

TRÍ TUỆ NHÂN TẠO VÀ QUẢN TRỊ NHÀ NƯỚC: NÂNG CAO HIỆU QUẢ, MINH BẠCH HAY NGUY CƠ LẠM QUYỀN?

NGUYỄN BÁ CHIẾN*

Trí tuệ nhân tạo (AI) đang tác động sâu rộng đến quản trị công, hứa hẹn nâng cao hiệu quả, minh bạch hóa quy trình và cải thiện chất lượng ra quyết định chính sách, tăng cường khả năng phản ứng trong khủng hoảng và minh bạch hóa thông tin cho công dân. Tuy nhiên, bên cạnh các lợi ích, AI cũng mang đến các rủi ro tiềm ẩn. Bài viết xây dựng trên khung lý thuyết về quản trị công tốt và đạo đức công nghệ, kết hợp với phân tích tình huống quốc tế và Việt Nam để đánh giá cả cơ hội cũng như thách thức của AI trong quản trị nhà nước. Từ đó, đề xuất một số chính sách: xây dựng khung pháp lý toàn diện, thành lập tổ chức giám sát độc lập, ban hành chuẩn mực đạo đức công nghệ, nâng cao năng lực số của công chức, người dân và bảo đảm minh bạch, trách nhiệm giải trình.

Từ khóa: Trí tuệ nhân tạo; quản trị nhà nước; hiệu quả; minh bạch; lạm quyền.

Artificial intelligence (AI) is creating profound changes in public governance, opening up opportunities to improve operational efficiency, increasing process transparency, enhancing decision-making quality, strengthening responsiveness to crises, and expanding citizens' access to information. However, alongside these benefits, AI also poses many challenges related to ethics, transparency, and accountability. Based on the theoretical framework of good governance and technology ethics, this article analyzes both the opportunities and risks of applying AI in state governance, while also proposing several key policies: perfecting the legal framework, establishing an independent oversight body, issuing technological ethics standards, developing digital capacity for civil servants and citizens, ensuring transparency, and strengthening accountability in public governance.

Keywords: Artificial intelligence; state governance; efficiency; transparency; power abuse.

NGÀY NHẬN: 15/8/2025

NGÀY PHẢN BIỆN, ĐÁNH GIÁ: 26/9/2025

NGÀY DUYỆT: 16/10/2025

DOI: <https://doi.org/10.59394/qlnn.357.2025.1306>

1. Giới thiệu khung lý thuyết và phương pháp nghiên cứu

Cuộc cách mạng số và kỷ nguyên dữ liệu mở ra những khả năng mới cho Chính phủ trong việc cải thiện hiệu quả và chất lượng dịch vụ công. Trong bối cảnh đó, trí tuệ nhân tạo (AI) trở thành một công cụ chiến lược của

chính phủ số, với mục tiêu tối ưu hóa thủ tục hành chính, ra quyết định chính sách dựa trên dữ liệu lớn và nâng cao khả năng giám sát, phản ứng trước khủng hoảng. Tuy nhiên,

* PGS. TS, Học viện Hành chính và Quản trị công

sự kết hợp giữa AI và quyền lực nhà nước cũng đặt ra các câu hỏi về công bằng, minh bạch và tính pháp quyền của chính quyền. Bài viết khai thác các lý thuyết quản trị công tốt (good governance) theo các quan điểm của Ngân hàng Thế giới (WB), UNDP, OECD về minh bạch, trách nhiệm, hiệu quả và quan điểm về ảnh hưởng của công nghệ đối với quyền lực, như: “Thuyết giám sát” của Foucault, nền kinh tế giám sát của Zuboff để phân tích về cơ hội và nguy cơ khi áp dụng AI trong quản trị nhà nước.

Lý thuyết quản trị công tốt đóng vai trò nền tảng cho phân tích. Theo OECD và Liên Hợp quốc, Chính phủ hoạt động tốt phải bảo đảm các giá trị cơ bản, như: pháp quyền, minh bạch, trách nhiệm giải trình, hiệu quả, và sự tham gia của công dân. Trong ngữ cảnh AI, Jana Mišić và cộng sự (2025)¹ đề xuất một khung lý thuyết kết hợp giữa “giá trị trật tự” (good order) và “giá trị xã hội” (good society).

Nhìn từ góc độ công nghệ và quyền lực, một số lý thuyết cảnh báo công nghệ AI có thể tạo ra sự giám sát kỹ thuật số, như: mô hình panopticon của Foucault và kinh tế giám sát (Shoshana Zuboff 2019)². AI tích hợp với các hệ thống thu thập dữ liệu lớn có khả năng củng cố cơ chế quan sát và kiểm soát xã hội nếu không có cơ chế kiểm soát thích hợp. Ngược lại, công nghệ chỉ là công cụ trung lập; từ góc độ này, quyền lực vẫn nằm ở con người sử dụng công nghệ. Do đó, phần khung lý thuyết của bài viết kết hợp cả hai chiều: AI được xem như một công cụ hữu ích cho quản trị công nhưng cũng phải được quản lý chặt chẽ để ngăn ngừa lạm dụng quyền lực.

Về phương pháp, nghiên cứu sử dụng phân tích tài liệu định tính. Dữ liệu được thu thập từ các công bố học thuật, báo cáo chính sách quốc tế (OECD, UNESCO, Liên Hợp quốc...) và nguồn chính thức; đồng thời, so sánh các trường hợp thực tiễn về ứng dụng AI trong quản trị công ở một số nước phát

triển và Việt Nam. Căn cứ vào khung lý thuyết nêu trên, các minh chứng và khuyến nghị chính sách được sàng lọc, trích dẫn rõ nguồn và phân tích khách quan nhằm bảo đảm tính khoa học của những nhận định.

2. AI và khả năng nâng cao hiệu quả, minh bạch trong quản trị nhà nước

Thứ nhất, AI giúp tối ưu hóa quy trình hành chính. Một trong những ứng dụng rõ nét nhất của AI trong quản trị công là khả năng tự động hóa và tối ưu hóa quy trình hành chính. Nhiều công việc vốn đòi hỏi nhiều thời gian và nhân lực, như: tiếp nhận hồ sơ, phân loại dữ liệu, xử lý thủ tục, các tác nghiệp cụ thể... có thể được AI thay thế hoặc hỗ trợ giúp rút ngắn thời gian và giảm công sức cho công chức và người dân. Ở Estonia, hệ thống chính phủ điện tử X-Road, tích hợp AI để liên thông dữ liệu giữa các cơ quan. Kết quả có hơn 95% dịch vụ công có thể được truy cập trực tuyến, tiết kiệm hàng triệu giờ cho cả công dân và bộ máy hành chính (e-Estonia, 2022)³. X-Road giúp Estonia tiết kiệm thời gian làm việc mỗi năm hơn 1.345 lần, bằng cách: (1) Loại bỏ giấy tờ và các thủ tục không thực sự cần thiết, (2) Cho phép truy cập tức thời vào thông tin đáng tin cậy, (3) Công chức tập trung vào các nhiệm vụ có ý nghĩa đòi hỏi tư duy của con người, không phải thực hiện những tác nghiệp cụ thể.

Tại Các Tiểu vương quốc Ả Rập Thống nhất (UAE), trợ lý ảo “Rashid” (Virtual Assistant Rashid) được triển khai để phục vụ công dân 24/7, giải đáp thông tin và hỗ trợ thực hiện dịch vụ công, qua đó, giảm tải cho công chức (Government of UAE, 2021)⁴.

Tương tự, ứng dụng AI ở Trung Quốc cho phép đăng ký kinh doanh hay đăng ký phương tiện tự động hóa, giảm rủi ro thủ tục sai sót. Ở Việt Nam, một số cơ quan mới bắt đầu triển khai trợ lý ảo hỗ trợ dân (chatbot) tại cổng dịch vụ công và bộ phận một cửa, bước đầu chứng minh khả năng giải quyết câu hỏi đơn giản hơn.

Thứ hai, AI có khả năng xử lý và phân tích lượng dữ liệu khổng lồ (Big Data) từ các nguồn khác nhau. AI giúp chính phủ khai thác nguồn dữ liệu khổng lồ vốn được ví như “tài nguyên ngủ yên” (sleeping resources). Các công nghệ học máy (machine learning) và học sâu (deep learning) cho phép phát hiện các mẫu dữ liệu phức tạp, từ đó, nâng cao năng lực dự báo và hoạch định chính sách dựa trên bằng chứng.

Tại Singapore, Chương trình Smart Nation sử dụng AI để phân tích dữ liệu giao thông theo thời gian thực, từ đó, điều chỉnh tín hiệu đèn giao thông, dự báo ùn tắc và cải thiện vận tải công cộng (Kraemer-Mbula et al., 2021)⁵. Trong lĩnh vực y tế, AI được áp dụng để phát hiện sớm dịch bệnh thông qua phân tích bệnh án điện tử và dữ liệu mạng xã hội, giúp Chính phủ phản ứng kịp thời trong giai đoạn dịch Covid-19 (Whitelaw et al., 2020)⁶. Tại Việt Nam, Cục Y tế dự phòng kết hợp AI để quản lý thông tin dữ liệu bệnh truyền nhiễm nâng cao chất lượng giám sát, phát hiện, cảnh báo sớm và dự báo tình hình dịch bệnh; triển khai xây dựng hệ thống quản lý thông tin dữ liệu bệnh truyền nhiễm theo định hướng sử dụng nền tảng dùng chung tích hợp, kết nối các hệ thống liên quan.

Những ứng dụng này cho thấy, AI đã chuyển đổi cách Chính phủ sử dụng dữ liệu thay vì chỉ lưu trữ, dữ liệu trở thành nền tảng tri thức phục vụ dự báo và thiết kế chính sách. Đây là bước tiến quan trọng hướng tới Chính phủ dựa trên dữ liệu (data-driven government).

Thứ ba, AI hỗ trợ ra quyết định chính sách nhanh trong quản lý khủng hoảng và ứng phó tình huống khẩn cấp. AI cũng đóng vai trò như “trợ lý ảo” cho các nhà hoạch định chính sách. Các mô hình dự báo, mô phỏng kịch bản và phân tích xu hướng giúp Chính phủ đánh giá tác động của chính sách trước khi ban hành, từ đó, giảm thiểu rủi ro. AI có thể xử lý dữ liệu thời gian thực giúp chính quyền phản ứng kịp thời với thiên tai,

dịch bệnh hoặc sự kiện khẩn cấp. Hệ thống AI dự báo lũ quét, cháy rừng hoặc giám sát đường phố có thể cảnh báo trước cho cơ quan chức năng. Các công cụ học máy hỗ trợ phân tích tín hiệu cảnh báo sớm và lập kế hoạch ứng phó.

Tại Canada, AI được sử dụng trong hệ thống tư pháp để dự báo nguy cơ tái phạm của phạm nhân, hỗ trợ thẩm phán trong các quyết định bảo lãnh và quản lý án treo (Kehl & Kessler, 2017)⁷. Dù gây tranh luận về tính thiên lệch, ứng dụng này cho thấy tiềm năng của AI trong việc cung cấp dữ liệu hỗ trợ quyết định. Ở Hàn Quốc, AI được áp dụng để phân tích dữ liệu kinh tế vĩ mô, dự báo biến động thị trường lao động và tài chính, giúp Chính phủ kịp thời điều chỉnh chính sách (Lee, 2020)⁸. Liên minh châu Âu (EU) cũng thử nghiệm sử dụng AI để đánh giá tác động môi trường, mô phỏng các kịch bản phát thải CO2 nhằm lựa chọn giải pháp tối ưu (European Commission, 2021)⁹.

Tại Việt Nam, mặc dù còn mới nhưng ý tưởng xây dựng hệ thống giám sát, cảnh báo thiên tai thông qua ứng dụng AI đang được nghiên cứu, hứa hẹn cải thiện năng lực ứng phó thời tiết cực đoan.

Những trường hợp này khẳng định AI có thể cải thiện tính chính xác và khoa học trong hoạch định chính sách. AI đóng vai trò bổ trợ, nâng cao chất lượng quyết định và tăng cường trách nhiệm giải trình khi các quyết định dựa trên bằng chứng dữ liệu. AI không chỉ nâng cao hiệu quả quản trị trong điều kiện bình thường mà còn tăng khả năng ứng phó trước các khủng hoảng bất thường, một yếu tố cốt lõi của quản trị thích ứng và linh hoạt.

Thứ tư, AI tăng cường minh bạch và trách nhiệm giải trình. Ngoài hiệu quả, AI còn góp phần nâng cao minh bạch trong quản trị nhà nước. Tại Brazil, các cơ quan kiểm toán đã áp dụng AI để phân tích hàng triệu giao dịch tài chính công, phát hiện dấu hiệu bất thường và cảnh báo tham nhũng (OECD, 2020)¹⁰.

Một ví dụ tiêu biểu khác là Hàn Quốc, nơi Chính phủ phát triển nền tảng Government 24, tích hợp AI để cung cấp thông tin về hơn 5.000 dịch vụ công. Người dân có thể truy cập trực tuyến và theo dõi tiến trình xử lý hồ sơ, từ đó giảm thiểu tình trạng “xin - cho” và tăng cường giám sát của công dân (Lee, 2020)¹¹. Ở Phần Lan, AI được dùng để phân tích ý kiến của công dân trong các dự thảo luật, tạo điều kiện để Chính phủ tiếp nhận phản hồi một cách minh bạch và khách quan (Susha et al., 2022)¹².

3. Nguy cơ lạm quyền trong việc sử dụng AI vào quản trị nhà nước

Mặc dù AI hứa hẹn mang lại những cải tiến lớn cho hiệu quả và minh bạch, song các nghiên cứu chỉ ra rằng nó cũng có nguy cơ dẫn tới việc lạm quyền trong quản trị nhà nước (O’Neil, 2016)¹³.

Một là, AI có khả năng tạo ra nguy cơ lạm quyền trong quản trị nhà nước thông qua việc mở rộng quy mô giám sát xã hội và xâm phạm quyền riêng tư.

Hệ thống AI có thể kết hợp các nguồn dữ liệu công và tư tạo nên một “máy giám sát kỹ thuật số” khiến người dân bị theo dõi. Công nghệ nhận diện khuôn mặt và phân tích hành vi đã được nhiều thành phố lớn trên thế giới triển khai. Tại London (Anh), cảnh sát đô thị đã thử nghiệm hệ thống nhận diện khuôn mặt trực tiếp trong không gian công cộng, với lý do tăng cường an ninh. Tuy nhiên, báo cáo độc lập cho thấy, hệ thống này có tỷ lệ nhận diện sai cao và gây lo ngại về quyền riêng tư, quyền hội họp và quyền biểu đạt (Fussey & Murray, 2019)¹⁴. Tại Moscow, hơn 170.000 camera giám sát được tích hợp AI để nhận diện công dân trong thời gian thực, dẫn đến nhiều chỉ trích từ các tổ chức nhân quyền về việc xâm phạm quyền tự do cá nhân (Human Rights Watch, 2021)¹⁵ tức AI trong chính quyền phải có sự giám sát của con người và minh bạch.

Hai là, AI có khả năng tạo ra nguy cơ áp đặt và thiếu minh bạch trong thuật toán. Do được huấn luyện trên dữ liệu lịch sử, AI dễ dàng tái sản xuất các định kiến xã hội vốn có. Trường hợp hệ thống COMPAS tại Hoa Kỳ là minh chứng rõ nét: công cụ này được dùng để dự báo khả năng tái phạm của phạm nhân, nhưng bị phát hiện có xu hướng đánh giá rủi ro cao hơn đối với người da màu so với người da trắng, dù có cùng điều kiện (Angwin et al., 2016)¹⁶.

Các trường hợp này cho thấy, AI không hề “trung lập” như nhiều người lầm tưởng mà chịu ảnh hưởng sâu sắc từ dữ liệu đầu vào và cách thức thiết kế thuật toán của người đặt đề bài.

Ngoài thiên lệch, tính “hộp đen” của AI cũng là một thách thức nghiêm trọng. Nhiều hệ thống AI hoạt động với cơ chế ra quyết định khó giải thích, khiến công dân không thể hiểu hoặc phản biện quyết định hành chính ảnh hưởng trực tiếp đến quyền lợi của họ. Bởi lẽ AI dựa trên dữ liệu đã có (phần tĩnh) để đưa ra tiên đoán, nhưng khó có thể giải thích và thấu hiểu được nội tâm và sự sáng tạo của con người trong những cảnh huống cụ thể (phần động). Thiếu minh bạch về cách thức hoạt động của các thuật toán cũng đặt dấu hỏi về khả năng giải thích quyết định và trách nhiệm của chính quyền khi AI ra quyết định sai hoặc bất công. Khi không có khả năng giải thích, trách nhiệm giải trình của Nhà nước cũng bị suy giảm, đi ngược lại nguyên tắc minh bạch và công bằng trong quản trị công.

Ba là, tập trung quyền lực và độc quyền công nghệ. Khi dữ liệu cá nhân tập trung trong tay nhà nước và các công ty công nghệ, quyền lực chính trị và kinh tế có nguy cơ bị tập trung hóa. Điều này tạo ra sự bất cân xứng thông tin nghiêm trọng giữa nhà nước và công dân (Zuboff, 2019)¹⁷. Nếu không có kiểm soát, nhà nước có thể phụ thuộc hoặc cộng tác chặt chẽ với các doanh nghiệp gây nguy cơ lạm quyền và lũng đoạn quyền lực.

Thậm chí chính phủ các nước, trong lúc chạy đua công nghệ, có thể bỏ qua kiểm duyệt dân chủ. Như Kuziemski & Misuraca (2020)¹⁸ đã nhận định, công chức nhà nước đối mặt với nghịch lý: “Nghĩa vụ bảo vệ công dân khỏi tác hại của thuật toán song song với cám dỗ tăng hiệu quả qua chính thuật toán”. Nếu người ra quyết định coi AI là “công cụ vạn năng” để tăng hiệu suất mà không kiểm soát chặt, quyền lực của nhà nước dễ bị lợi dụng để thao túng, làm suy giảm tinh thần pháp quyền.

Bốn là, các hệ thống AI phức tạp thường khó giải thích (explicability) và kiểm toán. Nếu chính phủ triển khai AI nhưng không công bố dữ liệu đầu vào, thuật toán hoặc quy trình ra quyết định, công dân không thể biết lý do của quyết định ảnh hưởng trực tiếp đến họ (như từ chối phê duyệt đơn, xét thưởng, cấp phép). Điều này có thể dẫn đến khủng hoảng niềm tin và tính chính danh của chính phủ (Susha et al., 2022)¹⁹. Nhiều nghiên cứu cho thấy, việc công dân cảm nhận bị giám sát hoặc bị đối xử bất công bởi hệ thống AI có thể dẫn đến hiện tượng “tự kiểm duyệt hành vi”, làm suy yếu không gian công cộng và giảm mức độ tham gia xã hội (Foucault, 1995/1975)²⁰. Điều này đi ngược lại mục tiêu ban đầu của AI là tăng cường dân chủ và sự tham gia. Thay vì là công cụ phục vụ công dân, AI có nguy cơ trở thành rào cản cho sự phát triển của nền dân chủ bền vững.

Năm là, AI có thể tạo ra nguy cơ thao túng dư luận và suy giảm niềm tin xã hội. Helbing et al. (2019)²¹ cảnh báo việc sử dụng AI để kiểm soát và định hướng thông tin có thể ảnh hưởng đến hành vi chính trị của công dân. Vụ bê bối Cambridge Analytica năm 2018 là minh chứng điển hình, khi dữ liệu cá nhân từ hàng triệu người dùng Facebook bị khai thác nhằm mục tiêu quảng cáo chính trị, qua đó, thao túng quá trình bầu cử ở Hoa Kỳ và trưng cầu dân ý Brexit (Isaak & Hanna, 2018)²².

4. Một số định hướng nhằm quản lý AI trong quản trị nhà nước

Để AI thực sự trở thành động lực và công cụ thúc đẩy cải cách hành chính và quản trị nhà nước, các quốc gia cần xây dựng một chiến lược quản lý AI toàn diện. Chiến lược này phải vừa khai thác tối đa tiềm năng của công nghệ, vừa đặt ra giới hạn nhằm bảo vệ các giá trị căn bản của quản trị công: pháp quyền, minh bạch, trách nhiệm giải trình, công bằng và sự tham gia. Dưới đây là năm định hướng then chốt.

Thứ nhất, xây dựng khung pháp lý và nguyên tắc đạo đức. Điều kiện tiên quyết để kiểm soát AI là thiết lập một khung pháp lý toàn diện, rõ ràng và có tính khả thi. Liên minh châu Âu (EU) đã tiên phong trong lĩnh vực này với Đạo luật AI (AI Act), ban hành năm 2024, áp dụng mô hình quản trị dựa trên rủi ro. Theo đó, các hệ thống AI được phân loại theo mức độ rủi ro từ thấp, hạn chế, cao, đến bị cấm và mỗi nhóm phải tuân thủ các nghĩa vụ pháp lý khác nhau. Đây là mô hình đáng tham khảo cho nhiều quốc gia, đặc biệt ở những nước đang phát triển, nơi việc ứng dụng AI thường diễn ra nhanh hơn khả năng kiểm soát pháp lý.

Việt Nam ban hành *Luật Công nghệ thông tin* năm 2006, *Luật An ninh mạng* năm 2018, *Luật An toàn thông tin mạng* năm 2015... cũng đã bước đầu khái quát các nguyên tắc cơ bản liên quan tới AI... tuy nhiên, cần tiếp tục hoàn thiện khung pháp lý hoàn thiện để cụ thể hóa các quy định liên quan tới tiêu chuẩn kỹ thuật, đánh giá tác động của AI, bảo vệ quyền riêng tư, cân bằng quyền được giải thích, trách nhiệm giải trình của các bên liên quan.

Thứ hai, thiết lập cơ chế giám sát độc lập. Một khung pháp lý mạnh mẽ sẽ trở nên vô nghĩa nếu thiếu cơ chế giám sát độc lập. AI, với tính chất phức tạp và tốc độ thay đổi nhanh đòi hỏi sự giám sát liên tục và linh hoạt. Cần xác lập các cơ quan chức năng

chuyên trách về AI cấp quốc gia. Ngoài ra, cần sự phối hợp của các tổ chức độc lập và cơ chế phối hợp của nhiều chủ thể liên quan: Nhà nước, doanh nghiệp, các chuyên gia và các tổ chức xã hội... Ví dụ, Hiệp hội Phần mềm và Dịch vụ Công nghệ thông tin Việt Nam (VINASA) vừa thành lập Ủy ban Đạo đức trí tuệ nhân tạo²³ cho thấy, mô hình tư nhân - hợp tác nhà nước có thể đóng vai trò tư vấn chuyên môn. Việt Nam cũng có thể tham khảo mô hình “Ủy ban đánh giá tác động thuật toán” (AIA) của Canada, nơi mọi hệ thống AI công phải được thẩm định độc lập về quyền công dân và giá trị xã hội trước khi phê duyệt. Cơ quan giám sát này cần có tính độc lập, quyền yêu cầu thông tin và quyền đình chỉ dự án nếu phát hiện vi phạm, tương tự mô hình cơ quan bảo vệ dữ liệu cá nhân (Data Protection Authority) trong EU. Để bảo đảm cơ chế này hoạt động hiệu quả, cần quy định rõ vai trò và trách nhiệm (ví dụ trách nhiệm báo cáo định kỳ lên Quốc hội, Chính phủ hoặc cơ quan chủ quản) và quyền lực đủ mạnh (có thể xử phạt vi phạm về dữ liệu hoặc đạo đức công nghệ).

Thứ ba, tăng cường minh bạch và trách nhiệm giải trình. Chính phủ cần yêu cầu các cơ quan công khai thông tin về hệ thống AI họ sử dụng: dữ liệu đầu vào cơ bản, mục tiêu ứng dụng, tiêu chí ra quyết định... Như OECD khuyến nghị, “các bên tham gia AI phải cam kết minh bạch, cung cấp thông tin phù hợp và dễ hiểu về logic dẫn đến quyết định”. Điều này bao gồm việc công bố thuật toán (nơi có thể) hoặc các báo cáo giải trình cho phép người dân, chuyên gia giám sát được hoạt động AI. *Lưu ý* được ban hành năm 2024 và có hiệu lực từ ngày 01/7/2025 cần tiếp tục được cụ thể hóa để có khung pháp lý hoàn thiện quy định các vấn đề liên quan tới công khai, minh bạch các dữ liệu vì mục đích chung và bảo đảm quyền riêng tư. Thêm nữa, phải bảo đảm quyền “được giải thích” cho người dân khi bị ảnh hưởng bởi quyết định do AI hỗ trợ.

Thứ tư, xây dựng chuẩn mực đạo đức AI và đào tạo nhân lực số. Cần xây dựng quy chuẩn đạo đức AI dựa trên tham khảo các giá trị chung của quốc tế (như UNESCO khuyến nghị): tôn trọng nhân phẩm, minh bạch, công bằng, bảo đảm lợi ích xã hội, sự trung thực... Dựa trên các giá trị chuẩn mực chung, cần được cụ thể hóa và đưa vào giảng dạy về chuẩn mực đạo đức nghề nghiệp trong các ngành đào tạo công nghệ thông tin của các trường đại học.

Tính đến năm 2024, cả nước có khoảng 700 kỹ sư chuyên về AI nhưng chỉ có khoảng 300 người được công nhận là chuyên gia AI thực thụ. Theo báo cáo Thị trường công nghệ thông tin Việt Nam, trong năm 2025 cần khoảng 700.000 nhân lực công nghệ thông tin nhưng hệ thống đào tạo hiện mới đáp ứng được khoảng 500.000 người, tương đương với việc thiếu hụt tới 200.000 nhân lực²⁴. Do đó, Việt Nam cần có chiến lược tập trung vào đào tạo chuyên sâu, chuẩn bị thể hệ chuyên gia mới để giải quyết vấn đề thực tiễn. Ngoài ra, cần tăng cường nâng cao nhận thức số cho công dân, giúp họ hiểu được quyền lợi và nguy cơ khi tương tác với hệ thống AI công thông qua chiến dịch truyền thông, tài liệu hướng dẫn, bình dân học vụ số....

Thứ năm, thúc đẩy hợp tác quốc tế. Việt Nam cần tham gia tích cực vào việc xây dựng các chuẩn mực toàn cầu về AI, thông qua các tổ chức, như: OECD, UNESCO, Liên Hợp quốc, ASEAN... Việc chia sẻ kinh nghiệm, công nghệ và dữ liệu có thể giúp thu hẹp khoảng cách phát triển giữa các nước; đồng thời, hạn chế tình trạng “chủ nghĩa dân tộc công nghệ” - nơi một số cường quốc tận dụng AI để củng cố vị thế địa - chính trị của mình.

5. Kết luận

AI đang mở ra nhiều triển vọng cho quản trị nhà nước, từ tối ưu hóa thủ tục hành chính, khai thác dữ liệu lớn đến tăng cường minh bạch và sự tham gia của công dân. AI có thể giúp Chính phủ ra quyết định chính sách chính xác hơn và tăng cường phục vụ

công dân. Tuy nhiên, AI cũng ẩn chứa nguy cơ dễ bị lạm dụng - từ xâm phạm quyền riêng tư, thiên kiến xã hội cho đến nguy cơ tập trung quyền lực và mất kiểm soát. Các nghiên cứu cho thấy, nếu thiếu cơ chế điều phối chặt chẽ, AI có thể làm suy giảm niềm tin vào chính quyền và làm trầm trọng thêm bất công xã hội.

Đối với Việt Nam, kinh nghiệm quốc tế và chính sách trong nước đang dần hình thành một khuôn khổ quản trị AI ban đầu. Các khuyến nghị chính sách từ việc hoàn thiện khung pháp lý, thiết lập cơ chế giám sát độc lập và chuẩn mực đạo đức; đồng thời, tập trung phát triển nguồn lực con người và xây dựng niềm tin xã hội. Các đề xuất này nhằm bảo đảm việc triển khai AI trong chính quyền phải “đặt con người làm trung tâm”, tôn trọng các giá trị dân chủ và pháp quyền và vì hạnh phúc Nhân dân □

Chú thích:

1. Mišić, J., van Est, R., & Kool, L. (2025). *Good governance of public sector AI: a combined value framework for good order and a good society. AI and Ethics*.
- 2, 17. Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs*.
3. e-Estonia. (n.d.). X-Road: *The Backbone of Estonia's Interoperable Digital State. e-Estonia*.
4. Government of UAE. (2021). *Virtual assistant Rashid*. <https://u.ae/en/about-the-uae/digital-uae/virtual-assistant-rashid>.
5. Kraemer-Mbula, E., Wunsch-Vincent, S., & Zhou, Y. (Eds.). (2021). *The role of artificial intelligence in innovation: Insights from Singapore's Smart Nation initiative. World Intellectual Property Organization*.
6. Whitelaw, S., Mamas, M. A., Topol, E., & Van Spall, H. G. C. (2020). *Applications of digital technology in COVID-19 pandemic planning and response. The Lancet Digital Health, 2(8), e435-e440*. [https://doi.org/10.1016/S2589-7500\(20\)30142-4](https://doi.org/10.1016/S2589-7500(20)30142-4).
7. Kehl, D., & Kessler, S. (2017). *Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing. Berkman Klein Center for Internet & Society*.
- 8, 11. Lee, J. (2020). *Government use of artificial intelligence in Korea: Current practices and policy implications. Journal of Open Innovation: Technology, Market, and Complexity, 6(4), 166*. <https://doi.org/10.3390/joitmc6040166>.
9. European Commission. (2021). *The role of artificial intelligence in the European Green Deal. Publications Office of the European Union*. <https://doi.org/10.2777/09127>.
10. OECD. (2020). *The use of AI to detect fraud in public spending. OECD Publishing*.
- 12, 18, 20, 23. Susa, I., Janssen, M., & Verhulst, S. (2022). *Data collaboratives as a new frontier of cross-sector partnerships in the era of AI: A critical review. Government Information Quarterly, 39(1), 101-124*. <https://doi.org/10.1016/j.giq.2021.101624>.
13. O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing Group*.
14. Human Rights Watch. (2021). *Moscow's facial recognition system threatens rights*. <https://www.hrw.org/news/2021/05/21/moscow-facial-recognition-system-threatens-rights>.
15. Creemers, R. (2018). *China's social credit system: An evolving practice of control. SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3175792>.
16. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias: There's software used across the country to predict future criminals. And it's biased against Blacks. ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
19. Fussey, P., & Murray, D. (2019). *Independent report on the London Metropolitan Police Service's trial of live facial recognition technology. University of Essex Human Rights Centre*.
21. Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y.,... & Zwitter, A. (2019). *Will democracy survive big data and artificial intelligence? Scientific American, 25(2), 1-5*.
22. Isaak, J., & Hanna, M. J. (2018). *User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer, 51(8), 56-59*. <https://doi.org/10.1109/MC.2018.3191268>.
24. *Việt Nam thiếu 200.000 nhân sự CNTT, 70% sinh viên ra trường chưa đủ năng lực* <https://nguoiduatin.vn/viet-nam-thieu-200000-nhan-su-cntt-70-sinh-vien-ra-truong-chua-du-nang-luc-204250210135002572.htm>.