp cận, và không nên liên quan đến hoặc dẫn đến sự phân biệt đối xử đối với cá nhân, cộng đồng hoặc nhóm;
- **Bảo vệ quyền riêng tư và bảo mật:** Các hệ thống AI nên tôn trọng và duy trì quyền riêng tư và bảo vệ dữ liệu, đồng thời đảm bảo bảo mật dữ liệu;
- **Độ tin cậy và an toàn:** Các hệ thống AI nên hoạt động đáng tin cậy theo mục đích dự định của chúng;
- **Minh bạch và giải thích được:** Phải có sự minh bạch và công khai có trách nhiệm để mọi người có thể hiểu khi họ đang bị ảnh hưởng đáng kể bởi AI và có thể tìm hiểu khi một hệ thống AI đang tương tác với họ;
- **Khả năng phản đối, khiếu nại:** Khi một hệ thống AI ảnh hưởng đáng kể đến một cá nhân, cộng đồng, nhóm hoặc môi trường, cần có một quy trình kịp thời để cho phép mọi người phản đối, khiếu nại việc sử dụng hệ thống AI hoặc kết quả của hệ thống AI;
- **Trách nhiệm giải trình:** Những người chịu trách nhiệm cho các giai đoạn khác nhau trong vòng đời hệ thống AI nên được xác định và chịu trách nhiệm về kết quả của các hệ thống AI, và con người nên giám sát đối với các hệ thống AI.

Trung Quốc cũng xây dựng những nguyên tắc đối với AI của nước này; cụ thể như: Các quy tắc đạo đức cho AI thế hệ mới; Các nguyên tắc quản trị đối với AI thế hệ mới – phát triển AI có trách nhiệm; Hướng dẫn về nguyên tắc đạo đức trong khoa học, công nghệ<sup>39</sup>.

### 3.1.2. Khuôn khổ chính sách về AI

Tại thời điểm cuối năm 2024, theo tổng hợp chưa đầy đủ của OECD, có hàng chục văn bản chính sách cấp độ quốc gia trên thế giới điều chỉnh AI<sup>40</sup>. Còn Global AI Law and Policy Tracker, một sáng kiến tư nhân, đã ra báo cáo cập nhật tình hình xây dựng chính sách, pháp luật về AI ở 24 nước đến đầu năm 2024<sup>41</sup>. Về chính sách, nhiều nước đã ban hành Chiến lược quốc gia về AI hoặc văn bản tương tự, đưa ra các định hướng lớn phát triển AI trong một quãng thời gian nhất định. Canada ban hành chiến lược AI quốc gia đầu tiên vào tháng 3 năm 2017. Cho đến năm 2024, 75 chiến lược AI quốc gia đã được công bố; trong đó đỉnh cao là năm 2019 với 24 chiến lược; còn năm 2023, 8 chiến lược mới đã được bổ sung, từ các quốc gia ở Trung Đông, Châu Phi và Caribe, thể hiện sự mở rộng trên toàn thế giới chính sách quốc gia về AI<sup>42</sup>.

Trong số các nước, đáng chú ý là Singapore, quốc gia vẫn chưa ban hành quy định pháp luật về AI, đang tìm kiếm một mô hình quản trị AI kết hợp cách tiếp cận của Mỹ hạn chế tác động của quy định quá mức và ủng hộ thị trường dẫn dắt, khuyến khích đổi mới sáng tạo; Liên minh Châu Âu, nơi kiểm soát chặt rủi ro ở các cấp độ cao nhất; và Trung Quốc, quốc gia nhấn mạnh vào sự ổn định xã hội và sự kiểm soát của nhà nước<sup>43</sup>. Đảo quốc này xác định cần tránh cả hai thái cực: tình trạng quản lý không đầy đủ, khiến công dân gặp rủi ro và làm suy yếu lòng tin của công chúng; và quản lý quá mức, có thể khiến các doanh nghiệp cả trong và ngoài nước lo ngại, kìm hãm sự đổi mới.

### 3.1.3. Khuôn khổ pháp luật liên quan đến AI

#### 3.1.3.1. Tình hình xây dựng pháp luật về AI

Theo Oxford AI Index 2024, AI được đề cập ngày càng nhiều trong thảo luận lập pháp ở các nước; cụ thể là năm 2023 AI được đề cập 2.175 lần ở nghị viện của 49 quốc gia, tăng gần gấp đôi so với năm 2022; trong đó Vương quốc Anh đứng đầu bảng xếp hạng, tiếp theo là Hoa Kỳ, Úc, Ireland và Tây Ban Nha<sup>44</sup>. Từ năm 2016-2023, tổng cộng trên toàn thế giới có 148 đạo luật có liên quan đến AI được ban hành; trong đó Hoa Kỳ thông qua nhiều nhất với 23 đạo luật, tiếp theo là Bồ Đào Nha (15), Bỉ (12). Riêng năm 2023, Bỉ dẫn đầu với 5 đạo luật về AI được ban hành, tiếp theo là Pháp, Hàn Quốc, UK mỗi nước 3 đạo luật; còn ở Hoa Kỳ năm 2023 mặc dù có đến 181 dự luật được trình Quốc hội liên bang, nhưng chỉ có 1 đạo luật ra đời. Còn ở các tiểu bang của Hoa Kỳ, 30 đạo luật được ban hành năm 2023, trong đó California ban hành 7 đạo luật, Virginia ban hành 5 đạo luật<sup>45</sup>.

<sup>39</sup> Xem một số sáng kiến chính sách của Trung Quốc về AI tại đây, trong đó có một số quy tắc đạo đức về AI: *Policies - OECD.AI; China's New AI Governance Initiatives Shouldn't Be Ignored - Carnegie Endowment for International Peace | Carnegie Endowment for International Peace*

<sup>40</sup> *Emerging AI-related regulation AI Strategies and Policies - OECD.AI*

<sup>41</sup> Xem báo cáo tại đây: [global\\_ai\\_law\\_policy\\_tracker.pdf](https://global_ai_law_policy_tracker.pdf)

<sup>42</sup> *Oxford AI Index 2024*, [https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI\\_AI-Index-Report-2024\\_Chapter\\_7.pdf](https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_AI-Index-Report-2024_Chapter_7.pdf)

<sup>43</sup> <https://vneconomy.vn/hon-dao-ai-singapore-da-lam-the-nao-de-duoi-kip-hai-sieu-cuong-ai-my-trung.htm>

<sup>44</sup> *Oxford AI Index 2024* <https://aiindex.stanford.edu/report/>

<sup>45</sup> *Oxford AI Index 2024* <https://aiindex.stanford.edu/report/>

Điều đáng chú ý là ở nhiều nước, các văn bản quản lý hành chính dưới luật (regulations) về AI càng ngày càng tăng<sup>46</sup>. Các văn bản loại này do các bộ, các cơ quan hành quản lý ban hành để thực thi luật, quy định chi tiết cụ thể các nội dung của luật. Điều này cho thấy mỗi quan tâm ngày càng tăng đối với quy định quản lý về AI ở các nước, nhất là ở Mỹ và EU. Bảng dưới đây so sánh thực tế ban hành các văn bản loại này ở Mỹ và EU và cho thấy một số điểm đáng quan tâm. Trước hết, về số lượng văn bản được ban hành từ 2016-2023, EU khá vượt trội so với Hoa Kỳ cả về tổng số và số văn bản liên quan đến AI ở mức trung bình và cao, nhưng số văn bản trực tiếp quy định về AI ngang nhau ở hai khu vực này (7 văn bản). Về tính chất quy định (hỗ trợ phát triển AI hay giới hạn, kiểm soát rủi ro của AI), trái với suy nghĩ thông thường, tỷ lệ văn bản giới hạn, kiểm soát AI ở Hoa Kỳ cao hơn so với EU. Như vậy, mối lo ngại về các rủi ro, tác hại của AI ở Hoa Kỳ lớn không kém so với ở EU; và cách tiếp cận thúc đẩy, hỗ trợ phát triển AI, đổi mới, sáng tạo trong ban hành văn bản ở Hoa Kỳ không có nghĩa là tránh né ban hành các quy định để kiểm soát rủi ro.

**Bảng 3. So sánh tình hình xây dựng pháp luật về AI ở Hoa Kỳ và EU<sup>47</sup>**

Nội dung so sánh	Hoa Kỳ (cấp liên bang)	EU
<b>Số lượng văn bản được ban hành</b>	Từ 2016 – 2023: 85 văn bản. 2020: 16 văn bản; 2021: 12 văn bản; 2022: 16 văn bản; 2023: 25 văn bản.	Từ 2017 – 2023: 129 văn bản. 2020: 12 văn bản; 2021: 46 văn bản; 2022: 22 văn bản; 2023: 32 văn bản.
<b>Mức độ liên quan đến AI</b>	Từ 2016 – 2023: 50 văn bản ít liên quan; 26 văn bản liên quan vừa phải; 7 văn bản liên quan trực tiếp (2019: 2 văn bản; 2020: 1 văn bản; 2023: 4 văn bản).	Từ 2017 – 2023: 71 văn bản ít liên quan; 51 văn bản liên quan vừa phải; 7 văn bản liên quan trực tiếp (2017: 1 văn bản; 2021: 2 văn bản; 2022: 2 văn bản; 2023: 2 văn bản).
<b>Cơ quan ban hành</b>	Từ 2016 – 2023, Văn phòng Tổng thống, Bộ Y tế và Dịch vụ nhân sinh, Cơ quan quản lý bảo hiểm y tế quốc gia ban hành nhiều văn bản nhất liên quan đến AI. Năm 2023, số lượng cơ quan ban hành quy định có liên quan đến AI tăng lên 21 vào năm 2023 (so với 17 năm 2022); Văn phòng Tổng thống và Bộ Thương mại mỗi cơ quan ban hành 05 văn bản; Bộ Y tế và Dịch vụ nhân sinh, Cục Công nghiệp và an ninh: mỗi cơ quan ban hành 04 văn bản.	Từ 2017 – 2023, Hội đồng Liên minh Châu Âu và Nghị viện Châu Âu ban hành nhiều văn bản nhất liên quan đến AI, lần lượt là 69 và 43. Năm 2023, hai cơ quan này vẫn quy định nhiều nhất về AI: Hội đồng Liên minh Châu Âu – 13, Nghị viện Châu Âu – 9.
<b>Tính chất quy định</b>	Từ 2016 – 2023: 15 văn bản hỗ trợ phát triển AI; 28 văn bản giới hạn, kiểm soát AI. Năm 2023: 3 văn bản hỗ trợ phát triển AI; 10 văn bản giới hạn, kiểm soát AI.	Trong những năm gần đây, quy định liên quan đến AI của EU có xu hướng hỗ trợ phát triển hơn. Năm 2021, có 15 văn bản tập trung hỗ trợ so với 6 văn bản hạn chế; năm 2022: 10 và 5 văn bản; năm 2023: 12 văn bản và 8 văn bản.
<b>Lĩnh vực được quy định</b>	Năm 2023: Lĩnh vực có nhiều văn bản nhất liên quan đến AI là thương mại quốc tế và tài chính quốc tế (03). Y tế; thương mại; khoa học, công nghệ và truyền thông mỗi lĩnh vực 02 văn bản.	Năm 2023, chủ đề phổ biến nhất trong các văn bản liên quan đến AI ở EU là khoa học, công nghệ và truyền thông (5); hoạt động của chính phủ, chính trị như bầu cử (3).

<sup>46</sup> Oxford AI Index 2024 <https://aiindex.stanford.edu/report/>

<sup>47</sup> IPS tổng hợp từ Oxford AI Index 2024, [https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI\\_AI-Index-Report-2024\\_Chapter\\_7.pdf](https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_AI-Index-Report-2024_Chapter_7.pdf)

### 3.1.3.2. Các cách tiếp cận trong xây dựng pháp luật về AI

Trong xây dựng khuôn khổ pháp luật về AI trên thế giới, có hai cách tiếp cận chính: theo chiều rộng, tức là một khung chung, có thể là một đạo luật riêng bao quát về AI và theo chiều hẹp, tức là các văn bản pháp luật khác nhau điều chỉnh các nội dung cụ thể liên quan đến AI<sup>48</sup>.

Luật riêng về AI: Trên bình diện quốc tế, đã có sự thay đổi hướng đến điều chỉnh công nghệ AI bằng luật riêng, cụ thể như: Đạo luật AI của EU năm 2024; Đại hội đồng LHQ thông qua nghị quyết toàn cầu đầu tiên về AI vào tháng 3/2024; Hội đồng châu Âu vào tháng 5/2024 thông qua Công ước khung về trí tuệ nhân tạo với quyền con người, dân chủ và pháp quyền, một điều ước quốc tế đầu tiên có tính ràng buộc pháp lý về AI.

Ở cấp độ quốc gia, Chính phủ Anh Starmer đã công bố ý định xây dựng dự luật về AI; Quốc hội Canada tiếp tục xem xét Đạo luật Trí tuệ nhân tạo và dữ liệu (AIDA). Đáng chú ý là một số bang của Hoa Kỳ cũng đã hoặc đang xây dựng đạo luật có tính chất như vậy. Cụ thể, cơ quan lập pháp California đã thông qua Luật Đổi mới một cách an toàn và bảo mật các mô hình AI tiên tiến (nhưng Thống đốc bang lại phủ quyết dự luật)<sup>49</sup>. Bang Colorado của Hoa Kỳ đã ban hành Luật Bảo vệ người tiêu dùng trong quá trình tương tác với các hệ thống AI có hiệu lực từ 2/2026<sup>50</sup>. Đây là đạo luật toàn diện, điều chỉnh các rủi ro cao của AI đối với người dùng trong vòng đời của hệ thống AI; quy định trách nhiệm của bên phát triển và bên triển khai hệ thống AI; được đánh giá có cách tiếp cận khá giống với Đạo luật AI của EU<sup>51</sup>.

Các văn bản pháp luật khác nhau điều chỉnh AI: Ví dụ về cách tiếp cận theo lĩnh vực hẹp là Chính phủ Trung Quốc ban hành các văn bản pháp luật riêng lẻ để quản lý việc tạo nội dung từ AI tạo sinh<sup>52</sup>; tổng hợp sâu thông tin Internet<sup>53</sup>; dùng thuật toán đề xuất nội dung trên Internet<sup>54</sup>. Điều đáng chú ý là các văn bản này đều điều chỉnh việc tạo nội dung trên môi trường số, là một mối quan tâm rất lớn của Chính phủ Trung Quốc, nhằm bảo đảm thông tin được tổng hợp, tạo ra nhờ sự trợ giúp của AI không làm ảnh hưởng đến an ninh quốc gia, vị thế của chính quyền.

Tiểu bang California, nơi có trụ sở Silicon Valley, luôn thu hút sự chú ý của cả thế giới về các chính sách công nghệ, đã chứng tỏ vai trò đi đầu trong ban hành chính sách, pháp luật điều chỉnh các vấn đề khác nhau về AI. Chỉ riêng trong tháng 9-10/2024, Thống đốc bang đã ký ban hành hàng chục đạo luật điều chỉnh các vấn đề khác nhau liên quan đến AI<sup>55</sup>. Các tiểu bang khác của Hoa Kỳ cũng thông qua các luật liên quan đến AI, ví dụ về việc chính phủ sử dụng AI, hệ thống tự động; bảo vệ người tiêu dùng AI; sử dụng AI hoặc nội dung do AI tạo ra; ứng dụng AI trong quản lý nhân lực như tuyển dụng, đánh giá nhân viên ở các công ty<sup>56</sup>.

---

<sup>48</sup> *Lessons From the World's Two Experiments in AI Governance - Carnegie Endowment for International Peace | Carnegie Endowment for International Peace*

<sup>49</sup> SB-1047 *Safe and Secure Innovation for Frontier Artificial Intelligence Models Act.*

<sup>50</sup> [https://leg.colorado.gov/sites/default/files/2024a\\_205\\_signed.pdf](https://leg.colorado.gov/sites/default/files/2024a_205_signed.pdf)

<sup>51</sup> <https://insightplus.bakermckenzie.com/bm/data-technology/north-america-from-brussels-to-boulder-colorado-enacts-comprehensive-ai-law-on-the-heels-of-eus-ai-act-with-significant-obligations-for-business-and-employers>

<sup>52</sup> *Translation: Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment) – April 2023; How will China's Generative AI Regulations Shape the Future? A DigiChina Forum.*

<sup>53</sup> *Translation: Internet Information Service Deep Synthesis Management Provisions (Draft for Comment) – Jan. 2022*

<sup>54</sup> *Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022*

<sup>55</sup> *Governor Newsom announces new initiatives to advance safe and responsible AI, protect Californians | Governor of California*

<sup>56</sup> <https://www.connectontech.com/illinois-joins-colorado-and-nyc-in-restricting-generative-ai-in-hr-plus-a-quick-survey-of-the-legal-landscape-across-the-us-and-globally/>

**Bảng 4. Ưu, nhược điểm của từng cách tiếp cận**

Cách tiếp cận	Ưu điểm	Nhược điểm/yêu cầu
<b>Theo chiều rộng</b>	<p>Bao quát hơn, ổn định hơn, bởi lẽ bất chấp sự đa dạng của các ứng dụng AI, các biện pháp kiểm soát rủi ro mà chúng gây ra thường liên quan đến tính minh bạch, trách nhiệm giải trình, tránh phân biệt đối xử.</p> <p>Giúp các nhà phát triển, doanh nghiệp dự đoán tốt hơn; giúp các chính phủ tập trung nguồn lực hơn.</p>	<p>Cơ quan lập pháp cần khả năng sửa đổi hoặc bổ sung quy định để theo kịp công nghệ. Một khung quy định theo chiều rộng cho một công nghệ phát triển nhanh như AI khó có thể duy trì được lâu.</p> <p>Khi khung pháp luật chung ủy quyền việc thiết lập các yêu cầu tuân thủ chi tiết cho bộ, ngành, cần có cơ chế phòng ngừa, giảm thiểu khả năng thâm tóm, lạm dụng của ngành.</p>
<b>Theo chiều hẹp</b>	<p>Có tính chất cụ thể, rõ ràng, giảm thiểu các tác hại cụ thể do các ứng dụng AI khác nhau gây ra, không phải chịu sức ép phải khái quát hóa cho tất cả các ứng dụng của công nghệ.</p> <p>Khi các biện pháp can thiệp cụ thể chứng minh thành công, có thể kết hợp các công cụ hoặc yêu cầu đó vào luật trong tương lai nhằm vào các tác hại khác nhau.</p>	<p>Đòi hỏi một mức độ phối hợp nhất định giữa các cơ quan quản lý (regulators) để giảm thiểu chi phí cho cả cơ quan quản lý và doanh nghiệp. Nếu không, có nguy cơ “phát minh lại bánh xe” mỗi khi một bộ phận mới được giao nhiệm vụ quản lý một ứng dụng cụ thể của AI.</p> <p>Nếu các quy định không đi kèm với các nguồn lực kỹ thuật và tài chính cần thiết, các cơ quan quản lý ngành sẽ không thể đáp ứng yêu cầu kiểm soát rủi ro mang tính chuyên sâu của các ứng dụng AI hẹp.</p>

### 3.1.3.3. Điều chỉnh AI bằng khuôn khổ pháp luật hiện hành

Ở các nước, khuôn khổ pháp luật hiện hành đều có những quy định có thể điều chỉnh các khía cạnh khác nhau liên quan đến rủi ro do AI gây ra. Ngay cả khi nếu có một đạo luật riêng về AI hoặc các đạo luật khác nhau về AI như ở nhiều nước, thì các luật đó vẫn không thay thế, mà được tích hợp vào hệ thống pháp luật hiện hành.

Ví dụ như ở Úc, đó là Luật về quyền riêng tư năm 1988, Luật về bản quyền năm 1968, Bộ luật hình sự năm 1995, Luật Công ty năm 2001, Luật Việc làm công bằng năm 2009, Luật Cạnh tranh và bảo vệ người tiêu dùng năm 2010, pháp luật hành chính<sup>57</sup>. Một số nội dung trong các luật hiện hành của Úc có thể được áp dụng để điều chỉnh rủi ro của AI như: Quy định các điều kiện phải được đáp ứng trước khi một sản phẩm được đưa vào thị trường; loại trừ một số loại quy trình hoặc hành vi đối với các nhà cung cấp hàng hóa hoặc dịch vụ nhằm đảm bảo các sản phẩm đó an toàn, đáng tin cậy, ví dụ: cấm hành vi gây hiểu lầm trong Luật Tiêu dùng Úc; quy định đầu ra hoặc tiêu chuẩn mà cá nhân hoặc công ty phải đáp ứng, ví dụ: trách nhiệm của các nhà sản xuất đối với lỗi sản phẩm phát sinh khi hàng hóa không an toàn theo cách “mọi người thường có quyền mong đợi”<sup>58</sup>.

### 3.1.3.4. Cơ chế thử nghiệm có kiểm soát (Sandbox)

Sandbox đóng vai trò là tạo ra không gian nơi các công ty có thể thử nghiệm các sản phẩm, dịch vụ mới có tính chất vượt khỏi các khuôn khổ pháp lý hiện hành. Các công ty này nhận được sự hỗ

<sup>57</sup> Australian Government Department of Industry, Science and Resources, *Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings*, 9/2024.

<sup>58</sup> Jeannie Marie Paterson, *AI có trách nhiệm: Tiếp cận tổng thể và ứng dụng cụ thể, tài liệu tập huấn, Viện Nghiên cứu chính sách và phát triển truyền thông, Hà Nội, 9/2024.*

trợ, hướng dẫn của các cơ quan chức năng và được miễn trừ các quy định pháp lý hoặc quy trình tuân thủ có liên quan. Mặc dù các cách tiếp cận đối với sandbox quy định có thể khác nhau, nhưng chúng có một số đặc điểm chung:

- Có tính chất tạm thời, với thời hạn thử nghiệm nhất định, thường là từ 06 – 24 tháng.
- Miễn trừ các quy định pháp luật hiện hành và cung cấp hỗ trợ pháp lý phù hợp cho một dự án cụ thể, thường dựa trên thử và sai (trial-and-error).
- Các thông tin về kỹ thuật, thị trường và dữ liệu mà được thu thập trong quá trình thử nghiệm giúp các cơ quan quản lý đánh giá khung pháp lý cụ thể có phù hợp hay cần được điều chỉnh.

Tại thời điểm 2024, đã có khoảng 100 sáng kiến sandbox trên toàn thế giới, bao gồm sandbox về fintech và quyền riêng tư. Cơ quan Tiêu chuẩn Tài chính của Vương quốc Anh (FCA) là đơn vị tiên phong trong sandbox fintech đầu tiên vào năm 2015 và nhiều quốc gia khác đã làm theo<sup>59</sup>.

### 3.1.4. Hướng dẫn, tiêu chuẩn kỹ thuật

Các tài liệu hướng dẫn của chính phủ, các viện nghiên cứu, các quy chuẩn, tiêu chuẩn kỹ thuật là một trong những công cụ ngày càng được chú ý trong quản trị AI ở trên thế giới. Các tiêu chuẩn thường được xây dựng từ quá trình thảo luận, thống nhất giữa các chuyên gia kỹ thuật từ giới nghiên cứu, doanh nghiệp, chính phủ; bảo đảm tính chính xác, tin cậy về mặt kỹ thuật, ít chịu ảnh hưởng về chính trị<sup>60</sup>. Loại văn bản hướng dẫn này có ưu điểm linh hoạt, tự nguyện thực hiện, cho nên không tạo áp lực tuân thủ cho doanh nghiệp, cá nhân. Tuy nhiên, nếu không kết hợp với các nguồn lực, áp lực thực hiện theo từng lĩnh vực cụ thể, các văn bản này không được thực hiện.

Dưới đây là một số Bộ tiêu chuẩn AI đáng chú ý:

- Khung quản lý rủi ro AI (RMF) của Viện Tiêu chuẩn và công nghệ quốc gia Hoa Kỳ ban hành với mục đích “cải thiện khả năng xem xét độ tin cậy trong thiết kế, phát triển, sử dụng và đánh giá các sản phẩm, dịch vụ và hệ thống AI”.
- Các tiêu chuẩn kỹ thuật về AI có trách nhiệm đã được phát triển bởi Standards Australia và phối hợp với các nỗ lực quốc tế, chẳng hạn như ISO-IEC.
- Tiêu chuẩn 22989 của Tổ chức Tiêu chuẩn quốc tế (ISO), ban hành tháng 7 năm 2022, cung cấp tổng quan về thuật ngữ AI, thiết lập cơ sở chung cho các thuật ngữ và khái niệm AI.
- ISO 23894 ban hành tháng 2 năm 2023, hướng dẫn chiến lược cho các tổ chức trong mọi lĩnh vực để quản lý rủi ro liên quan đến việc phát triển và sử dụng AI.
- ISO 38507 ban hành tháng 4 năm 2022, cung cấp hướng dẫn cho các cơ quan quản lý như hội đồng quản trị tạo môi trường và quản lý việc sử dụng AI.
- IEEE P2863: Thực hành được khuyến nghị dành cho quản trị tổ chức về AI: Tạo các tiêu chí và các bước quản trị để phát triển hoặc sử dụng AI hiệu quả trong các tổ chức.
- ISO 42001 ra mắt năm 2023 và ISO 42005 nêu chi tiết các tiêu chuẩn về hệ thống quản lý rủi ro AI và đánh giá tác động AI.

### 3.1.5. Quy tắc ứng xử tự nguyện

Quy tắc tự nguyện là một hình thức tự điều chỉnh, được phát triển bởi các ngành công nghiệp thường là để đáp ứng các mối quan tâm của xã hội. Ví dụ:

- *Cam kết toàn cầu gần đây của các công ty công nghệ chống lại AI deepfakes trong các cuộc bầu cử, bao gồm phát triển các công cụ phát hiện, giải quyết việc phân phối và thúc đẩy nhận thức của công chúng, nâng cao năng lực đọc hiểu, đánh giá thông tin trên truyền thông (media literacy).*
- *Cam kết an toàn sản phẩm của Úc (Australia's Product Safety Pledge) – một sáng kiến tự nguyện cam kết các bên ký tên chịu trách nhiệm đối với các sản phẩm không an toàn được bán thông qua các thị trường trực tuyến của họ, bao gồm cam kết báo cáo kết quả cho chính phủ hàng năm.*

<sup>59</sup> OECD, *Regulatory sandboxes in artificial intelligence*, 6/2023.

<sup>60</sup> Human Technology Institute, *AI Governance in Australia*, 2023.

## 3.2. PHÁT TRIỂN HỆ SINH THÁI AI

### 3.2.1. Hạ tầng và dữ liệu

Trên thế giới đang tồn tại khoảng cách lớn về hạ tầng và dữ liệu cho AI giữa các nước phát triển và đang phát triển, các nền kinh tế thị trường mới nổi đã tụt hậu trong việc áp dụng các dịch vụ đám mây. Ví dụ, các quốc gia ở Châu Mỹ Latinh và Trung Đông đang tụt hậu so với các nước phát triển về tổng chi tiêu cho dịch vụ đám mây công cộng, mỗi khu vực chỉ khoảng 10 tỷ USD vào năm 2022, so với 100 tỷ USD chỉ riêng tại Hoa Kỳ; tỷ lệ sử dụng đám mây ở Châu Phi chỉ ở mức 15%, bằng 1/3 mức trung bình của OECD<sup>61</sup>.

Dù thực tiễn thế nào, chính sách nhiều nước coi hạ tầng, cụ thể là siêu máy tính, điện toán đám mây, năng lực tính toán là nền tảng không thể thiếu để nghiên cứu, phát triển, ứng dụng AI. Chẳng hạn, Chính phủ Anh dành khoảng 900 triệu bảng Anh đầu tư vào siêu máy tính exascale một nghìn tỷ phép tính mỗi giây và Trung tâm nghiên cứu AI chuyên dụng<sup>62</sup>. Anh sẽ trở thành một trong số ít quốc gia trên thế giới lưu trữ siêu máy tính exascale, qua đó có thể thu hút nhân lực AI giỏi nhất, đảm bảo các nhà nghiên cứu có thể truy cập vào cơ sở hạ tầng tiên tiến nhất.

Trên phương diện năng lực tính toán, Hoa Kỳ duy trì lợi thế trên thị trường chip máy tính AI thông qua Nvidia. GPU của Nvidia là sản phẩm cuối cùng của chuỗi cung ứng bán dẫn phức tạp, với nhiều thành phần khác nhau do một số ít các công ty nắm giữ, nổi tiếng nhất là ASML và TSMC. Sau Hoa Kỳ, năm quốc gia chiếm vị trí tiếp theo là Trung Quốc, Singapore, Nhật Bản, Đài Loan và Hàn Quốc. Tuy nhiên, thành công trong sản xuất chip không nhất thiết là công thức cho thành công AI trong những công đoạn khác. Cụ thể, các quốc gia dẫn đầu về cơ sở hạ tầng bán dẫn như Nhật Bản, Đài Loan và Hàn Quốc đều có hiệu suất AI kém hơn so với quy mô và sự giàu có của họ<sup>63</sup>.

Phân tích toàn cầu của Tortoise đối với việc trích dẫn về sử dụng GPU Nvidia trong các ấn phẩm nghiên cứu AI cho thấy, lệnh cấm xuất khẩu GPU của Hoa Kỳ sang Trung Quốc chỉ có hiệu quả một phần. Các doanh nghiệp, cơ sở nghiên cứu Trung Quốc đã lách lệnh cấm xuất khẩu để mua và sử dụng chip H100 và A100 mạnh của Nvidia. Bên cạnh đó, họ cũng chuyển sang dùng nhiều các GPU RTX-series của Nvidia không mạnh bằng, nhưng vẫn đáp ứng nhu cầu<sup>64</sup>. Điều này có thể gợi ý về việc khắc phục tình trạng thiếu GPU công suất mạnh của Nvidia.

Để thúc đẩy việc sử dụng dịch vụ đám mây trong nước nhiều hơn, chính sách được khuyến nghị đối với các chính phủ là ưu tiên sử dụng điện toán đám mây hơn là các hệ thống CNTT truyền thống<sup>65</sup>. Những câu chuyện thành công trong sử dụng điện toán đám mây để cung cấp dịch vụ công như việc xử lý hóa đơn điện tử của Cơ quan thuế Mexico, đơn xin cấp phép của Philippines chứng minh lợi ích của các giải pháp đám mây. Cụ thể, Cơ quan thuế Mexico sử dụng dịch vụ điện toán đám mây từ năm 2012, tiết kiệm 20% chi phí đầu tư; xử lý khoảng 180 ngàn hóa đơn điện tử mỗi phút theo thời gian thực<sup>66</sup>.

Các ví dụ khác về chính sách liên quan đến điện toán đám mây: Khung chính sách đám mây của Qatar, quy định việc đánh giá các giải pháp đám mây so với các giải pháp sử dụng trung tâm dữ liệu<sup>67</sup>; Dự án Nimbus của Israel, thiết lập một khuôn khổ toàn diện cho toàn bộ chính phủ để áp dụng các dịch vụ đám mây; Chiến lược quốc gia về AI 2.0 của Singapore đặt ra kế hoạch đầu tư vào cơ sở hạ tầng điện toán đám mây, cho phép công chúng tiếp cận điện toán đám mây để sử dụng, mang lại giá trị, xây dựng, nâng cao năng lực, đổi mới sáng tạo, vì lợi ích công cộng; Chương trình “Go Cloud First” của Thái Lan tạo nền tảng hiện đại hóa các dịch vụ công thông qua công nghệ AI, bắt đầu với giao thông công cộng, dịch vụ chính phủ điện tử và sử dụng dữ liệu lớn; Chiến lược quốc gia về AI 2020-2045 của Indonesia khuyến nghị thành lập Trung tâm siêu máy tính AI quốc gia (INAISCC), nơi sẽ đóng vai trò là trung tâm (Hub) cho các cơ sở nghiên cứu, các ngành và các cơ quan chính phủ cần các dịch vụ tính toán lớn<sup>68</sup>.

<sup>61</sup> Google, *AI\_Sprinters\_Report.pdf*

<sup>62</sup> Government commits up to £3.5 billion to future of tech and science - GOV.UK, 3/2023.

<sup>63</sup> Tortoise 's Global Artificial Intelligence Index 2024, *The Global Artificial Intelligence Index 2024* - Tortoise

<sup>64</sup> Tortoise 's Global Artificial Intelligence Index 2024, *The Global Artificial Intelligence Index 2024* - Tortoise

<sup>65</sup> Google, *AI\_Sprinters\_Report.pdf*

<sup>66</sup> How can harnessing tax data create value for all? | EY - Global | EY - Taiwan

<sup>67</sup> Cloud Policy Framework

<sup>68</sup> Google, *An AI Opportunity Agenda for ASEAN, 2023.*

Đồng thời, để tăng cơ hội, điều kiện tiếp cận dữ liệu – một nguồn lực cho nghiên cứu, phát triển, ứng dụng AI, rất cần đến chính sách về dữ liệu mở, luồng dữ liệu xuyên biên giới đáng tin cậy. Bốn nước Brunei Darussalam, Malaysia, Singapore và Việt Nam đã cam kết cho phép luồng dữ liệu xuyên biên giới thông qua Hiệp định CPTPP. Singapore cũng đã các cam kết tương tự về luồng dữ liệu xuyên biên giới với Úc, Vương quốc Anh, Hàn Quốc, Chile và New Zealand<sup>69</sup>.

### 3.2.2. Nhân lực

Sự tập trung các chuyên gia AI, trình độ, năng lực chuyên môn của họ vẫn là yếu tố quan trọng hàng đầu đối với việc phát triển AI. Không có gì ngạc nhiên khi Hoa Kỳ dẫn đầu, với 27% các nhà khoa học AI hàng đầu trên toàn cầu sinh sống, làm việc, chưa kể các nhà nghiên cứu ứng dụng, kỹ sư AI. Ấn Độ đứng thứ hai về nhân lực một phần là do nhân khẩu học, phần khác nhờ vào mạng lưới các tổ chức giáo dục đại học của quốc gia này, và vị thế như một cường quốc gia công phần mềm. Đáng chú ý là Ấn Độ có số lượng lớn các nhà phát triển hoặc chuyên gia làm việc về AI nói chung, nhưng các nhà khoa học nghiên cứu AI rất chuyên sâu của Ấn Độ xếp sau nhiều quốc gia nhỏ hơn (đứng thứ 24 theo khảo sát của Tortoise).

Trong các năm 2018-2024, trên toàn thế giới, số lượng các nhà khoa học chuyển từ học thuật sang làm việc ở các công ty, tổ chức ứng dụng AI đã tăng đáng kể, với hơn một nửa số nhà khoa học AI làm việc trong các ngành này; cao nhất là Pháp, Đức, Israel với tỷ lệ gần 80% ở mỗi nước. Ở Ấn Độ mặc dù cũng có xu hướng tương tự, nhưng một tỷ lệ tương đối cao các nhà khoa học AI vẫn làm việc trong giới học thuật do thị trường AI trong nước vẫn còn tương đối nhỏ<sup>70</sup>.

### 3.2.3. Giáo dục, đào tạo

Chính sách về giáo dục, đào tạo liên quan đến AI xuất phát từ quan niệm: Mọi công dân, bao gồm cả công chức, cần phải có kiến thức cơ bản về AI, hiểu cách AI hoạt động, các ứng dụng và hạn chế tiềm năng của nó, và khả năng sử dụng các công cụ đơn giản chạy bằng AI. Cụ thể, có những ví dụ về chính sách giáo dục AI như<sup>70</sup>:

- **Tích hợp AI vào lớp học:** Tích hợp các khái niệm về AI vào chương trình giảng dạy của trường ngay từ khi còn nhỏ, coi kiến thức về kỹ thuật số và AI là một thành phần cốt lõi của giáo dục. Trải nghiệm AI, một chương trình giáo dục trực tuyến dành cho thanh thiếu niên (từ 11 đến 14 tuổi) được Raspberry Pi Foundation phát triển, thiết kế để hỗ trợ giáo viên giảng dạy những kiến thức cơ bản về AI và ML, và giúp những người trẻ tuổi, bao gồm cả những người trên các thị trường mới nổi đam mê AI.
- **Mở rộng các nguồn tài nguyên có thể truy cập:** Phát triển và quảng bá các khóa học trực tuyến miễn phí hoặc chi phí thấp, hướng dẫn và hội thảo về các kiến thức cơ bản về AI, giúp mọi người đều có thể học tập.
- **Phát triển các chương trình đào tạo và hỗ trợ AI quốc gia** để cung cấp cho người lao động kinh nghiệm thực tế trong việc áp dụng AI, tạo ra nguồn nhân tài cho các vai trò theo nhu cầu trong các lĩnh vực như nông nghiệp, chăm sóc sức khỏe và sản xuất, nơi mà hiệu quả do AI thúc đẩy có thể có tác động đáng kể. Ví dụ, Cơ quan Dữ liệu và trí tuệ nhân tạo của Ả Rập Xê Út đào tạo về dữ liệu và AI cho hơn 25.000 phụ nữ.
- **Mở rộng việc học trực tuyến:** Cung cấp các chương trình đào tạo và cấp chứng chỉ trực tuyến có thể tiếp cận rộng rãi, cung cấp các lộ trình linh hoạt để phát triển kỹ năng AI. Chương trình Chứng chỉ chuyên nghiệp trực tuyến của Google được cung cấp ở một số nước cho những người không có kinh nghiệm trước đó trong lĩnh vực kỹ thuật số hoặc AI.
- **Các sáng kiến đào tạo AI cho các doanh nghiệp vừa và nhỏ và các vùng nông thôn** để đảm bảo rằng lợi ích của công nghệ này được chia sẻ trên toàn xã hội. Tại Israel, chương trình Hành trình đến công nghệ tập trung vào việc đa dạng hóa tài năng công nghệ của Israel trong 5 năm, ưu tiên phát triển các chuyên gia công nghệ AI từ các nhóm dân cư ít được chú ý.
- **Giáo dục STEM:** Đầu tư mạnh vào giáo dục STEM từ khi còn nhỏ, bồi dưỡng một thế hệ những người có tư duy về AI, giải quyết vấn đề là một chính sách được quan tâm. Ví dụ, ở Kenya, Bộ Giáo dục đã áp dụng chương trình giảng dạy dành cho nhà phát triển Android, bao gồm một số mô-đun AI, vào chương trình quốc gia cho các cơ sở giáo dục bậc đại học

<sup>69</sup> Tortoise 's Global Artificial Intelligence Index 2024, The Global Artificial Intelligence Index 2024 - Tortoise

<sup>70</sup> Google, AI\_Sprinters\_Report.pdf

và dạy nghề, với mục tiêu đào tạo ít nhất 5000 nhà phát triển hàng năm. Ở Thổ Nhĩ Kỳ, kể từ khi ra mắt vào năm 2017 của Học viện BTK thuộc Cơ quan Công nghệ thông tin và truyền thông, hơn 200 nghìn sinh viên đã được tiếp cận thông qua các khóa đào tạo cấp chứng chỉ trực tuyến và trực tiếp, với trọng tâm chính là dữ liệu lớn và gần đây hơn là AI và AI tạo sinh.

### 3.2.4. Nghiên cứu

Hoa Kỳ, Trung Quốc, Singapore, Vương quốc Anh, Israel và Thụy Sĩ là những nước có hệ sinh thái nghiên cứu AI phát triển. Có hai loại nghiên cứu AI: nghiên cứu 'nền tảng' - sự phát triển của các thuật toán và mô hình cơ bản trong khoa học máy tính và AI - và nghiên cứu AI 'ứng dụng', giải quyết việc áp dụng các công nghệ AI hiện có vào các lĩnh vực khác. Ấn Độ có tỷ lệ nghiên cứu ứng dụng tương đối lớn hơn, trong khi Pháp, Israel và Singapore có tỷ lệ nghiên cứu nền tảng tương đối lớn hơn<sup>71</sup>.

Theo Tortois, tại thời điểm 2024, Chiến lược quốc gia của trên dưới 20 nước trong số hơn 80 các nước được khảo sát đã đề ra mục tiêu, giải pháp phát triển các mô hình AI nền tảng<sup>72</sup>. Mặc dù vậy, trong nghiên cứu các mô hình AI quy mô lớn mang tính đột phá, Hoa Kỳ và Trung Quốc vẫn thống trị trong những năm qua, nhờ dễ dàng tiếp cận vốn, hạ tầng tính toán mạnh, dữ liệu lớn, cũng như nhân lực nghiên cứu xuất sắc. Vương quốc Anh từng là nước đóng góp nhiều hơn cho nghiên cứu, phát triển AI (chủ yếu thông qua DeepMind) nhưng hiện đang tụt hậu. Chính phủ Vương quốc Anh đã dành 3,5 tỷ bảng Anh dành cho khoa học, công nghệ; trong đó khoảng 1 tỷ bảng Anh được cam kết dành cho thế hệ siêu máy tính và nghiên cứu AI; 100 triệu tỷ bảng cho 26 dự án R&D; tăng mức giảm thuế cho các SMEs tiến hành R&D bị thua lỗ, các công ty đủ điều kiện sẽ nhận được 27 bảng cho mỗi 100 bảng Anh đầu tư vào R&D<sup>73</sup>. Có lẽ các chính sách này nhằm đưa Vương quốc Anh trở lại đường đua thành siêu cường khoa học và công nghệ.

Mặt khác, các mô hình nguồn mở đã nhanh chóng bắt kịp về khả năng trong vài năm qua, mở ra sự cạnh tranh cho các công ty không phải của Hoa Kỳ. Pháp đã nhanh chóng khẳng định mình về phát triển mô hình AI loại này (24 mô hình), sau Hoa Kỳ (65) và Trung Quốc (41) tính đến tháng 8/2024<sup>74</sup>. Điển hình là Mistral gần đây đã công bố các mô hình có khả năng cạnh tranh với các mô hình của Hoa Kỳ và Trung Quốc. Đây là kết quả của những nỗ lực to lớn của Pháp nhằm phát triển các LLM không phải tiếng Anh với hệ sinh thái các công ty khởi nghiệp AI.

Các nước thường có chính sách hỗ trợ nghiên cứu và phát triển AI thông qua nguồn tài trợ có mục tiêu, thành lập các trung tâm nghiên cứu đẳng cấp thế giới, hợp tác giữa các trường đại học và các ngành kinh tế. Ở Đông Nam Á, Bản Lộ trình AI quốc gia của Philippines đặt mục tiêu thành lập Trung tâm nghiên cứu AI quốc gia (N-CAIR) hỗ trợ các doanh nghiệp Philippines, bao gồm SMEs, trong việc áp dụng công nghệ AI. Quỹ Nghiên cứu quốc gia Singapore cấp nguồn tài chính nâng cao năng lực nghiên cứu AI cho các chủ thể khác nhau. Một số nước có chính sách khuyến khích, tạo cơ chế phối hợp công - tư trong nghiên cứu nền tảng và nghiên cứu ứng dụng AI. Chẳng hạn, Google hợp tác với Đại học Trí tuệ nhân tạo Mohammed bin Zayed của UAE nhằm mục đích cải thiện chức năng tiếng Ả Rập trong các LLMs<sup>75</sup>. Viện Khoa học máy tính, trí tuệ nhân tạo và công nghệ ở Sofia, Bulgaria được Chính phủ hỗ trợ với một quỹ tài trợ gần 100 triệu USD; Google, DeepMind và Amazon Web Services hỗ trợ đầu tư vào cơ sở hạ tầng đám mây để chạy các mô hình học máy hiệu suất cao; hợp tác với các trường đại học công nghệ Thụy Sĩ để thu hút các nhà nghiên cứu và kỹ sư AI hàng đầu<sup>76</sup>.

### 3.2.5. Phát triển thị trường AI; đầu tư, nguồn tài chính

Đầu tư công vào AI của chính phủ các nước trên toàn cầu ngày càng tăng, từ 1 tỷ USD vào năm 2017 lên đến 19 tỷ năm 2023, 64 tỷ USD năm 2024, gần bằng 68 tỷ USD đầu tư tư nhân vào AI năm 2024. Đặc biệt là Ả Rập Xê Út dẫn đầu với các cam kết chi tiêu công lớn cho AI thập niên tới, vượt xa cả Hoa Kỳ và Trung Quốc. Quỹ đầu tư công của Ả Rập Xê Út (PIF) đã công bố một quỹ mới

<sup>71</sup> Tortoise 's Global Artificial Intelligence Index 2024, The Global Artificial Intelligence Index 2024 - Tortoise

<sup>72</sup> Tortoise 's Global Artificial Intelligence Index 2024, The Global Artificial Intelligence Index 2024 - Tortoise

<sup>73</sup> Government commits up to £3.5 billion to future of tech and science - GOV.UK, 3/2023.

<sup>74</sup> Tortoise 's Global Artificial Intelligence Index 2024, The Global Artificial Intelligence Index 2024 - Tortoise

<sup>75</sup> AI\_Sprinters\_Report.pdf

<sup>76</sup> Investing in Eastern Europe's AI future



trị giá 100 triệu đô la cho ngành bán dẫn và các ngành liên quan, và có kế hoạch thành lập một quỹ trị giá 40 tỷ đô la để đầu tư vào AI<sup>77</sup>. Hoa Kỳ đứng thứ hai về chi tiêu của chính phủ cho AI được công bố, tiếp theo là Hàn Quốc, quốc gia đã thể hiện chiến lược rõ ràng về đầu tư cho AI; Pháp, quốc gia hiện đang chi tiêu nhiều hơn Vương quốc Anh 60%; Đức và Tây Ban Nha, cả hai đều đã dành ngân sách lớn cho các chiến lược AI mới được cập nhật của họ; và Canada, quốc gia đứng đầu bảng xếp hạng về đầu tư công cho cơ sở hạ tầng tính toán AI cao cấp.

Đầu tư vào nâng cao năng lực của doanh nghiệp nhỏ và vừa (SMEs) là rất quan trọng, nhất là khi họ chiếm đa số, thậm chí đại đa số doanh nghiệp ở các nước, nhưng lại thường tụt hậu trong việc áp dụng các công nghệ tiên tiến. Có những nước đã ban hành chính sách với các biện pháp, cơ chế hỗ trợ SMEs tiếp cận nguồn lực, hưởng lợi ích từ việc khai thác và triển khai AI.

### Chính sách hỗ trợ SMEs phát triển AI

Các nước có các chính sách cải thiện khả năng tiếp cận vốn, bao gồm thông qua các chương trình cho vay và tài trợ lãi suất thấp được thiết kế để hỗ trợ quá trình chuyển đổi bắt kịp với AI. Đào tạo về tích hợp AI dựa trên đám mây, cùng với các ưu đãi tài chính như hoàn thuế cho việc sử dụng điện toán đám mây, có thể tạo điều kiện cho các SMEs áp dụng AI như một phần trong chiến lược chuyển đổi kỹ thuật số của họ. Chương trình Ứng dụng AI (AI Adoption Programme) của chính phủ Úc có thể là một ví dụ để các nước khác tìm hiểu. Chương trình này có tổng ngân sách là 17 triệu đô la Úc, nhận đơn, xét duyệt tài trợ từ 3 – 5 triệu đô la Úc cho 5 Trung tâm AI thực hiện các hoạt động hỗ trợ SMEs áp dụng AI trong công việc của doanh nghiệp, cụ thể là tư vấn miễn phí cho doanh nghiệp áp dụng AI an toàn, hiệu quả; tập huấn các kỹ năng chuyên sâu về AI. Hoặc như chương trình Chuyển đổi số quốc gia của Chile cung cấp các khóa học nâng cao kỹ năng AI cho các doanh nghiệp nhỏ.

Về đầu tư vào AI của khu vực tư, Hoa Kỳ vẫn là quốc gia dẫn đầu với hơn 60% thị phần thị trường AI toàn cầu, phần lớn nhờ một số ít các công ty công nghệ lớn và các quỹ đầu tư mạo hiểm vào công nghệ<sup>78</sup>. AI tạo sinh đang chiếm tỷ trọng ngày càng lớn trong đầu tư tư nhân vào AI trên toàn cầu. Tuy nhiên, không có mô hình kinh doanh rõ ràng nào cho AI tạo sinh, và các chuyên gia hiện đang nêu lên mối lo ngại về lợi nhuận đầu tư và tính bền vững lâu dài.

**Bảng 5. Đầu tư vào AI trên toàn cầu từ 2017 – 2024<sup>79</sup>**

Năm	Đầu tư công	Đầu tư của tư nhân
2017	1 tỷ USD	36 tỷ USD
2018	2 tỷ USD	42 tỷ USD
2019	14 tỷ USD	42 tỷ USD
2020	10 tỷ USD	47 tỷ USD
2021	16 tỷ USD	105 tỷ USD
2022	14 tỷ USD	66 tỷ USD
2023	19 tỷ USD	73 tỷ USD
2024	64 tỷ USD	68 tỷ USD

<sup>77</sup> Tortoise 's Global Artificial Intelligence Index 2024, The Global Artificial Intelligence Index 2024 - Tortoise

<sup>79</sup> Tortoise 's Global Artificial Intelligence Index 2024, The Global Artificial Intelligence Index 2024 - Tortoise

<sup>79</sup> Tortoise 's Global Artificial Intelligence Index 2024, The Global Artificial Intelligence Index 2024 - Tortoise

Trong khi các công ty công nghệ lớn có nguồn lực mạnh về tài chính, công nghệ, con người, việc hợp tác của chính phủ với họ là một giải pháp chính sách được áp dụng ở một số nước, trong đó nổi bật là các hoạt động của chính phủ các nước Đông Nam Á với Google (xem hộp). Đây là một chính sách có thể cần được xem xét học hỏi, nhất là đối với các nước đang phát triển vốn gặp khó khăn về nguồn lực đầu tư cho nghiên cứu, phát triển, ứng dụng AI.

### Sáng kiến hợp tác công – tư AI Trailblazers<sup>80</sup>

AI Trailblazers là sáng kiến chung giữa một số cơ quan của Chính phủ Singapore với Google Cloud nhằm đẩy nhanh quá trình phát triển các giải pháp AI. Hai cơ chế thử nghiệm sáng tạo (Innovation Sandbox) đã được thành lập để cung cấp cho 100 cơ quan, doanh nghiệp, tổ chức ở Singapore quyền truy cập miễn phí trong 3 tháng vào các bộ công cụ của Google gồm: các GPU hiệu suất cao, nền tảng AI Vertex, các mô hình AI tạo sinh được đào tạo trước, các công cụ dành cho nhà phát triển chi phí thấp. Nhờ đó, các cơ quan, doanh nghiệp, tổ chức có thể xây dựng và thử nghiệm các giải pháp AI tạo sinh của riêng mình trong một môi trường dựa trên đám mây được kiểm soát và chuyên dụng trước khi đưa vào sử dụng hoặc thương mại hóa.

## 3.3. CHÍNH SÁCH KIỂM SOÁT RỦI RO CỦA AI: MỘT SỐ CÁCH TIẾP CẬN

Có thể nhóm các chính sách kiểm soát rủi ro của AI trên thế giới thành các cách tiếp cận sau: Xây dựng Khung quản trị rủi ro toàn diện; quản trị theo cấp độ rủi ro của AI; quản trị rủi ro theo vòng đời AI; dựa trên các chuẩn mực quốc tế về các giá trị chung của loài người; kiểm soát chặt chẽ rủi ro của AI trong khu vực công; đặc biệt chú trọng bảo đảm an ninh quốc gia.

### 3.3.1. Khung quản trị rủi ro toàn diện

Kiểm soát rủi ro từ việc phát triển, sử dụng AI đòi hỏi một cách tiếp cận quản trị toàn diện, mà nổi bật là Khung quản lý rủi ro của AI (AI RMF) do Viện Tiêu chuẩn và Công nghệ Quốc gia (NIST) Hoa Kỳ công bố đầu năm 2023; được cập nhật các nội dung về AI tạo sinh năm 2024<sup>81</sup>. AI RMF hướng dẫn và đưa ra các biện pháp thực hành tốt nhất trong phát triển, thử nghiệm, triển khai AI, ưu tiên tính minh bạch, đạo đức và trách nhiệm giải trình; áp dụng cho tất cả các bên tham gia vào vòng đời của hệ thống AI. AI RMF có hai nhóm nội dung lớn:

**Một là**, RMF đưa ra lộ trình để các tổ chức xác định rủi ro liên quan đến AI; đặc điểm chính của AI đáng tin cậy (an toàn, bảo mật, có thể giải thích được, thúc đẩy quyền riêng tư, minh bạch và công bằng, xác tín - reliable); các tác hại của AI đối với quyền tự do dân sự, an toàn, cơ hội kinh tế; tác hại đối với hoạt động kinh doanh, an toàn và danh tiếng của doanh nghiệp; tác hại đối với các hệ sinh thái như chuỗi cung ứng và các hệ thống toàn cầu.

**Hai là**, RMF giới thiệu các quy trình và hoạt động để quản lý rủi ro AI, được nhóm thành bốn "chức năng" gồm: quản trị; rà soát; đo lường; quản lý (Governance, Mapping, Measurement, Management), viết tắt là GMMM:

- **Quản trị:** Các chính sách, quy trình và tiêu chuẩn rõ ràng cho việc triển khai và vận hành hệ thống; trách nhiệm rõ ràng trong tổ chức đối với việc quản lý AI, đảm bảo AI được phát triển và triển khai theo các nguyên tắc pháp lý và đạo đức.
- **Rà soát:** Khuôn khổ để đánh giá rủi ro AI, trong đó có phân loại hệ thống AI, đánh giá lợi ích và chi phí sử dụng.
- **Đo lường:** Đo hiệu suất của các mô hình, bao gồm độ chính xác, độ tin cậy, và độ mạnh thông qua việc đánh giá và giám sát các rủi ro đã xác định. Chức năng này cũng tập trung vào việc thử nghiệm và xác thực các mô hình với các số liệu hiệu suất rõ ràng để tăng độ tin cậy, độ chính xác, độ xác tín của kết quả.
- **Quản lý:** Giám sát, đánh giá, quản lý liên tục các rủi ro AI, trong đó tập trung vào các hệ thống AI có rủi ro cao hơn.

<sup>80</sup> IPS lập bảng theo số liệu của Tortoise's Global Artificial Intelligence Index 2024.

<sup>81</sup> Artificial Intelligence Risk Management Framework (AI RMF 1.0) ; Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile

Khung này ngày càng được áp dụng nhiều ở Hoa Kỳ cả trong khu vực công và khu vực tư, cũng như ở các nước khác. Cụ thể, Sắc lệnh hành pháp về Phát triển và sử dụng AI an toàn, bảo mật và đáng tin cậy của Tổng thống Biden đã viện dẫn AI RMF; Sắc lệnh Hành pháp về AI năm 2023 của Thống đốc bang California chỉ đạo các cơ quan chính phủ của bang xây dựng các hướng dẫn về mua sắm công và sử dụng AI tạo sinh dựa trên AI RMF. Đạo luật Đổi mới an toàn và bảo mật cho AI tiên tiến của California (bị Thống đốc phủ quyết) yêu cầu các nhà phát triển AI và các trung tâm tính toán phải xem xét Khung của NIST. Tương tự như vậy, Luật Bảo vệ người tiêu dùng trong tương tác với AI của bang Colorado yêu cầu các bên triển khai hệ thống AI có rủi ro cao phải thực hiện chính sách, chương trình quản lý rủi ro dựa trên AI RMF; nếu tuân thủ AI RMF thì sẽ được áp dụng các biện pháp bào chữa giảm nhẹ (affirmative defenses) khi bị kiện dân sự hoặc truy tố hình sự. Khung của NIST cũng đã được áp dụng để đánh giá, điều chỉnh rủi ro của AI đối với quyền con người<sup>82</sup>; trong một công ty cung cấp ứng dụng trên đám mây, nhà phát triển AI<sup>83</sup>; trong phát triển, triển khai hệ thống AI của Google DeepMind. Những diễn biến này làm nổi bật xu hướng ngày càng tăng của các nhà lập pháp, hành pháp, các tổ chức, công ty coi AI RMF của NIST là khung tiêu chuẩn trong quản lý rủi ro AI.

### 3.3.2. Kiểm soát dựa trên mức độ rủi ro

Hiện nay trên thế giới nhiều nước hoặc tổ chức quốc tế đang có cách tiếp cận kiểm soát theo mức độ rủi ro của AI như Đạo luật AI của EU, Các nguyên tắc về AI của OECD, Khung quản trị rủi ro của Hoa Kỳ. Cách tiếp cận này đặt trọng tâm vào các biện pháp tiền kiểm phòng ngừa (ex ante), ngăn ngừa không để rủi ro lan rộng gây ra những hậu quả, tổn hại nặng nề; hơn là trông chờ vào các biện pháp áp đặt trách nhiệm thị trường về sau, hoặc khiếu kiện dân sự ở tòa án kéo dài, tốn kém. Các biện pháp phòng ngừa có thể gồm đánh giá rủi ro, các khung quản lý rủi ro, hoặc là cấm những hệ thống, công cụ AI có những rủi ro không chấp nhận được.

Cách xác định rủi ro ở cấp độ cao nhất: Đạo luật AI của EU và luật một số nước khác xác định các rủi ro ở cấp độ cao nhất (rủi ro không thể chấp nhận theo cách phân loại của EU) bằng cách đưa ra các dấu hiệu, nguyên tắc. Theo tài liệu của Chính phủ Canada, những yếu tố chính sau đây cần được xem xét để xác định hệ thống AI nào được coi là có tác động lớn (bao gồm cả rủi ro không chấp nhận được theo ngôn ngữ của Đạo luật EU về AI)<sup>84</sup>: (1) Bằng chứng về rủi ro gây hại cho sức khỏe và sự an toàn, hoặc rủi ro tác động tiêu cực đến quyền con người, dựa trên cả mục đích dự kiến và hậu quả tiềm ẩn không mong muốn; (2) Mức độ nghiêm trọng của tác hại tiềm ẩn; (3) Quy mô sử dụng; (3) Bản chất của tác hại hoặc tác động tiêu cực đã xảy ra; (4) Mức độ mà con người không thể lựa chọn chối bỏ hệ thống AI một cách hợp lý; (5) Sự mất cân bằng về hoàn cảnh kinh tế hoặc xã hội, hoặc độ tuổi của những người bị ảnh hưởng.

Còn tài liệu của Chính phủ Úc đang được tham vấn công chúng (từ 9/2024) cho rằng, cần xem xét mức độ nghiêm trọng và phạm vi của những tác động tiêu cực trên các phương diện sau để đánh giá một hệ thống AI có rủi ro cao: (1) Rủi ro tác động tiêu cực đến các quyền của cá nhân được công nhận trong pháp luật về quyền con người của Úc mà không có lý do chính đáng, cũng như đối với các quyền theo cam kết quốc tế của Úc; (2) Rủi ro tác động tiêu cực đến sức khỏe thể chất hoặc tinh thần hoặc sự an toàn của cá nhân; (3) Rủi ro tác động pháp lý bất lợi, phỉ báng hoặc các tác động đáng kể tương tự đối với một cá nhân; (4) Rủi ro tác động tiêu cực đến các nhóm cá nhân hoặc quyền tập thể của các nhóm dân cư chung nền văn hóa; (5) Rủi ro tác động tiêu cực đến nền kinh tế, xã hội, môi trường và pháp quyền rộng lớn hơn của Úc<sup>85</sup>.

Đạo luật AI của EU là ví dụ điển hình cho cách tiếp cận kiểm soát theo mức độ rủi ro; theo đó, ứng với từng mức độ rủi ro, đạo luật có những quy định đề ra các biện pháp kiểm soát tương ứng. Cụ thể, AI gây ra rủi ro không thể chấp nhận như hệ thống AI cho phép chính phủ hoặc doanh nghiệp chấm điểm xã hội (social scoring) bị cấm ở EU. AI gây ra rủi ro cao như hệ thống AI được dùng trong tuyển dụng nhân sự, thực thi pháp luật sẽ phải đáp ứng các yêu cầu nghiêm ngặt về chất lượng dữ liệu, ghi nhật ký hệ thống, vận hành với sự giám sát của con người. AI gây ra rủi ro giới hạn như chatbot phải được gắn nhãn để người dùng nhận biết được họ đang tương tác với máy móc. Và AI có rủi ro tối thiểu như hệ thống gợi ý hoặc bộ lọc thư rác sẽ không phải đáp ứng các yêu cầu cứng trong luật. Thay vào

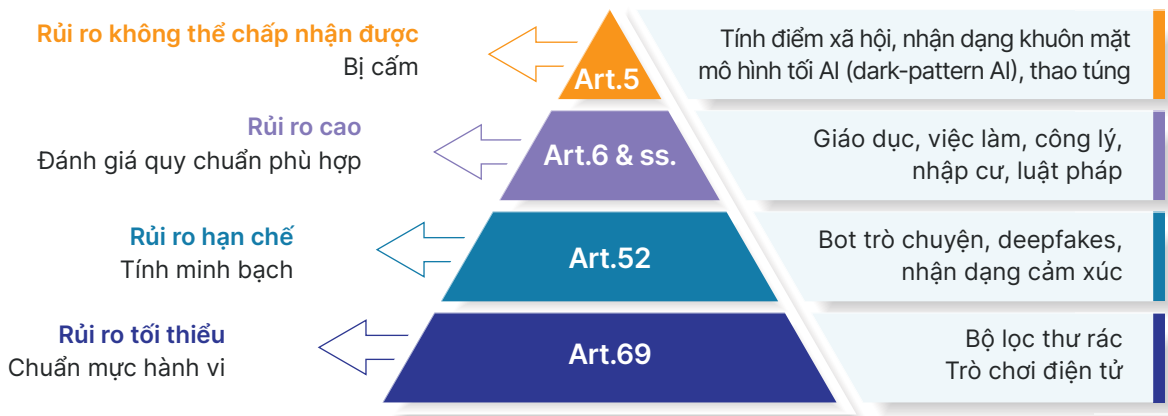
<sup>82</sup> Risk Management Profile for AI and Human Rights - United States Department of State

<sup>83</sup> Using the AI Risk Management Framework: Workday, 2024.

<sup>84</sup> The Artificial Intelligence and Data Act (AIDA) – Companion document

<sup>85</sup> Australian Government Department of Industry, Science and Resources, Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings, 9/2024.

đó, trên cơ sở tự nguyện, các doanh nghiệp vận hành hệ thống AI rủi ro tối thiểu có thể áp dụng các bộ quy tắc ứng xử (code of conduct) cho các hệ thống AI này<sup>86</sup>.



**Hình 5. Các mức độ rủi ro AI**

Tại thời điểm cuối năm 2024, một số nước khác đang theo mô hình kiểm soát dựa trên rủi ro này. Chẳng hạn, Dự luật về Trí tuệ nhân tạo và Dữ liệu (AIDA) của Canada đang được xem xét ở Hạ viện liên bang, đặt nền tảng cho việc phát triển và triển khai có trách nhiệm các hệ thống AI tại Canada. Đạo luật phân loại AI theo rủi ro và áp đặt một số nghĩa vụ đối với các hệ thống có tác động lớn, rủi ro cao bao gồm các biện pháp thử nghiệm, minh bạch và trách nhiệm giải trình<sup>87</sup>. Hoặc là Chính phủ Úc đang tham vấn công chúng về các cơ chế kiểm soát bắt buộc đối với những hệ thống AI rủi ro cao<sup>88</sup>. Cách tiếp cận này ở Úc tuân theo nguyên tắc kiểm soát phòng ngừa (precautionary principle) đối với những lĩnh vực như biến đổi khí hậu, an toàn thực phẩm, dược phẩm, hóa học, khi còn nhiều điều chưa rõ ràng về các rủi ro cao do công nghệ gây ra trên các mặt môi trường, đạo đức, sức khỏe, an toàn.

### 3.3.3. Kiểm soát rủi ro theo vòng đời hệ thống AI

Cách tiếp cận kiểm soát rủi ro theo vòng đời của hệ thống AI liên quan chặt chẽ với kiểm soát theo cấp độ rủi ro, ngày càng được các tổ chức quốc tế, các nước theo đuổi. Chẳng hạn, như đã đề cập, Công ước khung về AI với dân chủ, pháp quyền, quyền con người của Hội đồng châu Âu nhấn mạnh cần kiểm soát rủi ro của AI trong tất cả các giai đoạn vòng đời của hệ thống AI. Khung quản lý rủi ro AI của Viện NIST đã được đề cập cũng có các nội dung nhận diện rủi ro và quản lý rủi ro theo vòng đời của hệ thống AI.

Trong tài liệu về AI của Chính phủ Úc đang được tham vấn công chúng từ tháng 9/2024, có các cơ chế kiểm soát rủi ro áp đặt trách nhiệm phòng ngừa, giảm thiểu những tổn hại do AI gây ra đối với bên phát triển, triển khai, ứng dụng AI trong suốt vòng đời của AI và trong chuỗi cung ứng AI<sup>89</sup>. Tương tự, Dự luật AI và Dữ liệu của Canada ban đầu chỉ quy định trách nhiệm về kiểm soát rủi ro nói chung, nhưng sau khi nhận được nhiều ý kiến góp ý đã bổ sung các quy định cụ thể về trách nhiệm của các bên phải có các biện pháp kiểm soát rủi ro trong từng giai đoạn trong vòng đời AI (xem thêm trong Bảng dưới đây)<sup>90</sup>.

<sup>86</sup> Dựa theo 'The EU AI Act: a summary of its significance and scope' by Lilian Edwards, Professor of Law, Innovation and Society, Newcastle University (April 2022).

<sup>87</sup> Luật này là một phần trong đạo luật lớn hơn có tên gọi là Đạo luật Thực thi hiến chương số năm 2022 (Digital Charter Implementation Act, 2022), dự kiến có hiệu lực từ 2025. Xem ở đây: Government Bill (House of Commons) C-27 (44-1) - First Reading - Digital Charter Implementation Act, 2022 - Parliament of Canada; The Artificial Intelligence and Data Act (AIDA) - Companion document

<sup>88</sup> Australian Government Department of Industry, Science and Resources, Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings, 9/2024.

<sup>89</sup> Australian Government Department of Industry, Science and Resources, Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings, 9/2024.

<sup>90</sup> Government of Canada, and Parliament of Canada, 'Amendments to Bill C-27 (44-1)'; 28 November 2023.

**Bảng 6. Kiểm soát rủi ro theo vòng đời AI<sup>91</sup>**

Các hoạt động theo vòng đời AI	Ví dụ về các biện pháp kiểm soát rủi ro
<p><b>Thiết kế hệ thống:</b> bao gồm việc xác định mục tiêu của hệ thống AI và nhu cầu dữ liệu, phương pháp luận hoặc mô hình dựa trên các mục tiêu đó.</p>	<ul style="list-style-type: none"> <li>• Đánh giá ban đầu về các rủi ro tiềm ẩn liên quan đến việc sử dụng hệ thống AI đặt trong bối cảnh và quyết định xem việc sử dụng AI có phù hợp hay không.</li> <li>• Đánh giá và giải quyết các thiên kiến tiềm ẩn do lựa chọn tập dữ liệu.</li> <li>• Đánh giá mức độ khả năng diễn giải cần thiết và đưa ra quyết định thiết kế phù hợp.</li> </ul>
<p><b>Phát triển hệ thống:</b> bao gồm xử lý các tập dữ liệu, huấn luyện hệ thống bằng dữ liệu, sửa đổi các tham số của hệ thống, phát triển và sửa đổi phương pháp luận hoặc mô hình được sử dụng trong hệ thống, hoặc thử nghiệm hệ thống.</p>	<ul style="list-style-type: none"> <li>• Lập hồ sơ các tập dữ liệu và mô hình đã sử dụng</li> <li>• Thực hiện đánh giá và xác thực, bao gồm huấn luyện lại khi cần</li> <li>• Xây dựng cơ chế giám sát và theo dõi của con người</li> <li>• Lập hồ sơ cách sử dụng và hạn chế phù hợp</li> </ul>
<p><b>Đưa hệ thống vào sử dụng:</b> Triển khai một hệ thống đầy đủ chức năng, có thể là do bên phát triển hệ thống thực hiện, hoặc thông qua giao dịch thương mại, thông qua giao diện lập trình ứng dụng (API) hoặc bằng cách công bố hệ thống đã đi vào hoạt động.</p>	<ul style="list-style-type: none"> <li>• Lưu giữ tài liệu về cách thức đáp ứng các yêu cầu về thiết kế và phát triển</li> <li>• Cung cấp tài liệu phù hợp cho người dùng về các tập dữ liệu được sử dụng, các hạn chế và cách sử dụng phù hợp</li> <li>• Thực hiện đánh giá rủi ro về cách thức hệ thống được cung cấp</li> </ul>
<p><b>Quản lý hoạt động của hệ thống:</b> Giám sát hệ thống trong khi sử dụng, bao gồm bắt đầu hoặc dừng hoạt động, theo dõi và kiểm soát quyền truy cập vào đầu ra của hệ thống trong khi hệ thống đang hoạt động, thay đổi các thông số liên quan đến hoạt động của hệ thống trong bối cảnh cụ thể.</p>	<ul style="list-style-type: none"> <li>• Ghi nhật ký và giám sát đầu ra của hệ thống khi thích hợp trong bối cảnh cụ thể.</li> <li>• Đảm bảo theo dõi và giám sát của con người Can thiệp khi cần thiết dựa trên các thông số vận hành</li> </ul>

### 3.3.4. Dựa trên các chuẩn mực quốc tế về các giá trị chung của loài người

Đánh giá, điều chỉnh AI dựa trên các chuẩn mực quốc tế về các giá trị chung của loài người như dân chủ, pháp quyền, quyền con người là một cách tiếp cận được thống nhất rộng khắp, từ các tổ chức quốc tế như LHQ, OECD, EU, ASEAN, cho đến các nước riêng biệt như Hoa Kỳ, Úc, Canada<sup>92</sup>. Cụ thể, Tuyên ngôn quốc tế về quyền con người (UDHR), Công ước quốc tế về quyền dân sự và chính trị (ICCPR); Công ước quốc tế về quyền kinh tế, xã hội và văn hóa (ICESCR) và nhiều văn kiện luật quốc tế khác tạo nên chuẩn mực chung phổ quát về quyền con người. Khung này bao gồm: (1) Các giá trị cốt lõi của quyền con người phẩm giá, tự do, bình đẳng, đoàn kết/gắn kết; (2) các quyền con người cụ thể như quyền riêng tư; bình đẳng trước pháp luật; quyền tự do ngôn luận và bày tỏ ý kiến; quyền tự do hội họp và lập hội một cách hòa bình; quyền tiếp cận thông tin v.v...

Quyền con người được coi như cơ sở, nền tảng để đánh giá công nghệ như AI trong nhiều văn kiện liên quan đến công nghệ của Liên Hợp quốc và các tổ chức quốc tế khác. Điển hình là vào tháng 3 năm 2024, 193 quốc gia thành viên của Liên Hợp quốc đã thông qua nghị quyết nhấn mạnh: "quyền con người và các quyền tự do cơ bản phải được tôn trọng, bảo vệ và thúc đẩy trong suốt vòng đời của các hệ thống trí tuệ nhân tạo"; và "kêu gọi tất cả các quốc gia thành viên và các bên liên quan ngừng sử dụng các hệ thống trí tuệ nhân tạo không phù hợp với pháp luật quốc tế về

<sup>91</sup> Government of Canada, *The Artificial Intelligence and Data Act (AIDA) – Companion document*.

<sup>92</sup> Australian Government Department of Industry, Science and Resources, *Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings*, 9/2024.

quyền con người hoặc gây ra rủi ro quá mức đối với việc thực hiện các quyền con người<sup>93</sup>.

Vào tháng 9/2024, 46 quốc gia thành viên của Hội đồng châu Âu (EC) và 11 quốc gia ngoài EC đã ký kết tham gia Công ước khung về AI với quyền con người, dân chủ, pháp quyền. Công ước đặt ra mục tiêu và yêu cầu các quốc gia thành viên phải đảm bảo mọi hoạt động trong vòng đời của các hệ thống AI hoàn toàn phù hợp với quyền con người, dân chủ, pháp quyền<sup>94</sup>. Mặc dù không có yêu cầu cụ thể về các biện pháp đối với AI, đây là điều ước quốc tế đầu tiên có hiệu lực pháp lý về công nghệ mới này, một dấu mốc lớn trong quá trình xây dựng khung pháp lý toàn cầu về AI. Điều đáng nói là việc ký kết, tham gia Công ước để mở cho tất cả các quốc gia trên thế giới mà không chỉ dành cho các quốc gia thành viên của EC.

Trước đó, Các nguyên tắc về Kinh doanh và quyền con người của Liên Hợp quốc được Hội đồng Nhân quyền Liên hợp quốc thông qua theo một nghị quyết vào năm 2011 trong đó nêu ra một hệ thống các tiêu chuẩn cần phải áp dụng nhằm phòng chống và giải quyết các vi phạm quyền con người liên quan tới các hoạt động kinh doanh. Các nguyên tắc này được nhấn mạnh và khuyến cáo áp dụng trong các nghị quyết mới đây của LHQ về quyền trên môi trường số, quyền con người trong phát triển, sử dụng AI.

Tuyên bố Toronto về bảo vệ các quyền bình đẳng và chống phân biệt đối xử trong các hệ thống máy học được thông qua vào ngày 16/5/2018. Tuyên bố tái khẳng định vai trò của luật quốc tế về quyền con người trong việc bảo vệ các cá nhân và các nhóm khỏi sự phân biệt đối xử và bất bình đẳng khi áp dụng AI, đồng thời nêu ra yêu cầu xây dựng các khung quy định về mặt đạo đức cho ngành máy học. Tuyên bố khẳng định sự cần thiết phải tuân thủ các nguyên tắc pháp lý quốc tế về quyền con người trong mọi hoàn cảnh, và liệt kê trách nhiệm bảo vệ quyền con người của các quốc gia và các chủ thể tư nhân khi sử dụng các hệ thống máy học, bao gồm việc giảm thiểu các tác động mang tính phân biệt đối xử, sự minh bạch, việc cung cấp các giải pháp đền bù hiệu quả cho những người bị xâm hại<sup>95</sup>.

### **3.3.5. Kiểm soát rủi ro ở khu vực công; thúc đẩy đổi mới sáng tạo trong khu vực tư**

Tại thời điểm cuối năm 2024, Quốc hội liên bang của Hoa Kỳ chưa ban hành một khuôn khổ pháp lý chung về AI; đang xem xét một số dự luật có liên quan. Cụ thể, dự thảo Luật Nghiên cứu, đổi mới và trách nhiệm giải trình về trí tuệ nhân tạo (AIRIA) đề cập các trường hợp sử dụng AI có “tác động nghiêm trọng” và “tác động lớn”, là những trường hợp AI được sử dụng để ra quyết định có hiệu lực pháp lý gây rủi ro cho các quyền hiến định hoặc rủi ro đối với sự an toàn<sup>96</sup>. Hoặc một số dự luật nhằm điều chỉnh vấn đề lao động, việc làm trong bối cảnh phát triển AI, trong đó nhận diện rủi ro của AI đối với việc làm; phòng ngừa, giảm thiểu rủi ro đối với việc làm, người lao động trong quá trình các tổ chức, doanh nghiệp sử dụng AI<sup>97</sup>.

Còn nhánh hành pháp Hoa Kỳ tiếp tục theo đuổi thúc đẩy đổi mới sáng tạo trong khu vực tư, thể hiện ở cách tiếp cận điều tiết “mềm” trong kiểm soát rủi ro của AI thông qua các bộ hướng dẫn, quy tắc ứng xử theo ngành. Chẳng hạn, Bản hướng dẫn thiết kế, sử dụng, triển khai AI vì các quyền của công dân Mỹ (Blueprint for an AI Bill of Rights) được Nhà Trắng công bố vào tháng 10/2022. Sau đó, Viện Tiêu chuẩn và Công nghệ quốc gia thuộc Bộ Thương mại Hoa Kỳ xây dựng Khung quản trị rủi ro AI, hướng dẫn đánh giá, phân loại AI và phát triển các tiêu chuẩn dựa trên đồng thuận chung. Ngay sau khi nhậm chức vào cuối tháng 1/2025, Tổng thống Donald Trump đã ra Sắc lệnh yêu cầu các cơ quan Chính phủ lập kế hoạch đổi mới sáng tạo, phát triển AI.

Mặt khác, đối với khu vực công, Chính phủ liên bang và các tiểu bang Hoa Kỳ ngày càng có những văn bản áp đặt trách nhiệm chặt chẽ hơn. Sắc lệnh hành pháp của Biden về Phát triển và sử dụng AI an toàn, bảo mật và đáng tin cậy năm 2023 quy định riêng về trách nhiệm của các cơ quan

<sup>93</sup> UN General Assembly, Seventy-eighth session, Resolution 78/265, adopted on 21 March 2024, *Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development*.

<sup>94</sup> Xem toàn văn Công ước tại đây: <https://rm.coe.int/1680afae3c>; tóm tắt và bình luận về Công ước tại đây: [https://www.connecton-tech.com/over-fifty-coun-](https://www.connecton-tech.com/over-fifty-coun-tries-sign-on-to-the-council-of-europes-framework-convention-on-artificial-intelligence-the-first-legally-binding-international-treaty-on-ai/)

[tries-sign-on-to-the-council-of-europes-framework-convention-on-artificial-intelligence-the-first-legally-binding-international-treaty-on-ai/](https://www.connecton-tech.com/over-fifty-coun-tries-sign-on-to-the-council-of-europes-framework-convention-on-artificial-intelligence-the-first-legally-binding-international-treaty-on-ai/)

<sup>95</sup> Mark Latonero, *Governing Artificial Intelligence: Upholding Human Rights and Dignity*, 2018, p. 12.

<sup>96</sup> Congress of the United States, *Artificial Intelligence Research, Innovation, and Accountability Bill 2023*, U.S. Government, 15 November 2023.

<sup>97</sup> Xem Sắc lệnh tại đây: <https://www.whitehouse.gov/presidential-actions/2025/01/removing-riars-to-american-leadership-in-artificial-intelligence/>

chính phủ của Hoa Kỳ trong áp dụng, quản lý và sử dụng AI<sup>98</sup>. Tại thời điểm hoàn thành báo cáo này, chưa rõ chính quyền Trump có bãi bỏ các nội dung về trách nhiệm của các cơ quan chính phủ của Hoa Kỳ trong áp dụng, quản lý và sử dụng AI hay không. Ở cấp độ tiểu bang, Thống đốc California Gavin Newsom đã ban hành Sắc lệnh Hành pháp về AI vào tháng 9 năm 2023 nhằm thúc đẩy việc sử dụng AI có trách nhiệm trên khắp các hoạt động của khu vực công California. Cuối năm 2024, nghị viện bang New Hampshire ban hành Luật về AI trong khu vực công, trong đó cấm các cơ quan chính phủ của bang sử dụng một số công cụ, ứng dụng AI và đề ra các biện pháp để thực thi các quy định cấm.

### **3.3.6. Coi trọng an ninh quốc gia trong khi phát triển AI**

Trung Quốc xây dựng các biện pháp quản lý công nghệ số với trọng tâm bảo đảm an ninh quốc gia. Luật An ninh mạng, Luật Bảo mật dữ liệu, Luật Bảo vệ thông tin cá nhân do Cơ quan quản lý không gian mạng Trung Quốc (Cyberspace Administration of China - CAC) xây dựng và thực thi là ba trong số nhiều văn bản cho thấy cách tiếp cận này<sup>99</sup>. Chẳng hạn, yêu cầu dữ liệu phải được lưu trữ tại Trung Quốc đại lục (data localization), cấp phép đối với hoạt động chuyển dữ liệu cá nhân ra khỏi lãnh thổ Trung Quốc, để đảm bảo “tài nguyên số” nằm trong chủ quyền quốc gia.

Tiếp nối cách tiếp cận này, tháng 4/2023, CAC đã xây dựng Các biện pháp tạm thời quản lý dịch vụ AI tạo sinh (Interim Measures for the Management of Generative Artificial Intelligence Services) có hiệu lực từ ngày 15/8/2023. Văn bản này khẳng định nguyên tắc quản lý AI, dịch vụ AI của Trung Quốc là coi trọng an ninh và phát triển như nhau (国家坚持发展和安全并重). Các hoạt động triển khai dịch vụ AI tạo sinh được Đảng và Nhà nước theo dõi thận trọng để đảm không làm xói mòn các giá trị cốt lõi của chủ nghĩa xã hội đặc sắc Trung Quốc<sup>100</sup>. Sự thận trọng đối với dịch vụ AI tạo sinh tương tự với cách tiếp cận với thương mại điện tử, công nghệ tài chính, tiền mã hóa trước đây. Cách tiếp cận này dường như xuất phát từ việc chưa xác định rõ ràng rủi ro và cơ hội mà công nghệ mới tạo ra. Do đó, chính quyền vẫn để thị trường phát triển, nhưng phải đúng định hướng chủ nghĩa xã hội đặc sắc Trung Quốc. Mới đây nhất, nghị quyết của Đảng Cộng sản Trung Quốc đã nhấn mạnh nhiệm vụ thể chế hóa các hệ thống giám sát, kiểm soát bảo đảm sự an toàn đối với AI<sup>101</sup>.

<sup>98</sup> Xem Sắc lệnh tại đây: <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>

<sup>99</sup> Xem cụ thể hơn các viết, cũng như bản dịch tiếng Anh các chính sách, luật của Trung Quốc về các vấn đề liên quan đến AI tại đây: DigiChina.

<sup>100</sup> Xem thêm bình luận của các chuyên gia về văn bản này tại đây: *How will China's Generative AI Regulations Shape the Future? A DigiChina Forum*; DigiChina.

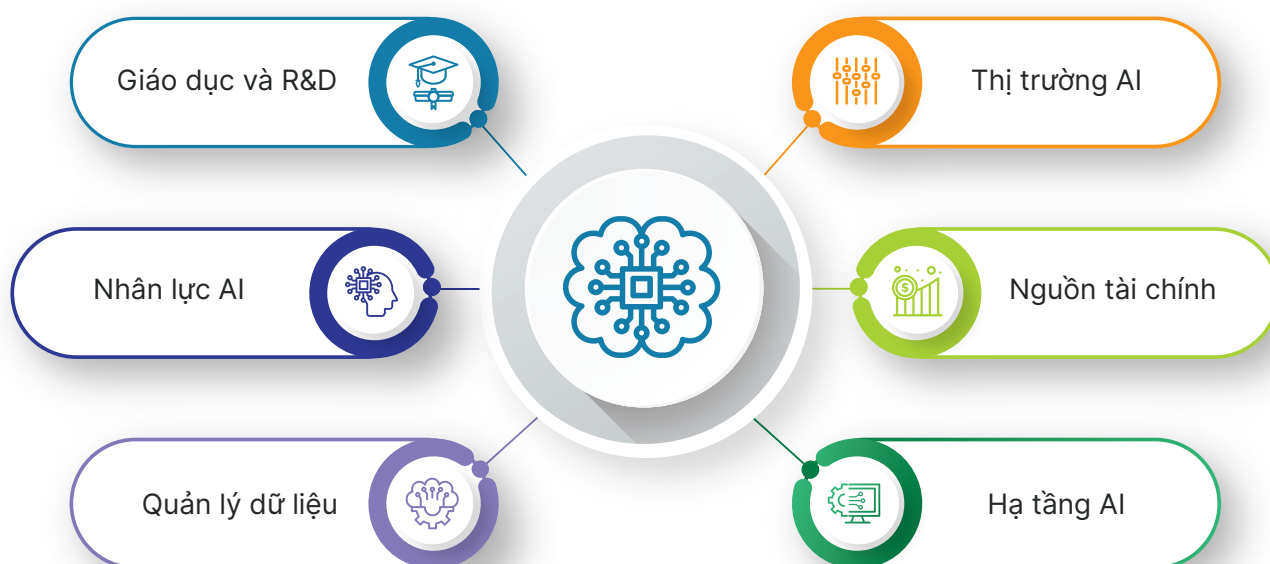
<sup>101</sup> *What does the Chinese leadership mean by "instituting oversight systems to ensure the safety of AI?"*; *AI Safety in China #16 - AI Safety in China*

# IV.

## HỆ SINH THÁI VÀ CHÍNH SÁCH VỀ TRÍ TUỆ NHÂN TẠO Ở VIỆT NAM

### 4.1. HIỆN TRẠNG HỆ SINH THÁI AI Ở VIỆT NAM

Báo cáo này đánh giá hệ sinh thái AI ở Việt Nam, các kết quả, cũng như hạn chế, khó khăn của hệ sinh thái trên các khía cạnh: Thị trường AI; nguồn và cơ chế tài chính; hạ tầng & dữ liệu; nhân lực AI; giáo dục, đào tạo và nghiên cứu về AI.



Hình 6. Hệ sinh thái AI

#### 4.1.1. Đầu tư vào thị trường AI ở Việt Nam: Sức nóng và thực tế

Để có cái nhìn toàn cảnh về sự phát triển của thị trường AI ở Việt Nam, cần đặt trong bối cảnh công nghiệp CNTT, công nghệ số đã trở thành một trong những ngành kinh tế - kỹ thuật có quy mô lớn, có tốc độ tăng trưởng nhanh. Theo Bộ Thông tin và truyền thông, năm 2024, doanh thu công nghiệp CNTT ước đạt 152 tỷ USD, tăng 35,71% so với năm 2019 (112 tỷ USD); số doanh nghiệp công nghệ số Việt Nam đang hoạt động ước khoảng 54.000 doanh nghiệp, tăng 28,16% so với năm 2019 (42.136 doanh nghiệp); tạo công ăn việc làm cho hơn 1,5 triệu người, tăng khoảng 50,44% so với năm 2019 (1.005.206 lao động)<sup>102</sup>. Như vậy, quy mô doanh thu, số lượng doanh nghiệp, việc làm từ CNTT, công nghệ số đều đã tăng trưởng với tỷ lệ khá cao, mặc dù còn khiêm tốn so với các ngành khác và so với các nước.

Về thị trường AI, trong những năm gần đây, các khảo sát khác nhau cho thấy mối quan tâm lớn của các doanh nghiệp Việt Nam đối với AI. Đơn cử như Chỉ số sẵn sàng AI (AI Readiness Index) năm 2024 của CISCO cho thấy, các công ty ở Việt Nam ngày càng cảm thấy cấp thiết trong việc áp dụng và triển khai công nghệ AI, với 100% các công ty cho biết nhu cầu triển khai AI ngày càng

<sup>102</sup> Cục ICT, Bộ Thông tin và truyền thông, Định hướng xây dựng thể chế về công nghiệp công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.



tăng trong năm qua; 48% công ty dành 10% đến 30% ngân sách công nghệ thông tin cho việc triển khai AI, cam kết một nguồn lực lớn cho AI<sup>103</sup>. Theo một khảo sát của PwC, tiềm năng chuyển đổi của GenAI ngày càng thu hút sự chú ý của các doanh nghiệp gia đình ở Việt Nam: 82% bày tỏ sự quan tâm trong việc tìm hiểu về GenAI, phản ánh sự công nhận rộng rãi về tầm ảnh hưởng của GenAI; 55% cho rằng họ am hiểu về GenAI (so với Toàn cầu là 53%)<sup>104</sup>. Điều này cho thấy sự tin tưởng vào khả năng của GenAI trong việc thúc đẩy sự chuyển đổi và đổi mới tích cực, phản ánh xu hướng chung về việc nắm bắt các công nghệ tiên tiến trong bối cảnh môi trường kinh doanh đang phát triển tại Việt Nam.

Nhu cầu ngày càng tăng của thị trường đã tạo ra một số kết quả tích cực đối với thị trường AI ở Việt Nam. Quy mô thị trường năm 2024 ước tính là 753,40 triệu đô la Mỹ; 3,4 tỷ đô la Mỹ vào năm 2030 với tốc độ tăng trưởng hàng năm (CAGR) 2024-2030 là 28,63%<sup>105</sup>. Một số ít doanh nghiệp nhà nước và tư nhân đã đầu tư vào nghiên cứu và ứng dụng AI với quy mô khá lớn, chẳng hạn như Viettel, FPT, Mobifone, Vingroup. AI đã được ứng dụng trong các tổ chức, doanh nghiệp ở các ngành như ngân hàng, thương mại điện tử, bán lẻ, sản xuất, y tế, giáo dục v.v...<sup>106</sup> Thị trường AI ở Việt Nam đang chứng kiến sự gia tăng khá lớn trong việc áp dụng các chatbot, trợ lý ảo và công nghệ nhận dạng giọng nói hỗ trợ AI. Xu hướng này được thúc đẩy bởi nhu cầu ngày càng tăng về tự động hóa và hiệu quả trong các ngành như ngân hàng, thương mại điện tử và chăm sóc sức khỏe. Khi công nghệ AI trở nên tiên tiến và dễ tiếp cận hơn, dự kiến sẽ có tác động đáng kể đến hoạt động kinh doanh và trải nghiệm của khách hàng<sup>107</sup>.

Các tập đoàn toàn cầu đã có các hoạt động đầu tư vào thị trường AI ở Việt Nam. Cụ thể như Google hợp tác với Trung tâm Đổi mới Sáng tạo Quốc gia Việt Nam (NIC) ra mắt các dự án: (1) "Xây dựng cho Tương lai AI": trao 40.000 học bổng với kỹ năng AI thiết yếu; (2) "Google for Staups Accelerator Đông Nam Á: tập trung vào AI" cung cấp cho các công ty khởi nghiệp quyền truy cập vào khoản tín dụng Google Cloud lên tới 350.000 USD. Còn Qualcomm bắt đầu triển khai từ năm 2019 "Thử thách đổi mới sáng tạo Qualcomm Việt Nam" nhằm xác định và nuôi dưỡng các doanh nghiệp khởi nghiệp đổi mới sáng tạo hoạt động trong lĩnh vực công nghệ đầy tiềm năng của Việt Nam. Chương trình bao gồm hỗ trợ kỹ thuật, huấn luyện kinh doanh và đào tạo kiến thức về sở hữu trí tuệ. Hoặc là "Chương trình Tăng tốc AI tạo sinh" của AWS bổ sung thêm 80 công ty khởi nghiệp giai đoạn đầu trên khắp thế giới đang sử dụng AI thể hệ để giải quyết các vấn đề phức tạp. Quỹ GenAI là quỹ trị giá 10 triệu USD đầu tư từ 50.000 USD đến 1 triệu USD cho mỗi công ty khởi nghiệp<sup>108</sup>.

Về khởi nghiệp AI, một khảo sát năm 2024 cho biết, trong số 250 công ty khởi nghiệp dựa trên GenAI ở các nước ASEAN, 44% có trụ sở tại Singapore, 27% tại Việt Nam, 13% tại Indonesia, 7% tại Thái Lan, 6% tại Malaysia và 3% tại Philippines, còn lại ở một số nước khác thuộc ASEAN; tỷ lệ khởi nghiệp GenAI của Việt Nam được dự báo sẽ còn tăng từ 27% lên 35% năm 2025<sup>109</sup>. Các công ty này chủ yếu tập trung vào B2B (92%); hoạt động trong các lĩnh vực chính gồm: Giải pháp nâng cao năng suất và kinh doanh (26%), Chăm sóc sức khỏe (13%), Dịch vụ tài chính (13%), Giải pháp công nghệ & AI (11%). Còn theo dữ liệu của PitchBook Data Inc., Việt Nam đứng thứ hai ở Đông Nam Á sau Singapore về số lượng AI startup năm 2024 với 765 công ty<sup>110</sup>.

Singapore dẫn đầu bảng về số lượng công ty khởi nghiệp GenAI là điều dễ hiểu do môi trường thân thiện với doanh nghiệp, cơ cấu thuế, khả năng tiếp cận nguồn tài trợ, nguồn nhân tài đa dạng, cơ sở hạ tầng mạnh và chính phủ hỗ trợ; nhưng sự vươn lên của Việt Nam lên vị trí số 2 đặc biệt đáng chú ý. Theo khảo sát của GenAI Fund, chi phí hoạt động thấp hơn và nguồn nhân lực đa dạng ở Việt Nam cho phép các công ty khởi nghiệp tạo ra sản phẩm chất lượng cao, trong khi vẫn giữ được chi phí trong tầm kiểm soát, biến Việt Nam trở thành lựa chọn ngày càng phổ biến để thành lập công ty khởi nghiệp GenAI.

<sup>103</sup> CISCO, Cisco 2024 AI Readiness Index (Vietnam): Hype Meets Reality [https://www.cisco.com/c/dam/m/en\\_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf](https://www.cisco.com/c/dam/m/en_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf)

<sup>104</sup> PwC Việt Nam, Thế hệ kế nghiệp (NextGen) Việt Nam: Thành công trong kỷ nguyên của trí tuệ nhân tạo, 2024.

<sup>105</sup> Statista, Artificial Intelligence – Vietnam, <https://www.statista.com/outlook/tmo/artificial-intelligence/vietnam>

<sup>106</sup> The Future of AI in Vietnam: Top Business Opportunities, <https://jdi.group/the-future-of-ai-in-vietnam-top-business-opportunities/#:~:text=AI%20In%20Vietnam%20E2%80%93%20Landscape%20Overview,-Vietnam%20AI%20Subsector&text=As%20a%20result%2C%20Vietnam%20has,Al%20startup%20investments%20in%202023>

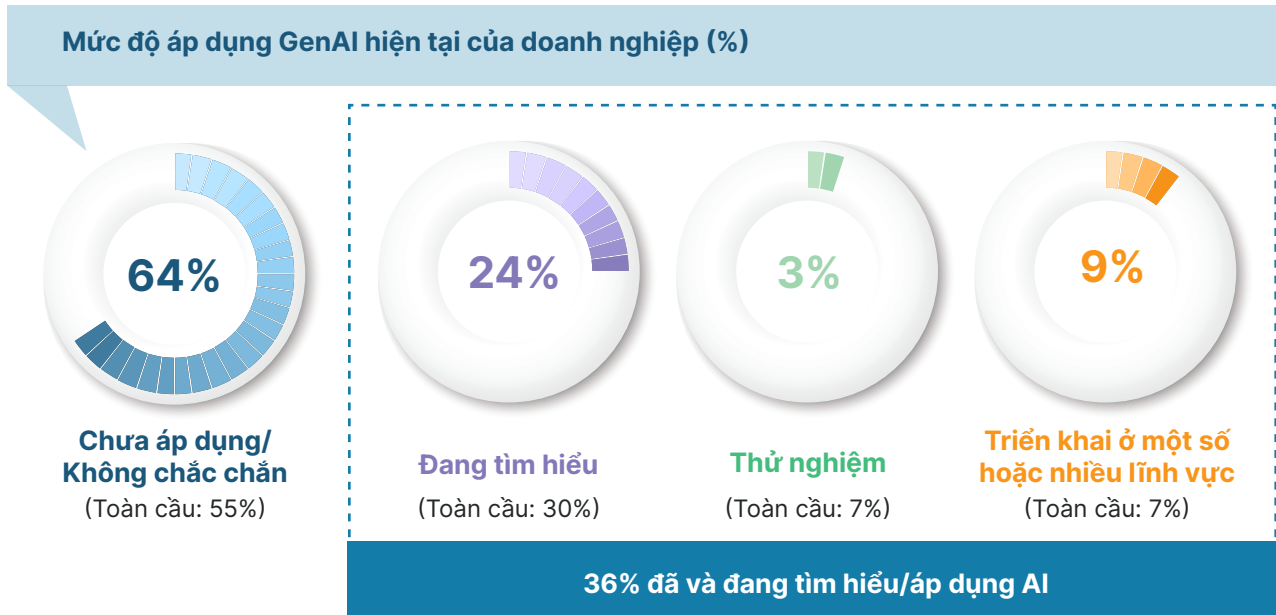
<sup>107</sup> Statista, Artificial Intelligence – Vietnam, <https://www.statista.com/outlook/tmo/artificial-intelligence/vietnam#analyst-opinion>

<sup>108</sup> TopDev, Vietnam IT & Tech Talent Landscape Repo, 2024.

<sup>109</sup> GenAI Fund, ASEAN GenAI Startup Report 2024, 9/2024.

<sup>110</sup> Ricky Hồ, Startup AI Việt Nam gặp cảnh 'voi nhiều, cỏ quá ít', Thời báo Kinh tế Sài Gòn, <https://thesaigontimes.vn/startup-ai-viet-nam-gap-canh-voi-nhieu-co-qua-it/>

Tuy nhiên, dữ liệu cũng cho thấy những điểm hạn chế trong đầu tư vào thị trường AI ở Việt Nam. Theo Chỉ số sẵn sàng AI năm 2024 của CISCO, chỉ có 22% các tổ chức tại Việt Nam đã hoàn toàn sẵn sàng triển khai và tận dụng các công nghệ AI, giảm so với mức 27% của năm 2023<sup>111</sup>. Các doanh nghiệp gia đình Việt Nam mới đang ở giai đoạn đầu áp dụng AI; với phần lớn (64%) vẫn chưa bắt đầu tìm hiểu về AI, gần một phần ba hiện đang tìm hiểu và thử nghiệm các sáng kiến liên quan đến AI (27%), tỷ lệ nhỏ hơn đã triển khai các sáng kiến này (9%); 27% (so với mức trung bình toàn cầu là 14%) đã có nhân sự chuyên trách về GenAI (xem hình dưới đây)<sup>112</sup>. Doanh nghiệp Nhà nước hoặc doanh nghiệp có vốn Nhà nước thường dè chừng trong việc thử nghiệm AI hơn so với doanh nghiệp tư nhân; thường phải đợi chủ trương hoặc quy trình đấu thầu gây tốn kém về thời gian; lo ngại về trách nhiệm và không có quy trình cho thử nghiệm AI.



**Hình 7. Mức độ áp dụng AI**

Các dự án đầu tư vào AI trong các lĩnh vực như kinh doanh, dịch vụ tiêu dùng, truyền thông, giáo dục có nguồn vốn huy động được còn nhỏ. Đơn cử như 765 startup AI và máy học tại Việt Nam chỉ gọi vốn được 47,3 triệu đô la trong chín tháng đầu năm 2024, trung bình mỗi startup chỉ nhận được gần 62.000 đô la<sup>113</sup>. Còn bảng dưới đây cho thấy quy mô đầu tư của các doanh nghiệp vào AI ở một số nước Đông Nam Á giai đoạn 2012 – 2021, theo đó đầu tư vào AI ở Việt Nam còn khiêm tốn so với các nước trong khu vực; cụ thể là năm 2021, chỉ có 31 triệu đô la Mỹ được đầu tư vào AI ở Việt Nam. Tại thời điểm 2024, đầu tư vào AI tại Việt Nam đã lớn hơn, đạt 95 triệu đô la Mỹ; nhưng vẫn thấp nhất trong sáu nền kinh tế chính tại ASEAN. Singapore vượt trội hơn các nước với 8,4 tỉ đô la Mỹ, chiếm 75% tổng vốn đầu tư cho AI của sáu nền kinh tế lớn ở ASEAN. Indonesia tiếp theo với 1,9 tỉ đô la, Malaysia 371 triệu đô la, Thái Lan 255 triệu đô la và Philippines 126 triệu đô la<sup>114</sup>.

<sup>111</sup> CISCO, Cisco 2024 AI Readiness Index (Vietnam): Hype Meets Reality [https://www.cisco.com/c/dam/m/en\\_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf](https://www.cisco.com/c/dam/m/en_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf)

<sup>112</sup> PwC Việt Nam, *Thế hệ kế nghiệp (NextGen) Việt Nam: Thành công trong kỷ nguyên của trí tuệ nhân tạo*, 2024.

<sup>113</sup> Ricky Hồ, *Startup AI Việt Nam gặp cảnh 'voi nhiều, cỏ quá ít'*, Thời báo Kinh tế Sài Gòn, <https://thesaigontimes.vn/startup-ai-viet-nam-gap-can-h-voi-nhieu-co-qua-it/>

<sup>114</sup> Albert Jehoshua Rapha, *Charting the future of Southeast Asian AI governance*, East Asia Forum, 5/2024 <https://eastasiaforum.org/2024/05/21/charting-the-future-of-southeast-asian-ai-governance/>

## Đầu tư vào trí tuệ nhân tạo tại một số quốc gia ở Đông Nam Á (triệu USD)

	Việt Nam	Malaysia	Thái Lan	Philippines	Indonesia	Singapore
2012		0.8		0.5		86
2013		3.1		1		33
2014		6.7		1.5		135
2015		1		1	0.2	188
2016	0	15	0.7	3	2.4	92
2017	0.5	23	0.7	5	6.5	286
2018	1.3	6.9	2.1	6	35	633
2019	3.2	8.3	2.9	7	224	669
2020	11	20	1.1	12	327	730
2021	31	114	162	34	344	2293

Nguồn: OECD.AI

### Hình 8. Đầu tư của doanh nghiệp vào AI ở một số nước Đông Nam Á (nguồn: OECD.AI)

Theo quan sát, các doanh nghiệp Việt Nam đầu tư AI với quy mô nhỏ, nhưng rải rác ở nhiều lĩnh vực. Mặt khác, theo một nghiên cứu, tại thời điểm năm 2021, các công ty đầu tư và ứng dụng AI chủ yếu có trụ sở tại Hà Nội (52,6%) và Thành phố Hồ Chí Minh (41,0%); chỉ có một số ít công ty có trụ sở tại Đà Nẵng (3,8%) và các tỉnh khác (2,6%)<sup>115</sup>. Điều này có thể là dấu hiệu cho thấy không có định hướng nào từ nhà nước trong việc phát triển các ngành công nghiệp AI quan trọng để có được khả năng cạnh tranh tập trung. Trong các văn bản hiện tại của Chính phủ Việt Nam, các lĩnh vực cụ thể chưa được xác định, cũng như không có phương tiện nào để thu hút và định hướng các nguồn đầu tư có đòn bẩy.

Theo khảo sát của CISCO, ở Việt Nam, dù đã đầu tư nhiều vào AI trong các lĩnh vực chiến lược như an ninh mạng, cơ sở hạ tầng công nghệ thông tin, phân tích và quản lý dữ liệu, nhưng các công ty cho biết, lợi nhuận từ những khoản đầu tư này không đáp ứng được kỳ vọng của họ. Ngoài ra, trung bình chưa đến một phần tư các công ty không thấy có lợi ích hoặc lợi ích không như kỳ vọng của họ trong việc tăng cường, hỗ trợ hoặc tự động hoá các quy trình hay việc vận hành hiện tại. Bên cạnh khó khăn về cơ sở hạ tầng, dữ liệu và quản trị, việc thiếu nhân tài có kỹ năng thực hiện các sáng kiến AI là thách thức lớn nhất mà các công ty đang phải đối mặt<sup>116</sup>.

Các công ty khởi nghiệp AI phải đối mặt với những thách thức như quá trình tích hợp AI vào doanh nghiệp chậm, sự phù hợp của sản phẩm với thị trường, quản lý dòng tiền, cạnh tranh với các đối thủ đã thành danh và thiếu đầu ra bán hàng. Họ thiếu vốn nhưng không thể gọi vốn vì chưa thể tiếp cận các quỹ hoặc không được nhà đầu tư sẵn lòng và rót vốn. Nhà đầu tư chỉ nhìn vào một lý do duy nhất: sản phẩm hay dịch vụ chưa đúng nhu cầu ngách của thị trường, không có khả năng tạo lợi nhuận cao; còn tình trạng “quá nhỏ lẻ và sản phẩm na ná nhau” của giới khởi nghiệp AI tại Việt Nam. Bên cạnh đó, chính sách hỗ trợ cho doanh nghiệp start-up ở Việt Nam thiếu hấp dẫn (ví dụ như không được hỗ trợ về thuế).

Phân tích thực trạng trên đây cho thấy, mục tiêu của Chiến lược quốc gia đến năm 2025 có “05 thương hiệu AI uy tín” dường như khó đạt được, khi mà chưa có thương hiệu AI nào được xác định cho đến thời điểm cuối năm 2024. Do đó, mục tiêu “10 thương hiệu AI” năm 2030 cũng sẽ không đạt được nếu không có sự điều chỉnh chính sách. Hơn nữa, trong các văn bản chính sách, pháp luật, cách thức xây dựng và bảo đảm nguồn lực cho các trung tâm AI quốc gia vẫn chưa rõ ràng, tác động đến các mục tiêu năm 2025 và 2030.

#### 4.1.2. Hạ tầng, dữ liệu dành cho AI

Theo nhận định của chuyên gia khi trao đổi với IPS, lợi thế để phát triển AI ở Việt Nam là chi phí hạ tầng bắt đầu giảm từ giai đoạn 2012 đến nay (2024); chip của Nvidia cho phép giảm thời gian

<sup>115</sup> Hien Thu Pham et al., *Artificial intelligence (AI) development in the Vietnam's energy and economic systems: A critical review*, *Journal of Cleaner Production* 438 (2024) 140692.

<sup>116</sup> CISCO, *Cisco 2024 AI Readiness Index (Vietnam): Hype Meets Reality*.

đào tạo AI so với chip truyền thống, trong khi có khả năng tiếp cận chip của Nvidia; đã có nhiều hơn dữ liệu mở, các mã nguồn mở. Một số doanh nghiệp ứng dụng AI đang sử dụng hạ tầng lưu trữ dữ liệu của doanh nghiệp Việt Nam, dữ liệu được lưu trữ tại Việt Nam, đáp ứng quy định về an ninh mạng.

Đồng thời, Việt Nam đang có sự tăng trưởng đáng kể trên thị trường trung tâm dữ liệu được thúc đẩy bởi nhu cầu ngày càng tăng về các giải pháp quản lý và lưu trữ dữ liệu, các cơ sở xử lý dữ liệu cục bộ. Đến cuối năm 2024, Việt Nam có 33 trung tâm dữ liệu với tổng số 48 nhà cung cấp và ước tính tổng công suất khoảng 150MW, xếp sau Singapore, Malaysia, Thái Lan, Indonesia và Philippines ở Đông Nam Á<sup>117</sup>. Trong đó, tập đoàn Viettel đã xây dựng 14 trung tâm dữ liệu với tổng công suất khoảng 87MW; FPT cùng với NVIDIA đầu tư 200 triệu USD để xây dựng “nhà máy” AI (AI Factory), một dạng trung tâm nghiên cứu và dữ liệu AI.

Ở khối doanh nghiệp, các “ông lớn” như Viettel, Tập đoàn Bưu chính Viễn thông Việt Nam (VNPT), FPT, các ngân hàng như Techcombank và VP Bank từ lâu đã chuẩn hóa dữ liệu của họ để tận dụng các nguồn dữ liệu nhằm tối ưu hóa hoạt động, chức năng và dịch vụ khách hàng của họ. Các tổ chức này đã đầu tư đáng kể vào số hóa dữ liệu và đã thiết lập các phương pháp quản lý dữ liệu để hỗ trợ các mục tiêu chiến lược và nâng cao hiệu quả hoạt động.

Mặc dù vậy, thực tế cho thấy thách thức để theo kịp các quốc gia khác trong việc thiết lập cơ sở hạ tầng AI và dữ liệu, đạt được tham vọng trở thành trung tâm AI của khu vực. Chi phí đầu tư cho hạ tầng AI rất lớn; đơn cử như tại thời điểm đầu 2024, ba công ty của VinGroup gồm Vin Brain, VinAI, Vin Bigdata sở hữu khoảng 100 chip GPU, trong khi một trung tâm dữ liệu nhỏ nhất của Meta có tới 16.000 chip GPU. FPT đã sở hữu chip H100 từ Nvidia, nhưng chỉ có thể dùng để huấn luyện mô hình AI nhỏ (khoảng 07-10 tỉ tham số), trong khi OpenAI có thể chạy mô hình AI lớn với 175 tỉ tham số. Hơn nữa, dòng đời hạ tầng phần cứng ngày càng ngắn (chỉ kéo dài 1-2 năm so với 3 – 5 năm trước đây), ví dụ như các dòng card GPU như NVIDIA H100 phát hành tháng 3/2023, đến giữa năm 2024 đã có thêm 3 version cao cấp hơn: H200, Blackwell B100, Blackwell B200 với hiệu suất cao, giá thành cạnh tranh<sup>118</sup>. Thực tế, hiện phần lớn các doanh nghiệp trên thế giới và Việt Nam không đủ nguồn lực để xây dựng LLM cho riêng mình, hoặc không có cloud riêng.

Ở Việt Nam, các chuyên gia về AI và đại diện của các doanh nghiệp, cơ quan Nhà nước cho rằng, Việt Nam đang gặp nhiều khó khăn liên quan đến hạ tầng tính toán. Doanh nghiệp Việt Nam chủ yếu vẫn đang phát triển, ứng dụng AI trên cơ sở hạ tầng tính toán, hạ tầng dữ liệu của doanh nghiệp nước ngoài. Nvidia cung cấp hạ tầng tính toán cho các công ty về AI của VinGroup và Viettel; FPT Cloud chủ yếu thuê hạ tầng điện toán đám mây từ Amazon Web Services; hoặc hạ tầng tính toán của một doanh nghiệp truyền thống ở Việt Nam chạy trên Google Lab. Theo một khảo sát năm 2024 đối với 250 công ty khởi nghiệp AI ở ASEAN (trong đó 27% ở Việt Nam), phần lớn thuê dịch vụ đám mây của các nhà cung cấp như AWS (tỷ lệ sử dụng 67%), GCP (47%) và Azure (41%); hoặc sử dụng mô hình nền tảng như của OpenAI (tỷ lệ sử dụng 76%), Anthropic (31%)<sup>119</sup>. Tình trạng phụ thuộc này có thể khiến Việt Nam đi sau 10 - 20 năm so với các quốc gia như Mỹ và Nhật. Bên cạnh đó, một lo ngại khác là các doanh nghiệp không đảm bảo rằng dữ liệu được lưu trữ tại Việt Nam. Vì vậy, chuyên gia nhấn mạnh, tương tự như điện, trong tương lai, Việt Nam cần phải sở hữu cơ sở hạ tầng cho AI chứ không thể “nhập khẩu” mãi. Đại diện của một số doanh nghiệp cho biết, nếu có một hạ tầng trong nước đủ tốt thì sẵn sàng sử dụng, người dùng không lo ngại về việc phải chuyển dữ liệu lên hạ tầng của Google hay Open AI.

Một vấn đề cấp thiết về hạ tầng cho AI là phát triển điện toán đám mây (cloud). Theo phản ánh của đại diện các doanh nghiệp trao đổi với IPS, cloud trong nước chất lượng không đủ để chạy hệ thống, dễ bị “chết” và mất nhiều thời gian để sửa lỗi. Hầu hết các doanh nghiệp cloud trong nước mới chỉ cung cấp dịch vụ cho thuê trung tâm dữ liệu (không phù hợp với doanh nghiệp vừa và nhỏ) mà chưa cung cấp dịch vụ tính tiền theo thời gian thực; hỗ trợ khách hàng vừa và nhỏ chưa tốt. Trong khi đó, doanh nghiệp Cloud nước ngoài cho phép tính tiền theo thời gian thực - phù hợp với doanh nghiệp vừa và nhỏ; có hệ thống quản trị tốt, dịch vụ hỗ trợ tốt hơn, giúp nuôi dưỡng doanh nghiệp vừa và nhỏ.

Chính vì chi phí đầu tư lớn, trên thế giới, theo nghiên cứu của Scale năm 2023, có tới 78% doanh

<sup>117</sup> <https://vneconomy.vn/techconnect/thi-truong-trung-tam-du-lieu-viet-nam-truoc-thach-thuc-kep-xanh-va-ai.htm>

<sup>118</sup> Trần Đăng Dũng, Viettel IDC, Hạ tầng công nghệ cho ứng dụng “AI” – động lực phát triển kinh tế số, hội thảo DataFest, Ninh Bình, 10/2024.

<sup>119</sup> GenAI Fund, ASEAN GenAI Startup Report 2024, 9/2024.

nghiệp được khảo sát lựa chọn LLM nguồn mở hoặc Cloud API để triển khai các giải pháp ứng dụng LLM<sup>120</sup>. Còn ở vùng châu Á – Thái Bình Dương, phần lớn các công ty đang tìm cách tận dụng các mô hình AI tạo sinh nguồn mở (41%) hoặc các mô hình tạo sinh API đám mây (37%), trong khi ít công ty muốn xây dựng các mô hình AI tạo sinh của riêng mình (22%)<sup>121</sup>.

Tuy nhiên, Việt Nam không tham gia phát triển các phần mềm nguồn mở trên thế giới và không tạo ảnh hưởng đến sự phát triển của các phần mềm nguồn mở (cá nhân người Việt có tham gia phát triển các phần mềm nguồn mở, nhưng được công ty nước ngoài thuê để làm việc). Trong khi đó, xây dựng phần mềm nguồn mở tạo ra tiêu chuẩn cho cộng đồng người dùng; giúp doanh nghiệp xây dựng phần mềm nguồn mở có thể “kiểm soát” cả thị trường, giảm thiểu chi phí marketing.

Về dữ liệu, doanh nghiệp Việt Nam kì vọng quá nhiều vào lợi ích kinh tế sinh ra từ dữ liệu, nhưng chưa đầu tư xứng đáng vào xây dựng hệ thống dữ liệu của doanh nghiệp. Điều này không chỉ cản trở khả năng của doanh nghiệp khai thác giá trị kinh tế từ dữ liệu của mình, mà còn gây khó khăn trong việc ứng dụng AI vào hoạt động quản trị, kinh doanh. Các doanh nghiệp vừa và nhỏ (SMEs) gặp nhiều thách thức trong quản lý dữ liệu. SMEs thiếu năng lực kỹ thuật để số hóa và tối ưu hóa sử dụng dữ liệu hiện có, khó tiếp cận các nguồn dữ liệu về thị trường, dữ liệu từ các cơ quan Nhà nước, cũng như không chắc chắn về chi phí-lợi ích của việc áp dụng các giải pháp kỹ thuật số ở giai đoạn này. Do đó, nhiều doanh nghiệp vừa và nhỏ vẫn chưa hoàn toàn sẵn sàng về dữ liệu để ứng dụng AI nội bộ, chưa nói đến phát triển các giải pháp AI cho các bên.

Một số chuyên gia được hỏi cho rằng, dữ liệu ở Việt Nam có thể là nhiều, nhưng dữ liệu đã xử lý, có cấu trúc vẫn đang thiếu, về lâu dài vẫn cần những doanh nghiệp lớn để xử lý dữ liệu. Dữ liệu còn nằm rải rác và không được sử dụng nhiều để phân tích, đưa ra quyết định. Ví dụ, không có thông tin dữ liệu về ngành chuẩn, nên không đo đếm được nhu cầu việc làm, không dự đoán được thị trường CNTT, cho nên đào tạo không sát với thị trường. Hoặc là trong ngành du lịch, dịp lễ 2/9 ở Sapa bị quá tải, do thiếu dữ liệu và cơ sở hạ tầng để giúp cảnh báo sớm những trường hợp như vậy.

Bên cạnh đó, muốn làm AI ở Việt Nam cần phải dựa nhiều vào dữ liệu mở vì nếu dữ liệu đóng - độc quyền thì rất khó để cạnh tranh. Trong khi đó, khu vực công Việt Nam sở hữu một lượng lớn dữ liệu chất lượng cao, tính chất chuyên môn sâu, phạm vi bao phủ rộng, lớn hơn so với dữ liệu của doanh nghiệp, nên được khai thác để đào tạo cho AI. Tuy nhiên, dữ liệu mở của Chính phủ, dữ liệu mở quốc gia, mức độ chuẩn hóa dữ liệu ở Việt Nam còn thấp. Thiếu dữ liệu mở đủ lớn, đáp ứng chất lượng, nhất là dữ liệu mở từ các cơ quan Nhà nước là một điểm bất cập lớn, làm hạn chế khả năng tiếp cận dữ liệu của doanh nghiệp, các viện, trường, bản thân các cơ quan nhà nước để phát triển các mô hình, giải pháp, ứng dụng AI.

### **Tiềm năng khai thác dữ liệu của doanh nghiệp**

*Một doanh nghiệp cho biết, họ có toàn bộ dữ liệu công khai về đấu thầu và có thể nhìn dữ liệu ở cấp độ quốc gia, ở từng địa phương và theo ngành; có thể khai thác dữ liệu thực để làm báo cáo về xu hướng đầu tư của Chính phủ, phân bổ ngân sách vào các lĩnh vực. Trường hợp của doanh nghiệp này có thể là một ví dụ tốt về sử dụng kết hợp cả cloud trong nước và cloud nước ngoài. Cloud trong nước dùng để lưu trữ dữ liệu vì khách hàng ở Việt Nam truy cập đến máy chủ ở Việt Nam sẽ nhanh hơn truy cập ra quốc tế và yêu cầu lưu trữ dữ liệu ở Việt Nam; còn cloud nước ngoài nhờ có các ưu thế được dùng để tăng hiệu năng hệ thống và dự phòng (back up).*

*Dữ liệu từ hệ sinh thái của một doanh nghiệp cung cấp dịch vụ cảng biển giúp xác định vị trí của container, giờ lấy hàng giúp doanh nghiệp dễ nhận biết container và quyết định giờ nào nên lấy hàng để tránh ùn đứ (doanh nghiệp thường có tâm lý là hàng cập cảng là lấy hàng luôn nhưng trên thực tế là không cần thiết như vậy). Đồng thời, dữ liệu được chia sẻ với cơ quan nhà nước để phục vụ hoạt động quản lý, ví dụ chia sẻ cho cảnh sát giao thông để điều phối giao thông trong giờ cao điểm.*

(Nguồn: Các cuộc phỏng vấn sâu do IPS tiến hành năm 2023-2024).

<sup>120</sup> Scale. (2023). Zeitgeist: 2023 AI Readiness Report.

<https://go.scale.com/hubfs/Scale-Zeitgeist-AI-Readiness-Report-2023.pdf>

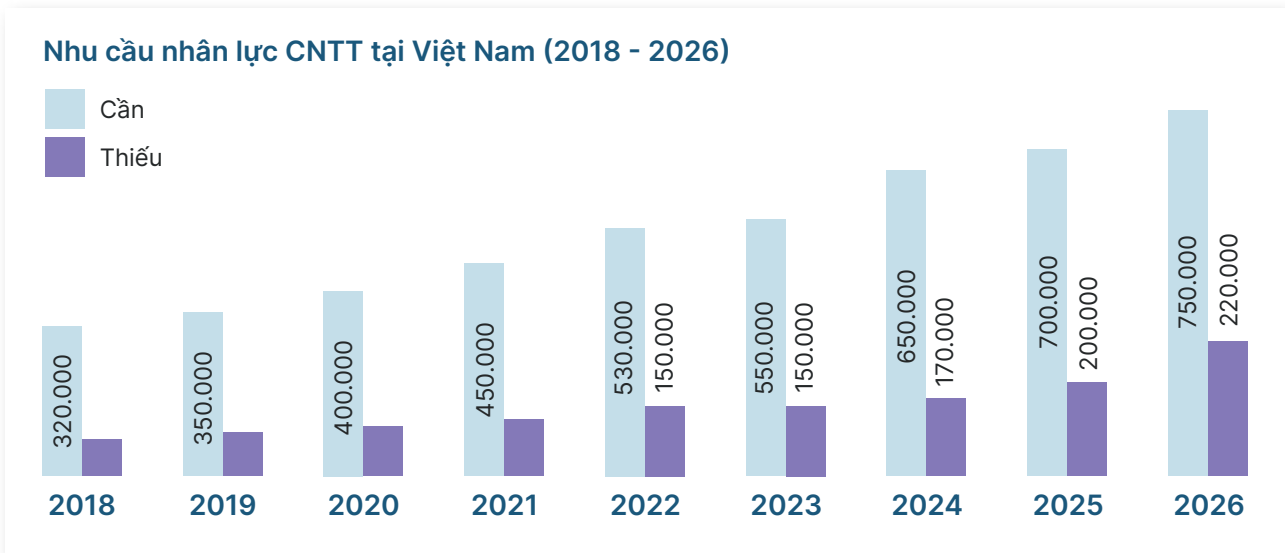
<sup>121</sup> Asia Pacific AI Readiness Index 2023, [salesforce\\_ai\\_readiness\\_index\\_2023.pdf](https://www.salesforce.com/content/dam/assets/ai-readiness-index-2023.pdf)

Việc tinh chỉnh các mô hình nền tảng (foundation models) yêu cầu bộ dữ liệu quy mô lớn, được tổ chức theo từng miền cụ thể, cùng với kho tri thức phong phú. Đối với các AI thế hệ trước, bằng cách tiếp cận theo hướng ứng dụng, doanh nghiệp vẫn có thể khai thác giá trị từ AI mà không cần hiện đại hóa kiến trúc dữ liệu. Tuy nhiên, điều này không còn phù hợp với AI tạo sinh. Các mô hình nền tảng cần một lượng lớn dữ liệu được quản lý và chuẩn hóa, do đó, giải quyết thách thức về dữ liệu trở thành ưu tiên cấp bách đối với mọi doanh nghiệp.

### 4.1.3. Nhân lực AI

Trong quá trình phát triển AI, nhân công giá rẻ với số lượng lớn không còn là lợi thế của Việt Nam, mà đòi hỏi nhân lực trình độ cao. Trên phương diện này, năng lực kỹ sư CNTT nói chung của Việt Nam trên phương diện nhất định được đánh giá khá cao, cụ thể như: theo khảo sát SkillValue, năng lực lập trình viên Việt Nam nằm trong TOP10 thế giới; theo UNESCO Statistics, Việt Nam nằm trong số 10 quốc gia đào tạo nhiều kỹ sư phần mềm nhất; hạng 6 trong số các quốc gia hàng đầu về gia công phần mềm<sup>122</sup>.

Lực lượng lao động CNTT ở Việt Nam vào năm 2024 khoảng 560.000, hàng năm có 50.000-60.000 sinh viên ra trường. Tuy nhiên, như biểu đồ dưới đây cho thấy, hàng năm Việt Nam vẫn thiếu hụt từ 150.000 đến 200.000 lập trình viên, kỹ sư có chuyên môn cao và các kỹ năng mềm, ngoại ngữ<sup>123</sup>.



Hình 9. Nhu cầu nhân lực CNTT

Theo một khảo sát, đa số các doanh nghiệp, chiếm 87,3% ưu tiên tuyển dụng nhân sự có kinh nghiệm, bằng cấp; chỉ có 5,4% nhà tuyển dụng sẵn lòng tuyển dụng ứng viên chưa có kinh nghiệm/không có nền tảng CNTT/IT<sup>124</sup>. Nhu cầu cấp thiết về chiến lược nhân sự trẻ nhằm giải quyết vấn đề thiếu hụt nhân tài trong lĩnh vực CNTT trong tương lai. Đối với doanh nghiệp công nghệ số, bài toán cần giải quyết là: Sinh viên học 4-5 năm nhưng vẫn cần đào tạo lại từ 6 tháng -1 năm tại doanh nghiệp mới đáp ứng được nhu cầu công việc; sinh viên ra trường chưa có định hướng rõ ràng; doanh nghiệp mất nhiều nỗ lực trong việc kèm cặp đào tạo cả chuyên môn và những kỹ năng mềm cơ bản nhất (viết email, làm việc nhóm, kỹ năng thuyết trình/trình bày,...); kỹ năng ngoại ngữ; kinh nghiệm làm việc, khả năng “thực chiến”. Chiến lược của doanh nghiệp: Hợp tác, đồng hành cùng các trường đại học tham gia thiết kế chương trình đào tạo, trang bị kỹ năng, huấn luyện thực chiến cho sinh viên từ năm thứ nhất.

Riêng về nhân lực cho AI, theo các chuyên gia, đại diện các doanh nghiệp trong lĩnh vực AI trao

<sup>122</sup> TopDev, Vietnam IT & Tech Talent Landscape Repo, 2024. Xem thêm: Đặng Trọng Hợp, Đào tạo nhân lực công nghệ số chất lượng cao đáp ứng nhu cầu của ngành công nghiệp công nghệ số, Định hướng xây dựng thể chế về công nghiệp công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.

<sup>123</sup> TopDev, Vietnam IT & Tech Talent Landscape Repo, 2024. Xem thêm: Phạm Thị Nhung, Nhu cầu nguồn nhân lực trong lĩnh vực công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.

<sup>124</sup> Dẫn theo: Phạm Thị Nhung, Nhu cầu nguồn nhân lực trong lĩnh vực công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.

đối với IPS, có thể xếp thành các nhóm sau: Các nhà khoa học AI có thể nghiên cứu, phát triển các mô hình AI; Kỹ sư hệ thống và kỹ sư lập trình; bộ phận back-end để duy trì lâu dài (nhóm nhân sự dán dữ liệu có thể là người khuyết tật, người tự kỉ). Tại Việt Nam, hiện có khoảng 2000 chuyên gia AI làm việc trong nước, chủ yếu sống tại Hà Nội, Thành phố Hồ Chí Minh và Đà Nẵng, nơi các công ty liên quan đến AI đặt trụ sở chính<sup>125</sup>. Chính phủ đang đặt ra mục tiêu đào tạo ít nhất 5.000 kỹ sư có chuyên môn AI vào năm 2030. Ngoài ra, Trung tâm Đổi mới sáng tạo quốc gia (NIC) có nhiệm vụ thành lập một trung tâm đào tạo, nghiên cứu và ứng dụng AI, với mục tiêu đào tạo 7.000 chuyên gia AI và hỗ trợ khoảng 500 công ty khởi nghiệp AI vào năm 2030. Mặc dù vậy, con số này còn có khoảng cách xa so với nhu cầu 100.000 nhân sự AI trong 5 năm tới ở Việt Nam<sup>126</sup>.

Theo ý kiến của những người tham gia khảo sát của IPS, các nhà khoa học về AI còn thiếu nhiều ở Việt Nam. Còn kỹ sư AI có thể tìm được, nhưng thiếu người có kinh nghiệm và làm tốt; đặc biệt là kỹ sư khoảng 40 tuổi đến 50 tuổi rất hiếm, là độ tuổi đã tích lũy được nhiều kiến thức, kỹ năng giỏi để truyền lại cho thế hệ sau. Việt Nam có số lượng du học sinh học về AI đáng kỳ vọng, nhưng thiếu kinh nghiệm thực tế. Mặt khác, kỹ sư AI ở Việt Nam hiện nay phần lớn có độ tuổi trẻ và có thể trả lương vừa phải, cho nên khách hàng nước ngoài sẵn sàng đầu tư thêm thời gian để phát triển sản phẩm, thay vì đầu tư vào kỹ sư ở các nước như Mỹ, Pháp.

Hơn nữa, hiện nay, các công ty đã tuyển dụng dựa trên kỹ năng của ứng viên thay vì chỉ tập trung vào bằng cấp, lịch sử công việc hoặc chức danh công việc trước đây của họ. Cách tiếp cận “Ưu tiên kỹ năng trước” (skills-first approach) tập trung xem xét một người có các kỹ năng, năng lực phù hợp đối với một vị trí, vai trò nào đó<sup>127</sup>. Điều này giúp các công ty tìm được những người lao động họ cần và giúp người lao động dễ dàng thích nghi hơn với thị trường việc làm thay đổi nhanh chóng. Mặt khác, xu hướng này đặt ra thách thức, bài toán cho các trường đại học, doanh nghiệp cần phải giải trong đào tạo nhân lực liên quan đến AI.

Đồng thời với những kỹ năng đặc thù liên quan đến AI, người lao động trong kỷ nguyên AI rất cần đến những kỹ năng mềm khác như kỹ năng phân tích. Kỹ năng phân tích, bao gồm các khả năng như quản lý dự án, tư duy phản biện và giải quyết vấn đề, được nhân viên Việt Nam coi là quan trọng nhất (62%), cao hơn mức trung bình của khu vực (49%). Các chuyên gia công nghệ được hỏi cho biết, giải quyết vấn đề và tư duy thuật toán đang trở nên quan trọng hơn, vì AI có thể giúp lập trình, nhưng lập trình viên cần các kỹ năng phân tích như cách tiếp cận vấn đề, phân tích vấn đề, thiết kế nội dung cần lập trình.

Thực tế hoạt động của doanh nghiệp cho thấy, năm 2024 đã có sự chuyển dịch rõ rệt trong yêu cầu kỹ năng, khi nhiều doanh nghiệp cần nhân lực có năng lực phát triển toàn diện (full-stack) và chuyên sâu vào các công nghệ mới như: Khả năng xây dựng hoặc ứng dụng các mô hình học máy và sử dụng các framework; hiểu biết và kinh nghiệm với các dịch vụ cloud của AWS, Google Cloud, Azure; nắm vững các công cụ và phương pháp phát hiện, ngăn chặn tấn công, đặc biệt là trong bảo mật hệ thống đám mây<sup>128</sup>. Đây đều là những kiến thức, kỹ năng liên quan đến phát triển, ứng dụng AI.

Đồng thời, các ngành không phải là CNTT đang thiếu nhân sự có kiến thức, kỹ năng biết cách ứng dụng AI vào quản trị doanh nghiệp, kinh doanh, cung cấp dịch vụ. Để ứng dụng được AI trong các ngành, đòi hỏi phải có kiến thức về AI, khoa học máy tính, khoa học dữ liệu và kiến thức chuyên ngành như kinh doanh, y khoa, luật. Theo khảo sát doanh nghiệp gia đình ở Việt Nam năm 2024, có sự đồng thuận đáng kể về việc GenAI sẽ là yếu tố thúc đẩy lực lượng lao động của họ phát triển các kỹ năng mới (73%); các nhân viên cần phải phát triển các kỹ năng cần thiết để khai thác tối đa lợi ích từ công nghệ.

Trên thực tế, các doanh nghiệp đang đầu tư dưới mức đáng kể trong việc trang bị cho người lao động những kiến thức và kỹ năng cần thiết để làm việc cùng AI. Bên cạnh đó, theo một chuyên gia AI hiện đang là một chủ doanh nghiệp cung cấp giải pháp AI, tầm nhìn của doanh nghiệp Việt

<sup>125</sup> Hien Thu Pham et al., *Artificial intelligence (AI) development in the Vietnam's energy and economic systems: A critical review*, *Journal of Cleaner Production* 438 (2024) 140692.

<sup>126</sup> <https://opengovasia.com/2024/08/20/vietnams-ambition-to-lead-ai-innovation-in-southeast-asia/#:~:text=The%20government%20is%20proposing%20a,500%20AI%20startups%20by%202030.>

<sup>127</sup> World Economic Forum, *Putting Skills First: Opportunities for Building Efficient and Equitable Labour Markets*, 2024.

<sup>128</sup> Phạm Thị Nhung, *Nhu cầu nguồn nhân lực trong lĩnh vực công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số*, Hà Nội, 15/11/2024.

Nam về phát triển sản phẩm dựa trên còn hạn chế. Nhân sự AI ở các ngành hiện hữu không cần có kĩ năng về AI sẵn, mà chú trọng vào tư duy ứng dụng AI. Trong khi đó, sinh viên Việt Nam có khoảng cách xa về tư duy so với sinh viên nước ngoài, mới chỉ mạnh về kỹ năng kỹ thuật, chỉ giúp sinh viên Việt Nam ra trường đi làm thuê.

#### 4.1.4. Giáo dục, đào tạo; nghiên cứu khoa học về AI

Nói chung, giáo dục đào tạo các ngành CNTT ở Việt Nam có xu hướng tăng lên. Trong giai đoạn 2019-2023, số lượng sinh viên nhập học mới các ngành thuộc lĩnh vực Máy tính và công nghệ thông tin tăng 1,65 lần; tổng số sinh viên nhập học mới các ngành thuộc lĩnh vực công nghệ số khoảng 90 nghìn, chiếm 14,5%. Tại thời điểm tháng 04/2024, trong tổng số 1248 chương trình đào tạo đã được kiểm định theo tiêu chuẩn trong nước thì khối STEM chiếm 312 chương trình (25%); trong 519 chương trình đào tạo được kiểm định theo tiêu chuẩn khu vực hoặc quốc tế thì khối STEM chiếm 272 chương trình (52,4%); chương trình đào tạo chất lượng cao: chiếm 208/430, quy mô khoảng 14.000/40.000 sinh viên<sup>129</sup>.

Các ngành đào tạo về công nghệ số	Tuyển mới 2023		Đang học 2023		Tốt nghiệp 2023	
	Số lượng	%	Số lượng	%	Số lượng	%
Máy tính và CNTT	66.599	10.8	231.385	10.5	50.172	10.5
Các ngành kỹ thuật liên quan	10.481	1.7	45.987	2.1	9.976	2.1
Cách ngành Công nghệ kỹ thuật liên quan	11.23	1.8	40.163	1.8	7.924	1.7
Các ngành Toán và Thống kê liên quan	1.588	0.3	4.762	0.2	905	0.2
<b>Tất cả</b>	<b>89.691</b>	<b>14.5</b>	<b>89.691</b>	<b>14.6</b>	<b>68.977</b>	<b>14.4</b>

**Hình 10. Tình hình đào tạo đại học ngành CNTT tại thời điểm 2023**

Về giáo dục - đào tạo AI nói riêng, hiện tại Việt Nam đã có những trường đại học thuộc khối CNTT đã có các chương trình nhằm đáp ứng sự phát triển của công nghệ AI như đào tạo khoa học máy tính, kho học dữ liệu, trí tuệ nhân tạo. Nổi bật là Đại học Bách khoa Hà Nội đã công bố triển khai chương trình đào tạo chuyên ngành Trí tuệ nhân tạo tạo sinh (GenAI) kéo dài 1,5 năm, gồm 48 tín chỉ. Chỉ những sinh viên có bằng cử nhân Công nghệ thông tin và truyền thông mới có thể nộp đơn. Sinh viên sẽ được cung cấp kiến thức nền tảng về AI và kiến thức chuyên sâu về GenAI, chẳng hạn như: học máy, học sâu; vận hành, triển khai và giám sát các mô hình học máy; xây dựng các ứng dụng GenAI; tạo văn bản, hình ảnh, video, âm thanh; áp dụng GenAI trong nhiều lĩnh vực khác nhau; đạo đức AI và xây dựng trí tuệ nhân tạo đáng tin cậy<sup>130</sup>. Bên cạnh đó, một số trường đại học khác không thuộc khối CNTT đã đưa AI vào chương trình đào tạo của mình, như Đại học Ngoại thương đang xây dựng chương trình kết hợp khoa học máy tính, khoa học dữ liệu và kinh doanh ở bậc cử nhân, tạo ra lực lượng lao động mới.

Tại Việt Nam, chỉ một số ít trường đại học có các hoạt động nghiên cứu trong lĩnh vực AI, chủ yếu là các trường đại học kỹ thuật, công nghệ. Mặc dù các công nghệ AI được áp dụng để giải quyết các vấn đề ở các ngành, nhưng nghiên cứu AI chưa thu hút được sự chú ý của các trường đại học kinh tế và kinh doanh tại Việt Nam. Trong số này, Đại học Quốc gia Thành phố Hồ Chí Minh dẫn đầu về số lượng ấn phẩm trong tất cả các năm từ 2010 đến 2021, đạt 170 vào năm 2021. Đại học Khoa học và Công nghệ Hà Nội, Đại học Tôn Đức Thắng và Đại học Duy Tân đã theo sau với 60-80

<sup>129</sup> Đặng Trọng Hợp, Đào tạo nhân lực công nghệ số chất lượng cao đáp ứng nhu cầu của ngành công nghiệp công nghệ số, Định hướng xây dựng thể chế về công nghiệp công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.

<sup>130</sup> Hanoi university Vietnam's first to train GenAI engineers, VnExpress, 18/6/2024, <https://e.vnexpress.net/news/news/education/hanoi-university-vietnam-s-first-to-train-genai-engineers-4759305.html>



ấn phẩm mỗi năm. Các trường đại học tập trung vào kinh tế, tài chính và kinh doanh như Đại học Kinh tế Thành phố Hồ Chí Minh và Đại học Ngoại thương, có số lượng tương đối ít các ấn phẩm về AI trong giai đoạn trên. Về các ấn phẩm trên các tạp chí có hệ số tác động lớn hơn 1,5, Đại học Duy Tân dẫn đầu với khoảng 50 ấn phẩm mỗi năm<sup>131</sup>.

So với các quốc gia khác trong khu vực Đông Nam Á, các chỉ số liên quan đến AI như đầu tư và nghiên cứu của Việt Nam thấp hơn nhiều so với Singapore, Malaysia, Thái Lan và Indonesia. Mục tiêu lọt vào top 5 ở Đông Nam Á vào năm 2025 có thể khả thi, nhưng cần bước tiến rất lớn để đảm bảo vị trí top 4 khu vực vào năm 2030 với mức đầu tư hiện tại ở Singapore, Malaysia, Thái Lan và Indonesia.

Ở cấp độ doanh nghiệp, theo những người được phỏng vấn thuộc nghiên cứu này, nhân sự hiện tại của nhiều doanh nghiệp Việt Nam chưa được đào tạo bài bản về AI (kiến thức, kỹ năng về AI của họ có thường do yêu thích công nghệ, hoặc đáp ứng do làm trong môi trường cần AI). Bên cạnh đó, đầu tư vào nghiên cứu và phát triển (R&D) của các doanh nghiệp Việt Nam còn yếu. Thường R&D để làm ra các sản phẩm, dịch vụ có giá trị thương mại hóa cao nằm ở các doanh nghiệp, nhưng doanh nghiệp Việt Nam có mức độ phát triển công nghiệp thấp nên chưa quan tâm đến R&D.

Trên phương diện đầu tư vào R&D cho AI, Trung tâm Nghiên cứu và đào tạo AI và mạch tích hợp (IC) tại Đà Nẵng là trung tâm đầu tiên ở Việt Nam, được hy vọng góp phần thu hút các nhà nghiên cứu, nhà đổi mới và nhà đầu tư chất lượng cao về AI. Điều quan trọng là tạo ra một mạng lưới phối hợp hơn của các trung tâm nghiên cứu như vậy và thúc đẩy sự phối hợp hiệu quả giữa giới nghiên cứu và các doanh nghiệp. Việc đầu tư vào năng lực tính toán AI cũng rất cần thiết cho R&D, để đào tạo và chạy các mô hình AI. Đại học FPT là đơn vị đầu tiên sử dụng siêu máy tính để nghiên cứu AI tại Việt Nam, giúp FPT đẩy nhanh quá trình nghiên cứu tập trung vào xử lý hình ảnh và ngôn ngữ trên nền tảng FPT.AI và phát triển sản phẩm. Kế hoạch của FPT và Nvidia mở AI Factory với hệ thống siêu máy tính sẽ giúp tăng cường năng lực R&D AI của Việt Nam.

Tuy nhiên, các doanh nghiệp, trường đại học có nguồn lực như FPT để đầu tư mạnh vào R&D không nhiều ở Việt Nam. Trong khi đó, còn thiếu sự hỗ trợ của Chính phủ về hạ tầng tính toán, dữ liệu cho đào tạo, cho R&D đối với AI. Hơn nữa, các chuyên gia được phỏng vấn đều cho biết, quy trình thủ tục phê duyệt đề tài nghiên cứu của nhà nước hiện nay phức tạp, kéo dài, khiến các phòng thí nghiệm kỹ thuật ở trường đại học không bắt kịp với tốc độ phát triển của công nghệ trên thế giới. Trên thực tế, việc phân bổ và sử dụng ngân sách nhà nước, cấp kinh phí cho các đề tài, dự án nghiên cứu khoa học và công nghệ được thực hiện tương tự như đối với dự án đầu tư xây dựng cơ bản. Muốn nghiên cứu phải lập dự toán, lập thuyết minh, phải được các cấp có thẩm quyền thẩm định, phê duyệt, sau đó được lập thành danh mục, tổng hợp và đưa vào dự toán ngân sách nhà nước hàng năm thì năm sau mới được cấp kinh phí.

Mặc dù cách tiếp cận này nhằm mục đích đảm bảo giảm thiểu thất thoát ngân sách Nhà nước, tránh đầu tư vào các dự án nghiên cứu kém khả thi, nhưng cũng hạn chế nghiên cứu “thử và sai”. Trong khi đó, nghiên cứu “thử và sai” là nền tảng cho việc thử nghiệm và khám phá, một bước quan trọng trong việc phát triển và thử nghiệm các sản phẩm công nghệ, đặc biệt là các sản phẩm AI. Trên thế giới, các công nghệ AI đột phá, chẳng hạn như ChatGPT, đã trải qua các giai đoạn thử nghiệm và thất bại, thường được gọi là "giai đoạn đốt tiền". Mặc dù tốn kém nhưng giai đoạn này rất cần thiết để thúc đẩy đổi mới, khuyến khích sáng tạo và đạt được các hệ thống AI hiệu suất cao.

Hơn nữa, đầu tư cho việc nghiên cứu và phát triển các lĩnh vực khoa học, công nghệ và đổi mới sáng tạo chủ chốt, trong đó có AI ở Việt Nam còn hạn hẹp. Tổng chi cho nghiên cứu và phát triển của Việt Nam chỉ đạt khoảng 0.67% GDP<sup>132</sup>. Mặc dù có sự gia tăng so với những năm trước đó, (0.42% vào năm 2021 với tỷ lệ chi từ doanh nghiệp vào khoảng 62%)<sup>133</sup>, nhưng con số này vẫn còn hạn chế nếu so sánh với mức trung bình của thế giới là 1,93%, mà chưa so với các nước phát triển với trình độ khoa học, công nghệ cao là từ 3-5% GDP<sup>134</sup>.

<sup>131</sup> Hien Thu Pham et al., *Artificial intelligence (AI) development in the Vietnam's energy and economic systems: A critical review*, *Journal of Cleaner Production* 438 (2024) 140692.

<sup>132</sup> Vương Trần, “Bố trí ít nhất 15% ngân sách chi cho sự nghiệp khoa học,” *Báo Lao Động*, tháng 1-2025, truy cập tại <https://laodong.vn/thoi-su/-bo-tri-it-nhat-15-ngan-sach-chi-cho-su-nghiep-khoa-hoc-1449258.ldo>.

<sup>133</sup> *Sách Khoa học và công nghệ thế giới 2023, Cục Thông tin Khoa học và công nghệ quốc gia, Bộ Khoa học và công nghệ*, tr.20.

<sup>134</sup> Vương Trần, “Bố trí ít nhất 15% ngân sách chi cho sự nghiệp khoa học,” *Báo Lao Động*, tháng 1-2025, truy cập tại <https://laodong.vn/thoi-su/-bo-tri-it-nhat-15-ngan-sach-chi-cho-su-nghiep-khoa-hoc-1449258.ldo>.

#### 4.1.5. Nguồn tài chính và cơ chế tài chính cho AI

Theo ý kiến của các chuyên gia, nhà nước sẽ là bên “mua” lớn nhất các sản phẩm, dịch vụ dựa trên AI. Trong khi đó, nguồn tài chính, cơ chế tài chính cho AI hiện nay từ ngân sách nhà nước còn gặp nhiều vướng mắc. Đầu tư công cho AI được thực hiện theo cơ chế tài chính chung đối với công nghệ thông tin theo 3 nhóm lớn: Phần cứng; dịch vụ; và phần mềm. Nhóm phần cứng & nhóm dịch vụ có thể định lượng được (ví dụ: thuê đường truyền internet, thuê dịch vụ lưu trữ dữ liệu có thể tính bằng dung lượng băng thông bit/s; dung lượng lưu trữ byte/đơn vị thời gian) và không gặp khó khăn khi dự toán, quyết toán. Phần mềm về bản chất là tài sản trí tuệ vô hình, khó định giá do liên quan đến yếu tố “chất xám”, liên quan đến cung – cầu trên thị trường. Nhất là từ khi điện toán đám mây trở nên phổ biến, phương thức cung cấp dịch vụ SaaS (software as a service) tạo ra thay đổi lớn, theo đó phần mềm triển khai trên đám mây, được cung cấp theo tài khoản, mà không “đóng gói” trong đĩa cứng và cài đặt ở máy tính cá nhân. Nhưng phương thức thuê dịch vụ qua tài khoản thì phần giá định giá lại phụ thuộc vào thị trường (mức độ cạnh tranh của các bên cung cấp dịch vụ: nhiều hay ít nhà cung cấp; sản phẩm phổ biến hay sản phẩm chuyên dụng).

Vì vậy, tiềm ẩn rủi ro lớn khi cơ quan nhà nước mua sắm gói thầu phần mềm, do các đơn vị cung cấp dịch vụ chào với các mức giá khác nhau; nếu bị thanh tra, kiểm tra thì khó giải trình; dẫn đến cơ quan nhà nước sợ sai, không dám mua sắm. Đồng thời tình trạng này cũng có thể tạo ra thất thoát, lợi dụng sự không thống nhất để móc ngoặc, trục lợi. Trên thực tế, trong 6 tháng đầu năm 2024, một thành phố lớn không giải ngân được ngân sách cho chuyển đổi số. Nhiều địa phương xây dựng Trung tâm điều hành thông minh, và chỉ thanh toán được một phần chi phí (cho dịch vụ thuê đường truyền dữ liệu), còn dịch vụ khác không thanh toán được.

#### **Ngành y tế khó chi tiền cho công nghệ**

*Các cơ sở y tế phải đầu tư ban đầu rất lớn cho hạ tầng công nghệ, điều kiện cần và rất quan trọng để ứng dụng AI trong khám, chữa bệnh; cần phải huy động từ nhiều nguồn lực mới có khả năng cân đối chi phí. Tuy nhiên, hiện tại vẫn chưa có hướng dẫn cụ thể về cơ chế tài chính cho ứng dụng công nghệ y tế nói chung và triển khai, ứng dụng AI nói riêng. Cụ thể, còn thiếu các quy định về giá dịch vụ khám bệnh, chữa bệnh bao gồm chi phí CNTT; chưa có mục chi cho công nghệ, ví dụ chi trả kinh phí khi ứng dụng phần mềm lưu trữ và truyền tải hình ảnh y tế (PACS) thay cho in phim; trong giá thành dịch vụ y tế chưa có thành phần công nghệ; các bệnh viện chủ yếu sử dụng quỹ đầu tư phát triển của đơn vị để ứng dụng công nghệ khiến cho các hoạt động khác cũng chịu ảnh hưởng.*

Đặc biệt, do bản chất đặc thù và độ phức tạp của các ứng dụng AI, khiến việc đảm bảo nguồn vốn cần thiết cho các dự án AI trở nên cấp thiết trong giai đoạn nghiên cứu, thiết kế và thử nghiệm. Tuy nhiên, hiện chưa có căn cứ, hướng dẫn cụ thể nào về lập dự toán, quản lý, sử dụng, quyết toán kinh phí từ ngân sách Nhà nước đối với việc nghiên cứu, xây dựng, ứng dụng các mô hình, hệ thống AI. Nhất là việc phát triển các mô hình nền tảng khác với phần mềm, thường không thể xác định chính xác các chức năng và độ phức tạp cụ thể của chúng cho mục đích định giá. Hoặc là khó có thể lập dự toán, chi cho việc tổng hợp, làm sạch dữ liệu, sử dụng dữ liệu để huấn luyện AI. Điều này sẽ gây nhiều vướng mắc trong giai đoạn nghiên cứu và phát triển, nơi tư duy định giá chưa được cập nhật kịp thời để phản ánh bản chất của các nghiên cứu liên quan đến AI.

Về đầu tư của tư nhân vào AI, theo khảo sát của GenAI Fund, trong các nguồn tài chính của 250 công ty khởi nghiệp GenAI tại ASEAN, 50% là do tự lực hoặc được tài trợ bởi các nhà đầu tư thiên thần (angel-funded), 41% đã đảm bảo được nguồn tài trợ tiền hạt giống (pre-seed funding) hoặc hạt giống (seed funding). Đây là cơ hội đầu tư tốt do định giá thấp hơn so với các đối tác Hoa Kỳ, tập trung vào các mô hình kinh doanh bền vững và tiềm năng tận dụng lợi thế đi sau của ASEAN trong việc áp dụng AI<sup>135</sup>. Với khoảng 70 trong số 250 công ty này có trụ sở ở Việt Nam, những con số đó cho thấy những hàm ý chính sách để thúc đẩy, phát triển các nguồn tài chính cho doanh nghiệp khởi nghiệp AI ở Việt Nam.

<sup>135</sup> GenAI Fund, ASEAN GenAI Startup Report 2024, 9/2024.

## 4.2. CHÍNH SÁCH, PHÁP LUẬT VỀ AI Ở VIỆT NAM

Sự phát triển của AI đã đi nhanh hơn các quy định và hướng dẫn đạo đức về AI trên toàn thế giới. Tương tự, hệ thống chính sách, pháp luật của Việt Nam vẫn chưa theo kịp nhu cầu và thực tế triển khai, ứng dụng AI để đảm bảo sự phát triển và vận hành suôn sẻ của công nghệ này.

### 4.2.1. Chính sách, pháp luật hiện có về AI

Trong những năm qua, Đảng Cộng sản Việt Nam và Chính phủ Việt Nam đã ban hành các văn bản chính sách mang tính định hướng không chỉ trực tiếp về AI, mà còn về các nội dung liên quan đến AI. Đó là các văn bản như: (1) Nghị quyết số 52-NQ/TW về một số chủ trương, chính sách chủ động tham gia cuộc Cách mạng Công nghiệp lần thứ 4, (2) Quyết định số 127/QĐ-TTg về Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030, (3) Quyết định số 569/QĐ-TTg về việc Ban hành chiến lược phát triển khoa học, công nghệ và đổi mới sáng tạo đến năm 2030, (4) Nghị quyết số 57-NQ/TW về Đột phá phát triển Khoa học, Công nghệ, Đổi mới sáng tạo và Chuyển đổi số quốc gia, (5) Nghị quyết số 03/NQ-CP về việc ban hành Chương trình hành động của Chính phủ thực hiện Nghị quyết số 57-NQ/TW. Báo cáo này trình bày cụ thể hơn về Nghị quyết 57-NQ/TW và Quyết định 127/QĐ-TTg.

**Nghị quyết 57-NQ/TW:** Cuối năm 2024, Đảng Cộng sản Việt Nam đã ban hành Nghị quyết 57-NQ/TW về đột phá phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia (Nghị quyết 57). Nghị quyết đặt mục tiêu Việt Nam từng bước làm chủ một số công nghệ chiến lược, trong đó ưu tiên đầu tư phát triển trí tuệ nhân tạo; đến năm 2030 nằm trong nhóm 3 nước dẫn đầu khu vực Đông Nam Á về nghiên cứu và phát triển trí tuệ nhân tạo. Có thể hiểu nội hàm của “làm chủ” là Việt Nam chuyển từ người sử dụng sang người sáng tạo trong lĩnh vực AI; không chỉ đáp ứng nhu cầu nội địa, mà còn xuất khẩu các giải pháp công nghệ cao, nâng cao năng lực cạnh tranh quốc gia. Các yếu tố “làm chủ” gồm: (1) các hoạt động nghiên cứu và phát triển AI, từ các công nghệ cốt lõi như học máy (machine learning), xử lý ngôn ngữ tự nhiên (NLP) đến các giải pháp cụ thể; (2) Hạ tầng AI, bao gồm trung tâm dữ liệu, siêu máy tính và các phòng thí nghiệm chuyên sâu sẽ được xây dựng và nâng cấp; (3) hệ sinh thái AI nội địa, nơi các viện nghiên cứu, trường đại học và doanh nghiệp công nghệ hợp tác chặt chẽ; trong đó đáng chú ý là mục tiêu xây dựng một số trường, trung tâm đào tạo tiên tiến chuyên sâu về trí tuệ nhân tạo; (4) các startup AI trong nước phát triển mạnh mẽ hơn. Đồng thời, Nghị quyết 57 đưa ra những chính sách lớn để tạo đột phá trong khoa học, công nghệ nói chung, trong đó có những điều kiện để nghiên cứu, phát triển, ứng dụng AI gồm cơ sở hạ tầng; dữ liệu; nguồn nhân lực; giáo dục; nguồn lực, cơ chế cho nghiên cứu khoa học; đầu tư, tài chính cho khoa học, công nghệ.

**Chiến lược quốc gia về AI:** Tại thời điểm 2024, trong số những văn bản chính sách đã ban hành trực tiếp hoặc gián tiếp có nội dung điều chỉnh AI ở Việt Nam, nổi bật nhất là Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030. Chiến lược đặt ra một số mục tiêu và định hướng chính để phát triển công nghệ AI, đưa Việt Nam trở thành một quốc gia phát triển công nghệ, tận dụng những thay đổi về công nghệ để phát triển kinh tế.

**Các chính sách có liên quan:** Bên cạnh những văn bản nói trên, Việt Nam hiện đã có những văn bản khác của Đảng, Chính phủ có những nội dung có liên quan, tạo cơ sở cho phát triển, ứng dụng AI như Chương trình chuyển đổi số quốc gia 2021 – 2025, tầm nhìn đến 2025; Chiến lược quốc gia về dữ liệu.

### Khuôn khổ pháp luật liên quan đến AI

**Dự thảo luật Công nghiệp công nghệ số:** Dự thảo Luật Công nghiệp công nghệ số được thảo luận tại kỳ họp Quốc hội vào cuối tháng 11/2024, dự kiến thông qua vào kỳ họp giữa năm 2025. Luật theo cách tiếp cận kiểm soát dựa trên rủi ro để phát triển và triển khai AI một cách an toàn và có trách nhiệm. Tại thời điểm tháng 1/2025, dự thảo Luật nêu một số nguyên tắc trong phát triển, cung cấp, sử dụng AI; phân loại hệ thống AI có rủi ro cao và tác động cao; trách nhiệm của các bên liên quan đến hoạt động phát triển, cung cấp, triển khai và sử dụng hệ thống AI. Dự thảo yêu cầu sản phẩm công nghệ số tạo ra bởi hệ thống AI phải có dấu hiệu nhận dạng rõ ràng về điều này. Dự thảo giao cho Chính phủ quy định chi tiết về tiêu chí, thẩm quyền, trình tự, thủ tục xác định hệ thống AI rủi ro cao, hệ thống AI có khả năng tác động cao; trách nhiệm của tổ chức, cá nhân có liên quan đến phát triển, cung cấp và sử dụng hệ thống AI.

**Dự luật có những chính sách tạo điều kiện để phát triển, ứng dụng AI như:** ưu đãi cho các viện, trung tâm đổi mới sáng tạo, nghiên cứu và phát triển công nghệ số (trong đó có AI); ưu đãi cho nhân tài công nghệ số, thuế thu nhập cá nhân cho chuyên gia AI, chuyên gia công nghệ người nước ngoài; đặc biệt ưu đãi về đầu tư, thuế, đất đai đối với sản xuất sản phẩm công nghệ số ứng dụng AI, xây dựng trung tâm dữ liệu AI; dự án xây dựng trung tâm dữ liệu trí tuệ nhân tạo có quy mô vốn đầu tư từ 6.000 tỷ đồng trở lên, thực hiện giải ngân tối thiểu 3.000 tỷ đồng trong thời hạn 03 năm kể từ ngày được cấp Giấy chứng nhận đăng ký đầu tư hoặc chấp thuận chủ trương đầu tư thuộc đối tượng áp dụng ưu đãi và hỗ trợ đầu tư đặc biệt theo quy định của Luật Đầu tư.

**Các nguyên tắc về đạo đức AI:** Vào giữa năm 2024, Bộ trưởng Bộ Khoa học và công nghệ ban hành Quyết định 1290/QĐ-BKHHCN hướng dẫn về các nguyên tắc nghiên cứu và phát triển hệ thống trí tuệ nhân tạo có trách nhiệm; có tính chất tự nguyện, không phải bắt buộc; có thể cập nhật thường xuyên. Quyết định nhấn mạnh cách tiếp cận lấy con người làm trung tâm, cân bằng lợi ích kinh tế của AI với các cân nhắc về đạo đức và tuân thủ pháp luật. Các nguyên tắc chính được nêu trong Hướng dẫn kèm theo Quyết định gồm: Khuyến khích hợp tác giữa các nhà phát triển AI; các hệ thống AI phải minh bạch và các quyết định của chúng phải có thể giải thích được; khả năng kiểm soát, an toàn và bảo mật; bảo vệ quyền riêng tư của cá nhân, không phân biệt đối xử; trách nhiệm giải trình và hỗ trợ người dùng. Nhìn chung, văn bản hướng dẫn này nhằm xây dựng một hệ sinh thái AI có trách nhiệm và đạo đức tại Việt Nam, nơi các công nghệ AI được nghiên cứu, phát triển để mang lại lợi ích cho xã hội, đồng thời giảm thiểu các rủi ro và tác hại tiềm ẩn.

**Tiêu chuẩn kỹ thuật về AI:** Hiện nay Việt Nam đã xây dựng một số bộ tiêu chuẩn, quy định kỹ thuật cho các sản phẩm và công nghệ AI, cụ thể là: Tiêu chuẩn quốc gia TCVN 13902:2023 (ISO/IEC 22989:2022) về Công nghệ thông tin - Trí tuệ nhân tạo<sup>136</sup>; Tiêu chuẩn quốc gia TCVN 13903:2023 (ISO/IEC TR 24028:2020) về Công nghệ thông tin - Trí tuệ nhân tạo - Tổng quan tính đáng tin cậy trong trí tuệ nhân tạo<sup>137</sup>. Tuy nhiên, nếu so với yêu cầu, vẫn còn nhiều bộ tiêu chuẩn kỹ thuật về chưa được ban hành ở Việt Nam. Ủy ban Kỹ thuật chung về Trí tuệ nhân tạo của ISO và IEC đã công bố 11 Bộ tiêu chuẩn ISO và đang xây dựng thêm 26 Bộ tiêu chuẩn trong lĩnh vực AI, bao gồm cả việc xử lý sự thiên vị, cũng như các mối quan tâm về đạo đức và xã hội liên quan.

#### 4.2.2. Khoảng cách giữa chính sách và nhu cầu thực tế

Việt Nam vẫn đang trong những bước đầu xây dựng khung chính sách, pháp luật và quy tắc ứng xử đối với AI, cho nên các chính sách hiện có đang còn khoảng cách lớn so với yêu cầu. Khoảng cách này có thể làm cho các hệ thống AI được triển khai mà không có sự giám sát đầy đủ, làm tăng khả năng rủi ro, nguy cơ của AI đối với xã hội, con người.

##### (a) Hạn chế của các chính sách hiện có

Mặc dù một số văn bản đã ban hành hoặc đang xây dựng, nhưng khuôn khổ chính sách, pháp luật về AI của Việt Nam vẫn còn hạn chế so với nhu cầu thực tế. Chiến lược quốc gia về AI của Việt Nam còn chung chung, không có các nội dung chi tiết hơn để tạo ra lộ trình rõ ràng phía trước. Chiến lược chưa xác định rõ trọng tâm là nghiên cứu, phát triển các mô hình AI hay ứng dụng chúng, hay cả hai; các ngành cần ưu tiên phát triển, ứng dụng AI. Chiến lược hiện tại nhấn mạnh toàn diện đến cả AI lý thuyết và nghiên cứu ứng dụng, mặc dù không sai, nhưng có thể không phải là cách tiếp cận tối ưu nhất cho Việt Nam trong bối cảnh hiện tại. Là một quốc gia tầm trung ở khu vực Châu Á - Thái Bình Dương với nguồn lực tài chính và nhân lực hạn chế, có thể không thực tế khi áp dụng một cách tiếp cận toàn diện đối với AI, giải quyết cả nghiên cứu lý thuyết và ứng dụng trải dài trên nhiều lĩnh vực kinh tế. Điều này càng trở nên rõ ràng hơn trong bối cảnh AI vừa có những bước phát triển nhanh chóng, mạnh mẽ trong 2 năm qua như AI tạo sinh, tác nhân AI (AI Agents).

Bên cạnh đó, trong khi đề ra nhiệm vụ cho 16 bộ và Viện Hàn lâm khoa học và công nghệ Việt Nam cần làm trong 10 năm tới, Quyết định số 127/QĐ-TTg hoàn toàn không đề cập ngoại giao công nghệ vốn đóng vai trò rất quan trọng trong bối cảnh toàn cầu hóa, hội nhập hoàn toàn, và cạnh tranh công nghệ giữa các quốc gia. Trong khi giao nhiệm vụ cho các bộ, ngành, Viện Hàn lâm khoa học tự nhiên Việt Nam, Chiến lược không nêu vai trò, nhiệm vụ của Bộ Ngoại giao là đầu mối không thể thiếu trong việc kết nối, hợp tác về công nghệ giữa chính phủ Việt Nam với chính phủ

<sup>136</sup> <https://thuvienphaplu->

<at.vn/tcvn/Linh-vuc-khac/TCVN-13902-2023-ISO-IEC-22989-2022-Cong-nghe-thong-tin-Tri-tue-nhan-tao-920969.aspx>

<sup>137</sup> <https://luatvietnam.vn/thong-tin/tieu-chuan-quoc-gia-tcvn-13903-2023-cong-nghe-thong-tin-tri-tue-nhan-tao-310542-d3.html>

các nước, giữa chính phủ với các công ty lớn trên thế giới, chính quyền địa phương, các doanh nghiệp Việt Nam với các trung tâm công nghệ lớn ở các nước.

Chiến lược quốc gia về AI của Việt Nam không đề cập vai trò của các bên tham gia khác như doanh nghiệp, các công ty khởi nghiệp, các tổ chức xã hội, những người hưởng lợi từ AI. Các nội dung về phát triển bao trùm trong nghiên cứu, ứng dụng AI chưa được làm rõ, cụ thể là chưa nhấn mạnh việc tham vấn quan điểm, ý kiến và bảo đảm quyền, lợi ích của các nhóm khác nhau trong chính sách, pháp luật về AI, nhất là từ các nhóm dễ tổn thương như người khuyết tật, đồng bào dân tộc thiểu số, người nghèo, phụ nữ, trẻ em.

Khuôn khổ chính sách, pháp luật chưa chú trọng làm rõ khái niệm, chưa nói đến ứng dụng thực tế, của đạo đức AI ở tầm quốc gia. Chiến lược quốc gia về AI không có các nội dung về kiểm soát rủi ro, duy trì đạo đức trong phát triển, triển khai, ứng dụng AI như quyền riêng tư, sự thiên vị, phân biệt đối xử, vai trò của con người. Việc tích hợp các chuẩn mực quốc tế về đạo đức, pháp lý, quyền con người để giảm thiểu các rủi ro liên quan đến AI hầu như chưa được chú ý. Hướng dẫn của Bộ KHCN về các nguyên tắc đạo đức AI chỉ áp dụng cho các công đoạn nghiên cứu, phát triển hệ thống AI, mà không đề cập việc áp dụng các nguyên tắc trong triển khai, ứng dụng AI. Hơn nữa, chưa rõ cơ chế, hành động để đưa các nguyên tắc vào nghiên cứu, phát triển hệ thống AI ở các viện, trường đại học, các doanh nghiệp trên thực tế như thế nào.

Dự thảo Luật Công nghiệp công nghệ số mặc dù dường như có yếu tố kiểm soát theo cấp độ rủi ro của AI như EU, nhưng các nội dung của dự thảo định nghĩa về cấp độ rủi ro cao và tác động cao chưa rõ ràng; đồng thời chưa nêu rõ biện pháp tương ứng cần phải thực hiện để kiểm soát các hệ thống AI có mức độ rủi ro và tác động cao. Hơn nữa, dự thảo chưa nhận diện cấp độ rủi ro cao nhất (theo ngôn ngữ của Luật AI EU là rủi ro không chấp nhận được), cũng như chưa nêu các biện pháp kiểm soát nghiêm ngặt nhất đối với các hệ thống AI có rủi ro như vậy. Quy định về dán nhãn cho các sản phẩm AI cũng gây băn khoăn về tính khả thi, cũng như phạm vi áp dụng có thể kéo theo chi phí thực thi cho doanh nghiệp, tạo kẽ hở lạm dụng.

Chiến lược quốc gia về AI mới chỉ có một số nội dung rải rác giao nhiệm vụ cho một số bộ, ngành áp dụng AI trong hoạt động của mình, nhưng chưa có định hướng chung về sử dụng AI trong các cơ quan Nhà nước từ trung ương đến địa phương (ví dụ như ứng dụng để thực hiện chức năng, nhiệm vụ của cơ quan Nhà nước, phục vụ lợi ích công, lợi ích người dùng; lựa chọn ưu tiên trong ứng dụng AI vào cơ quan Nhà nước; đảm bảo các nguyên tắc quản trị rủi ro như minh bạch, giải trình v.v...). Đặc biệt, các cơ quan Nhà nước có trách nhiệm xử lý dữ liệu nhạy cảm, phục vụ các nhóm dân cư đa dạng và đảm bảo an toàn công cộng, bao gồm các lĩnh vực quan trọng như chăm sóc sức khỏe, dịch vụ xã hội, thực thi pháp luật và cơ sở hạ tầng. Việc áp dụng AI trong những lĩnh vực này có thể có tác động xã hội đáng kể, nên phải đòi hỏi phải đánh giá rủi ro kỹ lưỡng và triển khai cẩn thận để tránh những hậu quả tiêu cực. Mặc dù vậy, Chiến lược quốc gia về AI, cũng như các văn bản chính sách khác không xác định rõ nhiệm vụ và trách nhiệm của các cơ quan Nhà nước trong việc phát triển, ứng dụng AI một cách an toàn, minh bạch, vì lợi ích công cộng.

## **(b) Một số vấn đề còn bỏ ngỏ**

Khuôn khổ chính sách, pháp luật hiện hành vẫn còn những hạn chế hoặc chưa điều chỉnh đối với một số vấn đề liên quan đến thúc đẩy phát triển AI và kiểm soát rủi ro của AI. Cụ thể như:

- Các văn bản quy phạm pháp luật hiện hành không có quy định nào đủ mạnh tạo cơ chế hỗ trợ, ưu đãi đối với nhân tài, cơ sở nghiên cứu, đào tạo, doanh nghiệp, cơ chế tài chính để khuyến khích, thúc đẩy sự phát triển của AI, cả trong khu vực Nhà nước và tư nhân. Các quy định về các nội dung này trong dự thảo Luật Công nghiệp công nghệ số đều phải dẫn chiếu đến các quy định có liên quan về thuế, đất đai, đầu tư, ngân sách Nhà nước. Trong khi đó, như các thành viên Ủy ban Thường vụ Quốc hội có lưu ý, ưu đãi đầu tư cho lĩnh vực khoa học, công nghệ còn gặp nhiều khó khăn trong giải ngân do cơ chế chi tiêu theo Luật Ngân sách nhà nước, từ phân bổ, quản lý, đến quyết toán<sup>138</sup>. Tương tự, Chiến lược cải cách thuế đến năm 2030 của Chính phủ nêu rõ, các chính sách thuế chỉ được quy định trong các văn bản pháp luật về thuế. Như vậy, các quy định về ưu đãi thuế trong dự thảo Luật Công nghiệp công nghệ số sẽ không có hiệu lực, vì trong các luật thuế không thể hiện, ngay cả trong Luật Thuế thu nhập doanh nghiệp trình Quốc hội tại kỳ họp thứ 8 cuối năm 2024.

<sup>138</sup> Phiên họp của Ủy ban Thường vụ Quốc hội cho ý kiến về việc giải trình, tiếp thu, chỉnh lý Dự án Luật Công nghiệp công nghệ số, 6/1/2025.

- Hệ thống pháp luật vẫn thiếu các quy định rõ ràng về địa vị pháp lý của các thực thể AI trong các mối quan hệ xã hội. Pháp luật dân sự quy định, chủ thể pháp lý phải là cá nhân hoặc thể nhân, chứ không phải máy móc hoặc chương trình máy tính. Tuy nhiên, việc xác định trách nhiệm pháp lý trong một số trường hợp xảy ra lỗi, vi phạm khi triển khai, ứng dụng các hệ thống AI trong y tế, giáo dục, nông nghiệp....trở nên khó khăn hơn vì tính chất “hộp đen” trong quá trình AI đưa thông tin đầu ra cho con người.

- Liên quan đến vấn đề sở hữu trí tuệ, câu hỏi về trách nhiệm pháp lý đối với hành vi vi phạm SHTT trong bối cảnh nội dung do AI tạo ra vẫn là một vấn đề gây tranh cãi, đặc biệt là đối với các nền tảng cho phép người dùng tải lên tài liệu. Luật SHTT hiện hành tại Việt Nam không đề cập cụ thể đến vấn đề nội dung do AI tạo ra xâm phạm quyền SHTT. Luật nghiêm cấm nhiều hành vi trái phép, nhưng những hạn chế này được thiết kế để áp dụng đối với các cá nhân hoặc tổ chức trực tiếp tham gia vào các hoạt động, thay vì được áp dụng cho các thuật toán AI hoặc người tạo ra chúng. Hoặc là việc sử dụng nội dung được bảo vệ bởi bản quyền trong dữ liệu dùng để huấn luyện AI có vi phạm bản quyền hay không vẫn chưa có câu trả lời. Các trường hợp các hệ thống AI sao chép trực tiếp tác phẩm có bản quyền cũng là một vấn đề cần làm rõ. Tương tự, cần trả lời câu hỏi: người dùng có sở hữu tác quyền đối với các đầu ra của các hệ thống AI mà họ sử dụng để tạo ra tác phẩm.

- Pháp luật Việt Nam còn thiếu những quy định cụ thể điều chỉnh việc kiểm soát rủi ro đối với quyền riêng tư, sử dụng dữ liệu cá nhân, bảo mật trong phát triển, triển khai, ứng dụng AI, ví dụ như giám sát, theo dõi ở các địa điểm công cộng; công nghệ nhận dạng khuôn mặt hoặc công nghệ sinh trắc học khác (xem thêm trong hộp dưới đây).

### **Kiểm soát rủi ro của nhận diện khuôn mặt trong ATM gạo như thế nào?**

*Trong đại dịch COVID-19, máy phát gạo tự động đã được triển khai tại một số thành phố để cung cấp gạo miễn phí cho những người thu nhập thấp mà không cần tiếp xúc trực tiếp. Sau đó, “ATM gạo” được cải tiến thành các máy phát khẩu trang khi các ca mắc COVID-19 tăng vọt. Để ngăn ngừa tình trạng quá tải, giảm thiểu tiếp xúc bề mặt, giúp xác minh danh tính, các máy phát gạo và khẩu trang dựa vào công nghệ nhận dạng khuôn mặt được liên kết với thông tin cá nhân thu thập được từ việc đăng ký trước. Trường hợp này đặt ra những vấn đề pháp lý chưa được điều chỉnh như: tính hợp lệ, phạm vi ứng dụng công nghệ nhận diện khuôn mặt dựa trên AI mà không vi phạm quyền riêng tư, tính bảo mật và việc sử dụng dữ liệu cá nhân; tính minh bạch, khả năng giải thích, trách nhiệm giải trình về công nghệ này.*

- Dự định sử dụng các công cụ AI trong “chủ động ngăn chặn tội phạm” đòi hỏi phải thảo luận thêm về những tác động của công nghệ giám sát, xử lý thuật toán và chất lượng của các tập dữ liệu cho mục đích dự đoán tội phạm. Bất chấp tiềm năng về hiệu quả của chúng, các báo cáo về sự thiên vị, phân biệt đối xử dẫn đến sai sót trong tư pháp hình sự ở Hoa Kỳ, Vương quốc Anh là những bài học cảnh báo cho Việt Nam.

- Luật An ninh mạng năm 2018 yêu cầu chủ sở hữu nền tảng phải chủ động ngăn chặn, phát hiện và xóa nội dung có hại theo yêu cầu của các cơ quan an ninh mạng. Nội dung này bao gồm nội dung tấn công chính phủ, kích động bạo lực, phát tán thông tin sai lệch, gây hại cho nền kinh tế hoặc gây hoang mang cho công chúng. Một số văn bản quy phạm pháp luật khác điều chỉnh nội dung trên Internet cũng có yêu cầu tương tự. Tuy nhiên, không rõ các nền tảng tạo nội dung với sự hỗ trợ của AI có thể bị cáo buộc vi phạm hay không, vì việc tạo nội dung dựa trên thuật toán chủ yếu là tự động.

- Pháp luật lao động, việc làm chưa có quy định về một số vấn đề mới nổi liên quan đến sử dụng AI tại nơi làm việc như: tuyển dụng lao động; theo dõi hành vi của nhân viên nhằm đo lường và đảm bảo năng suất; giảm thiểu rủi ro mất việc làm do công nghệ AI.

- Cơ chế khiếu nại, khiếu kiện, giải quyết tranh chấp nói chung, cũng như khiếu kiện về thiệt hại, vi phạm do việc triển khai, sử dụng AI gây ra là một khoảng trống lớn ở Việt Nam. Tòa án các cấp ở Việt Nam chưa xem xét một vụ kiện nào liên quan đến AI; và dường như chưa sẵn sàng đối mặt với lĩnh vực mới mẻ, phức tạp này.

# KHUYẾN NGHỊ CHÍNH SÁCH VỀ AI Ở VIỆT NAM

Để tận dụng lợi ích của AI đồng thời giảm thiểu các rủi ro đối với xã hội, con người, Việt Nam nên xem xét các nhóm giải pháp sau:

## 5.1. CÁCH TIẾP CẬN CHUNG

### 5.1.1. Cân bằng giữa kiểm soát rủi ro với phát triển AI

Các quy định, biện pháp kiểm soát, giảm thiểu rủi ro xã hội và bảo vệ quyền con người nhưng không làm cản trở hay đình trệ các tiến bộ và cải cách sáng tạo; cân bằng được yêu cầu vừa tạo ra môi trường và độ linh hoạt cho sự phát triển AI, đồng thời kiểm soát, giảm thiểu các mối nguy cơ, rủi ro với xã hội.

Việc ban hành các quy định mới cần lưu ý, các nguyên lý và trật tự pháp luật luôn bền vững theo thời gian, cho nên trước hết cần xem xét, khuôn khổ pháp luật hiện tại đã đủ khả năng để xử lý các vấn đề liên quan đến AI hay chưa; nếu không đủ mới cân nhắc ban hành quy định mới. Chẳng hạn, theo các chuyên gia pháp luật, vấn đề pháp lý dân sự cốt yếu là xác định AI là chủ thể tham gia quan hệ pháp luật hay là tài sản. Nếu AI không là chủ thể mới tham gia quan hệ pháp luật thì đó là tài sản, cho nên việc điều chỉnh các mối quan hệ pháp lý liên quan đến AI cần theo quy định về tài sản của pháp luật dân sự hiện tại. Đối với nhiều vấn đề liên quan đến AI, có thể lựa chọn cách tiếp cận “luật mềm” như các nguyên tắc đạo đức, bộ quy tắc ứng xử, hướng dẫn kỹ thuật. Nhánh lập pháp Việt Nam có thể cần can thiệp, nhưng hiện chưa phải thời điểm quy định chặt chẽ, để dành không gian phát triển AI.

Trên phương diện quản trị rủi ro của AI, có thể đạt mục tiêu cân bằng bằng cách thức sau (xem chi tiết hơn trong mục về chính sách kiểm soát rủi ro ở phần trước):

- Đánh giá, điều chỉnh AI theo các mức độ rủi ro để có phương án phù hợp với mức độ rủi ro đã được xác định; mỗi cấp độ rủi ro có những biện pháp kiểm soát phù hợp.
- Các công cụ chính sách cần được áp dụng tương thích với mức độ rủi ro: đối với rủi ro mức độ cao nhất, những quy định pháp luật khắt khe, có tính bắt buộc thực hiện là cần thiết; đối với rủi ro ở mức độ tiếp theo, các yêu cầu pháp lý nên “nhẹ nhàng” hơn; và đối với các rủi ro thấp, chỉ cần áp dụng hướng dẫn, tiêu chuẩn kỹ thuật, bộ quy tắc ứng xử tự nguyện.
- Sự cân bằng với thúc đẩy, phát triển AI đòi hỏi kiểm soát rủi ro theo phạm vi, mức độ nắm giữ quyền lực của các chủ thể tham gia vào vòng đời hệ thống AI; theo đó, chủ thể nào nắm giữ quyền lực nhiều hơn thì phải chịu sự kiểm soát chặt chẽ hơn và ngược lại.
- Trong quá trình xây dựng, thực thi chính sách về AI, cần cân bằng giữa các giá trị đối lập, chẳng hạn như quyền riêng tư và tính minh bạch. Cách tiếp cận phù hợp là tùy theo ngữ cảnh, trong đó các bên liên quan ưu tiên các giá trị dựa trên ứng dụng cụ thể của hệ thống AI; thích ứng các loại hệ thống AI với bối cảnh khác nhau.

## **Cân bằng giữa quyền riêng tư và tính minh bạch**

*Làm sao để tận dụng lợi ích của AI để phát triển kinh tế mà vẫn đảm bảo sự riêng tư, tự do cá nhân trong xử lý lượng lớn dữ liệu cá nhân? Liệu nguyên tắc đồng thuận và tối thiểu trong sử dụng dữ liệu cá nhân có tạo ra rào cản nhất định cho việc tận dụng công nghệ để phát triển kinh tế - xã hội? Trả lời câu hỏi này, có thể ưu tiên áp dụng nguyên tắc hợp lý và tương thích: cho phép sử dụng lại dữ liệu cá nhân vào mục đích khác mục đích ban đầu trong trường hợp không ảnh hưởng đến lợi ích hợp pháp, chính đáng của chủ thể dữ liệu và sử dụng biện pháp bảo mật tương thích để bảo vệ dữ liệu cá nhân. Bên cạnh đó, có thể thiết kế thủ tục, công cụ đơn giản, dễ dàng để nhận được sự đồng thuận của cá nhân về việc sử dụng dữ liệu của họ.*

### **5.1.2. Kết hợp các công cụ chính sách khác nhau**

Như kinh nghiệm các nước cho thấy, để điều chỉnh các vấn đề liên quan đến AI, trong đó có kiểm soát rủi ro trong phát triển, triển khai, ứng dụng AI, chỉ riêng pháp luật là không đủ. Cần áp dụng các công cụ chính sách khác nhau, gồm có: Các nguyên tắc lớn về AI như Các nguyên tắc của OECD, UNESCO, ASEAN; các văn kiện chính sách như Chiến lược quốc gia về AI, Chiến lược quốc gia về dữ liệu; khuôn khổ pháp luật liên quan đến AI gồm những quy định tổng quát về AI (trong một đạo luật riêng hoặc một chương riêng trong đạo luật về công nghệ); những quy định trong một số đạo luật riêng biệt điều chỉnh các vấn đề cụ thể về AI; sửa đổi, áp dụng các văn bản quy phạm pháp luật hiện hành.

Cụ thể, cần sửa đổi, bổ sung Chiến lược quốc gia về AI cho phù hợp với những diễn biến mới trong quá trình phát triển của công nghệ AI như AI tạo sinh; đề ra tầm nhìn, mục tiêu lộ trình phát triển AI; chú trọng đồng thời cả thúc đẩy, hỗ trợ phát triển AI cùng với quản trị rủi ro của AI; bảo đảm AI có trách nhiệm, đáng tin cậy, tuân theo các giá trị của con người, quyền con người trong suốt vòng đời của hệ thống AI; ưu tiên phát triển AI theo lĩnh vực hẹp trên cơ sở ứng dụng các mô hình AI đã được phát triển trên thế giới (xem cụ thể hơn về hướng ưu tiên này ở mục về lựa chọn thị trường đầu tư AI).

Chiến lược cũng cần giao trách nhiệm cụ thể, rõ ràng, hợp lý hơn cho các bộ, ngành; vai trò của các bên liên quan trong phát triển AI ở Việt Nam. Trong bối cảnh cạnh tranh công nghệ, hợp tác quốc tế đóng vai trò hết sức quan trọng, Chiến lược cần có những nội dung rõ ràng về ngoại giao công nghệ giữa chính phủ với chính phủ, giữa chính phủ với các công ty công nghệ lớn, giữa các công ty với nhau, và giữa các trường, viện nghiên cứu trong với ngoài nước.

Khung pháp luật trực tiếp điều chỉnh về AI ở Việt Nam trong thời gian tới nên theo cả hai cách tiếp cận tổng quát, theo chiều rộng và theo những lĩnh vực hẹp. Cụ thể là Luật Công nghiệp Công nghệ số có những nội dung khái quát về các chính sách khuyến khích, hỗ trợ phát triển AI; kiểm soát rủi ro AI ở Việt Nam (xem cụ thể hơn ở mục về giải pháp lập pháp kiểm soát rủi ro của AI). Các nội dung chi tiết, mang tính kỹ thuật về AI nói chung, cũng như AI trong những lĩnh vực hẹp nên được điều chỉnh trong nghị định, thông tư.

Cần sớm xây dựng khung pháp lý để triển khai cơ chế thử nghiệm có kiểm soát (sandbox) đối với các ý tưởng, sản phẩm, dịch vụ công nghệ mới; nếu an toàn, hiệu quả, có thể mở rộng, nếu không, phải điều chỉnh kịp thời để tránh rủi ro và lãng phí nguồn lực. Thời hạn thử nghiệm 2 năm trở xuống; phạm vi thử nghiệm được xác định phù hợp với từng trường hợp; thẩm quyền phê duyệt tham gia thử nghiệm thuộc các bộ, ngành; phạm vi, mức độ miễn trừ trách nhiệm pháp lý cho các doanh nghiệp, tổ chức tham gia thử nghiệm.

Cần tiếp tục xây dựng các Bộ tiêu chuẩn, quy chuẩn kỹ thuật liên quan đến AI có tính chất tự nguyện, theo các tiêu chuẩn quốc tế của ISO như các tiêu chuẩn về quản trị AI trong doanh nghiệp; về vòng đời của hệ thống AI; về đạo đức AI; các nội dung cần chú ý về xã hội, môi trường liên quan đến AI.

Bên cạnh đó, tổ chức, doanh nghiệp, cá nhân có thể theo cơ chế tuân thủ tự nguyện các bộ quy tắc ứng xử, hoặc các bộ tiêu chuẩn, quy chuẩn công nghệ. Dựa trên các nguyên tắc được nêu



trong Quyết định số 1290/QĐ-BKHCN, các cơ sở nghiên cứu, đào tạo, các doanh nghiệp Việt Nam có thể thiết lập các nguyên tắc cho tổ chức mình trong nghiên cứu, phát triển AI có trách nhiệm. Điều này bao gồm thúc đẩy tính minh bạch, trách nhiệm giải trình và công bằng trong các hệ thống AI, cũng như thúc đẩy lòng tin và sự hiểu biết của công chúng về các công nghệ AI. Có thể xây dựng bộ quy tắc ứng xử AI trong từng ngành nghề có sử dụng trí tuệ nhân tạo. Các bộ nguyên tắc này sẽ hướng dẫn cụ thể, đáp ứng theo đặc thù của ngành nghề. Nhờ đó, doanh nghiệp sẽ giảm chi phí thiết kế tiêu chuẩn, chính sách nội bộ khi sử dụng trí tuệ nhân tạo vì đã có sẵn bộ quy tắc chung của ngành.

### **5.1.3. Tiếp cận tổng thể; thu hút các bên liên quan**

Xét đến bản chất xuyên suốt của AI, cần tránh các cách tiếp cận phân mảnh trong điều chỉnh AI. Mặc dù chúng ta cần câu trả lời cụ thể cho từng trường hợp đối với các vấn đề riêng biệt của từng lĩnh vực, nhưng thường các vấn đề về AI liên quan đến nhiều bên khác nhau. Đơn cử như ngay trong chính phủ, các cơ quan chịu trách nhiệm về quyền riêng tư, an ninh mạng, tăng trưởng kinh tế, thương mại, thực thi pháp luật, y tế, tài chính đều có liên quan. Để phối hợp các cơ quan khác nhau, có thể cân nhắc phương án thiết lập cơ chế liên ngành có tính đại diện, hiệu quả và cân bằng các quan điểm, góc nhìn khác nhau để chính sách phù hợp với lợi ích rộng lớn hơn của công chúng.

Các bên liên quan trong xã hội đa dạng cần được tham gia từ đầu trong vòng đời của hệ thống AI; không chỉ các chuyên gia kỹ thuật, doanh nghiệp, mà còn cả các chuyên gia pháp lý, các chuyên gia về quyền con người, đại diện từ cộng đồng, các tổ chức xã hội, và các nhóm khác có thể bị ảnh hưởng bởi hệ thống AI. Trong đó, cần đặc biệt tham vấn ý kiến của các cơ sở nghiên cứu, thực hành về quyền con người, cụ thể ở Việt Nam là Viện Nghiên cứu quyền con người thuộc Học viện Chính trị quốc gia Hồ Chí Minh, Trung tâm Nghiên cứu quyền con người thuộc Trường Luật, Đại học quốc gia Hà Nội v.v... Bên cạnh đó, nâng cao năng lực về công nghệ số, trong đó có AI, nhận thức, kỹ năng sử dụng AI an toàn cho tất cả các chủ thể có thể góp phần đảm bảo quyền con người.

Các bên liên quan cần được tạo cơ hội tham gia tích cực vào việc tạo ra và thử nghiệm các mô hình, ứng dụng AI. Quá trình lặp đi lặp lại này cho phép phản hồi liên tục và điều chỉnh, đảm bảo rằng thiết kế cuối cùng phù hợp với các giá trị và nhu cầu của các bên liên quan. Cần thu thập, tổng hợp thông tin từ các bên liên quan để hiểu quan điểm, nhu cầu và cách họ diễn giải các giá trị mà một hệ thống AI cần đạt được trong bối cảnh của họ. Các phương pháp như khảo sát, phỏng vấn và hội thảo được sử dụng để thu thập dữ liệu này. Mục tiêu là đảm bảo rằng thiết kế AI phản ánh trải nghiệm và kỳ vọng của những người sẽ chịu ảnh hưởng.

Việc ban hành chính sách điều chỉnh AI cần hướng tới công bằng xã hội, bảo đảm quyền, lợi ích, có sự tham gia của các nhóm người dễ tổn thương như người khuyết tật, dân tộc thiểu số, người nghèo, phụ nữ, trẻ em. Nhà nước cần tham gia thúc đẩy, ưu tiên ứng dụng AI để bảo đảm, nâng cao quyền con người như ứng dụng AI trong dịch vụ hành chính công, các ứng dụng giúp người khuyết tật tiếp cận giáo dục, chăm sóc sức khỏe từ xa. Các ứng dụng AI luôn phải đảm bảo yếu tố thân thiện với mọi người dùng để tránh hậu quả không những không giúp gì cho người dễ tổn thương mà còn tạo ra một nhóm dễ tổn thương mới là yếu tố về công nghệ. Nhà nước cần điều tiết để phân bổ lại thu nhập, cơ hội trong xã hội như bảo đảm thực hiện trách nhiệm xã hội của doanh nghiệp, trách nhiệm đào tạo lại, đào tạo nâng cao cho người lao động.

### **5.1.4. Vai trò tập trung và cầu nối của chính quyền trung ương**

Cần nhấn mạnh hơn nữa vai trò chỉ đạo, điều hành tập trung từ trên xuống của chính quyền trung ương trong việc thiết lập tầm nhìn và giám sát việc thực hiện các kế hoạch hành động. Cụ thể, một mặt, Chính phủ sẽ xác định một số lĩnh vực ưu tiên để đầu tư phát triển, ứng dụng AI, từ đó mang lại hiệu ứng lan tỏa trong toàn bộ nền kinh tế (xem cụ thể hơn ở các mục sau).

Đồng thời, Chính phủ đóng vai trò là cầu nối giữa các bên liên quan khác nhau, đảm bảo sự phối hợp và hợp tác hiệu quả giữa họ. Chính phủ có thể thực hiện các chính sách cầu nối lớn sau đây (xem thêm ở các mục sau):

- Thiết lập sandbox cho phép các doanh nghiệp lớn và nhỏ cùng nhau làm việc, tập trung vào những thách thức chính của các lĩnh vực ưu tiên. Các sandbox như đã đề cập trong tài liệu này có thể là một nơi an toàn để thử nghiệm những ý tưởng và sáng kiến mới.

- Hỗ trợ thúc đẩy R&D trong các doanh nghiệp vừa và nhỏ. Ví dụ, NIC có thể đóng vai trò tương tự như InnovateUK tại Vương quốc Anh. Thông qua NIC, chính phủ có thể trợ cấp phần lớn chi phí cho các hoạt động R&D có rủi ro cao để cho phép các doanh nghiệp vừa và nhỏ chấp nhận rủi ro nhiều hơn. Sự hỗ trợ này có thể được thực hiện thông qua một số quỹ của chính phủ như ở các nước Anh, Úc để giúp các doanh nghiệp vừa và nhỏ áp dụng các công nghệ R&D tiên tiến.
- Thành lập Trung tâm nghiên cứu AI quốc gia: Thay vì mong đợi nhiều trường đại học đạt được đột phá trong nghiên cứu, sẽ hiệu quả hơn về mặt tài chính nếu có một trung tâm nghiên cứu quốc gia có thể thực hiện nghiên cứu sáng tạo nhất. Các trường đại học sau đó có thể hợp tác với trung tâm này để tiến hành nghiên cứu. Các kế hoạch/chiến lược đào tạo và giáo dục tài năng quốc gia cũng có thể được điều phối thông qua trung tâm này để đảm bảo sự phối hợp hiệu quả trên toàn quốc. Vì vậy, việc học hỏi từ những sai lầm như của Viện Turing ở Anh là điều đáng giá nếu chính phủ quyết định thành lập trung tâm nghiên cứu AI quốc gia.

## 5.2. CHÍNH SÁCH HỖ TRỢ, THÚC ĐẨY PHÁT TRIỂN AI

Cần sửa đổi, bổ sung các luật hiện hành để bắt kịp với sự phát triển của công nghệ AI như: một số quy định về sở hữu trí tuệ, cụ thể là bản quyền; pháp luật về lao động, việc làm; các quy định về hạ tầng, dữ liệu; cơ chế tài chính, đầu tư công cho khoa học, công nghệ, trong đó có AI; phát triển nguồn nhân lực, đào tạo, giáo dục về AI v.v...

### 5.2.1. Đầu tư; lựa chọn ưu tiên phân khúc thị trường AI

Theo các chuyên gia, đại diện các doanh nghiệp mà IPS trao đổi, phát triển AI ở Việt Nam có thể lựa chọn để đi theo các hướng lớn sau: (1) Các startup gọi được nguồn vốn lớn (như OpenAI), các công ty lớn sẽ xây dựng mô hình AI lớn; (2) Các doanh nghiệp chuyên về AI tối ưu hóa mô hình AI để cung cấp giải pháp công nghệ cụ thể cho khách hàng; (3) Các doanh nghiệp vừa và nhỏ trong các ngành hiện hữu tập trung vào việc ứng dụng mô hình AI, giải pháp AI vào quản trị doanh nghiệp, kinh doanh, dịch vụ.

Trong ba hướng trên, hướng thứ nhất gặp nhiều khó khăn, bởi lẽ chi phí đầu tư để phát triển mô hình AI quá lớn so với chi phí cho con người, nhất là đối với Việt Nam là nước đang phát triển, chưa có những doanh nghiệp đủ nguồn lực tài chính, nhân lực để thực hiện. Các BigTech như Google, Microsoft, AWS có nguồn tài chính rất lớn để đầu tư vào hạ tầng tính toán, R&D, đội ngũ kỹ sư thuộc loại giỏi bậc nhất thế giới để triển khai giải pháp công nghệ và tích hợp AI vào hệ sinh thái của mình để cung cấp dịch vụ cho khách hàng trên toàn cầu. Vì lý do này, Việt Nam chưa có tiềm năng để phát triển ra một mô hình hoàn toàn mới như LLM của ChatGPT do hạn chế về cơ sở hạ tầng, chuẩn bị dữ liệu và xử lý dữ liệu: các mô hình LLM ở Việt Nam chỉ xoay quanh 7 tỉ, 13 tỉ, 33 tỉ tham số. Hơn nữa, như một số chuyên gia trao đổi với IPS, mô hình GPT tiếng Việt gặp khó khăn vì kho dữ liệu tiếng Việt không đủ để tạo ra một mô hình ngôn ngữ lớn có chất lượng; hơn nữa đối với một mô hình ngôn ngữ lớn, việc huấn luyện từ nhiều dữ liệu với các loại ngôn ngữ khác nhau thì tốt hơn là một loại ngôn ngữ.

Mặt khác, có những ví dụ cho thấy, hướng khai thác tiếng Việt vẫn có thể mang lại thành công như Vbee AI Voice API của Vbee, Olli Maika nói tiếng Việt cho hệ thống loa thông minh của Olli Technology, AI Hay<sup>139</sup>. Theo báo cáo của ZaloAI, tại thời điểm 2024 có 45 LLMs tiếng Việt được gửi đến để đánh giá, cho thấy mặc dù kết quả này còn khiêm tốn so với các nước phát triển, nhưng xu hướng LLMs tiếng Việt ngày càng tăng lên<sup>140</sup>. Bên cạnh đó, hiện đang có những nỗ lực xây dựng bộ dữ liệu tiếng Việt, trong đó nổi bật như Dự án ViGen hợp tác ba bên giữa Meta, NIC, tổ chức AI For Vietnam Foundation, hướng đến việc tạo ra bộ dữ liệu tiếng Việt mã nguồn mở chất lượng cao - giúp AI hiểu và sử dụng tiếng Việt tự nhiên hơn<sup>141</sup>. Đây được coi là bước tiến quan trọng trong việc tăng cường sự hiện diện của tiếng Việt trong AI, đồng thời thúc đẩy tăng trưởng kinh tế và phát triển công nghệ.

<sup>139</sup> Ricky Ho, Startup AI Việt Nam gặp cảnh 'vòi nhiều, cò quá ít', Kinh tế Sài Gòn Online, <https://thesaigontimes.vn/startup-ai-viet-nam-gap-canh-voi-nhieu-co-qua-it/>

<sup>140</sup> Zalo AI, Báo cáo tình hình phát triển LLMs tiếng Việt, 2024.

<sup>141</sup> <https://vietnamnet.vn/viet-nam-phat-trien-bo-du-lieu-tieng-vi-et-dung-cho-ai-tro-ly-ao-2380624.html?fbclid=IwY2xjawJTKuBleHRuA2FibQlXMAABHd658pJZ8klG-HgcSNhMG9mxtBOn1aJxrleB92gRIGZiYF>

WKbn15GEYhUA\_aem\_UU95pjA0XC01IsvcVud9xA

Chính vì vậy, một số chuyên gia cho rằng, cần đầu tư cả nghiên cứu, phát triển các mô hình AI. Điều này còn hướng tới để người Việt nắm được công nghệ cốt lõi về AI (các thuật toán vắn của người nước ngoài nắm giữ), cũng như đầu tư về cơ sở hạ tầng (phải có các phòng thí nghiệm trọng điểm, khi có hạ tầng tính toán thì người nghiên cứu mới có khả năng đầu tư tạo ra một mô hình thuật toán mới). Những trường hợp thành công trong xây dựng các mô hình LLM tiếng Việt ở Việt Nam trong năm 2023-2024 tạo cơ sở cho hướng đi này<sup>142</sup>. Trong đó, có thể tham khảo xu hướng giảm độ lớn mô hình LLM AI và nâng cao hiệu suất. Ví dụ mô hình của Mistral ở Pháp với 64 tỉ tham số, nhưng hiệu suất (performance) gần bằng ChatGPT; hoặc mô hình ViGPT quy mô 1,6 tỉ tham số, vì chỉ cho tiếng Việt, mà có thể vẫn giữ nguyên hiệu suất.

Về việc đầu tư vào AI theo hướng thứ hai và thứ ba, đại diện một doanh nghiệp chia sẻ, ứng dụng AI vào hoạt động kinh doanh của doanh nghiệp chưa chứng minh được hiệu quả kinh tế: bỏ bao nhiêu tiền ra đầu tư vào AI, sẽ thu lại được bao nhiêu tiền. Do đó, doanh nghiệp này chưa đi theo hướng ứng dụng AI vào hoạt động kinh doanh của mình (tại thời điểm 2024). Còn theo một chủ doanh nghiệp khác, sản phẩm dựa trên AI đòi hỏi tốn rất nhiều chi phí và công sức để cải thiện sản phẩm. Cụ thể, ứng dụng Elsa do người Việt làm ra với rất nhiều nguồn vốn đầu tư và nhiều thời gian cải thiện chất lượng sản phẩm mới có tính cạnh tranh toàn cầu. Hoặc là một doanh nghiệp đã thử nghiệm ứng dụng AI trong lĩnh vực đấu thầu nhưng vẫn chưa thấy khả quan: AI chatbot trả lời về luật đấu thầu vẫn chưa ở mức chấp nhận được, chưa nói đến AI tư vấn tình huống đấu thầu trên thực tế. Vinades đầu tư cho các nhóm sinh viên, nhóm cá nhân xây dựng ứng dụng AI nhưng chưa nhận lại được bất cứ sản phẩm nào nhưng vẫn chỉ cho đầu tư phát triển. Một người trả lời phỏng vấn của IPS đánh giá, trong khi doanh nghiệp nước ngoài mạnh dạn đầu tư hơn các sản phẩm dài hạn, nhiều doanh nghiệp Việt Nam vẫn dè dặt trong việc đầu tư vào các giải pháp dựa trên AI, vì họ muốn nhìn thấy kết quả ngay.

Tuy nhiên, hầu hết những người được hỏi cho rằng, đối với Việt Nam, nên theo hướng phát triển AI hệ chuyên gia trong từng lĩnh vực thì phù hợp hơn, vì ít tốn kém hơn; dữ liệu cấu trúc có sẵn hơn (phần lớn dữ liệu của cơ quan nhà nước là dữ liệu có cấu trúc); con người hiểu rõ dữ liệu và thuật toán thực hiện, dễ phân biệt đúng - sai và dễ kiểm soát AI hơn. Doanh nghiệp ứng dụng AI để giải quyết bài toán sát thực tế có khả năng trở thành doanh nghiệp dẫn đầu tại Việt Nam. Doanh nghiệp Việt Nam đang trong quá trình chuyển đổi số, vì vậy kho dữ liệu của doanh nghiệp có thể là điều kiện để ứng dụng AI. Đồng thời, theo ý kiến của một số chuyên gia AI mà IPS trao đổi, ở Việt Nam hiện nay, ứng dụng AI truyền thống dạng rule-based đem lại nhiều giá trị hơn tạo ra các mô hình Gen AI như Chat GPT.

### Chọn “ngách nhỏ” trong thị trường AI

*Nguyễn Thọ Chương, trưởng bộ phận AI thuộc AI Hay, một doanh nghiệp khởi nghiệp cho biết, trọng tâm hiện tại của doanh nghiệp này là tiếng Việt. “Dù chúng tôi có tham vọng vươn ra quốc tế, nhưng chúng tôi không muốn thâm nhập vào các thị trường cạnh tranh cao như tiếng Anh, nơi các công ty lớn như OpenAI và Google thống trị. Thay vào đó, chúng tôi có kế hoạch tập trung vào các ngôn ngữ có nguồn lực trung bình đến thấp, đặc biệt là ở Đông Nam Á và các khu vực lân cận. Đây là thị trường chúng tôi có thể cung cấp giá trị độc đáo và tránh cạnh tranh trực tiếp với các công ty lớn hơn”. Kể từ khi ra mắt, AI Hay đã xử lý khoảng 10.000 câu hỏi mỗi tháng, mỗi ngày hơn 300 câu, từ các yêu cầu dữ liệu rất cụ thể, như thông tin lũ lụt, đến các truy vấn xã hội hàng ngày. AI Hay hiện đang phục vụ nhu cầu của người dùng Việt Nam nhưng đặt mục tiêu mở rộng ra thị trường quốc tế trong tương lai.*

*Ngay cả như doanh nghiệp lớn ở Việt Nam như FPT tập trung giải quyết bài toán cụ thể về AI cho các doanh nghiệp khác từ năm 2015, xây dựng server để cung cấp dịch vụ nền tảng cho doanh nghiệp nội địa. Cụ thể, trong ngành sản xuất, FPT triển khai các giải pháp dựa trên AI tự động hóa trong nhà máy. Dữ liệu bắt buộc phải lưu trữ tại doanh nghiệp sản xuất do liên quan đến bí mật kinh doanh. Trong lĩnh vực ô tô, FPT tập trung phát triển phần mềm trợ lái thông minh và liên tục cải tiến hàng năm để tối ưu và thân thiện với người dùng. FPT đã mở văn phòng ở Okinawa (Nhật Bản) và Slovakia để triển khai các dịch vụ giải pháp công nghệ cho doanh nghiệp tại Nhật và EU. (Nguồn: Các cuộc phỏng vấn sâu do IPS thực hiện năm 2024).*

<sup>142</sup> Zalo AI, Báo cáo tình hình phát triển LLMs tiếng Việt, 2024.

Bên cạnh đó, các doanh nghiệp Việt Nam vẫn có cơ hội cung cấp các dịch vụ dựa trên AI cho thị trường toàn cầu, mặc dù nhiều doanh nghiệp vẫn thiếu năng lực về tài chính, quản trị và kinh nghiệm kinh doanh ở các thị trường khác. Sự kết hợp của nhóm học tập và làm việc ở nước ngoài với nhóm nhân sự ở Việt Nam sẽ có hội tạo ra thành công, vì nhân sự ở Việt Nam giá rẻ kết hợp với người Việt Nam sinh sống, du học ở nước ngoài có vốn hiểu biết về thị trường nước ngoài. Trong bối cảnh này, Nhà nước có thể trở thành kênh để kết nối hai nguồn nhân lực AI.

Việt Nam có thể nâng cao chuỗi giá trị thành công của quá trình bước ra thị trường với các giải pháp GenAI của mình (thay vì chỉ làm gia công phần mềm cho bên ngoài – outsourcing) bằng cách thúc đẩy hỗ trợ các công ty khởi nghiệp hơn trong trung hạn. Một trong những thách thức lớn của Việt Nam là sự tiến bộ nhanh chóng của GenAI về khả năng lập trình ngôn ngữ tự nhiên (Natural Language Programming) và tạo mã AI (AI Code Generation). Điều này có tác động làm xói mòn lợi thế coding của các kỹ sư, doanh nghiệp Việt Nam, trong khi thúc đẩy lợi thế của các quốc gia khác. Do đó, Việt Nam cần phải nhanh chóng tận dụng lợi thế hiện tại trong coding, đồng thời nâng cao năng lực cho lực lượng lao động theo kịp GenAI. Khuyến nghị này không chỉ hướng đến các doanh nghiệp, mà cho cả các trường đại học, cơ sở nghiên cứu, các chương trình R&D, và cả chính sách tầm quốc gia của chính phủ.

Với nguồn lực khan hiếm và hạn chế của Việt Nam, việc lựa chọn ưu tiên tập trung một số lĩnh vực để nghiên cứu và tạo điều kiện thuận lợi cho đầu tư vào AI sẽ phù hợp, thu hút và thúc đẩy đầu tư từ các khu vực công và tư nhân, từ các doanh nghiệp nước ngoài. Cơ hội nằm ở việc phát triển các giải pháp chuyên biệt cho các thị trường “ngách” và tận dụng nhân lực chuyên môn trong nước để tạo ra các ứng dụng AI phù hợp. Điều này sẽ cải thiện khả năng cạnh tranh của quốc gia bằng cách tập trung vào các đầu ra chính và cốt lõi trong thị trường AI, cải thiện hiệu quả và hiệu suất đầu tư, tận dụng những lợi thế và thế mạnh vốn có của quốc gia. Các lĩnh vực như vậy có thể là:

- **Quân sự:** Thông thường ở nhiều nước R&D trong quân sự là động lực thực sự cho sự đổi mới chứ không phải kinh doanh. Điển hình như Cơ quan về các dự án nghiên cứu quốc phòng tiên tiến của Mỹ (Defense Advanced Research Projects Agency – DARPA) là cơ quan đã đóng vai trò quan trọng trong nhiều phát minh như internet, AI, robot, v.v...; các chương trình tài trợ nghiên cứu của họ trong giai đoạn 2010-2020 đã góp phần vào sự gia tăng của AI tạo sinh và các mô hình ngôn ngữ lớn. Tuy nhiên, cần phòng ngừa và tránh rủi ro biến AI trở thành vũ khí giết người trong tay kẻ xấu.
- **Giáo dục:** Chiến lược hàng đầu của chính phủ là đảm bảo AI cho giáo dục và giáo dục/ng-hiên cứu học thuật cho AI được hỗ trợ ở mức cao nhất. Đây là cách duy nhất để duy trì nguồn nhân tài và đạt được, duy trì kiến thức chất lượng cao trong thời gian dài.
- **Nông nghiệp:** AI có thể giúp Việt Nam ứng phó với các vấn đề do biến đổi khí hậu gây ra trên khắp đồng bằng sông Cửu Long và các khu vực quan trọng về mặt chiến lược khác. Đồng thời, khử cacbon và các công nghệ sạch để hỗ trợ mục tiêu phát thải ròng bằng không, các ứng dụng trong lĩnh vực nông nghiệp và thực phẩm để đảm bảo an ninh lương thực, vì các lĩnh vực này đã trở thành xu hướng toàn cầu, đặc biệt là trong những năm và thập kỷ tới. Hơn nữa, Việt Nam có lợi thế so sánh trong lĩnh vực này, là một quốc gia nông nghiệp với khoảng 40% diện tích đất được sử dụng cho sản xuất nông nghiệp.

Bên cạnh đó, Việt Nam cần tăng quy mô đầu tư đáng kể vào công nghệ AI ở giai đoạn sớm. Ở Đông Nam Á, Singapore, Indonesia, Malaysia và Thái Lan đã đầu tư vào AI vừa nhiều hơn, vừa sớm hơn nhiều so với Việt Nam. Để đạt được mục tiêu top 5 ở Đông Nam Á vào năm 2025 và top 4 vào năm 2030, Việt Nam cần thúc đẩy càng sớm càng tốt các khoản đầu tư tương đương với mức mà các nước láng giềng đã phân bổ. Khoản đầu tư như vậy sẽ giúp Việt Nam đạt được nền tảng liên quan để phát triển các ứng dụng AI khi thị trường phát triển. Đầu tư nhiều hơn ở giai đoạn đầu sẽ tạo ra nhiều sản phẩm sáng tạo hơn và nắm giữ thị phần lớn trong một số thị trường AI nhất định, cũng như vai trò đối tác trong các hệ thống chuỗi cung ứng AI.

### 5.2.2. Chính sách phát triển hạ tầng, dữ liệu cho AI

Việc tối ưu hóa hơn nữa dữ liệu công khai để đào tạo các mô hình AI sẽ thúc đẩy đổi mới có ý nghĩa, phù hợp với bối cảnh Việt Nam. Chiến lược quốc gia về AI của Việt Nam nêu rõ mục tiêu

phát triển “50 bộ dữ liệu mở, liên kết và kết nối trong các khu vực kinh tế, lĩnh vực kinh tế - xã hội khác nhau” vào năm 2030 để thúc đẩy R&D về AI và ứng dụng AI vào năm 2030. Cần thực hiện kết hướng tới tối ưu hóa các bộ dữ liệu mở của Nhà nước, tạo điều kiện tiếp cận dữ liệu tiếng Việt có giá trị và chất lượng cao để sử dụng đào tạo các mô hình AI.

Một công thức cơ bản để phát triển hệ sinh thái AI thành công là đầu tư vào cơ sở hạ tầng đám mây, nghiên cứu và công nghệ cơ bản và ứng dụng (như bộ xử lý đồ họa và siêu máy tính), và các tập dữ liệu mở của chính phủ – và sau đó đưa ra các chính sách khuyến khích đổi mới và phát triển sản phẩm dựa trên các sáng kiến nền tảng này<sup>143</sup>. Kế hoạch tổng thể về cơ sở hạ tầng thông tin và truyền thông của Việt Nam giai đoạn 2021- 2030, tầm nhìn đến năm 2050 đặt ra một số mục tiêu trong việc phát triển cơ sở hạ tầng kỹ thuật thiết yếu cho sự phát triển AI như internet tốc độ cao, các trung tâm dữ liệu xanh và quy mô lớn, các KCN CNTT chuyên biệt, các trung tâm nghiên cứu và đổi mới. Chiến lược quốc gia về Nghiên cứu, phát triển và ứng dụng AI cũng nhấn mạnh, việc tăng cường điện toán đám mây là chìa khóa để khai thác hết tiềm năng của AI tại Việt Nam.

Dựa trên những chính sách này, tối đa hóa lợi ích của AI, cần áp dụng chính sách “Điện toán đám mây công cộng trên hết”, ưu tiên cơ sở hạ tầng và dịch vụ CNTT dựa trên đám mây công cộng và các phương án sử dụng đám mây khác so với dùng cơ sở hạ tầng tại chỗ. Vì đám mây công cộng sử dụng cơ sở hạ tầng dùng chung và tập hợp tài nguyên trong các trung tâm dữ liệu phân tán, nên nó có xu hướng mang lại hiệu quả hoạt động cao hơn nhiều trong các trường hợp sử dụng so với đám mây riêng. Việc ưu tiên các mô hình đám mây công cộng sẽ cho phép các doanh nghiệp tự động mở rộng quy mô tài nguyên điện toán và lưu trữ của mình trong khi vẫn đảm bảo tính bảo mật, điều này rất cần thiết cho khối lượng công việc AI. Điều này cũng sẽ cho phép các doanh nghiệp Việt Nam tận dụng việc triển khai AI mà không cần phải đầu tư nhiều vốn mà nếu không sẽ cần phải xây dựng hoàn chỉnh cơ sở hạ tầng AI từ đầu.

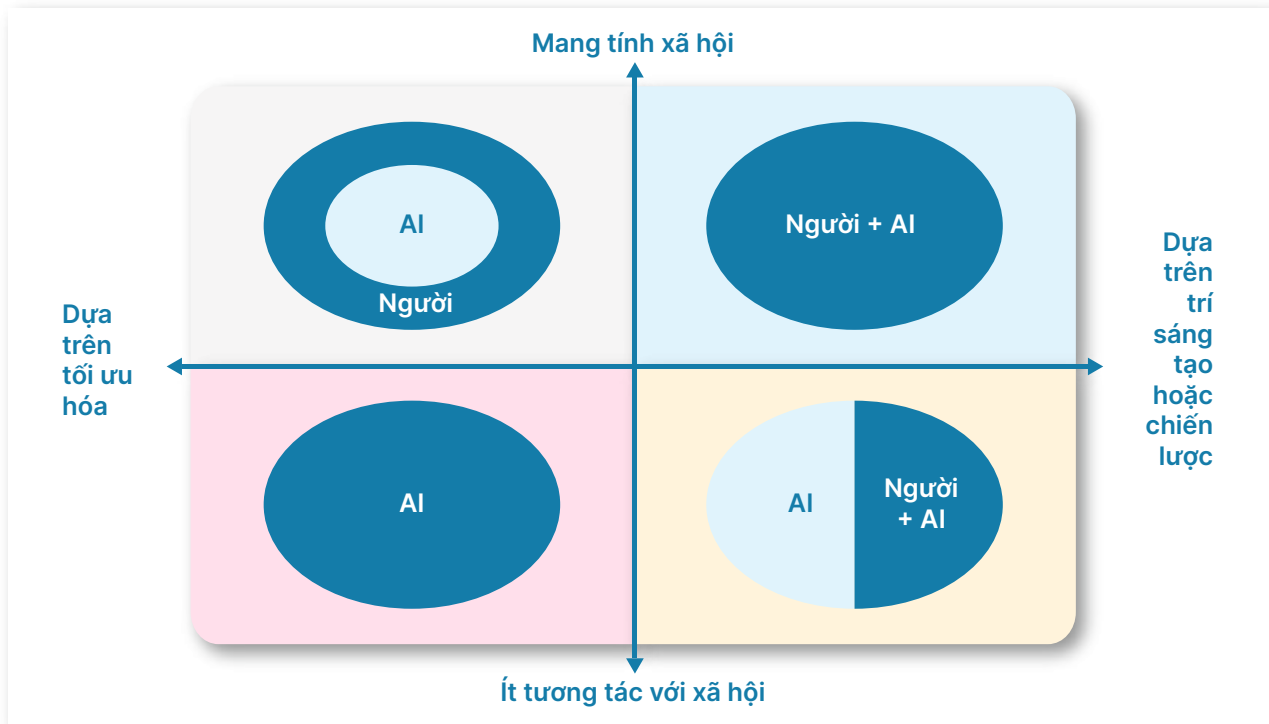
Tận dụng các giải pháp được triển khai trên cloud của các nhà cung cấp dịch vụ bên thứ ba được cho là con đường dễ dàng và nhanh chóng nhất nhằm ứng dụng LLM. Tuy nhiên, cần có các biện pháp hạn chế rủi ro bảo mật, tăng quyền kiểm soát hơn đối với các mô hình lõi, tăng hiệu suất làm việc ở các nghiệp vụ đặc thù. Các mô hình nguồn mở, mặt khác, mang lại cho doanh nghiệp nhiều quyền kiểm soát hơn, với chi phí thấp hơn, nhưng đòi hỏi phải có đội ngũ nội bộ có chuyên môn đặc thù để triển khai và tinh chỉnh. Còn việc tự xây dựng LLM hay AI tạo sinh cho phép doanh nghiệp kiểm soát chặt chẽ mô hình công nghệ, nhưng ngược lại, phải đánh đổi bằng chi phí rất lớn từ việc thu thập dữ liệu, tính toán, thuê các chuyên gia học máy để đào tạo và triển khai chúng. Chính sách quốc gia và bản thân doanh nghiệp có thể khuyến khích sử dụng mã nguồn mở để phát triển, ứng dụng AI, bởi hướng đi này giúp phát triển nhanh và tận dụng được các tài nguyên có sẵn, ví dụ khi cần phát triển AI để giải quyết vấn đề trong y tế, học thuật, giáo dục, có thể dùng DeepSeek, Llama, Gemma. Mặt khác, việc sử dụng mã nguồn mở là cách Việt Nam hay các nước nhỏ có thể đóng góp cho thế giới, tạo ra được cộng đồng có tiếng tăm và thu hút nhân tài để làm việc cho các dự án. Chẳng hạn, các trường đại học, doanh nghiệp Việt Nam có mô hình AI, ý tưởng đột phá, các bộ dữ liệu có thể phát triển dưới dạng mã nguồn mở để đóng góp cho nhân loại. Tuy nhiên, cần có các biện pháp kiểm soát rủi ro tiềm ẩn trong khi sử dụng mã nguồn mở.

### **5.2.3. Chính sách việc làm, phát triển nguồn nhân lực AI**

Để giải quyết vấn đề việc làm trong bối cảnh phát triển nhanh chóng của AI, câu hỏi mấu chốt ở đây là: AI làm được gì và không làm được gì. Khi đã xác định rõ câu trả lời, có thể định hình phương hướng chung về việc làm kể cả từ cấp độ cá nhân người lao động, tổ chức/doanh nghiệp, và chính sách quốc gia. Dựa trên việc phân tích bốn vùng tác động của AI đến việc làm, tác giả Kai Fu Lee đưa ra mô hình cộng sinh giữa con người với AI trong việc làm<sup>144</sup>.

<sup>143</sup> Google, *An AI Opportunity Agenda for Vietnam*, 2023.

<sup>144</sup> Kai-Fu Lee, *Các siêu cường AI: Trung Quốc, thung lũng Silicon và trật tự thế giới mới (sách dịch)*, NXB Trẻ, 2020; Chen Qiufan và Kai-Fu Lee, *AI 2041 - 10 viễn cảnh cho tương lai (sách dịch)*, NXB Thế Giới, 2024.



**Hình 11. Cộng sinh giữa con người với AI trong việc làm**

Dựa vào mô hình của Kai-Fu Lee trên đây, có thể thấy rằng ở vùng góc trái bên dưới, cần chấp nhận thực tế AI có thể thay thế hoàn toàn con người thực hiện các công việc. Ở các phần còn lại, việc con người cộng sinh với AI ở các mức độ khác nhau là cần thiết để phát triển. AI có thể thay thế một vài nhiệm vụ trong một ngành nghề nào đó của con người nhưng sẽ không thay thế hoàn toàn vị trí đó. Ví dụ như AI có thể hỗ trợ con người trong lĩnh vực y tế, chẩn đoán hình ảnh, phẫu thuật,... những không thể thay thế hoàn toàn sự tham gia của con người trong những công việc này. Việc cộng sinh có thể bao gồm việc con người và AI cùng hoạt động, hoặc con người có thể làm việc dựa vào những nền tảng đã được xử lý bởi AI. Đây là điều cần thiết và hợp lý khi con người cũng phải dần thích nghi với sự phát triển của xã hội bởi sự ra đời của những công nghệ mới, tương tự như ba cuộc cách mạng công nghiệp trong lịch sử nhân loại.

Như đã trình bày, AI thay đổi những yêu cầu về mặt kỹ năng, cho nên nhân sự cần thay đổi và đáp ứng được nhu cầu của thị trường lao động bằng cách tham gia đào tạo và cập nhật kiến thức. Trong một khảo sát, người lao động trong một khảo sát coi người sử dụng lao động có vai trò chính trong việc hỗ trợ nâng cao và đào tạo lại kỹ năng của họ trong bối cảnh công nghệ phát triển: 60% dựa vào người sử dụng lao động để có được hiểu biết sâu sắc về các kỹ năng và vai trò công việc khác nhau; 58% xác định người sử dụng lao động đóng vai trò quan trọng trong việc cung cấp các ưu đãi tài chính để nâng cao kỹ năng<sup>145</sup>. Các doanh nghiệp, tổ chức có thể áp dụng cách tiếp cận “ưu tiên kỹ năng trước” đối với những nhân viên hiện tại, giúp cả doanh nghiệp và người lao động thích ứng với quá trình chuyển đổi AI; đồng thời mở rộng nguồn nhân lực linh hoạt hơn cho các công việc của tương lai.

Nhiều công ty trên thế giới và Việt Nam đã tổ chức các chương trình đào tạo nội bộ về AI, như Google yêu cầu tất cả các kỹ sư phần mềm phải đăng ký vào chương trình giảng dạy về máy học. Nhưng xét đến tác động chuyển đổi của AI trên tất cả các lĩnh vực của nền kinh tế, các nỗ lực của từng công ty sẽ không đủ – các công ty sẽ cần phải thiết lập các quan hệ đối tác đào tạo AI liên ngành mới để đảm bảo rằng người lao động trong mọi ngành đều sẵn sàng khai thác AI.

Tuy nhiên, các vấn đề về việc làm vẫn cần sự can thiệp và hỗ trợ của chính phủ trong việc xây dựng các chính sách công để giải quyết những sự thay đổi lớn về công nghệ, cũng như phân phối thu nhập trong xã hội. Trong một khảo sát, theo 43% người trả lời, trong khi người sử dụng lao động đóng vai trò quan trọng trong việc nâng cao kỹ năng, chính phủ có thể tạo điều kiện cho người lao động tiếp cận các chương trình đào tạo kỹ năng đa dạng<sup>146</sup>. Trong đó phổ biến nhất là

<sup>145</sup> Ecomist Impact, Bridging the skills gap: Fuelling careers and the economy in Vietnam, 2023; [skilling\\_in\\_asia-vietnam\\_eng.pdf](#)

<sup>146</sup> Ecomist Impact, Bridging the skills gap: Fuelling careers and the economy in Vietnam, 2023; [skilling\\_in\\_asia-vietnam\\_eng.pdf](#)

tăng cường giáo dục, đào tạo, hỗ trợ người lao động; nhất là đối với những người có hoàn cảnh khó khăn, có thể hỗ trợ họ tìm lại việc làm, từ đó giảm thiểu xu hướng mất việc làm.

Nhà nước cần đóng vai trò tích cực trong việc giáo dục các kỹ năng mới, đào tạo lại người lao động để sử dụng AI hiệu quả hơn trong công việc và tạo điều kiện chuyển giao công việc suôn sẻ hơn khi công nghệ phát triển. Chiến lược AI quốc gia 2.0 của Singapore, chẳng hạn, tập trung vào nhu cầu đào tạo kỹ năng cho các loại năng lực khác nhau: Người sáng tạo AI (nhân tài AI hàng đầu), Người thực hành AI (nhân viên công nghệ) và Người dùng AI (doanh nghiệp và lực lượng lao động nói chung). Bên cạnh đó, Nhà nước có điều kiện tổng hợp thông tin về các rủi ro và cơ hội mà AI mang lại cho các nhóm người lao động cụ thể, qua đó đề ra các biện pháp hỗ trợ sát thực khác nhau, nắm bắt lợi ích của AI (năng suất tăng và tăng trưởng kinh tế) mà không làm tăng bất bình đẳng và sự phản kháng của xã hội đối với tiến bộ công nghệ.

Chương trình hỗ trợ cho những người lao động bị ảnh hưởng nhiều bởi AI cần được triển khai, với các bộ kỹ năng được thiết kế riêng, có thể thích ứng với các nhu cầu khác nhau của người lao động ở các khu vực địa lý khác nhau. Cần đặc biệt hỗ trợ người lao động có trình độ thấp và trung bình, có mức lương thấp, cộng đồng ở nông thôn, miền núi phải đối mặt với những trở ngại đáng kể khi tái gia nhập thị trường việc làm. Những nỗ lực hỗ trợ các chương trình có thể đảm bảo cho các cộng đồng ở mọi khu vực trên khắp Việt Nam có thể tận dụng các cơ hội kinh tế lớn hơn. Để thực hiện được điều này, chính phủ có thể hợp tác với khu vực tư nhân, các trường đại học để thiết kế các chương trình đào tạo phù hợp với nhu cầu của từng ngành; khuyến khích các doanh nghiệp phát triển các chương trình cấp chứng chỉ nghề nghiệp và học nghề làm việc liên ngành, phản ánh toàn bộ các kỹ năng cần thiết.

Thực tế, các tài liệu nghiên cứu trên thế giới và Việt Nam, cũng như ý kiến qua khảo sát của IPS cho thấy, xu hướng áp dụng AI để hỗ trợ một phần của công việc dường như phổ biến hơn so với việc thay thế hoàn toàn công việc bằng AI. Ví dụ: Một nhà kinh tế khi phải dự báo xu hướng kinh tế có thể dùng AI để tổng hợp dữ liệu (việc này đòi hỏi sự chính xác của đầu vào dữ liệu), đồng thời tự mình viết báo cáo và đưa ra hướng dẫn (đòi hỏi kỹ năng lý luận phức tạp, trừu tượng). Nhìn chung, không phải tất cả nhiệm vụ trong các ngành nghề đều có thể được thực hiện bởi AI. Như vậy, AI là công cụ đắc lực hỗ trợ để một ngành, một doanh nghiệp, một cá nhân tiến nhanh hơn chứ không thay thế con người. AI chỉ thay thế những người không thích ứng, còn vẫn tạo ra cho con người những không gian để làm việc và sáng tạo hơn. Chính vì vậy, các tổ chức, doanh nghiệp rất cần những người có khả năng sử dụng AI, biết cách làm việc cùng với AI để làm ra sản phẩm nhanh hơn, chất lượng tốt hơn, với giá rẻ hơn. Chính sách quốc gia cũng cần thích ứng để đào tạo, đào tạo lại, nâng cao năng lực cho người lao động để phù hợp với cấu trúc phân bổ việc làm đã thay đổi do AI.

Các nhà hoạch định chính sách có thể hỗ trợ người lao động bằng đào tạo/đào tạo lại và lưới an toàn, và định hình hệ thống giáo dục để giúp chuẩn bị cho người lao động cho thời đại AI, trong đó tư duy phản biện, sáng tạo và khả năng thích ứng có thể là những kỹ năng chính. Cuối cùng, các nhà hoạch định chính sách có thể nỗ lực để đảm bảo rằng sự thịnh vượng ngày càng tăng từ việc áp dụng AI được chia sẻ rộng rãi.

Bản thân các cơ quan Nhà nước cũng cần nhiều chuyên môn về AI hơn để khai thác hiệu quả AI. Chính phủ nên xây dựng và mở rộng quy mô kỹ năng AI “nội bộ” cho lực lượng lao động CNTT và các công chức khác. Có thể tổ chức các chương trình nâng cao năng lực AI cho cán bộ, công chức như tích hợp AI vào đào tạo, bồi dưỡng cán bộ, công chức tại các trường chính trị, hành chính; các khóa ngắn, dài hạn về nội dung này; các hoạt động trao đổi kinh nghiệm như hội thảo, tọa đàm, mạng lưới AI<sup>147</sup>. Chính phủ Singapore có Chương trình chuyển đổi ngắn hạn (Conversion Programme) đào tạo cho công chức tất cả các ngành kiến thức, kỹ năng áp dụng công nghệ số vào công việc của mình. Google.org đã tài trợ xây dựng Government AI Campus, bao gồm các khóa học, sự kiện và nội dung tập trung vào việc giúp các công chức toàn cầu nâng cao kỹ năng về AI. Đây là những mô hình có thể học hỏi để điều chỉnh, áp dụng tích hợp vào các chương trình đào tạo, bồi dưỡng cán bộ, công chức hiện nay ở Việt Nam.

<sup>147</sup> Xem thêm: Viện Nghiên cứu chính sách và phát triển truyền thông (IPS) & Chương trình phát triển LHQ (UNDP), Báo cáo Đánh giá nhanh ứng dụng AI trong các cơ quan Nhà nước ở Việt Nam: Mức độ sẵn sàng và khuyến nghị, Hà Nội, 2025.

#### 5.2.4. Chính sách giáo dục, nghiên cứu AI

Đào tạo nhân lực AI cần sát với nhu cầu của các cơ quan, tổ chức, doanh nghiệp. Các chương trình đào tạo có liên quan đến AI không chỉ tập trung vào công nghệ, mà cần xây dựng năng lực thích ứng tốt với giai đoạn chuyển đổi. Các kiến thức, kỹ năng cần được trang bị hướng tới hai nhóm nhân lực chính: nhóm phát triển mô hình AI, LLM (nhân sự trong lĩnh vực công nghệ, với năng lực kỹ thuật đặc thù) và nhóm ứng dụng mô hình AI, LLM (nhân sự trong các lĩnh vực kinh doanh, marketing, chăm sóc khách hàng,... biết cách sử dụng hiệu quả các công cụ công nghệ mới). Ngay cả những chuyên gia trong lĩnh vực công nghệ và dữ liệu cũng cần liên tục trau dồi kiến thức về cách các mô hình này hoạt động để tự tin ứng dụng chúng trong công việc.

Sẽ có những vị trí việc làm mới mà doanh nghiệp cần các trường đại học đào tạo để tuyển dụng, chẳng hạn như chuyên gia về ngôn ngữ học, kiểm soát chất lượng AI, chỉnh lý AI (AI editors), kỹ sư (prompt engineers). Ở những lĩnh vực mà AI có tiềm năng phát triển lớn nhất, mức độ mà AI có thể ảnh hưởng đến từng nhóm công việc của tổ chức, doanh nghiệp với mức độ khác nhau - hoàn toàn tự động hóa, ảnh hưởng một phần, hoặc không bị ảnh hưởng. Như vậy, chương trình đào tạo nhân lực AI cho các tổ chức, doanh nghiệp cũng cần xuất phát từ đặc điểm này.

Việt Nam cần một hệ thống giáo dục và đào tạo chuẩn bị cho người lao động có một phương pháp học tập suốt đời, trang bị cho tất cả học sinh và người lao động các kỹ năng AI cơ bản trong suốt sự nghiệp của họ. Điều này cũng có nghĩa là coi AI là một thành phần cốt lõi của hệ thống giáo dục và phát triển chuyên môn tại Việt Nam. Cần cập nhật khuôn khổ chương trình giảng dạy, tăng cường giáo dục STEM với trọng tâm vào kiến thức về AI (đồng thời tránh cách tiếp cận hạn hẹp như 'học cách viết mã', bởi lẽ AI tạo sinh có các kỹ năng viết mã cơ bản) và nhấn mạnh vào các mô hình học tập dựa trên kỹ năng, bao gồm các chương trình học nghề. Cần tạo cơ hội học hỏi kỹ năng cho những người lao động trẻ tuổi để tạo điều kiện chuyển đổi suôn sẻ sang các công việc ứng dụng AI.

Cần xây dựng cơ sở hạ tầng để hỗ trợ nghiên cứu và phát triển AI, với những điểm cụ thể về nguồn vốn, quy mô, vị trí và cơ sở vật chất để hỗ trợ các hoạt động nghiên cứu AI tập trung. Để tham khảo, như một chuyên gia AI người Việt đang giảng dạy ở Pháp, đồng thời là chủ một doanh nghiệp khởi nghiệp AI cho biết, Chính phủ Pháp có cơ sở hạ tầng tính toán lớn, mở cho doanh nghiệp, lab nghiên cứu có thể tiếp cận, sử dụng. Doanh nghiệp phối hợp với lab ở trường đại học làm R&D được Chính phủ cấp tiền và giảm thuế, cụ thể là doanh nghiệp trả lương cho nhà khoa học làm nghiên cứu, phần trả lương sau đó được nhà nước giảm vào thuế. Công việc mà nhà khoa học AI làm gắn với kinh doanh của doanh nghiệp; Chính phủ chỉ đầu tư tiền, còn nội dung công việc do doanh nghiệp quyết định để đảm bảo sát nhu cầu, chất lượng và hiệu quả. Tuy nhiên, cần lưu ý, không phải tất cả các nghiên cứu AI đều yêu cầu các cơ sở vật chất giống nhau. Do đó, bất kỳ chiến lược cơ sở hạ tầng AI nào cũng cần phản ánh các lĩnh vực trọng tâm, giúp Việt Nam hỗ trợ hiệu quả các công nghệ AI cụ thể theo cách hiệu quả hơn, có mục tiêu hơn, quản lý các nguồn lực một cách hiệu quả hơn.

Để hỗ trợ một hệ sinh thái R&D hiệu quả cho AI trong nước, cần sửa đổi, bổ sung các quy định hiện hành để tạo điều kiện thuận lợi hơn cho đầu tư vào nghiên cứu khoa học trong Luật Khoa học và công nghệ, cũng như các nghị định, thông tư có liên quan. Trong đó, cần đảm bảo rằng giai đoạn thử nghiệm và sai sót được khuyến khích trong cả ngành công nghiệp và giới học thuật, nghiên cứu, và cả trong đặt hàng của cơ quan Nhà nước. Cần sự hỗ trợ từ các quỹ của chính phủ, cũng như các khoản đầu tư tư nhân, tạo điều kiện cho các cơ quan Nhà nước yên tâm đầu tư vào nghiên cứu, mua, thuê các sản phẩm, dịch vụ AI phục vụ hoạt động của các cơ quan Nhà nước. Cần tăng chi tiêu công cho nghiên cứu và phát triển các lĩnh vực khoa học công nghệ chủ chốt, trong đó có trí tuệ nhân tạo; khuyến khích các hoạt động chi cho nghiên cứu và phát triển trong khu vực tư, hỗ trợ tiếp cận với các nguồn vốn cho nghiên cứu phát triển khoa học công nghệ cả trong nước và quốc tế.

Cần một khuôn khổ chính sách, pháp luật về bản quyền hỗ trợ sự đổi mới và sáng tạo, trong đó có các quy định về ngoại lệ cho phép các nhà phát triển dùng dữ liệu có sẵn công khai đào tạo các mô hình AI. Để các hệ thống AI có thể học hỏi và tương tác với nhiều nguồn thông tin và tập dữ liệu khác nhau, các khuôn khổ bản quyền phải cho phép sử dụng rộng rãi các dữ liệu đầu vào. Một ví dụ tham khảo là Singapore đã cập nhật Luật Bản quyền vào năm 2021, quy định ngoại lệ đối với phân tích dữ liệu tính toán, tạo cơ sở pháp lý chắc chắn cho các nhà nghiên cứu, nhà đổi mới và các công ty AI. Các quy định về quyền riêng tư, bảo vệ dữ liệu cá nhân cần hài hòa, cân bằng với bản quyền.



Để giải quyết các khía cạnh toàn diện và như đánh giá về sự phát triển và ứng dụng của công nghệ AI đối với nền kinh tế và xã hội rộng lớn hơn của Việt Nam (và môi trường), các trường đại học và tổ chức kinh tế và xã hội cũng cần tham gia. Chính phủ Việt Nam cần tạo cơ chế ở cấp Bộ để tạo điều kiện liên kết giữa tất cả các trường đại học để tham gia vào các hoạt động nghiên cứu AI rộng rãi hơn. Việt Nam cũng cần cung cấp nhiều kinh phí hơn cho các trường đại học và viện để tiến hành nghiên cứu AI và tạo điều kiện trao đổi kiến thức và chuyên gia giữa các tổ chức trong nước và nước ngoài. Điều này sẽ mở rộng mạng lưới giữa các nhà nghiên cứu Việt Nam và các chuyên gia nước ngoài để tiến hành nghiên cứu với các chủ đề rộng hơn và có liên quan.

Quan hệ đối tác công - tư có thể có hiệu quả trong việc thúc đẩy nghiên cứu, tận dụng nguồn lực chung cho hệ sinh thái AI. Cả Nhà nước và khu vực tư đều có thể hỗ trợ các nhà nghiên cứu thông qua các chương trình như khuôn khổ chuyển giao công nghệ, học bổng và hỗ trợ trực tiếp cho nghiên cứu. Các sáng kiến công - tư như AI Trailblazers giữa Chính phủ Singapore và Google Cloud nên được học hỏi, áp dụng một cách phù hợp với thực tế ở Việt Nam.

Bên cạnh đó, như thực tế cho thấy, cần khuyến khích các hoạt động, chương trình của khu vực tư, doanh nghiệp, tổ chức, cá nhân phổ biến kiến thức, kỹ năng AI. Cụ thể như Google (Google for Startups Accelerator), hợp tác với Trung tâm đổi mới sáng tạo quốc gia (NIC) thuộc Bộ Kế hoạch - đầu tư để phổ cập AI trên toàn quốc trong giáo viên, học sinh và nhà phát triển; cung cấp 40.000 học bổng Chứng chỉ nghề nghiệp của Google. Nỗ lực của Viện Blockchain và AI thuộc Hiệp hội Blockchain đưa kiến thức, kỹ năng AI đến 1 triệu học viên là học sinh, sinh viên, giáo viên, người lao động, công chức. Các diễn đàn trên mạng xã hội về AI như “Bình dân học AI” cũng góp phần phổ cập AI đối với cộng đồng.

Bên cạnh nhân lực AI trong nước, người Việt Nam đi học các ngành liên quan về AI ở các nước, sau đó ở lại trường đại học để nghiên cứu hoặc đi làm ở các doanh nghiệp tại nước sở tại là đầu mối tốt để liên kết cộng đồng các nhà nghiên cứu, sinh viên Việt Nam với các trường đại học, doanh nghiệp ở nước ngoài.

### **5.2.5. Khuyến nghị về cơ chế tài chính**

Khuôn khổ chính sách, pháp luật về đầu tư, cơ chế tài chính cho khoa học, công nghệ, trong đó có AI cần được sửa đổi, bổ sung một số nội dung sau đây:

Cần có những quy định pháp luật cụ thể, riêng biệt về quản lý đầu tư cho AI từ kinh phí ngân sách Nhà nước, từ việc lập dự toán, quản lý, sử dụng, quyết toán kinh phí; phương pháp định giá, định mức cho những mục chi như chi cho thu thập, tổng hợp, kiểm tra, làm sạch dữ liệu; huấn luyện, tinh chỉnh mô hình/ứng dụng AI; vận hành giải pháp AI trong công việc v.v... Trong đó, chú ý tách biệt chi phí cho việc huấn luyện mô hình AI (vốn hoàn toàn khác xây dựng phần mềm) với chi phí xây dựng hệ thống AI như hệ thống ra quyết định, UI, UX, quản trị dữ liệu v.v... Đồng thời, cần phải quy định rõ ràng, cụ thể hơn về nguồn kinh phí thường xuyên hàng năm đối với việc vận hành, bảo trì hệ thống, huấn luyện tối ưu các mô hình, ứng dụng AI.

Cần chuyển từ khuôn khổ mua sắm công truyền thống sang cơ chế linh hoạt, phù hợp hơn với bản chất thay đổi nhanh chóng của AI. Sự thay đổi này bao gồm việc hợp lý hóa các quy trình mua sắm thông qua các quy định và điều chỉnh được cập nhật đối với khuôn khổ mua sắm công, xây dựng các hướng dẫn bao gồm các thông lệ tốt và các cân nhắc về đạo đức, và thúc đẩy sự hợp tác với các chuyên gia AI và các bên liên quan trong ngành.

Cần chú ý lựa chọn và đánh giá trước các nhà cung cấp để đảm bảo họ đủ điều kiện cung cấp giải pháp AI cho các cơ quan Nhà nước. Ví dụ, lập danh sách tham khảo về các nhà cung cấp đủ điều kiện có thể cung cấp các dịch vụ, giải pháp và sản phẩm AI có trách nhiệm và hiệu quả. Danh sách này không chỉ có các doanh nghiệp Nhà nước, các doanh nghiệp lớn, mà cả những doanh nghiệp nhỏ và vừa, miễn là họ đáp ứng đầy đủ các yêu cầu về chuyên môn. Như ở Singapore, cơ quan chính phủ có nhu cầu xây dựng giải pháp, ứng dụng AI cần phải lựa chọn nhà thầu từ danh sách này trước, nếu không được thì mới được lựa chọn từ các nguồn khác<sup>148</sup>. Cách làm này vừa đảm bảo hiệu quả mua sắm công, chất lượng nhà cung cấp; đồng thời hỗ trợ phát triển các doanh nghiệp AI nội địa có thêm khách hàng từ các cơ quan Nhà nước.

<sup>148</sup> Public Sector AI Playbook, 2024.

Bên cạnh đó, Chính phủ nên thiết lập một số chương trình hỗ trợ các cơ quan Nhà nước đặt hàng các cơ sở nghiên cứu, doanh nghiệp phát triển các giải pháp AI cho các nhiệm vụ cụ thể của cơ quan, ngành mình. Cụ thể hơn, có thể nghiên cứu, học hỏi Chương trình 100 Experiments (100E) và Chương trình Tài trợ chuyển đổi R&D cho các cơ quan Nhà nước của Chính phủ Singapore (the Public Sector Translational R&D Grant - TRANS Grant)<sup>149</sup>. Theo đó, chương trình này có ngân sách cố định; các cơ quan Nhà nước gửi yêu cầu phát triển các giải pháp AI cụ thể; chương trình sẽ cấp kinh phí cho doanh nghiệp hoặc cơ sở nghiên cứu có quan tâm, đủ năng lực phát triển giải pháp AI, nếu đáp ứng đúng yêu cầu về hiệu quả, tính khả thi trong quá trình thử nghiệm (testing), giải pháp sẽ được triển khai trong thực tế, kinh phí sẽ được thanh toán cho doanh nghiệp/cơ sở nghiên cứu.

### 5.3. CHÍNH SÁCH KIỂM SOÁT AI

Cần xem xét rủi ro, tác động cụ thể của AI đối với xã hội, con người, quyền con người trong toàn bộ quá trình thiết kế, phát triển, triển khai, ứng dụng các hệ thống AI<sup>150</sup>. Điều này nhằm đảm bảo rằng AI không chỉ hoạt động hiệu quả về mặt kỹ thuật, mà còn phù hợp với các giá trị xã hội và thúc đẩy, bảo vệ các quyền con người.

#### 5.3.1. Đánh giá, điều chỉnh rủi ro theo vòng đời của hệ thống AI

Như đã được khuyến cáo hoặc thừa nhận trong các tổ chức quốc tế, chính phủ các nước, các chuyên gia, cơ hội và rủi ro của AI đối với xã hội, con người, quyền con người có thể được đánh giá, điều chỉnh, ứng phó theo vòng đời của các hệ thống AI. Trong từng công đoạn của vòng đời này, một hệ thống AI có thể mang lại cơ hội thúc đẩy hoặc đặt ra thách thức, rủi ro cụ thể khác nhau, đòi hỏi biện pháp, đối sách phù hợp (xem ví dụ về cách đánh giá rủi ro cụ thể đối với quyền con người trong từng giai đoạn của vòng đời hệ thống AI trong phần Phụ lục).

Đồng thời, trong vòng đời của hệ thống AI, cần xác định rõ ràng, cụ thể trách nhiệm của từng chủ thể: Nhà phát triển AI, nhà triển khai triển khai/ứng dụng, người dùng cuối. Theo các chuyên gia về AI mà IPS đã trao đổi, nhà phát triển mô hình AI có khả năng kiểm soát dữ liệu - vấn đề cốt lõi trong việc tạo ra các mô hình AI. Nhà phát triển có khả năng đảm bảo mô hình AI được xây dựng với mục đích đúng đắn, dữ liệu được dùng để huấn luyện AI có chất lượng, không thiên lệch, lọc các nội dung được cho là độc hại (mặc dù khó kiểm soát nội dung được tạo ra từ mô hình AI vì cùng một câu hỏi nhưng mô hình AI có thể trả lời khác nhau ở những thời điểm khác nhau); kiểm soát được chất lượng và tính năng liên quan đến mô hình. Trong khi đó, bên triển khai/ứng dụng có khả năng kiểm soát cách sử dụng mô hình AI, nhưng không thể kiểm soát chất lượng mô hình. Chẳng hạn họ có khả năng đánh giá tỉ lệ sai của mô hình AI, nhưng không thể chỉnh sửa mô hình - ví dụ như thêm dữ liệu đầu vào hay xây dựng thêm một mô hình khác để lọc nội dung. Do đó, kiểm soát việc nhà phát triển tạo ra mô hình AI an toàn, cũng như kiểm soát việc triển khai, ứng dụng mô hình một cách đúng đắn là hoàn toàn phù hợp, khả thi.

#### 5.3.2. Đánh giá, điều chỉnh theo mức độ rủi ro của AI

Các hệ thống hoặc ứng dụng AI có tác động khác nhau đối với quyền con người tùy theo mức độ rủi ro của từng hệ thống/ứng dụng. Vì vậy, cần có đánh giá để có sự điều chỉnh AI phù hợp với từng mức độ rủi ro của AI. Có thể tham khảo cách phân loại bốn mức độ rủi ro trong Đạo luật về AI của Liên minh châu Âu (EU). Theo đó, ở mức độ rủi ro không thể chấp nhận được, các ứng dụng AI được coi là vi phạm các giá trị cốt lõi của loài người như quyền con người bị cấm hoàn toàn; các hệ thống AI rủi ro cao đối với con người phải tuân theo một hệ thống xác nhận chi tiết; đối với AI có mức độ rủi ro hạn chế, các nhà sản xuất phải minh bạch, công khai nội dung chịu tác động; AI có mức độ rủi ro tối thiểu được điều chỉnh bởi các quy tắc ứng xử tự nguyện. Điều chỉnh theo mức độ rủi ro có ưu điểm là xác định được biện pháp phù hợp đối với từng mức độ, tránh được cả hai thái cực là quá chặt chẽ và quá lỏng lẻo, đảm bảo được yêu cầu cân bằng, dung hòa như đã đề cập ở trên. Đồng thời, theo chuyên gia về AI mà IPS đã phỏng vấn, cách tiếp cận này khả thi về mặt kỹ thuật, công nghệ.

#### 5.3.3. Hướng đến AI có trách nhiệm và đáng tin cậy

<sup>149</sup> Public Sector AI Playbook, 2024.

<sup>150</sup> Evgeni Aizenberg and Jeroen van den Hoven, *Designing for Human Rights in AI*; Bureau of Cyberspace and Digital Policy, *Risk Management Profile for Artificial Intelligence and Human Rights*, 7/2024, <https://www.state.gov/risk-management-profile-for-ai-and-human-rights/>.

Kỳ vọng trở thành một nền kinh tế phát triển đòi hỏi Việt Nam xem xét áp dụng các tiêu chuẩn, yêu cầu về AI trách nhiệm, đáng tin cậy đã được các quốc gia trưởng thành hơn về công nghệ đề xuất và áp dụng.

**Thúc đẩy phát triển AI có đạo đức:** Cần cân nhắc, thảo luận kỹ lưỡng những vấn đề đạo đức trong phát triển và triển khai các hệ thống AI; trong đó đảm bảo rằng các thuật toán AI được thiết kế giảm thiểu thiên kiến và thúc đẩy sự công bằng. Các doanh nghiệp cần tuân thủ các chính sách đạo đức nghiêm ngặt khi triển khai công nghệ AI. Tuy nhiên, Việt Nam nên thực hiện từng bước để tích hợp đạo đức vào phát triển, triển khai, ứng dụng AI, bắt đầu bằng việc nâng cao nhận thức về chủ đề này và trao đổi nhiều hơn các bên trong nước và khu vực. Hợp tác với các nước láng giềng ASEAN để điều chỉnh bộ nguyên tắc AI hiện có sao cho phù hợp nhất với nhu cầu trong nước và khu vực là một phương án phù hợp.

**Minh bạch và trách nhiệm giải trình:** Minh bạch là yếu tố quan trọng trong việc đảm bảo rằng công chúng có thể giám sát và truy cứu trách nhiệm các hệ thống AI và những người phát triển chúng. Công chúng cần được thông tin về cách thức và nơi AI được sử dụng; về các thuật toán được sử dụng trong các quy trình ra quyết định; giúp người dân hiểu và đánh giá các quyết định do hệ thống AI đưa ra có ảnh hưởng đến quyền lợi của họ. Việc sử dụng công nghệ AI bởi chính phủ cần tuân theo các tiêu chuẩn cao về minh bạch, trách nhiệm giải trình và khả năng giải thích của hệ thống AI. Cần tạo các cơ chế để các nhà phát triển và người sử dụng AI chịu trách nhiệm về các vi phạm quyền con người. Điều này bao gồm việc thành lập các cơ quan quản lý để giám sát các ứng dụng AI và đảm bảo tuân thủ các tiêu chuẩn về quyền con người.

**Giám sát sử dụng AI:** Cần thực hiện các biện pháp giám sát việc sử dụng AI, ví dụ áp đặt các hạn chế đối với việc triển khai công nghệ nhận dạng khuôn mặt ở các không gian công cộng để bảo vệ quyền riêng tư của cá nhân. Cần xem xét thành lập thiết chế giám sát, theo dõi việc triển khai AI và đảm bảo tuân thủ các hướng dẫn đạo đức và yêu cầu pháp lý. Chẳng hạn, ở khá nhiều nước có các Ủy ban đạo đức AI xem xét việc sử dụng AI trong các dịch vụ công, đưa ra các khuyến nghị để đảm bảo rằng các hệ thống AI không vi phạm quyền lợi của cá nhân, giải quyết các khiếu nại liên quan đến việc triển khai AI.

**Khắc phục hậu quả do vi phạm:** Cần có các biện pháp khắc phục dành cho những chủ thể có quyền lợi bị vi phạm do việc triển khai, sử dụng các hệ thống AI. Có thể lồng ghép nội dung về AI vào các cơ chế khiếu nại, khiếu kiện hiện hành ở Việt Nam, đảm bảo rằng có các quy trình rõ ràng để các cá nhân có thể thách thức các quyết định do AI đưa ra, đặc biệt là trong các lĩnh vực như thực thi pháp luật, dịch vụ xã hội, tuyển dụng, sử dụng lao động.

#### 5.3.4. Giải pháp lập pháp

Trong xây dựng khung pháp luật kiểm soát AI, không có công thức duy nhất nào có thể áp dụng cho tất cả các quốc gia. Nhưng có thể áp dụng những bài học kinh nghiệm từ những cách tiếp cận trên thế giới vào bối cảnh Việt Nam để thực hiện những bước đi nhằm giảm thiểu tác hại do các hệ thống AI gây ra. Cách tiếp cận hoàn toàn theo chiều rộng không thể đặt ra các yêu cầu cụ thể cho tất cả các công cụ AI. Ngược lại, nếu chỉ quy định cụ thể nhằm mục tiêu vào từng công cụ AI có thể sẽ tạo ra sự trùng lặp, chông chéo cho cả cơ quan quản lý và doanh nghiệp, gây gánh nặng tuân thủ cho các doanh nghiệp, tổ chức. Cách tiếp cận hiệu quả nhất là kết hợp cả các yếu tố theo chiều rộng và chiều hẹp, tức là vừa ban hành khung pháp luật chung về AI, vừa có những quy định cho các công cụ AI cụ thể. Cách tiếp cận theo chiều rộng điều chỉnh được những vấn đề chung; còn theo chiều hẹp điều chỉnh những vấn đề đặc thù của từng loại AI.

Trong đó, khung pháp luật chung về AI cần có các nội dung lớn sau đây: đưa ra khái niệm rõ ràng về AI, hệ thống AI, mô hình AI để làm điểm xuất phát cho các nội dung điều chỉnh khác về công nghệ này; xác định các nguyên tắc cốt lõi trong quản trị AI gồm: lấy con người, các giá trị của con người, quyền con người làm căn cứ cho toàn bộ vòng đời của hệ thống AI, minh bạch, trách nhiệm giải trình, đáng tin cậy; xác định các mức độ rủi ro và các biện pháp thích ứng với mức độ rủi ro; xác định trách nhiệm cụ thể của các chủ thể liên quan trong vòng đời của hệ thống AI gồm bên phát triển, bên triển khai/ứng dụng, người dùng AI. Trong đó, cần quy định trách nhiệm của các bộ, ngành về việc xây dựng, thực hiện các bộ tiêu chuẩn kỹ thuật, bộ quy tắc ứng xử về đạo đức AI có tính chất tự nguyện.

Đặc biệt, khung pháp luật cần quy định rõ ràng, chặt chẽ đối với các trường hợp sau đây:

- Các hệ thống AI phải bị hạn chế trong những trường hợp chúng gây ra mối đe dọa đáng kể và sắp xảy ra đối với tính mạng hoặc sự an toàn của cá nhân hoặc nhóm người.
- Cấm các hệ thống AI gây ra phân biệt đối xử, đảm bảo quyền bình đẳng cho tất cả mọi người theo Điều 16.1 Hiến pháp 2013. Điều này bao gồm cả phân biệt đối xử trực tiếp (ví dụ: dựa trên độ tuổi, giới tính, tôn giáo) và phân biệt đối xử gián tiếp (ví dụ: thông qua điểm số xã hội dẫn đến đối xử bất lợi). Khung này cần yêu cầu tính minh bạch về thuật toán trong quy trình ra quyết định của AI để ngăn ngừa sự thiên vị và kết quả phân biệt đối xử, ví dụ thuật toán tuyển dụng và bắt buộc kiểm toán để ngăn ngừa các kết quả thiên vị trong các quyết định tuyển dụng.
- Bảo vệ quyền riêng tư và dữ liệu cá nhân: Các hệ thống AI không được xâm phạm đời tư, bí mật cá nhân hoặc bí mật gia đình của cá nhân mà không có sự đồng ý rõ ràng. Việc thu thập, lưu trữ, sử dụng và phổ biến dữ liệu cá nhân, bao gồm dữ liệu sinh trắc học và hình ảnh cá nhân, phải tuân theo các quy định và giám sát chặt chẽ. Cần xây dựng các hướng dẫn rõ ràng, thủ tục đơn giản và minh bạch để có được sự đồng ý của cá nhân khi sử dụng dữ liệu của họ trong các hệ thống AI; với các điều khoản về việc rút lại sự đồng ý bất cứ lúc nào.
- Bảo vệ danh dự, phẩm giá và uy tín: Các hệ thống AI phải bị cấm thực hiện các hành động xâm phạm danh dự, phẩm giá và uy tín của cá nhân, phù hợp với các nguyên tắc văn hóa và pháp lý của Việt Nam. Khung này cần giải quyết các tác động về mặt đạo đức của công nghệ AI, ngăn chặn việc sử dụng sai mục đích có thể gây tổn hại đến danh tiếng của cá nhân hoặc dẫn đến kỳ thị xã hội.
- Việc sử dụng AI trong điều tra tội phạm phải được quản lý chặt chẽ, chỉ cho phép truy cập dữ liệu cá nhân theo sự cho phép của tòa án, với các điều khoản dành các tình huống khẩn cấp.
- Bảo vệ các nhóm dễ bị tổn thương: Cần có các biện pháp bảo vệ đặc biệt dành cho các nhóm yếu thế, chẳng hạn như người già, người khuyết tật hoặc người có hoàn cảnh kinh tế khó khăn. Các hệ thống AI khai thác những điểm yếu này, có khả năng gây ra tác hại về mặt tâm lý hoặc hành vi, cần được quản lý chặt chẽ. Điều này bao gồm đảm bảo rằng các công nghệ AI không thao túng hoặc ép buộc những cá nhân dễ bị tổn thương hành động trái với lợi ích của họ.

Đồng thời, cần ban hành các văn bản quy phạm pháp luật, có thể là nghị định, thông tư quy định cụ thể hơn về các vấn đề kỹ thuật, chuyên sâu liên quan đến AI, chẳng hạn như điều chỉnh thuật toán để đảm bảo sự minh bạch, công bằng, tránh phân biệt đối xử; sử dụng công nghệ nhận diện khuôn mặt; sử dụng AI trong một số lĩnh vực tiềm ẩn nhiều rủi ro như tuyển dụng, sử dụng, đánh giá lao động v.v....

### **5.3.5. Tích hợp quyền con người trong kiểm soát rủi ro của AI**

**Cơ sở pháp lý để tích hợp quyền con người trong kiểm soát rủi ro của AI:** Việt Nam đã ký kết, tham gia hầu hết các điều ước quốc tế có nội dung liên quan đến quyền con người; Hiến pháp năm 2013 dành một chương riêng (Chương II) về quyền con người, quyền công dân; nhiều đạo luật có các nội dung cụ thể hóa quy định của Hiến pháp về quyền con người (xem cụ thể hơn trong Danh mục các văn bản quy phạm pháp luật có liên quan đến điều chỉnh rủi ro của AI). Chính vì vậy, càng có cơ sở chính trị - pháp lý vững chắc để đánh giá, điều chỉnh rủi ro của AI đối với quyền con người ở Việt Nam.

Công ước khung về AI với quyền con người, dân chủ và pháp quyền nhấn mạnh tại Điều 4: Quốc gia thành viên cần phải thực hiện các biện pháp nhằm đảm bảo rằng các hoạt động trong vòng đời các hệ thống AI nhất quán với các nghĩa vụ bảo vệ quyền con người được quy định trong các văn kiện luật quốc tế có hiệu lực và pháp luật quốc gia. Mặc dù Việt Nam chưa ký kết, tham gia Công ước này, nhưng có thể xem xét áp dụng cách tiếp cận dựa trên quyền con người này nhằm đảm bảo rằng các hệ thống AI được thiết kế, triển khai, ứng dụng gắn với bảo vệ quyền con người. Cần xác định một nguyên tắc là không đánh đổi một bước lùi dù nhỏ nhất về quyền con người cho bất cứ một phát triển nào của công nghệ.

**Tuân thủ các yêu cầu về giới hạn quyền con người:** Giới hạn quyền (limitation of rights) được công nhận trong Tuyên ngôn thế giới về nhân quyền 1948 (UDHR) và những điều ước quốc tế về quyền con người, nhưng đều phải đáp ứng một số yêu cầu. Khoản 2, Điều 29 UDHR qui định: “Mỗi người trong khi thực hiện các quyền và tự do cho cá nhân chỉ phải chịu những hạn chế do luật định nhằm mục đích duy nhất là bảo đảm việc thừa nhận và sự tôn trọng đối với các quyền và tự do của những người khác và phù hợp với những đòi hỏi chính đáng về đạo đức, trật tự công cộng và sự phồn vinh chung trong một xã hội dân chủ”. Tương tự, theo Điều 14.2, Hiến pháp năm 2013 của Việt Nam, quyền con người và quyền công dân chỉ có thể bị hạn chế theo luật trong những trường hợp cần thiết cụ thể, chẳng hạn như quốc phòng, an ninh, trật tự xã hội và phúc lợi công cộng.

Nguyên tắc trên đây của luật pháp quốc tế và quy định của Hiến pháp 2013 về giới hạn quyền con người phải là trọng tâm trong các quy định và thực thi pháp luật về AI của Việt Nam. Việc triển khai, ứng dụng AI nếu có mục đích hoặc hệ quả giới hạn quyền con người cũng cần phải tuân theo nguyên tắc này. Cụ thể hơn, bất cứ biện pháp nào nhằm hạn chế bất kỳ quyền con người nào đều cần phải tuân thủ các điều kiện sau đây<sup>151</sup>:

- Việc giới hạn quyền phải được qui định rõ ràng trong luật. Không chỉ vậy, những qui định của luật đặt ra giới hạn đối với việc thực hiện quyền cần phải: Công khai với người dân và có chỉ dẫn thích hợp để mọi người có thể hiểu sự hạn chế luật định đối với các quyền của họ; qui định về giới hạn quyền trong luật phải chính xác, rõ ràng để người dân có thể hiểu rõ và tự điều chỉnh hành vi của họ; có những biện pháp thích hợp để phòng ngừa việc lạm dụng qui định giới hạn quyền, hoặc tùy tiện đặt ra các giới hạn mới.
- Giới hạn đặt ra không được trái với bản chất của các quyền bị giới hạn. Việc giới hạn quyền phải nhằm những mục đích chính đáng/hợp pháp (legitimate objective); biện pháp hạn chế phải hợp lý (reasonable), cần thiết (necessary) và tương xứng (proportion).
- Chỉ được đặt ra giới hạn quyền nếu điều đó là cần thiết và nhằm mục đích thúc đẩy phúc lợi chung của cộng đồng hoặc để bảo vệ an ninh quốc gia, an toàn của cộng đồng, trật tự công cộng, sức khỏe cộng đồng, đạo đức xã hội hoặc quyền tự do của người khác.

Các điều ước quốc tế về quyền con người, Hiến pháp, các luật đã ban hành là cơ sở pháp lý để thừa nhận, bảo vệ, bảo đảm các quyền con người nói chung, cũng như quyền con người trong vòng đời AI. Mặt khác, cần khắc phục những khoảng trống về chính sách, pháp luật có thể dẫn đến rủi ro AI được sử dụng theo cách gây hại cho cá nhân hoặc toàn xã hội, ảnh hưởng đến quyền con người.

### **Thiết kế hệ thống AI duy trì các giá trị chung của loài người**

Cần chuyển đổi các giá trị cốt lõi của con người, quyền con người thành các đặc điểm kỹ thuật và tính năng kỹ thuật trong hệ thống AI như: thiết kế các thuật toán ưu tiên công bằng, phát triển giao diện người dùng tôn trọng quyền riêng tư, hoặc tạo ra các cơ chế cho tính minh bạch và trách nhiệm. Ví dụ, nếu công bằng là một giá trị quan trọng, quá trình thiết kế sẽ liên quan đến việc tạo ra các thuật toán được kiểm tra để loại trừ thiên kiến và đảm bảo rằng kết quả của hệ thống AI đảm bảo tính công bằng.

**Duy trì phẩm giá:** Hệ thống AI cần được thiết kế tập trung vào trải nghiệm của con người, đảm bảo rằng các tương tác mang tính tôn trọng và cá nhân hóa. Ví dụ, trong y tế, AI nên hỗ trợ thay vì thay thế việc chăm sóc con người, với các tính năng ưu tiên phẩm giá và quyền tự chủ của bệnh nhân. AI nên hoạt động một cách minh bạch, cho phép người dùng hiểu cách các quyết định được đưa ra. Sự minh bạch này giúp đảm bảo rằng các cá nhân được đối xử như những người tham gia trong quá trình thay vì chỉ là những đối tượng bị động.

**Duy trì tự do:** Hệ thống AI cần tích hợp các cơ chế cho phép người dùng quyết định cách dữ liệu của họ được sử dụng; cung cấp các công cụ để các cá nhân kiểm soát và quản lý dữ liệu của họ giúp bảo vệ quyền tự do của họ; tránh các hành vi thao túng và đảm bảo rằng các ứng dụng AI tôn trọng quyền được đưa ra các quyết định độc lập và có thông tin.

<sup>151</sup> Nguyễn Hoàng Anh (Chủ biên), *Pháp luật về quyền tự do lập hội, hội họp hoà bình trên thế giới và của Việt Nam*, Nxb. Hồng Đức, 2015, tr.27.

**Duy trì bình đẳng:** Các hệ thống AI nên được thiết kế với các biện pháp phát hiện và giảm thiểu thiên vị. Điều này bao gồm việc sử dụng các bộ dữ liệu đào tạo đa dạng, thường xuyên kiểm tra thuật toán để phát hiện thiên vị, đảm bảo tính công bằng.

Các công nghệ AI cần được thiết kế để có thể truy cập và mang lại lợi ích cho tất cả mọi người, bao gồm cả các nhóm bị thiệt thòi và thiếu đại diện. Điều này có thể bao gồm việc tạo ra các giải pháp AI có giá cả phải chăng, các dịch vụ dựa trên AI có sẵn bằng nhiều ngôn ngữ và định dạng khác nhau.

**Duy trì gắn kết xã hội:** Hệ thống AI nên được thiết kế để thúc đẩy sự gắn kết xã hội bằng cách khuyến khích quan điểm đa dạng và sự tham gia của cộng đồng. Điều này có thể đạt được thông qua các thuật toán thúc đẩy việc tiếp xúc với các quan điểm khác nhau và ưu tiên nội dung khuyến khích sự hiểu biết và hợp tác. Nhu cầu của các nhóm dễ bị tổn thương được xem xét trong quá trình phát triển và triển khai các giải pháp AI; cải thiện khả năng tiếp cận các dịch vụ thiết yếu cho các cộng đồng này.

## PHỤ LỤC:

# RỦI RO VI PHẠM QUYỀN CON NGƯỜI THEO VÒNG ĐỜI CỦA HỆ THỐNG AI

Bảng dưới đây nêu các ví dụ về những rủi ro tiềm ẩn trực tiếp và gián tiếp đối với quyền con người trong vòng đời của hệ thống AI<sup>152</sup>.

Bảng 7. Ví dụ rủi ro tiềm ẩn

Các giai đoạn vòng đời AI	Hành động có nguy cơ, rủi ro đối với quyền con người	Tác động tiêu cực đối với quyền con người
<b>Thiết kế</b>	<p>Các nhà vận hành hệ thống và nhà thiết kế AI lập kế hoạch hoặc thiết kế hệ thống mà không tính đến những tác hại do các chế độ lỗi tiềm ẩn, và/hoặc do sử dụng vượt ra khỏi phạm vi dự kiến ban đầu.</p> <p>Các nhà vận hành và nhà thiết kế cố ý thiết kế một hệ thống, hoặc lựa chọn dữ liệu và thuật toán vi phạm quyền con người ngay từ khâu thiết kế, ví dụ như cho phép theo dõi tùy tiện và bất hợp pháp.</p>	<p>Các mô hình và ứng dụng AI có thể được sử dụng theo những cách dẫn đến phân biệt đối xử, không an toàn hoặc có hại, vượt ngoài ý định của nhà phát triển. Ví dụ, một công cụ xử lý hình ảnh không xử lý đúng các tông màu da sẫm có thể tạo ra sự thiên vị; hoặc một hệ thống được thiết kế để giám sát chuyển động của xe hoặc người có thể bị lạm dụng để xâm phạm quyền riêng tư hoặc quấy rối.</p> <p>Các tác động có thể xảy ra như: giám sát và theo dõi trái phép và/hoặc không chính xác (bao gồm quản lý bằng thuật toán không phù hợp và giám sát nơi làm việc do AI hỗ trợ); bắt giữ và giam giữ sai; tổn hại do tạo tài liệu lạm dụng tình dục trẻ em hoặc hình ảnh riêng tư không có sự đồng ý. Có thể xảy ra việc hạn chế quyền tự do ngôn luận, tự do hội họp và lập hội một cách hòa bình.</p>
<b>Xây dựng, kiểm tra, phê duyệt mô hình</b>	<p>Các nhà thiết kế hệ thống triển khai các biện pháp bảo vệ kỹ thuật không đủ để ngăn chặn rò rỉ dữ liệu, tiết lộ trái phép, hoặc hủy ẩn danh thông tin nhận dạng cá nhân, hoặc các dữ liệu nhạy cảm khác như dữ liệu sinh trắc học, sức khỏe, chỗ ở.</p> <p>Các nhà phát triển hoặc các tác nhân khác không tiến hành thử nghiệm và đánh giá để phát hiện các kết quả đầu ra không chính xác, bao gồm cả các phản hồi thiên lệch hoặc bịa đặt.</p>	<p>Các công cụ AI có thể suy luận thông tin nhận dạng cá nhân vi phạm quyền riêng tư, bao gồm các thuộc tính nhạy cảm như vị trí, giới tính, tuổi tác, khuynh hướng tình dục và niềm tin chính trị. Người dùng có thể bị tái nhận dạng, có thể bằng cách kết hợp dữ liệu ẩn danh với các điểm dữ liệu khác, và có thể bị theo dõi qua các vị trí vật lý và không gian trực tuyến.</p> <p>Người dùng có thể trích xuất dữ liệu huấn luyện từ các mô hình cho phép tái dựng thông tin định dạng cá nhân mà không có sự đồng ý của họ.</p>

<sup>152</sup> Risk Management Profile for AI and Human Rights - United States Department of State

	<p>Tương tác giữa các mô hình AI và con người được thiết kế theo cách mà các câu trả lời của mô hình được cung cấp cho người dùng mà không kiểm tra đầy đủ, bao gồm việc không cung cấp và đánh giá các lời giải thích hoặc lý do cho các đầu ra của hệ thống "hộp đen".</p>	<p>Các mô hình AI có thể không thực hiện được chức năng dự kiến, gây hại cho những người mà việc tận hưởng quyền con người của họ phụ thuộc vào chức năng đó (ví dụ: dịch tài liệu liên quan đến việc xin tị nạn, hoặc đưa ra các quyết định có ảnh hưởng quan trọng đến cuộc sống).</p> <p>Các dịch vụ hoặc tài nguyên có thể bị từ chối hoặc thu hồi dựa trên các quyết định không có căn cứ hoặc dữ liệu không chính xác.</p> <p>Những người bị ảnh hưởng bởi các đầu ra hoặc quyết định của AI có thể gặp khó khăn trong việc xác định khi nào và cách thức họ bị tổn hại, dẫn đến giảm trách nhiệm giải trình và khả năng giảm thiểu hoặc khắc phục các tổn hại.</p>
<b>Triển khai, sử dụng</b>	<p>Các doanh nghiệp, cơ quan công bố hoặc triển khai một mô hình AI mới với thiếu sót nêu trên chưa được giải quyết, hoặc họ đưa AI vào ứng dụng mà không có hướng dẫn và biện pháp bảo vệ để đảm bảo sử dụng có trách nhiệm.</p> <p>Người dùng cuối vượt qua các biện pháp bảo vệ để sử dụng hệ thống AI vi phạm hoặc lạm dụng quyền con người.</p>	<p>Những rủi ro được nêu trên có thể trở thành hiện thực. Ví dụ như các mô hình AI có thể bị lạm dụng để tạo ra hình ảnh thân mật không có sự đồng thuận, hoặc tài liệu lạm dụng tình dục trẻ em, có thể được sử dụng để kiếm lợi, đe dọa nạn nhân; hoặc để tiến hành giám sát trái phép hoặc kiểm duyệt.</p> <p>Thông tin sai lệch được hỗ trợ bởi AI có thể được sử dụng để quấy rối và đe dọa nhà báo, đối thủ chính trị, hoặc những người bảo vệ nhân quyền, buộc họ phải tự kiểm duyệt. Điều này có thể làm suy yếu khả năng thực thi quyền tự do tìm kiếm và nhận thông tin cần thiết để hình thành quan điểm, cũng như quyền tự do hội họp và lập hội một cách hòa bình.</p>
<b>Vận hành, giám sát</b>	<p>Các cơ quan, doanh nghiệp không cung cấp cách thức cho mọi người báo cáo và biện pháp khắc phục đối với các lạm dụng, sai sót, hoặc sự cố liên quan đến hệ thống AI.</p>	<p>Các hệ thống AI có thể tiếp tục làm trầm trọng hơn bất bình đẳng bằng cách tiếp tục tạo ra các kết quả không chính xác, dẫn đến việc gia tăng bất lợi đối với các nhóm dân cư bị thiệt thòi. Các hệ thống AI có thể bị sử dụng cho bạo lực giới, khiến các cá nhân phải tránh né không gian công cộng. Các quyết định mang tính phân biệt đối xử hoặc không công bằng có thể không được giải quyết và lặp lại trong tương lai. Nạn nhân có thể không thể tiếp cận biện pháp khắc phục sau khi quyền con người của họ bị vi phạm hoặc lạm dụng.</p>



## TÀI LIỆU THAM KHẢO

1. Goldman Sachs. (2023, April 5). Generative AI could raise global GDP by 7%. Goldmansachs.com. <https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent>
2. Stinson, J. (2024, June 6). The Ipsos AI Monitor 2024: Changing attitudes and feelings about AI and the future it will bring. Ipsos. <https://www.ipsos.com/en-uk/ipsos-ai-monitor-2024-changing-attitudes-and-feelings-about-ai-and-future-it-will-bring>
3. OECD. (2024). Assessing potential future artificial intelligence risks, benefits and policy imperatives. OECD. [https://www.oecd.org/en/publications/assessing-potential-future-artificial-intelligence-risks-benefits-and-policy-imperatives\\_3f4e3dfb-en.html](https://www.oecd.org/en/publications/assessing-potential-future-artificial-intelligence-risks-benefits-and-policy-imperatives_3f4e3dfb-en.html)
4. ACS. (2023). Rapid Response Information Report: Generative AI | Chief Scientist. Chiefscientist.gov.au. <https://www.chiefscientist.gov.au/GenerativeAI>
5. PWC. (2024). AI Jobs Barometer. PwC. <https://www.pwc.com/gx/en/issues/artificial-intelligence/ai-jobs-barometer.html>
6. PricewaterhouseCoopers. (2023). Vietnam Hopes and Fears Survey 2023. PwC. <https://www.pwc.com/vn/en/publications/vietnam-publications/workforce-hopes-fears-vietnam-2023.html>
7. World Economic Forum. (2023, April 30). The Future of Jobs Report 2023. World Economic Forum. <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>
8. Small Business Now: An AI Awakening 1. (n.d.). [https://assets.ctfassets.net/t21gix3kzulv/4DyMfIRpJEdiLaWb73-JioX/603078e4bafca3bf9d9c8f09dca7053a/Constant\\_Contact\\_Small\\_Business\\_Now\\_Report\\_-\\_An\\_AI\\_Awakening\\_-\\_August\\_2023.pdf](https://assets.ctfassets.net/t21gix3kzulv/4DyMfIRpJEdiLaWb73-JioX/603078e4bafca3bf9d9c8f09dca7053a/Constant_Contact_Small_Business_Now_Report_-_An_AI_Awakening_-_August_2023.pdf)
9. PricewaterhouseCoopers. (2024). PwC's NextGen Survey 2024 - Vietnam report. PwC. <https://www.pwc.com/vn/en/publications/vietnam-publications/nextgen-survey-2024.html>
10. Các cuộc phỏng vấn sâu của IPS với đại diện của một số doanh nghiệp Việt Nam, 2023-2024
11. IPS, Tọa đàm về phát triển AI ở Việt Nam, Hà Nội, 11/2023
12. Doan, T. N., Nguyen, Q. K., & Taylor-Robinson, A. W. (2023). Healthcare in Vietnam: Harnessing Artificial Intelligence and Robotics to Improve Patient Care Outcomes. Cureus. <https://doi.org/10.7759/cureus.45006>
13. Bộ NNPTNT, tọa đàm về ứng dụng AI trong ngành nông nghiệp, Hà Nội, 18/10/2024.
14. Assessing potential future artificial intelligence risks, benefits and policy imperatives. (2025). OECD. [https://www.oecd.org/en/publications/assessing-potential-future-artificial-intelligence-risks-benefits-and-policy-imperatives\\_3f4e3dfb-en.html](https://www.oecd.org/en/publications/assessing-potential-future-artificial-intelligence-risks-benefits-and-policy-imperatives_3f4e3dfb-en.html)
15. PricewaterhouseCoopers. (2024). PwC's NextGen Survey 2024 - Vietnam report. PwC. <https://www.pwc.com/vn/en/publications/vietnam-publications/nextgen-survey-2024.html>
16. OECD. (2019). The OECD Artificial Intelligence (AI) Principles. Oecd.ai; OECD. <https://oecd.ai/en/ai-principles>
17. Solomon, L., & Davis, N. (2023). Human Technology Institute AI Governance The State of in Australia. <https://www.uts.edu.au/globalassets/sites/default/files/2023-05/hti-the-state-of-ai-governance-in-australia---31-may-2023.pdf>

18. Nazanin Andalibi. (2024, March 6). Emotion-tracking AI on the job: Workers fear being watched – and misunderstood. *The Conversation*. <https://theconversation.com/emotion-tracking-ai-on-the-job-workers-fear-being-watched-and-misunderstood-222592> , Greenhouse, S. (2024, January 7). “Constantly monitored”: the pushback against AI surveillance at work. *The Guardian*. <https://www.theguardian.com/technology/2024/jan/07/artificial-intelligence-surveillance-workers>
19. Who will be the workers most affected by AI? (2025). OECD. [https://www.oecd.org/en/publications/who-will-be-the-workers-most-affected-by-ai\\_14dc6f89-en.html](https://www.oecd.org/en/publications/who-will-be-the-workers-most-affected-by-ai_14dc6f89-en.html)
20. Carbonero, F., Davies, J., Ernst, E., Fossen, F., Samaan, D., & Sorgner, A. (2021). The Impact of Artificial Intelligence on Labor Markets in Developing Countries: A New Method with an Illustration for Lao PDR and Viet Nam. <https://docs.iza.org/dp14944.pdf>
21. Carbonero, F., Davies, J., Ernst, E., Fossen, F., Samaan, D., & Sorgner, A. (2021). The Impact of Artificial Intelligence on Labor Markets in Developing Countries: A New Method with an Illustration for Lao PDR and Viet Nam. <https://docs.iza.org/dp14944.pdf>
22. Lee, K.-F. (2018). *AI superpowers China, Silicon Valley, and the new world order*. Boston: Houghton Mifflin Harcourt.
23. The Plight of Platform Workers Under Algorithmic Management in Southeast Asia. (2024). Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2024/04/the-plight-of-platform-workers-under-algorithmic-management-in-southeast-asia?lang=en>
24. Kai Fu Lee, *AI Powers*, 2018; Kai Fu Lee, *Ten Perspectives for AI 2041*, 2024; *Advancing a More Global Agenda for Trustworthy Artificial Intelligence*. (2024). Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2024/04/advancing-a-more-global-agenda-for-trustworthy-artificial-intelligence?lang=en>
25. General, UN. (2024, July 5). Enhancing international cooperation on capacity-building of artificial intelligence :: resolution /: adopted by the General Assembly. United Nations Digital Library System; UN,. <https://digitallibrary.un.org/record/4054005?ln=en&v=pdf>
26. Charles, L., Xia, S., & Coutts, A. (2022). *Digitalization and Employment A Review*. [https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed\\_emp/documents/publication/wcms\\_854353.pdf](https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed_emp/documents/publication/wcms_854353.pdf)
27. Kai Fu Lee, *AI Powers*, 2018; Kai Fu Lee, *Ten Perspectives for AI 2041*, 2024; *Advancing a More Global Agenda for Trustworthy Artificial Intelligence*. (2024). Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2024/04/advancing-a-more-global-agenda-for-trustworthy-artificial-intelligence?lang=en>
28. Charles, L., Xia, S., & Coutts, A. (2022). *Digitalization and Employment A Review*. [https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed\\_emp/documents/publication/wcms\\_854353.pdf](https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@ed_emp/documents/publication/wcms_854353.pdf)
29. *Advancing a More Global Agenda for Trustworthy Artificial Intelligence*. (2024). Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2024/04/advancing-a-more-global-agenda-for-trustworthy-artificial-intelligence?lang=en>
30. General, UN. (2024, July 5). Enhancing international cooperation on capacity-building of artificial intelligence :: resolution /: adopted by the General Assembly. United Nations Digital Library System; UN,. <https://digitallibrary.un.org/record/4054005?ln=en&v=pdf>
31. Majumdar, D., & Chattopadhyay, H. K. (2020). AI and Human Rights: From Business and Policy Perspectives. *ResearchGate*, 5(11), 51–60. [https://www.researchgate.net/publication/345902156\\_AI\\_and\\_Human\\_Rights\\_From\\_Business\\_and\\_Policy\\_Perspectives](https://www.researchgate.net/publication/345902156_AI_and_Human_Rights_From_Business_and_Policy_Perspectives)

32. General, UN. (2023, December 22). Promotion and protection of human rights in the context of digital technologies :: resolution /: adopted by the General Assembly. United Nations Digital Library System; UN,. <https://digitallibrary.un.org/record/4032837?ln=en&v=pdf>
33. Nguyen, Q. (2019). ỨNG DỤNG TRÍ TUỆ NHÂN TẠO VÀ TÁC ĐỘNG TỚI QUYỀN CON NGƯỜI: GÓC NHÌN TỪ MÔ HÌNH ĐÁNH GIÁ TÍN NHIỆM XÃ HỘI (SCS) Ở TRUNG QUỐC. NHÀ NƯỚC và PHÁP LUẬT. [https://www.academia.edu/40461456/ỨNG\\_DỤNG\\_TRÍ\\_TUỆ\\_NHÂN\\_TẠO\\_VÀ\\_TÁC ĐỘNG\\_TỚI QUYỀN\\_CON\\_NGƯỜI\\_GÓC\\_NHÌN\\_TỪ\\_MÔ\\_HÌNH\\_ĐÁNH GIÁ\\_TÍN\\_NHIỆM\\_XÃ\\_HỘI\\_SCS\\_Ở\\_TRUNG\\_QUỐC](https://www.academia.edu/40461456/ỨNG_DỤNG_TRÍ_TUỆ_NHÂN_TẠO_VÀ_TÁC ĐỘNG_TỚI QUYỀN_CON_NGƯỜI_GÓC_NHÌN_TỪ_MÔ_HÌNH_ĐÁNH GIÁ_TÍN_NHIỆM_XÃ_HỘI_SCS_Ở_TRUNG_QUỐC)
34. Advancing a More Global Agenda for Trustworthy Artificial Intelligence. (2024b). Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2024/04/advancing-a-more-global-agenda-for-trustworthy-artificial-intelligence?lang=en>
35. Advancing a More Global Agenda for Trustworthy Artificial Intelligence. (2024b). Carnegie Endowment for International Peace. <https://carnegieendowment.org/research/2024/04/advancing-a-more-global-agenda-for-trustworthy-artificial-intelligence?lang=en>
36. Có thể xem tổng hợp về các công cụ chính sách trên thế giới tại đây: OECD's live repository of AI strategies & policies - OECD.AI, accessed on 27/10/2024; các sáng kiến chính sách cấp quốc gia tại đây: Policies - OECD.AI
37. Xem giới thiệu và tải Các nguyên tắc về AI của OECD tại đây: OECD Legal Instruments, AI Principles Overview - OECD.AI; báo cáo về thực tiễn các nước áp dụng Các nguyên tắc ở đây: How countries are implementing the OECD Principles for Trustworthy AI - OECD.AI
38. Department of Industry, Science and Resources, Australia. Converlens - Engagement data insight platform for surveys, consultations and text. (2022). Industry.gov.au. <https://consult.industry.gov.au/supporting-responsible-ai>
39. Xem một số sáng kiến chính sách của Trung Quốc về AI tại đây, trong đó có một số quy tắc đạo đức về AI: Policies - OECD.AI; China's New AI Governance Initiatives Shouldn't Be Ignored - Carnegie Endowment for International Peace | Carnegie Endowment for International Peace
40. Emerging AI-related regulation AI Strategies and Policies - OECD.AI. (2025). Oecd.ai. [https://oecd.ai/en/dashboards/policy-instruments/Emerging\\_technology\\_regulation](https://oecd.ai/en/dashboards/policy-instruments/Emerging_technology_regulation)
41. Global AI Law and Policy Tracker. (n.d.). [https://iapp.org/media/pdf/resource\\_center/global\\_ai\\_law\\_policy\\_tracker.pdf](https://iapp.org/media/pdf/resource_center/global_ai_law_policy_tracker.pdf)
42. (2019). Stanford.edu. <https://hai.stanford.edu/ai-index>
43. Ngô Huyền. (2024, August 5). “Hòn đảo AI”: Singapore đã làm thế nào để đuổi kịp hai siêu cường AI Mỹ – Trung? Nhịp Sống Kinh Tế Việt Nam & Thế Giới. <https://vneconomy.vn/hon-dao-ai-singapore-da-lam-the-nao-de-duoi-kip-hai-sieu-cuong-ai-my-trung.htm>
44. (2019). Stanford.edu. <https://hai.stanford.edu/ai-index>
45. (2019). Stanford.edu. <https://hai.stanford.edu/ai-index>
46. (2019). Stanford.edu. <https://hai.stanford.edu/ai-index>
47. PS tổng hợp từ Oxford AI Index 2024, (2019). Stanford.edu. <https://hai.stanford.edu/ai-index>
48. Lessons From the World's Two Experiments in AI Governance. (2023). Carnegie Endowment for International Peace. <https://carnegieendowment.org/posts/2023/02/lessons-from-the-worlds-two-experiments-in-ai-governance?lang=en>
49. SB 1047: Safe and Secure Innovation for Frontier Artificial Intelligence Models Act. (2023). Digitaldemocracy.org. [https://calmatters.digitaldemocracy.org/bills/ca\\_202320240sb1047](https://calmatters.digitaldemocracy.org/bills/ca_202320240sb1047)

50. SENATE BILL 24-205. (n.d.). [https://leg.colorado.gov/sites/default/files/2024a\\_205\\_signed.pdf](https://leg.colorado.gov/sites/default/files/2024a_205_signed.pdf)
51. North America: From Brussels to Boulder - Colorado enacts comprehensive AI law on the heels of European Union's AI Act with significant obligations for businesses and employers. (2024). Bakermckenzie.com. <https://insightplus.bakermckenzie.com/bm/data-technology/north-america-from-brussels-to-boulder-colorado-enacts-comprehensive-ai-law-on-the-heels-of-eus-ai-act-with-significant-obligations-for-business-and-employers>
52. Translation: Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment) – April 2023 - DigiChina. (2023, April 19). DigiChina. <https://digichina.stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/>
53. Translation: Internet Information Service Deep Synthesis Management Provisions (Draft for Comment) – Jan. 2022 - DigiChina. (2023, April 12). DigiChina. <https://digichina.stanford.edu/work/translation-internet-information-service-deep-synthesis-management-provisions-draft-for-comment-jan-2022/>
54. Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022 - DigiChina. (2022, February 28). DigiChina. <https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/>
55. Governor Newsom announces new initiatives to advance safe and responsible AI, protect Californians | Governor of California. (2024, September 29). Governor of California. <https://www.gov.ca.gov/2024/09/29/governor-newsom-announces-new-initiatives-to-advance-safe-and-responsible-ai-protect-californians/>
56. Aft, A. (2024, August 27). Illinois Joins Colorado and NYC in Restricting Generative AI in HR (Plus a Quick Survey of the Legal Landscape Across the US and Globally). Connect on Tech. <https://connectontech.bakermckenzie.com/illinois-joins-colorado-and-nyc-in-restricting-generative-ai-in-hr-plus-a-quick-survey-of-the-legal-landscape-across-the-us-and-globally/>
57. Australian Government Department of Industry, Science and Resources Converlens - Engagement data insight platform for surveys, consultations and text. (2022b). Industry.gov.au. <https://consult.industry.gov.au/ai-mandatory-guardrails>
58. Jeannie Marie Paterson, AI có trách nhiệm: Tiếp cận tổng thể và ứng dụng cụ thể, tài liệu tập huấn, Viện Nghiên cứu chính sách và phát triển truyền thông, Hà Nội, 9/2024.
59. Regulatory sandboxes in artificial intelligence. (2025). OECD. [https://www.oecd.org/en/publications/regulatory-sandboxes-in-artificial-intelligence\\_8f80a0e6-en.html](https://www.oecd.org/en/publications/regulatory-sandboxes-in-artificial-intelligence_8f80a0e6-en.html)
60. Solomon, L., & Davis, N. (2023). Human Technology Institute AI Governance The State of in Australia. <https://www.uts.edu.au/globalassets/sites/default/files/2023-05/hti-the-state-of-ai-governance-in-australia---31-may-2023.pdf>
61. AI Sprinters Capturing the economic opportunity of AI in emerging markets A Digital Sprinters report. (2024). [https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/AI\\_Sprinters\\_Report.pdf](https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/AI_Sprinters_Report.pdf)
62. GOV.UK. (2023, March 16). Government commits up to £3.5 billion to future of tech and science. GOV.UK. <https://www.gov.uk/government/news/government-commits-up-to-35-billion-to-future-of-tech-and-science>
63. Alexi Mostrous, Cesareo, S., & White, J. (2024, September 19). The Global Artificial Intelligence Index 2024. Tortoise Media. <https://www.tortoisemedia.com/2024/09/19/the-global-artificial-intelligence-index-2024>

64. Alexi Mostrous, Cesareo, S., & White, J. (2024, September 19). The Global Artificial Intelligence Index 2024. Tortoise Media. <https://www.tortoisemedia.com/2024/09/19/the-global-artificial-intelligence-index-2024>
65. AI Sprinters Capturing the economic opportunity of AI in emerging markets A Digital Sprinters report. (2024). [https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/AI\\_Sprinters\\_Report.pdf](https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/AI_Sprinters_Report.pdf)
66. Bertrand, A., & Phillips, A. (2025). How can harnessing tax data create value for all? | EY – Global. Ey.com. [https://www.ey.com/en\\_tw/insights/government-public-sector/how-can-harnessing-tax-data-create-value-for-all](https://www.ey.com/en_tw/insights/government-public-sector/how-can-harnessing-tax-data-create-value-for-all)
67. Cloud Policy Framework. (2020). Cra.gov.qa. <https://www.cra.gov.qa/en/document/-cloud-policy-framework>
68. Ure, A. (2024, January 29). An AI opportunity agenda for ASEAN. Google. <https://blog.google/around-the-globe/google-asia/an-ai-opportunity-agenda-for-asean/>
69. Ure, A. (2024, January 29). An AI opportunity agenda for ASEAN. Google. <https://blog.google/around-the-globe/google-asia/an-ai-opportunity-agenda-for-asean/>
70. Alexi Mostrous, Cesareo, S., & White, J. (2024, September 19). The Global Artificial Intelligence Index 2024. Tortoise Media. <https://www.tortoisemedia.com/2024/09/19/the-global-artificial-intelligence-index-2024>
71. AI Sprinters Capturing the economic opportunity of AI in emerging markets A Digital Sprinters report. (2024). [https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/AI\\_Sprinters\\_Report.pdf](https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/AI_Sprinters_Report.pdf)
72. Alexi Mostrous, Cesareo, S., & White, J. (2024, September 19). The Global Artificial Intelligence Index 2024. Tortoise Media. <https://www.tortoisemedia.com/2024/09/19/the-global-artificial-intelligence-index-2024>
73. Alexi Mostrous, Cesareo, S., & White, J. (2024, September 19). The Global Artificial Intelligence Index 2024. Tortoise Media. <https://www.tortoisemedia.com/2024/09/19/the-global-artificial-intelligence-index-2024>
74. GOV.UK. (2023, March 16). Government commits up to £3.5 billion to future of tech and science. GOV.UK. <https://www.gov.uk/government/news/government-commits-up-to-35-billion-to-future-of-tech-and-science>
75. Alexi Mostrous, Cesareo, S., & White, J. (2024, September 19). The Global Artificial Intelligence Index 2024. Tortoise Media. <https://www.tortoisemedia.com/2024/09/19/the-global-artificial-intelligence-index-2024>
76. AI Sprinters Capturing the economic opportunity of AI in emerging markets A Digital Sprinters report. (2024). [https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/AI\\_Sprinters\\_Report.pdf](https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/AI_Sprinters_Report.pdf)
77. Petrov, S. (2022, April 11). Investing in Eastern Europe's AI future. Google. <https://blog.google/technology/ai/investing-in-eastern-europes-ai-future/>
78. Alexi Mostrous, Cesareo, S., & White, J. (2024, September 19). The Global Artificial Intelligence Index 2024. Tortoise Media. <https://www.tortoisemedia.com/2024/09/19/the-global-artificial-intelligence-index-2024>
79. Alexi Mostrous, Cesareo, S., & White, J. (2024, September 19). The Global Artificial Intelligence Index 2024. Tortoise Media. <https://www.tortoisemedia.com/2024/09/19/the-global-artificial-intelligence-index-2024>

80. IPS lập bảng theo số liệu của Tortoise 's Global Artificial Intelligence Index 2024.
81. NIST. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. <https://doi.org/10.6028/nist.ai.600-1>. NIST. (2023). AI Risk Management Framework. Artificial Intelligence Risk Management Framework (AI RMF 1.0). <https://doi.org/10.6028/nist.ai.100-1>
82. Risk Management Profile for AI and Human Rights - United States Department of State. (2024, July 25). United States Department of State. <https://2021-2025.state.gov/risk-management-profile-for-ai-and-human-rights/>
83. Using the AI Risk Management Framework: Workday, 2024.
84. The Artificial Intelligence and Data Act (AIDA) – Companion document. (2025, January 31). Canada.ca. <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>
85. Converlens - Engagement data insight platform for surveys, consultations and text. (2022). Industry.gov.au. <https://consult.industry.gov.au/ai-mandatory-guardrails>
86. Edwards, L. (2022). The EU AI Act: a summary of its significance and scope Expert explainer. <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf>
87. Luật này là một phần trong đạo luật lớn hơn có tên gọi là Đạo luật Thực thi hiến chương số năm 2022 (Digital Charter Implementation Act, 2022), dự kiến có hiệu lực từ 2025. Xem ở đây: Government Bill (House of Commons) C-27 (44-1) - First Reading - Digital Charter Implementation Act, 2022 - Parliament of Canada; The Artificial Intelligence and Data Act (AIDA) – Companion document
88. Converlens - Engagement data insight platform for surveys, consultations and text. (2022). Industry.gov.au. <https://consult.industry.gov.au/ai-mandatory-guardrails>
89. Converlens - Engagement data insight platform for surveys, consultations and text. (2022). Industry.gov.au. <https://consult.industry.gov.au/ai-mandatory-guardrails>
90. Government of Canada, and Parliament of Canada, 'Amendments to Bill C-27 (44-1)', 28 November 2023.
91. The Artificial Intelligence and Data Act (AIDA) – Companion document. (2025, January 31). Canada.ca. <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>
92. Converlens - Engagement data insight platform for surveys, consultations and text. (2022). Industry.gov.au. <https://consult.industry.gov.au/ai-mandatory-guardrails>
93. United Nations. A/RES/78/265 General Assembly. (n.d.). <https://documents.un.org/doc/undoc/gen/n24/087/83/pdf/n2408783.pdf>
94. Xem toàn văn Công ước tại đây: <https://rm.coe.int/1680afae3c>; tóm tắt và bình luận về Công ước tại đây: <https://www.connectontech.com/over-fifty-countries-sign-on-to-the-council-of-europes-framework-convention-on-artificial-intelligence-the-first-legally-binding-international-treaty-on-ai/>
95. Mark Latonero. (n.d.). [https://datasociety.net/wp-content/uploads/2018/10/DataSociety\\_Governing\\_Artificial\\_Intelligence\\_Upholding\\_Human\\_Rights.pdf](https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf)

96. R-SD, J. (2023). S.3312 - 118th Congress (2023-2024): Artificial Intelligence Research, Innovation, and Accountability Act of 2024. Congress.gov. <https://www.congress.gov/bill/118th-congress/senate-bill/3312>
97. Federal AI legislation: An evaluation of existing proposals and a road map forward. (2024). Economic Policy Institute. <https://www.epi.org/publication/federal-ai-legislation/>
98. Xem Sắc lệnh tại đây: <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/>
99. Xem cụ thể hơn các bài viết, cũng như bản dịch tiếng Anh các chính sách, luật của Trung Quốc về các vấn đề liên quan đến AI tại đây: DigiChina.
100. Xem thêm bình luận của các chuyên gia về văn bản này tại đây: How will China's Generative AI Regulations Shape the Future? A DigiChina Forum; DigiChina.
101. AI safety. (2024, August 2). What does the Chinese leadership mean by "instituting oversight systems to ensure the safety of AI?" Substack.com; AI Safety in China. <https://aisafetychina.substack.com/p/what-does-the-chinese-leadership>
102. Cục ICT, Bộ Thông tin và truyền thông, Định hướng xây dựng thể chế về công nghiệp công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.
103. Cisco AI Readiness Index Hype Meets Reality. (n.d.). Retrieved April 3, 2025, from [https://www.cisco.com/c/dam/m/en\\_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf](https://www.cisco.com/c/dam/m/en_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf)
104. PricewaterhouseCoopers. (2024). Khảo sát Thế hệ kế nghiệp (NextGen) toàn cầu năm 2024 | Thành công trong kỷ nguyên của trí tuệ nhân tạo. PwC. <https://www.pwc.com/vn/vn/publications/vietnam-publications/nextgen-survey-2024.html>
105. Artificial Intelligence - Vietnam | Market Forecast. (2025). Statista. <https://www.statista.com/outlook/tmo/artificial-intelligence/vietnam>
106. marketingadmin. (2024, November 7). The Future of AI in Vietnam: Top Business Opportunities JDI Group. JDI Group. <https://jdi.group/the-future-of-ai-in-vietnam-top-business-opportunities/#:~:text=AI%20In%20Vietnam%20%E2%80%93%20Landscape%20Overview,-Vietnam%27s%20AI%20Subsector&text=As%20a%20result%2C%20Vietnam%20has,AI%20startup%20investments%20in%202023>
107. Artificial Intelligence - Vietnam | Market Forecast. (2025). Statista. <https://www.statista.com/outlook/tmo/artificial-intelligence/vietnam#analyst-opinion>
108. Top IT Jobs For Developers TECH HIRING LANDSCAPE IN VIETNAM: VIETNAM OVERVIEW: SOCIO-ECONOMICS INDICATORS IT JOBS MARKET IN VIETNAM: SALARIES & DEMANDS Developers' salaries and demands. (n.d.). [https://shiftasia.com/wp-content/uploads/2024/12/Top-Dev\\_VietnamITMarketReport2024\\_2025\\_VietnamITTechTalentLandscape.pdf](https://shiftasia.com/wp-content/uploads/2024/12/Top-Dev_VietnamITMarketReport2024_2025_VietnamITTechTalentLandscape.pdf)
109. GenAI Fund. (2025). Content outline. <https://genaifund.ai/wp-content/uploads/2024/09/ASEAN-GenAI-Startup-Report-2024-by-GenAI-Fund.pdf>
110. Hồ, R. (2024, November 14). Startup AI Việt Nam gặp cảnh "voi nhiều, cỏ quá ít." Tạp Chí Kinh Tế Sài Gòn. <https://thesaigontimes.vn/startup-ai-viet-nam-gap-canhh-voi-nhieu-co-qua-it/>
111. Cisco AI Readiness Index Hype Meets Reality. (n.d.). [https://www.cisco.com/c/dam/m/en\\_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf](https://www.cisco.com/c/dam/m/en_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf)

112. PricewaterhouseCoopers. (2024). Khảo sát Thế hệ kế nghiệp (NextGen) toàn cầu năm 2024 | Thành công trong kỷ nguyên của trí tuệ nhân tạo. PwC.
113. Hồ, R. (2024, November 14). Startup AI Việt Nam gặp cảnh “voi nhiều, cỏ quá ít.” Tạp Chí Kinh Tế Sài Gòn. <https://thesaigontimes.vn/startup-ai-viet-nam-gap-canh-voi-nhieu-co-qua-it/>
114. Charting the future of Southeast Asian AI governance. (2024, May 21). East Asia Forum. <https://eastasiaforum.org/2024/05/21/charting-the-future-of-southeast-asian-ai-governance/>
115. Pham, H., Nong, D., Simshauser, P., Nguyen, G., & Duong Trung Kien. (2024). Artificial intelligence (AI) development in the Vietnam’s energy and economic systems: A critical review. *Journal of Cleaner Production*, 140692–140692. <https://doi.org/10.1016/j.jclepro.2024.140692>
116. Cisco AI Readiness Index Hype Meets Reality. (n.d.). [https://www.cisco.com/c/dam/m/en\\_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf](https://www.cisco.com/c/dam/m/en_us/solutions/ai/readiness-index/2024-m11/documents/cisco-ai-readiness-index-vn.pdf)
117. Hoàng Hà. (2025, January 10). Thị trường trung tâm dữ liệu Việt Nam trước thách thức kép “xanh và AI.” *Nhịp Sống Kinh Tế Việt Nam & Thế Giới*. <https://vneconomy.vn/techconnect/thi-truong-trung-tam-du-lieu-viet-nam-truoc-thach-thuc-kep-xanh-va-ai.htm>
118. Trần Đăng Dũng, Viettel IDC, Hạ tầng công nghệ cho ứng dụng “AI” – động lực phát triển kinh tế số, hội thảo DataFest, Ninh Bình, 10/2024.
119. GenAI Fund. (2025). Content outline. <https://genaifund.ai/wp-content/uploads/2024/09/ASEAN-GenAI-Startup-Report-2024-by-GenAI-Fund.pdf>
120. AI Readiness Report. (2023). <https://go.scale.com/hubfs/Scale-Zeitgeist-AI-Readiness-Report-2023.pdf>
121. Asia Pacific AI Readiness Index. (2023). [https://www.salesforce.com/content/dam/web/en\\_sg/www/documents/pdf/salesforce\\_ai\\_readiness\\_index\\_2023.pdf](https://www.salesforce.com/content/dam/web/en_sg/www/documents/pdf/salesforce_ai_readiness_index_2023.pdf)
122. TopDev, Vietnam IT & Tech Talent Landscape Repo, 2024. Xem thêm: Đặng Trọng Hợp, Đào tạo nhân lực công nghệ số chất lượng cao đáp ứng nhu cầu của ngành công nghiệp công nghệ số, Định hướng xây dựng thể chế về công nghiệp công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.
123. TopDev, Vietnam IT & Tech Talent Landscape Repo, 2024. Xem thêm: Phạm Thị Nhung, Nhu cầu nguồn nhân lực trong lĩnh vực công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.
124. Dẫn theo: Phạm Thị Nhung, Nhu cầu nguồn nhân lực trong lĩnh vực công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.
125. Pham, H., Nong, D., Simshauser, P., Nguyen, G., & Duong Trung Kien. (2024). Artificial intelligence (AI) development in the Vietnam’s energy and economic systems: A critical review. *Journal of Cleaner Production*, 140692–140692. <https://doi.org/10.1016/j.jclepro.2024.140692>
126. Vietnam’s Ambition to Lead AI Innovation in Southeast Asia – OpenGov Asia. (2024, August 20). [Opengovasia.com. https://opengovasia.com/2024/08/20/vietnams-ambition-to-lead-ai-innovation-in-southeast-asia/#:~:text=The%20government%20is%20proposing%20a,500%20AI%20startups%20by%202030](https://opengovasia.com/2024/08/20/vietnams-ambition-to-lead-ai-innovation-in-southeast-asia/#:~:text=The%20government%20is%20proposing%20a,500%20AI%20startups%20by%202030)
127. Putting Skills First Opportunities for Building Efficient and Equitable Labour Markets. (n.d.). [https://www3.weforum.org/docs/WEF\\_Putting\\_Skills\\_First\\_2024.pdf](https://www3.weforum.org/docs/WEF_Putting_Skills_First_2024.pdf)
128. Phạm Thị Nhung, Nhu cầu nguồn nhân lực trong lĩnh vực công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.



129. Đặng Trọng Hợp, Đào tạo nhân lực công nghệ số chất lượng cao đáp ứng nhu cầu của ngành công nghiệp công nghệ số, Định hướng xây dựng thể chế về công nghiệp công nghệ số, bài tham luận tại hội thảo Giải pháp thúc đẩy phát triển ngành công nghiệp công nghệ số, Hà Nội, 15/11/2024.
130. Tam, D. (2024, June 17). Hanoi university Vietnam's first to train GenAI engineers. VnExpress International – Latest News, Business, Travel and Analysis from Vietnam; VnExpress International. <https://e.vnexpress.net/news/news/education/hanoi-university-vietnam-s-first-to-train-genai-engineers-4759305.html>
131. Pham, H., Nong, D., Simshauser, P., Nguyen, G., & Duong Trung Kien. (2024). Artificial intelligence (AI) development in the Vietnam's energy and economic systems: A critical review. Journal of Cleaner Production, 140692–140692. <https://doi.org/10.1016/j.jclepro.2024.140692>
132. Vương Trần. (2025, January 13). Bố trí ít nhất 15% ngân sách chi cho sự nghiệp khoa học. Laodong.vn; Báo Lao Động. <https://laodong.vn/thoi-su/-bo-tri-it-nhat-15-ngan-sach-chi-cho-su-nghiep-khoa-hoc-1449258.lido>
133. Sách Khoa học và công nghệ thế giới 2023, Cục Thông tin Khoa học và công nghệ quốc gia, Bộ Khoa học và công nghệ, tr.20.
134. Vương Trần. (2025, January 13). Bố trí ít nhất 15% ngân sách chi cho sự nghiệp khoa học. Laodong.vn; Báo Lao Động. <https://laodong.vn/thoi-su/-bo-tri-it-nhat-15-ngan-sach-chi-cho-su-nghiep-khoa-hoc-1449258.lido>
135. GenAI Fund. (2025). Content outline. <https://genaifund.ai/wp-content/uploads/2024/09/ASEAN-GenAI-Startup-Report-2024-by-GenAI-Fund.pdf>
136. thuvienphapluat.vn. (2025, February 10). Tiêu chuẩn quốc gia TCVN 13902:2023 (ISO/IEC 22989:2022) về Công nghệ thông tin - Trí tuệ nhân tạo - Các khái niệm và thuật ngữ trí tuệ nhân tạo. THƯ VIỆN PHÁP LUẬT; thuvienphapluat.vn. <https://thuvienphapluat.vn/tcvn/Linh-vuc-khac/TCVN-13902-2023-ISO-IEC-22989-2022-Cong-nghe-thong-tin-Tri-tue-nhan-tao-920969.aspx>
137. LuậtVietnam. (2024, April 5). Tiêu chuẩn quốc gia TCVN 13903:2023 Công nghệ thông tin - Trí tuệ nhân tạo. LuậtVietnam. <https://luatvietnam.vn/thong-tin/tieu-chuan-quoc-gia-tcvn-13903-2023-cong-nghe-thong-tin-tri-tue-nhan-tao-310542-d3.html>
138. Phiên họp của Ủy ban Thường vụ Quốc hội cho ý kiến về việc giải trình, tiếp thu, chỉnh lý Dự án Luật Công nghiệp công nghệ số, 6/1/2025.
139. Hồ, R. (2024, November 14). Startup AI Việt Nam gặp cảnh “voi nhiều, cỏ quá ít.” Tạp Chí Kinh Tế Sài Gòn. <https://thesaigontimes.vn/startup-ai-viet-nam-gap-canhh-voi-nhieu-co-qua-it/>
140. Mô hình ngôn ngữ lớn (LLM) của Zalo AI bứt phá bảng xếp hạng. (2025, January 11). TechSignin.com. [https://www.techsignin.com/110125-mo-hinh-llm-zalo-ai/#:~:text=Zalo%20AI%20c%C3%B4ng%20b%E1%BB%91%20B%C3%A1o,Suite%20for%20Large%20Language%20Models\).](https://www.techsignin.com/110125-mo-hinh-llm-zalo-ai/#:~:text=Zalo%20AI%20c%C3%B4ng%20b%E1%BB%91%20B%C3%A1o,Suite%20for%20Large%20Language%20Models).)
141. Trọng Đạt. (2025). Việt Nam phát triển bộ dữ liệu tiếng Việt dùng cho AI, trợ lý ảo. VietNamNet News; Vietnamnet.vn. [https://vietnamnet.vn/viet-nam-phat-trien-bo-du-lieu-tieng-viet-dung-cho-ai-tro-ly-ao-2380624.html?fbclid=IwY2xjawJTKuBleHRuA2FibQIxMAABHd658pJZ8kIG-HgcSNhMG9mxtBOn1aJxrleB92gRIGZIYFWKbnl5GEYhUA\\_aem\\_UU95pjAOXC01svcVud9xA](https://vietnamnet.vn/viet-nam-phat-trien-bo-du-lieu-tieng-viet-dung-cho-ai-tro-ly-ao-2380624.html?fbclid=IwY2xjawJTKuBleHRuA2FibQIxMAABHd658pJZ8kIG-HgcSNhMG9mxtBOn1aJxrleB92gRIGZIYFWKbnl5GEYhUA_aem_UU95pjAOXC01svcVud9xA)
142. Mô hình ngôn ngữ lớn (LLM) của Zalo AI bứt phá bảng xếp hạng. (2025, January 11). TechSignin.com. [https://www.techsignin.com/110125-mo-hinh-llm-zalo-ai/#:~:text=Zalo%20AI%20c%C3%B4ng%20b%E1%BB%91%20B%C3%A1o,Suite%20for%20Large%20Language%20Models\).](https://www.techsignin.com/110125-mo-hinh-llm-zalo-ai/#:~:text=Zalo%20AI%20c%C3%B4ng%20b%E1%BB%91%20B%C3%A1o,Suite%20for%20Large%20Language%20Models).)

143. An AI Opportunity Agenda for Vietnam. (2023). Retrieved April 3, 2025, from [https://static.googleusercontent.com/media/publicpolicy.-google/vi/resources/vietnam\\_ai\\_opportunity\\_agenda\\_en.pdf](https://static.googleusercontent.com/media/publicpolicy.-google/vi/resources/vietnam_ai_opportunity_agenda_en.pdf)
144. Lee, K.-F. (2018). AI superpowers China, Silicon Valley, and the new world order. Boston Houghton Mifflin Harcourt.
145. Bridging the skills gap: Fuelling careers and the economy in Vietnam SUPPORTED BY. (n.d.). [https://impact.economist.com/perspectives/sites/default/files/skilling\\_in\\_asia-vietnam\\_eng.pdf](https://impact.economist.com/perspectives/sites/default/files/skilling_in_asia-vietnam_eng.pdf)
146. Bridging the skills gap: Fuelling careers and the economy in Vietnam SUPPORTED BY. (n.d.). [https://impact.economist.com/perspectives/sites/default/files/skilling\\_in\\_asia-vietnam\\_eng.pdf](https://impact.economist.com/perspectives/sites/default/files/skilling_in_asia-vietnam_eng.pdf)
147. Xem thêm: Viện Nghiên cứu chính sách và phát triển truyền thông (IPS) & Chương trình phát triển LHQ (UNDP), Báo cáo Đánh giá nhanh ứng dụng AI trong các cơ quan Nhà nước ở Việt Nam: Mức độ sẵn sàng và khuyến nghị, Hà Nội, 2025.
148. Public Sector AI Playbook - Digital Government Hub. (2024, February 3). Digital Government Hub. <https://digitalgovernmenthub.org/library/public-sector-ai-playbook/>  
Public Sector AI Playbook - Digital Government Hub. (2025, February 3). Digital Government Hub. <https://digitalgovernmenthub.org/library/public-sector-ai-playbook/>
149. Public Sector AI Playbook - Digital Government Hub. (2024, February 3). Digital Government Hub. <https://digitalgovernmenthub.org/library/public-sector-ai-playbook/>  
Public Sector AI Playbook - Digital Government Hub. (2025, February 3). Digital Government Hub. <https://digitalgovernmenthub.org/library/public-sector-ai-playbook/>
150. Risk Management Profile for AI and Human Rights - United States Department of State. (2024, July 25). United States Department of State. <https://2021-2025.state.gov/risk-management-profile-for-ai-and-human-rights/>
151. Nguyễn Hoàng Anh (Chủ biên), Pháp luật về quyền tự do lập hội, hội họp hoà bình trên thế giới và của Việt Nam, Nxb. Hồng Đức, 2015, tr.27.
152. Risk Management Profile for AI and Human Rights - United States Department of State. (2024, July 25). United States Department of State. <https://2021-2025.state.gov/risk-management-profile-for-ai-and-human-rights/>



Viện Nghiên cứu Chính sách  
và Phát triển Truyền thông

Tầng 18, Toà nhà VTC Online,  
18 Tam Trinh, Hai Bà Trưng, Hà Nội  
Tel: (84) 969 520 220  
Email: [contact@ips.org.vn](mailto:contact@ips.org.vn)  
[ips.org.vn](http://ips.org.vn)