Chính sách quản lý agent economy trong thời đại số: bài học kinh nghiệm cho Việt Nam

Lê Hải Hà

Viện Kinh doanh quốc tế và Logistics, Trường Đại học Thương Mại

Email: ha.lh@tmu.edu.vn

Tóm tắt

Sự trỗi dậy của trí tuệ nhân tạo (AI) đang hình thành một "nền kinh tế tác nhân" (Agent Economy - AE) mới, nơi các tác nhân AI tự trị đại diện con người thực hiện nhiều nhiệm vụ phức tạp. Agent Economy hứa hẹn mang lại những lợi ích to lớn về năng suất và đổi mới, song cũng đặt ra các thách thức quản lý mới về riêng tư dữ liệu, an ninh, đạo đức và nguy cơ mất việc làm. Vấn đề cấp thiết là làm sao để con người có thể giám sát và kiểm soát một cách hiệu quả nền kinh tế do tác nhân AI vận hành khi hệ thống này có thể mở rộng vượt ngoài năng lực quản lý truyền thống. Nghiên cứu này sử dụng phương pháp tổng hợp tài liệu và so sánh chính sách quốc tế, qua đó đề xuất một khung chính sách gồm 6 nhóm giải pháp cho Việt Nam nhằm vừa thúc đẩy phát triển AI vừa kiểm soát rủi ro trong thời đại số. Khung chính sách đề xuất góp phần tạo nền tảng cho Việt Nam chủ động đón nhận Agent Economy một cách an toàn và hiệu quả.

Từ khóa: *Agent economy; Agentic AI; chính sách; quản lý; Việt Nam*

Abstract

The rapid emergence of Artificial Intelligence (AI) is ushering in a new "Agent Economy" (AE), in which autonomous AI agents act on behalf of humans to perform complex tasks. While the Agent Economy promises substantial gains in productivity and innovation, it simultaneously presents profound governance challenges related to data privacy, security, ethics, and job displacement. A pressing concern is how humans can effectively oversee and regulate an AI-driven economy that may expand beyond the reach of traditional governance mechanisms. This study adopts a literature synthesis and comparative policy analysis approach to propose a six-pillar policy framework tailored for Vietnam. The framework aims to simultaneously foster AI development and mitigate associated risks in the digital age, thereby laying a strategic foundation for Vietnam to engage proactively, safely, and effectively with the Agent Economy.

Keywords*: Agent economy; Agentic AI, policy, governance, Vietnam*

https://doi.org/10.65153/e9w49h66

Journal of Science and Technology of East Asia University of Technology

1. MỞ ĐẦU

Năm 2025 được gọi là "năm của các tác nhân AI" cho thấy sự chuyển đổi căn bản trong công nghệ: AI không còn chỉ là công cụ hỗ trợ mà đã bắt đầu tự động ảnh hưởng đến môi trường thực và đưa ra các quyết định độc lập [1][2]. Điều này đòi hỏi phải có khung quản lý tương ứng nhằm đảm bảo những hệ thống tự động này mang lại lợi ích cho kinh tế - xã hội. Các quốc gia tiên phong trên thế giới đã bắt đầu xây dựng chính sách quản lý hiện tượng AE theo những cách tiếp cận khác nhau như Mỹ ưu tiên thúc đẩy đổi mới và giảm thiểu rào cản pháp lý, Liên minh Châu Âu (EU) ban hành luật AI chặt chẽ dựa trên phân loại mức độ rủi ro, Trung Quốc quản lý nghiêm ngặt các dịch vụ AI tạo sinh để bảo vệ lợi ích quốc gia và Singapore thử nghiệm cơ chế sandbox nhằm vừa khuyến khích sáng tạo vừa đảm bảo an toàn. Đối với Việt Nam, một quốc gia đang trong quá trình tăng tốc chuyển đổi số, việc chủ động nghiên cứu AE và xây dựng khung chính sách quản lý phù hợp có ý nghĩa cấp thiết. Bài viết này nhằm phân tích cơ sở lý luận về AE, đánh giá thực trạng chính sách tại một số quốc gia, từ đó rút ra bài học và đề xuất khung chính sách cho Việt Nam.

2. PHƯƠNG PHÁP NGHIÊN CỨU

Bài nghiên cứu tổng hợp các tài liệu khoa học giai đoạn 2020–2025 và các văn bản chính sách cập nhật liên quan đến AE. Các trường hợp chính sách của Mỹ, EU, Trung Quốc và Singapore được chọn làm đối tượng so sánh đối chiếu do tính đại diện cho những cách tiếp cận quản lý khác nhau. Tiêu chí chọn lọc tài liệu dựa trên mức độ mới, độ tin cậy và sự liên quan trực tiếp đến chủ đề nghiên cứu. Trên cơ sở phân tích lý thuyết và bối cảnh Việt Nam, nghiên cứu đề xuất khung chính sách quản lý AE phù hợp với điều kiện trong nước, đảm bảo các khuyến nghị đưa ra có luận cứ khoa học vững chắc.

3. CƠ SỞ LÝ LUẬN VỀ AGENT ECONOMY VÀ CHÍNH SÁCH QUẢN LÝ

Agent Economy (AE) được hiểu là một hệ sinh thái kinh tế số mới, trong đó các "tác nhân" AI tự trị (autonomous AI agents) đại diện cho con người hoặc tổ chức để tương tác và giao dịch với nhau, thực hiện các hoạt động kinh tế một cách tự động. Khác với mô hình kinh tế số hiện nay dựa trên tương tác trực tiếp giữa người và hệ thống, AE hướng tới tự động hóa giao tiếp, tức là mỗi người dùng có thể có một trợ lý AI tác nhân riêng truyền đạt nhu cầu và thông tin của mình tới vô số dịch vụ, trong khi mỗi doanh nghiệp triển khai các tác nhân AI dịch vụ để tiếp nhận yêu

cầu và phục vụ khách hàng [3][4]. Các tác nhân AI này đối thoại và thương lượng với nhau thay mặt cho người dùng và doanh nghiệp, giúp thu xếp giao dịch hiệu quả hơn cho các bên mà không cần tương tác thủ công lặp lại từ phía con người [3]. Về lý thuyết, điều này có thể tái cấu trúc cách thức vận hành của thị trường, làm giảm chi phí chuyển đổi nhà cung cấp, đồng thời thúc đẩy sự ra đời của những sản phẩm và mô hình kinh doanh hoàn toàn mới [3][4]. Nói cách khác, AE mở ra một mô hình kinh tế mới trong kỷ nguyên số với một "mạng lưới các tác nhân" hoạt động song song (và dần hòa nhập) với nền kinh tế do con người trực tiếp vận hành.

Trong mô hình AE, các tác nhân AI giữ vai trò trung tâm, được trang bị nhiều năng lực vượt trội so với các hệ thống phần mềm truyền thống. Theo Hosseini & Seilani (2025), Agentic AI hội tụ sáu đặc điểm chính: (i) **Tự chủ** là khả năng vận hành độc lập, tự đưa ra quyết định và hành động không cần đến chỉ thị trực tiếp từ con người [1]; (ii) **Hướng mục tiêu** là Agentic AI luôn theo đuổi mục tiêu xác định và tối ưu hóa hành động để đạt kết quả mong muốn; (iii) **Tương tác môi trường** là khả năng có thể quan sát, nhận thức và thích nghi với sự thay đổi của môi trường xung quanh theo thời gian thực; (iv) **Khả năng học hỏi** là Agentic AI tự cải thiện hiệu suất qua thời gian nhờ học máy, rút kinh nghiệm từ dữ liệu và phản hồi để ra quyết định ngày càng tốt hơn; (v) **Tối ưu quy trình** là khả năng tích hợp các kỹ năng ngôn ngữ, lập luận, lập kế hoạch và ra quyết định nhằm tự động hóa và phối hợp hiệu quả các quy trình công việc phức tạp; và (vi) **Hệ đa tác tử** là khả năng giao tiếp, hợp tác với các tác nhân AI khác. Nhờ các năng lực này, mỗi tác nhân AI có thể đảm nhận nhiều vai trò khác nhau trong nền kinh tế: từ trợ lý cá nhân đến điều phối chuỗi cung ứng, giao dịch tài chính tự động, dịch vụ khách hàng, hỗ trợ giáo dục, y tế và thậm chí tác nhân chính phủ điện tử giúp xử lý thủ tục cho người dân. Nhờ đó, AE tạo ra một mức độ phân công lao động mới, khi các tác nhân AI có thể được coi là "lao động số", thực hiện độc lập nhiều công việc trí óc mà trước đây do con người đảm nhiệm [1].

Lợi ích và rủi ro kinh tế - xã hội của AE: *Về lợi ích*, AE được kỳ vọng sẽ nâng cao hiệu quả hoạt động thị trường thông qua việc giảm chi phí giao dịch và chi phí thông tin giữa các chủ thể kinh tế [1]. Người tiêu dùng có thể dễ dàng tìm được sản phẩm hoặc dịch vụ phù hợp với nhu cầu nhờ tác nhân AI sàng lọc vô số lựa chọn, vượt qua giới hạn về thời gian và khả năng tìm kiếm của con người. Doanh nghiệp cũng có thể tiếp cận lượng khách hàng lớn thông qua mạng lưới tác nhân, giảm sự phụ thuộc vào quảng cáo trả phí theo mô hình "nền kinh tế chú ý" hiện nay. Sự phổ biến của tác nhân AI có thể thúc đẩy mô hình giao dịch vi mô phát triển khi các tác vụ mua bán

được tự động hóa hoàn toàn, ví dụ người dùng chỉ trả phí cho từng nội dung số như bài hát, bài báo,… mà tác nhân của họ truy cập, thay vì đăng ký thuê bao trọn gói, đồng thời, tác nhân AI có khả năng tạo ra sản phẩm cá nhân hóa cao. Tóm lại, AE có tiềm năng thúc đẩy đổi mới mô hình kinh doanh, gia tăng trải nghiệm cá nhân hóa cho khách hàng, đồng thời phân phối lại giá trị giữa các tác nhân tham gia chuỗi cung ứng số [3][4].

Tuy nhiên, đi kèm lợi ích là *những rủi ro và thách thức mới*. ***Thứ nhất***, rủi ro hệ thống: khi các tác nhân AI được kết nối rộng khắp trong một "mạng lưới tác nhân mở", những bất ổn hoặc hành vi sai lệch của một nhóm tác nhân có thể nhanh chóng lan truyền và gây hiệu ứng dây chuyền lên toàn hệ thống kinh tế số. Khác với môi trường thử nghiệm khép kín, một AE thẩm thấu vào nền kinh tế thực sẽ khó cô lập hoàn toàn các sự cố kỹ thuật trước khi chúng ảnh hưởng đến con người, ví dụ một đợt bán tháo tự động do tác nhân giao dịch lỗi có thể gây sụp đổ thị trường tài chính trước khi con người kịp can thiệp. ***Thứ hai,*** vấn đề an ninh, an toàn: tác nhân AI có thể bị lợi dụng bởi các đối tượng xấu như hacker, tội phạm mạng để tự động tạo ra thông tin sai lệch hoặc tấn công mạng trên quy mô lớn. Việc tác nhân AI tương tác trực tiếp với con người đòi hỏi phải có cơ chế đảm bảo chúng tuân thủ các tiêu chuẩn đạo đức, không đưa ra quyết định gây hại hay phân biệt đối xử. ***Thứ ba***, sự riêng tư của dữ liệu người dùng: để hoạt động hiệu quả, tác nhân AI cần truy cập lượng lớn dữ liệu cá nhân của người dùng, làm dấy lên lo ngại về quyền riêng tư và nguy cơ lạm dụng dữ liệu. Nếu không có khung pháp lý rõ ràng, các tác nhân có thể thu thập, chia sẻ thông tin nhạy cảm của người dùng mà họ không hề hay biết. ***Thứ tư***, tác động đến thị trường lao động: khi AE phát triển, một phần công việc trí óc sẽ được tự động hóa bởi những "lao động số" này. Mặc dù về dài hạn, công nghệ mới thường tạo thêm việc làm ở các lĩnh vực khác nhưng trong ngắn hạn có thể xảy ra nguy cơ nguy cơ thất nghiệp nếu lực lượng lao động không kịp thích nghi [1]. Điều này đòi hỏi phải có chiến lược chủ động đào tạo, nâng cao kỹ năng cho người lao động để họ chuyển sang các vai trò bổ trợ cho AI thay vì bị thay thế. ***Thứ năm***, trách nhiệm và pháp lý: trong AE, khi một tác nhân AI gây thiệt hại như tư vấn sai dẫn đến tổn thất tài chính, hoặc lỗi của phương tiện tự lái gây tai nạn, việc xác định ai chịu trách nhiệm pháp lý trở nên phức tạp hơn, là người dùng, nhà phát triển, hay chính tác nhân AI phải chịu trách nhiệm. Trách nhiệm giải trình của các hệ thống AI tự trị do đó là nội dung trọng tâm cần được chú ý trong chính sách quản lý.

Trước những lợi ích và rủi ro nói trên, các nhà nghiên cứu đã đề xuất áp dụng cách tiếp cận chủ động và linh hoạt trong quản lý AE. Một phương án được đề xuất nhiều là xây dựng "sandbox kinh tế tác nhân", đây là môi trường kiểm soát để tác nhân AI vận hành thử nghiệm, nhằm quan sát ảnh hưởng của chúng trước khi tích hợp vào nền kinh tế thực [5]. Hiện tại, xu hướng công nghệ cho thấy AE đang hình thành một cách tự phát, các tác nhân AI dần tham gia trực tiếp vào nền kinh tế con người mà không trải qua giai đoạn thử nghiệm biệt lập rõ rệt. Thực tế này đặt ra thách thức lớn cho nhà quản lý phải nhanh chóng thiết kế khung quy định và cơ chế hướng dẫn hành vi tác nhân kịp thời trước khi hệ thống vượt khỏi tầm kiểm soát [2]. Nhiều nhà nghiên cứu và chuyên gia đã nhấn mạnh tính cấp thiết của việc xây dựng **kiến trúc thị trường cho tác nhân AI** ngay từ bây giờ, khi mà những lựa chọn chính sách ở hiện tại sẽ quyết định cấu trúc thị trường và ai sẽ hưởng lợi trong làn sóng công nghệ mới [3].

Tóm lại, cơ sở lý luận cho thấy AE có tiềm năng chuyển đổi sâu sắc nền kinh tế số, nhưng đi kèm những rủi ro đòi hỏi cách tiếp cận quản lý mới. Để quản lý AE một cách hiệu quả, cần kết hợp đồng bộ giữa biện pháp kỹ thuật và chính sách. Các nhà nghiên cứu gợi ý phát triển các tiêu chuẩn mở để các tác nhân AI tương tác an toàn, xây dựng hạ tầng xác thực và hệ thống tín nhiệm (ví dụ cơ chế chứng chỉ cho tác nhân) song song với các công cụ kinh tế và pháp lý (cơ chế khuyến khích hành vi an toàn, quy định trách nhiệm rõ ràng) ngay từ khâu thiết kế hệ thống. Mục tiêu là tạo ra một AE "trong tầm kiểm soát": vừa phát huy tối đa lợi ích tự động hóa, vừa đảm bảo an toàn, minh bạch và phục vụ lợi ích chung của xã hội

4. CHÍNH SÁCH QUẢN LÝ AE TẠI MỘT SỐ NƯỚC TRÊN THẾ GIỚI

Mặc dù AE còn ở giai đoạn đầu phát triển nhưng nhiều quốc gia trên thế giới đã sớm ý thức và đề cập hiện tượng này thông qua khung quản trị AI chung, cũng như thử nghiệm những chính sách điều phối tác nhân AI. Nghiên cứu tổng hợp một số cách tiếp cận của Mỹ, EU, Trung Quốc và Singapore là những đại diện tiêu biểu cho các mô hình quản lý AE khác nhau, qua đó rút ra kinh nghiệm với Việt Nam.

Hiện Mỹ chưa có luật liên bang toàn diện dành riêng cho AI, Mỹ chủ yếu là khuyến khích đổi mới và hạn chế can thiệp sớm. Chính quyền liên bang nhấn mạnh mục tiêu duy trì vị thế dẫn đầu về AI, coi AI là động lực tăng trưởng kinh tế và an ninh quốc gia. Năm 2025 Nhà Trắng ban hành Chiến lược AI Quốc gia tập trung 3 trụ cột: đổi mới, hạ tầng, ngoại giao và ký sắc lệnh gỡ bỏ một số quy định bị cho là cản trở sáng tạo. Thay vì xây dựng luật AI mới, Mỹ chủ trương sử

dụng các hướng dẫn và khung tiêu chuẩn tự nguyện do chính phủ hoặc tổ chức chuyên môn ban hành, ví dụ Khung quản lý rủi ro AI của NIST và tận dụng luật hiện hành như FTC, EEOC, v.v. giám sát hành vi cạnh tranh, phân biệt đối xử để quản lý AI. Các bang cũng ban hành nhiều luật liên quan đến AI, tạo nên bức tranh quy định chắp vá [11]. Về khía cạnh AE, Mỹ chưa có chính sách riêng; các tập đoàn công nghệ tự xây dựng quy tắc an toàn cho tác nhân AI và chính phủ đưa ra hướng dẫn sử dụng AI có trách nhiệm trong khu vực công. Mô hình của Mỹ thiên về tự quản lý với can thiệp tối thiểu – giúp thúc đẩy đổi mới nhanh và thu hút đầu tư, nhưng tạo khoảng trống quản lý khi an toàn/đạo đức của tác nhân phần lớn phụ thuộc vào tự giác doanh nghiệp.

EU đi theo hướng quy định chặt chẽ dựa trên đánh giá rủi ro ngay từ sớm, nhằm định hướng phát triển AI một cách đáng tin cậy, lấy con người làm trung tâm. EU ban hành Đạo luật AI năm 2024 [6], đây là khung pháp lý toàn diện đầu tiên trên thế giới về AI. Đạo luật AI phân loại các hệ thống AI theo bốn mức độ rủi ro đối với an toàn, quyền lợi con người như: rủi ro tối thiểu là được tự do phát triển, khuyến khích tuân thủ tự nguyện tiêu chuẩn; rủi ro hạn chế yêu cầu phải có minh bạch cho người dùng, ví dụ chatbot phải tiết lộ "Tôi không phải người thật"; rủi ro cao bắt buộc phải tuân thủ hàng loạt yêu cầu nghiêm ngặt về quản trị rủi ro, quản lý dữ liệu, giám sát con người, độ chính xác, an toàn… và đăng ký với cơ quan chức năng trước khi đưa ra thị trường và rủi ro không chấp nhận là bị cấm hoàn toàn, như AI dùng cho hệ thống tín dụng xã hội, nhận diện gương mặt thời gian thực nơi công cộng phục vụ thực thi pháp luật [6]. Đạo luật cũng áp trách nhiệm xuyên suốt chuỗi giá trị AI, từ bên phát triển mô hình đến đơn vị triển khai và được xây dựng trên bốn trụ cột: quản trị rủi ro liên tục, minh bạch, độ tin cậy kỹ thuật và giám sát con người [6]. EU sẽ thành lập Ủy ban và Văn phòng AI châu Âu để điều phối thực thi luật [6]; các nước thành viên có thể ban hành luật bổ sung phù hợp bối cảnh, ví dụ Ý năm 2025 ban hành luật cấm nghiêm hành vi phát tán deepfake ác ý. Cách tiếp cận thận trọng của EU đặt AI và AE vào khuôn khổ pháp lý ngay từ đầu, ưu tiên kiểm soát rủi ro và bảo vệ giá trị nhân đạo. Cách làm này định hình chuẩn mực "AI đáng tin cậy" trên thế giới nhưng cũng khiến doanh nghiệp AI chịu chi phí tuân thủ cao và thủ tục phức tạp, có nguy cơ làm chậm triển khai tác nhân AI ở châu Âu.

Trung Quốc quản lý AI nói chung và AE nói riêng trong khuôn khổ ưu tiên an ninh chính trị và trật tự xã hội. Tháng 8/2023, Trung Quốc ban hành "Biện pháp Quản lý dịch vụ AI tạo sinh", đây là văn bản pháp quy cấp quốc gia đầu tiên trên thế giới chuyên về quản lý AI tạo sinh [9] có hiệu lực từ 15/8/2023, áp dụng cho mọi dịch vụ sử dụng công nghệ AI tạo sinh cung cấp cho công

chúng. Mục tiêu được nêu là "thúc đẩy phát triển lành mạnh và ứng dụng có trật tự của AI tạo sinh, bảo vệ an ninh quốc gia và lợi ích công cộng, đồng thời bảo vệ quyền và lợi ích hợp pháp của công dân và tổ chức", nó thể hiện rõ định hướng quản lý AI phục vụ ổn định chính trị, xã hội [9]. Cách tiếp cận của Trung Quốc mang tính "quản lý phân cấp, phân loại", cụ thể chính phủ giám sát AI dựa trên phân loại mức độ quan trọng và mục đích sử dụng, nhưng không công bố cụ thể thang phân loại như EU. Tuy nhiên, có thể thấy những dịch vụ AI tạo sinh có "thuộc tính dư luận hoặc khả năng huy động xã hội" sẽ chịu giám sát nghiêm ngặt hơn, ví dụ các chatbot có thể tương tác với lượng lớn người dùng sẽ bị kiểm soát chặt về nội dung. Văn bản quy định một loạt yêu cầu tuân thủ đối với nhà cung cấp dịch vụ AI tạo sinh như [9]: Sử dụng dữ liệu hợp pháp, tôn trọng sở hữu trí tuệ**;** Không thiên lệch hoặc phân biệt đối xử; Đảm bảo chất lượng và tính chính xác; Bảo vệ trẻ vị thành niên; An ninh và kiểm duyệt nội dung; Đăng ký và đánh giá an ninh. Đồng thời, các dịch vụ AI phải trải qua đánh giá an ninh trước khi ra mắt, tương tự yêu cầu đối với dịch vụ Internet theo Luật An ninh mạng. Tháng 11/2025, Trung Quốc ban hành thêm 3 tiêu chuẩn quốc gia về bảo mật và quản trị AI tạo sinh, nhằm cụ thể hóa các yêu cầu kỹ thuật đối với mô hình AI như tiêu chuẩn về quản lý truy xuất dữ liệu huấn luyện, đánh giá độ an toàn của mô hình AI. Có thể thấy Trung Quốc đang xây dựng khung quản lý AE theo hướng bảo hộ và chủ quyền, Nhà nước kiểm soát chặt chẽ sự phát triển của tác nhân AI, định hướng chúng phục vụ các mục tiêu quốc gia, đồng thời đầu tư mạnh mẽ để tự chủ công nghệ nhằm tránh phụ thuộc nước ngoài. Chính phủ Trung Quốc khuyến khích doanh nghiệp và địa phương thử nghiệm AI trong các sandbox có kiểm soát, ví dụ Thâm Quyến triển khai khu thử nghiệm xe tự lái, Bắc Kinh lập Khu thí điểm chính sách AI nhưng luôn kèm giám sát an ninh. Ưu điểm cách làm của Trung Quốc là tạo hàng rào chắn ngay từ đầu ngăn chặn tác hại của tác nhân AI như tin giả, bất ổn xã hội từ đó bảo vệ các giá trị văn hóa - chính trị nội địa; nhược điểm là môi trường đổi mới có thể bị gò bó, doanh nghiệp nhỏ khó khăn trong việc tuân thủ quy định phức tạpvà người dùng ít được tiếp cận các dịch vụ AI tiên tiến toàn cầu do rào cản kiểm duyệt.

Singapore được coi là hình mẫu về cách tiếp cận cân bằng giữa đổi mới và quản trị rủi ro trong kỷ nguyên AI. Singapore sớm xây dựng các khung hướng dẫn đạo đức AI dưới dạng mềm dẻo để doanh nghiệp "làm đúng ngay từ đầu". Năm 2019, Singapore ban hành Mô hình Quản trị AI hướng dẫn chi tiết giúp doanh nghiệp tích hợp các nguyên tắc AI có trách nhiệm như minh bạch, công bằng, đảm bảo quyền riêng tư và có thể giải thích trong vòng đời dự án AI và phát triển AI Verify, bộ công cụ đánh giá kỹ thuật giúp kiểm chứng một hệ thống AI có đáp ứng các tiêu chí

AI đáng tin cậy hay không. Về luật pháp, Singapore chưa có luật riêng về AI, nhưng các quy định ngành hiện hành có thể bao quát rủi ro AI như quy tắc quản lý rủi ro công nghệ tài chính, Luật An ninh mạng, Luật Bảo vệ Dữ liệu Cá nhân [9]. Thay vì áp đặt quy định cứng sớm như EU, Singapore chọn cách thí điểm linh hoạt. Năm 2022, chính phủ ra mắt AI Governance Testing Framework và chương trình AI Regulatory Sandbox, cho phép doanh nghiệp thử nghiệm giải pháp AI trong phạm vi giới hạn dưới sự giám sát của cơ quan quản lý. Đến tháng 7/2025, Singapore tiếp tục mở rộng nỗ lực này thông qua việc công bố "Global AI Assurance Sandbox" phiên bản mới, với trọng tâm bao gồm cả các hệ thống Agentic AI tiên tiến [7]. Theo Cơ quan Phát triển Truyền thông Infocomm (IMDA), sandbox mở rộng cho phép nhiều công ty trong và ngoài nước tham gia kiểm thử các ứng dụng AI gần với thực tiễn để đánh giá rủi ro và hiệu quả trước khi triển khai rộng rãi [7]. Sandbox mới bổ sung các kịch bản mới như tác nhân AI tự trị và rủi ro mới nổi như nguy cơ rò rỉ dữ liệu hoặc tấn công prompt injection đối với mô hình ngôn ngữ lớn [7]. Các cơ quan quản lý chuyên ngành ở Singapore cũng được mời tham gia sandbox nhằm cùng phát triển và kiểm chứng các hướng dẫn quản trị AI trong lĩnh vực của họ [7]. Song song với sandbox, Singapore ban hành Bộ công cụ Thử nghiệm An toàn LLM (LLM Safety Testing Starter Kit, 2025) cung cấp quy trình tiêu chuẩn để các bên thử nghiệm mức độ an toàn của ứng dụng AI dựa trên mô hình ngôn ngữ lớn [7]. Cách làm này thể hiện quan điểm "thực dụng, dựa trên rủi ro" của Singapore, cho phép đổi mới diễn ra trong môi trường được kiểm soát, thu thập thông tin thực tế để định hình chính sách, vừa tận dụng tối đa cơ hội AI vừa dựng sẵn "hành lang bảo vệ" nhằm xây dựng lòng tin của công chúng [7]. Bên cạnh AI sandbox, Singapore cũng rất chú trọng bảo vệ dữ liệu và quyền riêng tư, nền tảng của một nền kinh tế số đáng tin cậy. Năm 2020, Singapore sửa đổi Luật Bảo vệ Dữ liệu Cá nhân để tăng chế tài vi phạm và bổ sung yêu cầu chặt chẽ hơn về xử lý dữ liệu. Tháng 7/2025, Singapore nâng chuẩn bảo vệ dữ liệu lên tầm cao mới: chuẩn hóa Chương trình Chứng nhận Tín nhiệm Bảo vệ Dữ liệu thành Tiêu chuẩn quốc gia, khuyến khích doanh nghiệp đạt chứng nhận bảo vệ dữ liệu ở mức tương đương tiêu chuẩn quốc tế [9]. Tất cả các nỗ lực trên cho thấy Singapore xây dựng một hệ sinh thái chính sách khá toàn diện: vừa tạo không gian sandbox cho doanh nghiệp thử nghiệm và tinh chỉnh giải pháp AI, đặc biệt chú ý tác nhân AI, vừa hoàn thiện hạ tầng pháp lý về dữ liệu và an toàn làm hậu thuẫn cho triển khai AI một cách có trách nhiệm. Cách tiếp cận của Singapore phản ánh đặc thù quốc gia nhỏ, nguồn lực hạn chế: họ dựa nhiều vào hướng dẫn mềm dẻo và chứng nhận tiêu chuẩn để định hướng thị trường thay vì ban hành quá nhiều luật cứng; đồng thời chủ động tham gia định hình chuẩn mực quốc tế như thúc đẩy thảo luận về quản trị AI

trong ASEAN, hợp tác cùng OECD, WEF. Nhờ đó, Singapore duy trì được hình ảnh môi trường AI đổi mới có trách nhiệm, thu hút đầu tư quốc tế mà vẫn giữ được niềm tin của người dân đối với công nghệ mới.

Để làm rõ sự khác biệt giữa các quốc gia, **Bảng 1** dưới đây tóm tắt so sánh sơ lược cách tiếp cận chính sách AE (hoặc quản trị AI có liên quan) tại Mỹ, EU, Trung Quốc và Singapore:

Bảng 1. So sánh cách tiếp cận chính sách quản lý AI/AE ở một số quốc gia

Quốc gia/Khối	Cách tiếp cận quản lý AI	Chính sách và sáng kiến
Mỹ	Thị trường dẫn dắt, can thiệp tối thiểu. Ưu tiên duy trì vị thế dẫn đầu AI, gỡ bỏ rào cản để thúc đẩy đổi mới. Sử dụng luật hiện hành & hướng dẫn tự nguyện thay vì luật mới.	- Chiến lược AI quốc gia 2025 với 90 hành động (3 trụ cột: Đổi mới, Hạ tầng, Ngoại giao) [11]. - Sắc lệnh 2025 "Removing Barriers to American Leadership in AI" - Khung quản lý rủi ro AI hướng dẫn tự nguyện cho doanh nghiệp - Nhiều luật tiểu bang về AI về tuyển dụng, tài chính, minh bạch thuật toán… [11]. - Chưa có quy định riêng cho AE; tập trung khuyến khích doanh nghiệp tự quản lý AI có trách nhiệm.
Liên minh Châu Âu	Quy định toàn diện dựa trên đánh giá rủi ro. Đặt ra chuẩn mực pháp lý chi tiết để đảm bảo AI an toàn, nhân văn. Cập nhật hướng	- Đạo luật AI của EU (2024): phân loại AI theo 4 mức rủi ro; yêu cầu nghiêm ngặt với hệ thống rủi ro cao; cấm một số ứng dụng không chấp nhận [6]. - Yêu cầu minh bạch & giám sát con người: bắt buộc gắn nhãn nội dung AI tạo ra, đảm bảo con người chịu trách nhiệm cuối [8]. - Đối với tác nhân AI: áp dụng quy định cho GPAI (AI đa dụng) và quy tắc hệ thống rủi ro cao tương ứng [6];

Quốc gia/Khối	Cách tiếp cận quản lý AI	Chính sách và sáng kiến
	dẫn để bao quát công nghệ mới.	đang phát triển tiêu chuẩn kỹ thuật và hướng dẫn bổ sung chuyên biệt. - Cơ quan thực thi mới: thành lập Ủy ban và Văn phòng AI châu Âu để giám sát tuân thủ trên toàn khối [6]. - Phối hợp khối: khuyến khích các nước thành viên xây dựng chiến lược AI quốc gia; ví dụ Ý có luật AI 2025 bổ sung xử lý deepfake.
Trung Quốc	Quản lý tập trung, ưu tiên an ninh & kiểm soát. Ban hành nhanh quy định cụ thể để kiểm soát nội dung AI và định hướng AI phục vụ lợi ích nhà nước. Yêu cầu đăng ký, kiểm duyệt chặt chẽ các dịch vụ AI công cộng.	- Biện pháp quản lý AI tạo sinh (2023): quy định toàn diện đầu tiên về dịch vụ GenAI (hiệu lực 8/2023) [9]; yêu cầu kiểm duyệt đầu ra, chống thông tin "bất lợi" và tuân thủ giá trị XHCN. - Yêu cầu bắt buộc: đăng ký thuật toán với CAC, đánh giá an ninh trước triển khai; gắn nhãn deepfake; cấm dùng dữ liệu bất hợp pháp; ngăn thiên lệch và nội dung độc hại [9]. - Chế tài mạnh: vi phạm có thể phạt, đình chỉ dịch vụ hoặc truy cứu hình sự [9]. - Hệ sinh thái quy định bổ trợ: Quy định thuật toán cá nhân hóa (2022), Quy định deepfake (2023); Hướng dẫn đạo đức công nghệ (2023); Tiêu chuẩn quốc gia về AI (2025). -Thử nghiệm hạn chế: thiết lập các khu sandbox AI có giám sát (ví dụ: khu thí điểm AI ở Bắc Kinh, Thâm Quyến) để thử nghiệm công nghệ dưới sự kiểm soát liên ngành.

Quốc gia/Khối	Cách tiếp cận quản lý AI	Chính sách và sáng kiến
Singapore	Thực dụng, cân bằng đổi mới và an toàn. Đưa ra khung đạo đức mềm dẻo, khuyến khích tuân thủ tự nguyện; đồng thời dùng sandbox và chứng nhận để quản lý theo rủi ro. Tăng cường nền tảng pháp lý về bảo vệ dữ liệu và xây dựng lòng tin.	- Mô hình Quản trị AI (2019) & AI Verify: hướng dẫn tự nguyện, công cụ kỹ thuật giúp doanh nghiệp phát triển AI có trách nhiệm. - AI Regulatory Sandbox: thí điểm 2022, mở rộng 2025 thành Global AI Assurance Sandbox - môi trường thử nghiệm thực tế cho tác nhân AI và kiểm định hướng dẫn quản trị mới [7]. Mở cho cả cơ quan nhà nước và quốc tế tham gia phản hồi chính sách [7]. - Starter Kit thử nghiệm an toàn LLM (2025): bộ công cụ chuẩn giúp kiểm tra an toàn cho ứng dụng AI dựa trên mô hình ngôn ngữ lớn [7]. - Khung pháp lý dữ liệu mạnh: Luật PDPA (sửa 2020) và chứng nhận DPTM (2025) nâng cao tiêu chuẩn bảo vệ dữ liệu cá nhân [9]. Xuất bản Hướng dẫn PETs (2025) để ứng dụng công nghệ bảo vệ riêng tư trong AI [9]. - Hợp tác quốc tế: Singapore tích cực tham gia xây dựng tiêu chuẩn AI toàn cầu, dẫn dắt đối thoại ASEAN về AI.

5. DỰ BÁO TÁC ĐỘNG CỦA AE ĐỐI VỚI VIỆT NAM

Mặc dù Việt Nam hiện nay chưa xuất hiện rõ nét một hệ sinh thái AE, tuy nhiên các xu hướng công nghệ toàn cầu cho thấy sự chuyển đổi này sớm sẽ tác động sâu rộng đến kinh tế - xã hội nước ta. Việc chủ động đánh giá các cơ hội và thách thức từ AE sẽ giúp Việt Nam chuẩn bị hành lang chính sách phù hợp, tận dụng tốt lợi ích đồng thời giảm thiểu rủi ro.

Cơ hội cho phát triển kinh tế - xã hội. : AE có thể trở thành bệ phóng mới cho tăng trưởng kinh tế số Việt Nam. Nhờ tự động hóa giao dịch và quy trình, các tác nhân AI có thể giúp doanh

nghiệp Việt nâng cao năng suất, hiệu quả vận hành. Ví dụ, doanh nghiệp vừa và nhỏ có thể sử dụng trợ lý AI để tự động hóa khâu marketing, bán hàng trực tuyến như chatbot tư vấn khách hàng, tác nhân phân tích dữ liệu thị trường mà không cần thuê nhiều nhân viên, giúp giảm chi phí và mở rộng quy mô kinh doanh. Đối với người tiêu dùng: trợ lý AI cá nhân giúp họ sàng lọc và đàm phán với hàng loạt tác nhân doanh nghiệp để đề xuất dịch vụ tối ưu, giúp họ dễ dàng tiếp cận dịch vụ chất lượng với chi phí thấp hơn. Ở tầm vĩ mô, AE có thể đẩy nhanh chuyển đổi số các ngành. Trong chính phủ điện tử, các tác nhân AI có thể đảm nhiệm vai trò "công chức số" hỗ trợ xử lý thủ tục hành chính, trả lời câu hỏi cho người dân, giám sát chất lượng dịch vụ công, điều này là hữu ích khi Việt Nam đang nỗ lực cải thiện dịch vụ công, bởi tác nhân AI hoạt động 24/7, giúp giảm tải bộ máy và nâng cao mức độ hài lòng của người dân. Tác nhân AI còn thúc đẩy đổi mới sáng tạo: nhà nghiên cứu, kỹ sư có thể tận dụng tác nhân AI làm cộng sự để mô phỏng thí nghiệm, tổng hợp tài liệu, thiết kế sản phẩm mẫu, rút ngắn thời gian R&D trong các lĩnh vực như dược phẩm, kỹ thuật. Nhìn chung, nếu được khai thác đúng cách, AE sẽ góp phần hiện thực hóa mục tiêu kép của Việt Nam: vừa tăng năng suất lao động, vừa nâng cao năng lực đổi mới sáng tạo, hướng tới nền kinh tế số có giá trị gia tăng cao.

Một khía cạnh quan trọng khác là AE có thể giúp Việt Nam hội nhập sâu hơn vào kinh tế số toàn cầu. Hiện nay, các doanh nghiệp công nghệ Việt còn hạn chế nguồn lực so với tập đoàn lớn thế giới, nên khó cạnh tranh ở mảng dịch vụ nền tảng. Tuy nhiên, với sự phổ cập của các chuẩn mở cho tác nhân AI như giao thức Agent-to-Agent (A2A), các công ty của Việt Nam có thể phát triển dịch vụ chuyên biệt nhưng vẫn tham gia được mạng lưới tác nhân toàn cầu. Ví dụ, một startup Việt có thể tạo tác nhân AI cung cấp dịch vụ du lịch thông minh như tư vấn tour, đặt vé máy bay, khách sạn, kết nối qua A2A với hàng nghìn trợ lý AI của khách hàng quốc tế. Như vậy, doanh nghiệp Việt vẫn tiếp cận thị trường rộng lớn mà không cần phát triển ứng dụng riêng cho từng người dùng. Cơ hội khác là thu hẹp khoảng cách dịch vụ giữa đô thị và nông thôn: tác nhân AI có thể mang dịch vụ chuyên gia đến vùng sâu vùng xa như bác sĩ AI tư vấn sức khỏe cơ bản, giáo viên AI dạy ngoại ngữ, giúp giảm bất bình đẳng trong tiếp cận tri thức và dịch vụ. Ngoài ra, Việt Nam có cơ hội thu hút đầu tư nước ngoài vào các dự án AE. Các tập đoàn công nghệ có thể chọn Việt Nam làm nơi thí điểm triển khai tác nhân AI trong khu vực ASEAN, qua đó tạo việc làm trình độ cao và chuyển giao công nghệ cho nguồn nhân lực nội địa.

Thách thức và rủi ro: Bên cạnh cơ hội, Việt Nam sẽ đối mặt nhiều thách thức khi AE dần định hình trên thế giới và có thể du nhập vào trong nước.

Đầu tiên là khoảng trống pháp lý, Việt Nam chưa có văn bản pháp luật chuyên biệt về AI, một số quy định hiện hành có thể áp dụng phần nào, như Luật An ninh mạng 2018 có điều khoản xử lý thông tin sai lệch; Nghị định 13/2023 về bảo vệ dữ liệu cá nhân, nhưng chưa đủ một khung toàn diện cho những vấn đề đặc thù của tác nhân AI như trách nhiệm pháp lý, an toàn của hệ thống tự động, hay quyền và nghĩa vụ của các bên khi sử dụng AI. Nếu không sớm xây dựng khuôn khổ pháp lý rõ ràng, chúng ta có nguy cơ bị động trước các tình huống phát sinh: chẳng hạn, nếu một chatbot tự động gây hiểu lầm nghiêm trọng dẫn đến hậu quả xấu, hiện chưa có cơ chế quy trách nhiệm dứt khoát hay chế tài phù hợp. Ngày 14/6/2025, Quốc hội khóa XV, Kỳ họp thứ 9 đã thông qua Luật Công nghiệp công nghệ số 2025 (Luật số 71/2025/QH15), có hiệu lực từ ngày 01/01/2026. [12]. Luật này tạo khung pháp lý nền tảng cho nghiên cứu, phát triển, cung cấp và sử dụng AI - trong đó có nguyên tắc giữ con người ở vị trí ra quyết định cuối cùng và quản lý AI theo cách tiếp cận phân loại rủi ro tương tự EU [8]. Tuy vậy, thách thức sẽ nằm ở khâu thực thi: làm sao để các quy định không chỉ nằm trên giấy mà được tuân thủ nghiêm trong thực tế. Việt Nam có thể phải đối mặt hiện tượng "vùng xám pháp lý" khi công nghệ chạy trước luật pháp, gây khó khăn cho việc xử lý.

Thứ hai là năng lực quản trị và điều phối. AE có tính liên ngành rất cao, là sự giao thoa của công nghệ thông tin, kinh tế, pháp luật, an ninh..., điều này đòi hỏi bộ máy quản lý đủ năng lực và cơ chế phối hợp hiệu quả. Hiện nay, các cơ quan quản lý nhà nước của ta còn thiếu nhân sự am hiểu sâu về AI, các đầu mối phụ trách AI phân tán ở nhiều bộ ngành khác nhau, dẫn đến nguy cơ chồng chéo hoặc bỏ sót trách nhiệm. Nếu không thiết lập một đầu mối điều phối thống nhất ở tầm quốc gia và nâng cao năng lực quản trị AI, Việt Nam khó theo kịp diễn biến Agent Economy và phản ứng chính sách kịp thời. Ngoài ra, hạ tầng kỹ thuật phục vụ phát triển và kiểm soát AI tại Việt Nam còn hạn chế như năng lực tính toán, dữ liệu dùng huấn luyện mô hình lớn. Điều này có thể khiến Việt Nam phụ thuộc công nghệ nước ngoài và thụ động trong quản lý, giám sát tác nhân AI ngoại nhập.

Thứ ba, AE cũng đặt ra nguy cơ biến động thị trường lao động ở Việt Nam. Nhiều việc làm tại Việt Nam, nhất là trong các ngành dịch vụ hay văn phòng có thể bị tác nhân AI thay thế một phần, đặc biệt là lực lượng lao động trình độ thấp hoặc kỹ năng lặp lại sẽ chịu ảnh hưởng trước.

Nếu không có chính sách hỗ trợ đào tạo lại và chuyển đổi việc làm, tình trạng thất nghiệp cơ cấu ngắn hạn có thể gia tăng, kéo theo hệ lụy xã hội như bất bình đẳng thu nhập, gánh nặng an sinh xã hội. Mặt khác, doanh nghiệp Việt Nam đa phần quy mô vừa và nhỏ, năng lực tài chính, công nghệ hạn chế nên có thể bị tụt hậu trong cuộc đua AE, nhường thị phần cho các tác nhân, dịch vụ ngoại nhập. Người tiêu dùng Việt cũng có nguy cơ phụ thuộc vào tác nhân AI nước ngoài nếu doanh nghiệp nội không cung cấp được dịch vụ tương tự, điều này tiềm ẩn rủi ro về chủ quyền số và an ninh dữ liệu.

Cuối cùng, nhận thức xã hội về tác nhân AI ở Việt Nam còn chưa cao, nhiều người dùng có thể không biết mình đang tương tác với AI hay hiểu rõ quyền lợi, rủi ro liên quan. Điều này đòi hỏi nỗ lực lớn về tuyên truyền, giáo dục để người dân chuẩn bị tâm thế và kỹ năng cho kỷ nguyên mới.

Tóm lại, AE có thể mang đến cơ hội nâng tầm kinh tế số Việt Nam nhưng cũng đặt ra bài toán quản lý phức tạp. Cần nhận diện đầy đủ các thách thức trên cả phương diện pháp lý, quản trị nhà nước, kinh tế và xã hội để có giải pháp chủ động ứng phó.

6. BÀI HỌC KINH NGHIỆM

Kinh nghiệm của các quốc gia trên đem lại những bài học quý giá cho Việt Nam trong định hướng xây dựng chính sách quản lý AE. Mỗi mô hình đều có điểm mạnh và hạn chế riêng, song tựu trung có thể rút ra một số bài học chung như sau:

Thứ nhất, cần sớm xây dựng chiến lược và khung pháp lý về AI với tầm nhìn dài hạn. Kinh nghiệm EU và Trung Quốc cho thấy việc ban hành luật AI chuyên biệt giúp định hình hành lang an toàn ngay từ đầu. Dù cách tiếp cận khác nhau, cả hai đều nhất quán nguyên tắc con người kiểm soát cuối cùng. Việt Nam cần cập nhật Chiến lược quốc gia về AI để bổ sung định hướng về AE, không để bị chậm chân.

Thứ hai, thiết lập cơ chế điều phối liên ngành và nâng cao năng lực quản trị AI. Việt Nam nên thành lập cơ quan về AI và Chuyển đổi số với sự tham gia của lãnh đạo Chính phủ, các bộ ngành chủ chốt và chuyên gia.

Thứ ba, tập trung phát triển hạ tầng số và dữ liệu phục vụ AI và AE. Việt Nam cần tăng tốc triển khai các dự án hạ tầng số như xây dựng trung tâm tính toán hiệu năng cao quốc gia phục vụ nghiên cứu và phát triển AI.

Thứ tư, đảm bảo an toàn, an ninh và đạo đức trong môi trường AE. Đây là trụ cột sống còn để tạo lòng tin xã hội khi triển khai tác nhân AI, bài học từ Trung Quốc cho thấy cần có cơ chế ngắt kết nối khẩn cấp, cơ quan chức năng được quyền tạm dừng hoạt động của một nhóm tác nhân AI hoặc ngắt kết nối Internet của chúng nếu phát hiện hành vi bất thường nguy hiểm, tương tự cơ chế dừng khẩn cấp sàn chứng khoán khi xảy ra sự cố lớn. Việt Nam có thể đưa nguyên tắc này vào luật AI hoặc quy định dưới luật để có căn cứ thực hiện khi cần.

Thứ năm, chú trọng đào tạo nhân lực và hỗ trợ chuyển đổi việc làm trong kỷ nguyên tác nhân AI. Từ bài học quốc tế là nâng cao kỹ năng số cho lực lượng lao động, Việt Nam nên xây dựng các chương trình nâng cao kỹ năng số và kỹ năng AI cho người lao động.

Thứ sáu, tăng cường hợp tác quốc tế và chủ động định hình chuẩn mực khu vực về AI. AE là hiện tượng mang tính toàn cầu, do đó không quốc gia nào tự quản lý hiệu quả nếu thiếu hợp tác và học hỏi lẫn nhau. Bài học từ Singapore là chủ động đề xuất sáng kiến khu vực, Việt Nam có thể đề xuất một khung quản trị AI ASEAN, dựa trên các nguyên tắc chung phù hợp văn hóa khu vực.

Những bài học trên cho thấy tầm quan trọng của việc hành động sớm, linh hoạt và hợp tác trong quản trị AE. Việt Nam cần chọn lọc và áp dụng sáng tạo các kinh nghiệm phù hợp, tránh thái độ chủ quan hoặc chạy theo một mô hình cố định.

7. MỘT SỐ ĐỀ XUẤT

Dựa trên cơ sở lý luận và bài học quốc tế đã phân tích, phần này bài viết đề xuất một khung chính sách quản lý AE cho Việt Nam nhằm vừa thúc đẩy phát triển công nghệ, vừa bảo vệ lợi ích xã hội. Khung chính sách gồm sáu nhóm giải pháp chính sau: (1) Hoàn thiện thể chế pháp luật; (2) Thiết lập cơ chế điều phối và quản trị liên ngành; (3) Phát triển hạ tầng kỹ thuật số và dữ liệu; (4) Đảm bảo an toàn, an ninh và đạo đức trong Agent Economy; (5) Đào tạo nhân lực và hỗ trợ chuyển đổi việc làm; (6) Hợp tác quốc tế và thúc đẩy nghiên cứu triển khai.

(1) Hoàn thiện thể chế pháp luật về AI và Agent Economy: Việt Nam cần sớm ban hành Luật riêng về Trí tuệ nhân tạo làm nền tảng pháp lý cao nhất cho quản lý AI tại Việt Nam. Có thể dành một chương riêng quy định về hệ thống AI có tính tự chủ cao, trong đó nêu nguyên tắc quản lý AE như "AI phục vụ con người"; yêu cầu mọi triển khai tác nhân AI phải có cơ chế để con người giám sát và chịu trách nhiệm cuối cùng như luật đã nêu: "đảm bảo con người không bị thay thế hoàn toàn trong các quyết định quan trọngvà con người giữ trách nhiệm tối thượng" [12]. Luật

cũng nên quy định các hành vi bị nghiêm cấm liên quan đến AI, chẳng hạn: sử dụng tác nhân AI để thực hiện hành vi phạm tội; phát tán thông tin giả mạo gây hoang mang; xâm phạm đời tư cá nhân qua AI; phát triển sử dụng AI không qua kiểm định trong các lĩnh vực có điều kiện như tài chính.

(2) Thiết lập cơ chế điều phối liên ngành và tăng cường năng lực quản trị: cần thành lập cơ quan về AI và Chuyển đổi số, cơ quan này chịu trách nhiệm định hướng chiến lược AI quốc gia, các chương trình trọng điểm. Đồng thời, cơ quan theo dõi việc triển khai Luật AI, đề xuất cập nhật danh mục rủi ro, cấm và xử lý các vấn đề liên ngành nảy sinh như sự cố tác nhân AI diện rộng liên quan đến nhiều lĩnh vực.

Để nâng cao năng lực quản trị, Việt Nam cần tăng cường đào tạo đội ngũ "công chức AI". Thêm nữa, cần trang bị công cụ kỹ thuật cho cơ quan quản lý, như triển khai hệ thống quét và phát hiện nội dung deepfake trên Internet; hệ thống giám sát giao dịch tài chính tự động; xây dựng phòng thí nghiệm kiểm thử thuật toán AI để kiểm thử các mô hình AI xem có thiên lệch hay lỗ hổng không.

(3) Phát triển hạ tầng số và dữ liệu cho Agent Economy: Chính phủ cần xem AE là động lực để đầu tư hạ tầng số thế hệ mới như: hạ tầng tính toán, hạ tầng kết nối, hạ tầng dữ liệu. Đồng thời xúc tiến xây dựng các kho dữ liệu mở phục vụ cộng đồng AI để cả nước có thể dùng huấn luyện mô hình AI của Việt Nam. Khuyến khích các tổ chức, doanh nghiệp chia sẻ dữ liệu đã được ẩn danh lên nền tảng dữ liệu mở cơ quan chính phủ quản lý, kèm cơ chế khen thưởng nếu tích cực đóng góp dữ liệu và chế tài nếu không tuân thủ quy định.

(4) Đảm bảo an toàn, an ninh và đạo đức trong AE: Đây là trụ cột cốt lõi để tạo niềm tin xã hội khi triển khai AE. Nên thiết kế sẵn các "chốt chặn khẩn cấp" về kỹ thuật và pháp lý để ngăn chặn sự cố AI lan rộng. Bên cạnh đó, cần thiết lập hệ thống phân cấp độ tin cậy cho tác nhân AI, tác nhân AI muốn cung cấp dịch vụ rộng rãi phải được cấp phép hoặc chứng nhận an toàn bởi cơ quan quản lý. Khi tác nhân vi phạm luật, cơ quan chức năng thu hồi giấy phép và đưa tác nhân đó vào danh sách đen. Việt Nam có thể nghiên cứu ứng dụng công nghệ blockchain hoặc PKI cho việc cấp danh tính số và chứng chỉ cho tác nhân AI để quản lý.

Về an ninh mạng và phòng chống tội phạm AI: Cần nâng cao khả năng phát hiện sớm và xử lý các hành vi lợi dụng tác nhân AI. Cần trang bị công cụ AI cho chính cơ quan công an để dò

tìm loạt tài khoản ảo trên mạng xã hội lan truyền tin giả. Tăng cường phối hợp quốc tế để truy vết các vụ lừa đảo xuyên biên giới dùng AI. Về phòng chống gian lận, có thể xem xét yêu cầu bổ sung đối với nhà mạng viễn thông/ISP: giám sát và báo cáo nếu phát hiện lưu lượng dữ liệu bất thường nghi do mạng bot AI điều khiển, đồng thời, cần chuẩn bị kịch bản ứng phó sự cố AI khi cần thiết.

Về đạo đức và xã hội: Chính sách cần đảm bảo nguyên tắc không thiên vị, không phân biệt đối xử trong hệ thống AI [9]. Yêu cầu các nhà phát triển và triển khai tác nhân AI phải thực hiện kiểm thử và có biện pháp giảm thiểu thiên lệch nhằm tránh gây bất công cho nhóm yếu thế. Ngoài ra, cần bảo tồn văn hóa, ngôn ngữ: ưu tiên phát triển các mô hình AI sử dụng tiếng Việt, hiểu biết luật pháp và văn hóa Việt Nam.

(5) Đào tạo nhân lực và chuyển đổi việc làm trong kỷ nguyên tác nhân AI: Chính phủ nên triển khai một chương trình nâng cao kỹ năng số và kỹ năng AI cho người lao động gồm các hợp phần: đào tạo lại cho lao động trong những ngành dễ bị ảnh hưởng; bồi dưỡng nâng cao cho nhân viên văn phòng, kỹ sư về cách sử dụng công cụ AI tăng hiệu quả công việc; đào tạo bổ sung kiến thức AI cơ bản cho cán bộ quản lý, lãnh đạo doanh nghiệp để họ ra quyết định đúng đắn khi ứng dụng AI.

(6) Tham gia hợp tác quốc tế, định hình chuẩn mực khu vực: Việt Nam có thể đăng cai một sandbox AI chung cấp ASEAN - mời doanh nghiệp các nước ASEAN cùng thử nghiệm tác nhân AI trong một lĩnh vực có sự đồng thuận chung. Thông qua sandbox này, các nước sẽ cùng thống nhất một số quy tắc và tiêu chuẩn khu vực, tạo tiền đề cho hiệp định chung về AI sau này.

Trên diễn đàn quốc tế, Việt Nam có thể tham gia tích cực các chương trình như Đối tác Toàn cầu về AI, các nhóm công tác về AI của OECD, để vừa học hỏi chính sách, vừa đóng góp góc nhìn của nước đang phát triển. Khi Liên Hợp Quốc, UNESCO có các hoạt động về AI, Việt Nam cần nghiên cứu và nội luật hóa những khuyến nghị phù hợp.

Về hợp tác nghiên cứu, cần thúc đẩy các dự án nghiên cứu chung về AI giữa viện trường Việt Nam với đối tác phát triển. Tận dụng các quỹ như Horizon châu Âu, Quỹ VINIF để tài trợ các đề tài nghiên cứu về tác động xã hội của AI, phương pháp kiểm soát AI, v.v., tạo cơ sở khoa học cho chính sách. Khuyến khích trao đổi chuyên gia: mời các nhà khoa học Việt Kiều trong lĩnh vực AI về nước làm việc ngắn hạn, giúp đào tạo đội ngũ trẻ và chuyển giao tri thức.

Ngoài ra, hợp tác quốc tế rất cần thiết trong việc hình thành các hiệp định, quy ước quốc tế về AE. Việt Nam nên tích cực tham gia thể hiện tinh thần trách nhiệm và đồng hành cùng cộng đồng quốc tế hướng tới một tương lai AI an toàn và công bằng.

Khung chính sách đề xuất ở trên nhằm đảm bảo Việt Nam sẵn sàng về pháp lý, kỹ thuật và nguồn lực để đón nhận AE. Quan trọng hơn cả là tạo được đồng thuận xã hội: các bên từ Chính phủ, doanh nghiệp đến người dân đều hiểu vai trò của mình và hợp tác trong quá trình triển khai. Chính sách mềm dẻo, có tinh thần học hỏi và cải tiến liên tục bởi bản thân Agent Economy sẽ tiến hóa không ngừng. Như Rothschild và cộng sự nhận định: *"Lựa chọn chúng ta thực hiện hôm nay sẽ quyết định không chỉ cách thị trường này vận hành mà còn ai sẽ hưởng lợi từ làn sóng công nghệ mới"* [3]. Việt Nam cần những quyết sách đúng đắn ngay từ bây giờ để đảm bảo Agent Economy trong tương lai sẽ phục vụ lợi ích chung của dân tộc, đồng thời đóng góp vào sự phồn vinh của cộng đồng quốc tế trong thời đại số.

8. KẾT LUẬN

Agent Economy - nền kinh tế vận hành bởi các tác nhân AI tự trị – đang đặt nhân loại trước một bước ngoặt lịch sử. Những lợi ích to lớn về năng suất, hiệu quả, sáng tạo AE hứa hẹn mang lại đi kèm với những thách thức phức tạp về quản trị công nghệ, pháp lý và đạo đức. Đối với Việt Nam, một quốc gia đang tăng tốc chuyển đổi số, việc chủ động nắm bắt xu hướng AE và xây dựng khung chính sách quản lý phù hợp có ý nghĩa sống còn để không bị tụt hậu, đồng thời tránh sa vào mặt trái của công nghệ.

Bài nghiên cứu đã hệ thống hóa cơ sở lý luận về AE, nhấn mạnh đặc trưng của tác nhân AI và các tác động kinh tế - xã hội mà chúng có thể mang lại. Kinh nghiệm quốc tế được phân tích cho thấy nhiều cách tiếp cận đa dạng: từ Mỹ đề cao đổi mới đến EU thận trọng với luật lệ, từ Trung Quốc kiểm soát chặt đến Singapore linh hoạt cân bằng. Không có mô hình nào hoàn hảo tuyệt đối, nhưng tất cả đều chung nhận định rằng quản trị Agent Economy đòi hỏi tầm nhìn xa, phối hợp đa ngành và tinh thần cải cách liên tục.

Đối chiếu với điều kiện Việt Nam, bài viết đã đề xuất một khung chính sách toàn diện, bao quát từ hoàn thiện thể chế pháp luật, nâng cao năng lực quản lý, phát triển hạ tầng, đảm bảo an toàn - an ninh, đào tạo nguồn nhân lực đến hợp tác quốc tế. Một số đề xuất cụ thể đáng chú ý gồm: xây dựng Luật AI với cách tiếp cận quản lý theo rủi ro; thành lập cơ quan về AI để điều phối liên

ngành; triển khai cơ chế regulatory sandbox để thử nghiệm AI trong phạm vi kiểm soát; đầu tư Trung tâm HPC quốc gia và thúc đẩy dữ liệu mở; áp dụng hệ thống chứng chỉ và đăng ký tác nhân AI; lập quỹ hỗ trợ đào tạo lại lao động; và tham gia tích cực các sáng kiến quản trị AI toàn cầu.

Khung chính sách này hướng đến mục tiêu kép: thúc đẩy đổi mới sáng tạo trong lĩnh vực AI/tác nhân AI để phục vụ phát triển kinh tế - xã hội, đồng thời thiết lập các "làn ranh đỏ" và cơ chế kiểm soát hiệu quả để hạn chế rủi ro, bảo vệ người dân và ổn định xã hội. Việc cân bằng hai mục tiêu này không hề dễ dàng, đòi hỏi nghệ thuật quản lý tinh tế cũng như sự chung sức của mọi thành phần liên quan. Nhà nước cần đóng vai trò định hướng và điều phối, nhưng thành công sẽ phụ thuộc vào việc doanh nghiệp có tuân thủ triển khai AI một cách trách nhiệm và người dân có sẵn sàng đón nhận, nâng cao kỹ năng để cộng tác với AI hay không.

TÀI LIỆU THAM KHẢO

[1]. S. Hosseini & H. Seilani (2025). *The role of agentic AI in shaping a smart future: A systematic review*, Array, 26, 100399.

[2]. T. J. Chaffer (2025). *Can We Govern the Agent-to-Agent Economy?* arXiv preprint arXiv:2501.16606.

[3]. D. M. Rothschild *et al.* (2025). *The Agentic Economy*, arXiv preprint arXiv:2505.15799.

[4]. M. Yıldızoğlu (2023). *AI Agents: Transforming Economics and Beyond*, Lecture notes, University of Bordeaux.

[5]. S. Cecchini *et al.* (2025). *Virtual Agent Economies (Sandbox Economy)*, arXiv preprint arXiv:2509.10147.

[6]. The Future Society (2025). *Ahead of the Curve: Governing AI Agents under the EU AI Act* (Report). Website: https://thefuturesociety.org/aiagentsintheeu/

[7]. Infocomm Media Development Authority – IMDA Singapore (2025). *Singapore launches new tools to help businesses protect data and deploy AI in a trusted ecosystem* (Press release, 07/07/2025). Website: https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2025/singapore-launches-new-tools-to-help-businesses-protect-data-and-deploy-ai-in-a-trusted-ecosystem

[8]. E. Fincken (2025). *Vietnam unveils draft artificial intelligence law*, Global Legal Insights (GLI), 6 Oct 2025. Website: https://www.globallegalinsights.com/news/vietnam-unveils-draft-artificial-intelligence-law/

[9]. White & Case (2023). *AI Watch: Global regulatory tracker – China*, White & Case (online). Truy cập 2025, Website: https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china

[10]. A. Turrell (2020). *Agent-based models: understanding the economy from the bottom up*, Bank of England Quarterly Bulletin, Q4/2020.

[11]. Software Improvement Group (2025). *AI legislation in the US: A 2025 overview*, SIG Blog (online). Truy cập 2025, Website: https://www.softwareimprovementgroup.com/blog/us-ai-legislation-overview/

[12] Việt Nam (2025) Luật Công nghiệp Công nghệ số. Website: https://vanban.chinhphu.vn/?pageid=27160&docid=214609&classid=1&typegroupid=3