TRÍ TUỆ NHÂN TẠO VÀ QUYỀN LỰC CHÍNH TRỊ: MỘT GÓC NHÌN TRIẾT HỌC CHÍNH TRỊ ỨNG DỤNG

TS. PHAN DUY ANH *

Trong bối cảnh Cách mạng công nghiệp lần thứ tư, trí tuệ nhân tạo (AI) đang tái định hình quyền lực chính trị nói chung và quyền lực nhà nước nói riêng, xuất hiện những yêu cầu mới về nhận thức phương pháp luận cho triết học chính trị hiện đại. Bài nghiên cứu này khám phá mối quan hệ giữa AI và quyền lực chính trị qua lăng kính triết học chính trị C.Mác, Michel Foucault và Jürgen Habermas; phân tích cách AI củng cố giám sát, quyết định chính sách và phân tán quyền lực, đồng thời đặt ra những thách thức cho nền dân chủ và quyền bình đẳng của công dân.

AI là một khái niệm lý thuyết chỉ sự phát triển của những hệ thống máy tính có khả năng thực hiện các công việc đòi hỏi trí thông minh của con người. Nó sử dụng các hệ thống học máy để học hỏi và mô phỏng các hoạt động như suy nghĩ, lập luận và tự thích nghi. Mục tiêu cuối cùng là tạo ra các hệ thống thông minh có khả năng tương tác và đáp ứng linh hoạt như con người, từ đó mang lại giá trị to lớn trong nhiều lĩnh vực[1].

AI có thể được chia thành ba cấp độ chính dựa trên khả năng và phạm vi hoạt động. *Thứ nhất,* AI hẹp là loại phổ biến nhất hiện nay, tập trung vào việc thực hiện một nhiệm vụ cụ thể, chẳng hạn như nhận diện giọng nói trong trợ lý ảo Siri của Apple. *Thứ hai,* AI tổng quát có khả năng thực hiện bất kỳ nhiệm vụ trí tuệ nào mà con người có thể làm, với mức độ linh hoạt và thích nghi cao, nhưng hiện nay vẫn đang ở giai đoạn nghiên cứu. *Thứ ba,* siêu trí tuệ nhân tạo vượt trội hơn hẳn trí tuệ con người ở mọi lĩnh vực, có khả năng tự cải tiến và sáng tạo, nhưng vẫn mang tính giả định và đặt ra nhiều lo ngại đạo đức. Trong công trình *Phát minh cuối cùng - Trí tuệ nhân tạo và sự cáo chung của kỷ nguyên con người*, James Barrat đã miêu tả sinh động về loại AI này: "Trên một siêu máy tính chạy với tốc độ 36,8 petaflop, tương đương gấp hai lần tốc độ của não người, một AI đang cải tiến trí thông minh của nó… Lần đầu tiên, loài người chứng kiến sự hiện diện của một thứ thông

* *Trường Đại học Bách khoa, Đại học Quốc gia Thành phố Hồ Chí Minh*

minh hơn mình. *Siêu* trí tuệ nhân tạo (artificial superintelligence: ASI)"[2].

Sự phát triển của AI không chỉ là một vấn đề kỹ thuật thuần túy, mà còn đặt ra những thách thức sâu sắc đối với các nền tảng triết học chính trị. Nếu AI được coi là một công cụ đơn thuần, trọng tâm phân tích của triết học chính trị sẽ xoay quanh việc ai kiểm soát và sử dụng công cụ này. Phương pháp luận này tương tự như việc nghiên cứu vai trò của bất kỳ công nghệ nào trong lịch sử chính trị. Mục tiêu là làm rõ các mối quan hệ quyền lực đằng sau công nghệ: Ai thiết kế, ai sở hữu, và ai được lợi từ việc sử dụng nó. Quan điểm này yêu cầu xem xét các cơ chế quản lý, luật pháp và chính sách để bảo đảm công cụ được sử dụng một cách công bằng. Và nếu AI được xem là một tác nhân - một kịch bản hiện nay mang tính giả định nhưng đang dần được hiện thực hóa qua các mô hình như robot hình người Sophia[3] - triết học chính trị sẽ phải đối mặt với những câu hỏi phức tạp hơn về quyền công dân và quyền con người. Liệu một AI có ý thức có thể có các quyền chính trị như quyền bầu cử hay quyền được bảo vệ hay không? Quan điểm này đặt ra những thách thức sâu sắc về ranh giới giữa con người và máy móc.

Tháng 4-2025, Chính phủ Các Tiểu vương quốc Arập thống nhất (UAE) đã công bố "kế hoạch lập pháp do AI hỗ trợ" với ý tưởng chính là chủ động sử dụng AI để đưa ra đề xuất thay đổi các đạo luật hiện hành, như Phó Tổng thống kiêm Thủ tướng UAE Sheikh Mohammad bin Rashid Al Maktoum đã khẳng định: "Hệ thống lập pháp mới, với sự hỗ trợ của AI, sẽ thay đổi cách chúng tôi xây dựng pháp luật, giúp quy trình này trở nên nhanh hơn và chính xác hơn"[4]. Còn ở Thụy Điển, Thủ tướng Ulf Kristersson thừa nhận đã sử dụng công cụ AI để tham khảo trong quá trình đưa ra một số quyết sách: "Tôi sử dụng những công cụ này khá thường xuyên, nhưng chỉ để tham khảo góc nhìn thứ hai: Nước khác đã làm gì? Chúng ta có nên nghĩ ngược lại hoàn toàn không? Những câu hỏi kiểu như vậy"[5]. Thực tiễn này cho thấy, AI đã hiện hữu trong đời sống chính trị của con người. AI không chỉ ảnh hưởng trực tiếp đến việc bỏ phiếu và kiểm phiếu mà còn xuất hiện trong quy trình lập pháp; đồng thời, các chính trị gia xem AI như một "quân sư" cho việc ra các quyết định quyền lực của mình.

1. Chủ nghĩa tư bản công nghệ dùng AI như một công cụ quyền lực - theo góc nhìn của chủ nghĩa Mác

Để có thể hiểu mối quan hệ giữa AI và quyền lực chính trị qua lăng kính chủ nghĩa Mác, trước hết cần đặt AI vào bối cảnh kinh tế - xã hội đương đại mà giới học thuật gọi là chủ nghĩa tư bản công nghệ. Năm 2009, nhà khoa học Luis Suarez Villa đã đưa ra khái niệm "chủ nghĩa tư bản công nghệ" như một phiên bản mới của chủ nghĩa tư bản, trong đó tri thức, sáng tạo và đổi mới công nghệ trở thành nền tảng cho sự tích lũy tư bản, thay thế vai trò trung tâm của lao động công nghiệp và tư bản tài chính. Đặc trưng nổi bật của thời đại này là sự thương mại hóa khoa học và công nghệ như một dạng "tài sản trí tuệ"[6]. Trong khuôn khổ đó, AI không chỉ là một công cụ kỹ thuật mà còn là cốt lõi của hạ tầng kinh tế - xã hội mới, trở thành một lực lượng sản xuất trực tiếp chưa từng có tiền lệ.

Quay trở lại với các phân tích của C.Mác trong *Góp phần phê phán khoa kinh tế chính trị* (năm

1858), ông đã đề cập đến khái niệm "tri thức xã hội phổ biến" là sự tích lũy tri thức khoa học, kỹ thuật và công nghệ chung của xã hội, được vật hóa trong máy móc và hệ thống sản xuất. Ông cho rằng: "Sự phát triển của tư bản cố định là chỉ số cho thấy tri thức xã hội phổ biến [Wissen, knowledge] đã chuyển hóa đến mức độ nào đó thành *lực lượng sản xuất trực tiếp*, do đó nó cũng là chỉ số cho thấy những điều kiện của quá trình sống của xã hội đã phục tùng đến một mức độ nào sự kiểm soát của trí tuệ phổ biến và đã được cải tạo đến mức độ nào cho phù hợp với quá trình ấy; những lực lượng sản xuất xã hội đã được tạo ra đến mức độ nào không những dưới hình thức tri thức mà cả như là những cơ quan thực hành xã hội trực tiếp, những cơ quan trực tiếp của quá trình sống hiện thực"[7].

Với C.Mác, khoa học công nghệ trở thành lực lượng sản xuất trực tiếp khi tri thức khoa học được vật hóa thành máy móc, thành công cụ sản xuất của con người và được người lao động sử dụng trong quá trình sản xuất. Ông khẳng định: "… sự phát triển của hệ thống máy móc trên con đường ấy chỉ bắt đầu khi nền đại công nghiệp đã đạt được một trình độ phát triển cao hơn và tất cả các môn khoa học đều được đưa vào phục vụ tư bản, còn bản thân hệ thống máy móc hiện có thì có những nguồn lực to lớn. Như vậy, phát minh trở thành một nghề đặc biệt, và đối với nghề đó thì việc vận dụng khoa học vào nền sản xuất trực tiếp tự nó trở thành một trong những yếu tố có tính chất quyết định và kích thích"[8].

Ngày nay, AI chính là hiện thân sinh động và cao nhất của tri thức xã hội phổ biến. AI - một hệ thống máy tính có khả năng mô phỏng các chức năng nhận thức, lập luận và học hỏi của con người - là sự vật hóa của hàng tỷ giờ lao động trí tuệ, dữ liệu và thuật toán, được tích hợp vào các hệ thống sản xuất và quản trị. Việc AI có thể tự động hóa các quá trình ra quyết định, tối ưu hóa chuỗi cung ứng, và thậm chí tự thiết kế các thuật toán mới đã khẳng định nó là lực lượng sản xuất trực tiếp, làm thay đổi triệt để phương thức tạo ra giá trị.

Tuy nhiên, vấn đề cốt lõi theo C.Mác không nằm ở bản thân công nghệ mà ở quan hệ sản xuất chi phối nó. Trong chủ nghĩa tư bản công nghệ, AI và hạ tầng dữ liệu khổng lồ của nó bị tư nhân hóa và độc quyền hóa bởi một số ít tập đoàn công nghệ lớn. Các tập đoàn này sở hữu những công cụ tinh vi nhất của tri thức xã hội phổ biến và sử dụng chúng để củng cố quyền lực kinh tế và chính trị của mình, dẫn đến một hình thái bóc lột mới: Bóc lột giá trị thặng dư dữ liệu hành vi của con người.

Chính những lập luận trên của C.Mác trở thành nguồn cảm hứng cho nhà triết học Shoshana Zuboff phân tích sự chuyển đổi này như một sự phát triển tất yếu của logic tư bản. Bà cho rằng, chủ nghĩa tư bản hiện đại đã phát minh ra "dữ liệu hành vi dư thừa" - dữ liệu được thu thập từ mọi khía cạnh trong cuộc sống của con người vượt ra ngoài nhu cầu trực tiếp của sản phẩm hoặc dịch vụ và chuyển hóa chúng thành giá trị thặng dư nhờ công cụ AI. Cụ thể, thông qua các thuật toán học sâu và phân tích dữ liệu lớn, AI có khả năng xử lý lượng dữ liệu khổng lồ này để tạo ra các mô hình dự đoán hành vi tương lai. Các mô hình dự đoán này là sản phẩm mới được bán cho các nhà quảng cáo, bảo hiểm, hoặc các tổ chức chính trị. Lợi nhuận thu được từ việc bán các sản phẩm dự đoán

chính là giá trị thặng dư dữ liệu. Đây là nguồn lợi nhuận dựa trên lao động dữ liệu vô hình của người dùng, tương đương với lao động không được trả công mà C.Mác đã mô tả.

Do đó, AI không chỉ là công cụ giúp tư bản tiết kiệm chi phí lao động mà còn là công cụ phát minh ra một nguồn giá trị thặng dư hoàn toàn mới. Các tập đoàn Big Tech sử dụng quyền lực kinh tế có được từ AI để gây ảnh hưởng chính trị, can thiệp vào quy trình lập pháp, và định hình các chính sách quản trị, nhằm bảo vệ, củng cố đặc quyền độc quyền dữ liệu của họ. Quyền lực chính trị do đó trở thành một siêu kiến trúc phục vụ cho cơ sở hạ tầng kinh tế dữ liệu mới. Shoshana Zuboff gọi loại quyền lực này là "chủ nghĩa công cụ" - "là bộ công cụ và sự công cụ hóa hành vi để phục vụ cho mục đích điều chỉnh, dự đoán, kiếm tiền và kiểm soát"[9]. Bà cho rằng: "Trong công thức này, "bộ công cụ" nói đến con rối: Kiến trúc vật chất được kết nối có mặt khắp mọi nơi của điện toán có cảm giác chuyên chuyển đổi, diễn giải và kích thích trải nghiệm của con người. "Sự công cụ hóa" biểu thị các mối quan hệ xã hội hướng những người điều khiển rối đến trải nghiệm của con người khi tư bản giám sát sử dụng những cỗ máy để biến chúng thành các phương tiện phục vụ cho mục đích thị trường của người khác. Chủ nghĩa tư bản giám sát buộc chúng ta chuẩn bị đón nhận một hình thức chủ nghĩa tư bản chưa từng có trước đây. Giờ đây quyền lực công cụ, cái duy trì và mở rộng dự án tư bản giám sát, sẽ buộc chúng ta một lần nữa đối mặt với thứ chưa từng có"[10].

C.Mác cũng đã cảnh báo sự tha hóa của người lao động trên bốn khía cạnh: Tha hóa khỏi sản phẩm lao động, khỏi hoạt động lao động, khỏi bản chất loài người, và khỏi đồng loại. Trong chủ nghĩa tư bản công nghệ, AI và chủ nghĩa tư bản giám sát đã làm trầm trọng thêm sự tha hóa này, mở rộng nó từ phạm vi sản xuất vật chất sang toàn bộ đời sống xã hội. AI biến các quyết định hằng ngày của cá nhân (như mua gì, đọc gì, gặp ai …) thành dữ liệu thô để các thuật toán xử lý, qua đó tự động đưa ra các gợi ý và thúc đẩy hành vi. Con người dần bị tước bỏ khả năng tự chủ trong việc ra quyết định, bị thao túng bởi các thuật toán, bị kiểm soát và định hình bởi các hệ thống máy móc.

Mặc dù AI là sự vật hóa của tri thức xã hội phổ biến (tức là tri thức tập thể của nhân loại), nhưng tri thức này lại bị độc quyền sở hữu. Người lao động sử dụng các công cụ AI mà không hiểu hay kiểm soát được thuật toán (một "hộp đen" quyền lực), sẽ bị tách biệt khỏi chính nguồn tri thức mà về bản chất thuộc về họ. Sự tha hóa khỏi tri thức, hay chính là sự tha hóa khỏi bản chất sáng tạo của loài người, sẽ trở nên sâu sắc hơn bao giờ hết.

Như vậy, từ góc nhìn của chủ nghĩa Mác, mối quan hệ biện chứng giữa AI và quyền lực chính trị được thể hiện rõ nét: AI thúc đẩy lực lượng sản xuất, nâng cao năng suất xã hội, nhưng dưới quan hệ sản xuất tư bản chủ nghĩa, nó củng cố bất bình đẳng, bóc lột lao động và tập trung quyền lực vào tay tư bản. Nguy cơ AI sẽ trở thành công cụ duy trì sự thống trị của giai cấp tư sản dường như đang dần trở thành hiện thực.

2. Tri thức AI và quyền lực của nhà nước "Panopticon kỹ thuật số" theo góc nhìn triết học chính trị Foucault

Trong tác phẩm Giám sát và trừng phạt: Nguồn gốc nhà tù (*Surveiller et Punir: Naissance de la Prison*) năm 1975, nhà triết học người Pháp

Michel Foucault đã đưa ra quan điểm tri thức mãi mãi được kết nối với quyền lực, và thường viết theo cách: Tri thức/quyền lực. Ông khẳng định: "Tri thức gắn liền với quyền lực không chỉ mang trong mình thẩm quyền của "chân lý" mà còn có khả năng tự biến mình thành chân lý". Mọi tri thức, một khi được áp dụng vào thế giới thực, đều có tác động, và ít nhất theo nghĩa đó, "trở thành chân lý". Tri thức, một khi được sử dụng để điều chỉnh hành vi của người khác, đòi hỏi sự ràng buộc, điều chỉnh và kỷ luật hóa hành động. Do đó, "không có mối quan hệ quyền lực nào nếu không có sự cấu thành tương ứng của một lĩnh vực tri thức, cũng như không có tri thức nào không vừa giả định vừa cấu thành nên các mối quan hệ quyền lực"[11]. Và trong tư tưởng của Foucault, mỗi thời đại đều có một/một vài cấu trúc tư duy, thế giới quan, là một phần của các "hệ thống quyền lực/tri thức" khác nhau. Foucault đặt ra thuật ngữ "*episteme*" - từ tiếng Hy Lạp có nghĩa là kiến thức hay sự hiểu biết - để chỉ loại nền tảng này trong những điều kiện khả dĩ vốn luôn phản ánh các mối quan hệ quyền lực của một thời đại. Thuật ngữ này đặc trưng cho những cấu trúc có trật tự và thường không bị đặt thành vấn đề ("vô thức", như Foucault thường nói) làm nền tảng cho việc sản sinh ra tri thức khoa học tại một thời điểm và địa điểm nhất định, "lĩnh vực nhận thức luận" của nó[12].

Trong bối cảnh đương đại, AI chính là hình thái tri thức/quyền lực mang tính cách mạng nhất. AI, với khả năng xử lý, phân tích và dự đoán dựa trên dữ liệu lớn, đã trở thành một hệ thống tri thức có khả năng thực nghiệm hóa các quan niệm về con người, xã hội và hành vi. Những kết quả phân tích của thuật toán AI được coi là những "chân lý" kỹ thuật có khả năng "áp đặt" lên cá nhân và các nhóm xã hội. Đây là hình thái quyền lực không cần phải sử dụng sự áp đặt vật lý hay luật pháp rõ ràng, mà thông qua khả năng định hình nhận thức và kiểm soát thông tin của AI. Sự chuyển đổi này khiến quyền lực trở nên vô hình, tự động và hiệu quả hơn. Foucault đã chỉ ra rằng mục tiêu của quyền lực hiện đại là tạo ra những "cá thể ngoan ngoãn" - những cá nhân tự nguyện tuân thủ các quy tắc xã hội thông qua các cơ chế giám sát và kỷ luật. AI chính là công cụ hoàn hảo để thực hiện cơ chế kỷ luật này trên quy mô lớn.

Foucault nổi tiếng với việc sử dụng kỹ thuật/phương pháp điều tiết tri thức/quyền lực thông qua *mô hình giám sát toàn cảnh "Panopticon"* - mô hình nhà tù mà Jeremy Bentham đã đề xuất từ năm 1843. Ông mô tả "*Panopticon*" là một kiến trúc cho phép một người gác có thể quan sát tất cả tù nhân mà tù nhân không thể biết mình đang bị quan sát hay không. Cơ chế này hoạt động dựa trên nguyên lý: Giám sát vô hình tạo ra sự tuân thủ vô điều kiện. Trong tư tưởng của Foucault, *Panopticon* đưa ra một sự khuyến khích nội tâm mạnh mẽ và tinh vi, đạt được thông qua việc quan sát liên tục các tù nhân. Ngày nay, AI và hạ tầng dữ liệu đã biến mô hình Panopticon truyền thống thành "Panopticon kỹ thuật số" trên phạm vi toàn xã hội. "Panopticon kỹ thuật số" không còn là một công trình kiến trúc vật lý mà là một hệ thống mạng lưới thuật toán và cảm biến thu thập thông tin cá nhân trên quy mô lớn, liên tục và tự động. Sự chuyển đổi này được Foucault phân tích qua ba cơ chế rõ ràng, ngày nay đã được AI tăng cường.

Cơ chế *thứ nhất* là dữ liệu cá nhân bị thu thập liên tục và toàn diện. Trong "Panopticon kỹ

thuật số", việc thu thập dữ liệu thông qua các thiết bị kết nối internet diễn ra một cách liên tục và toàn diện. Mỗi hoạt động của cá nhân - từ tìm kiếm thông tin, giao dịch tài chính cho đến cảm xúc biểu hiện qua tương tác mạng xã hội - đều được hệ thống AI ghi lại, phân tích và lưu trữ. AI không chỉ ghi nhận những gì cá nhân đã làm, mà còn sử dụng dữ liệu đó để dự đoán những gì cá nhân sẽ làm. Dữ liệu này trở thành một "hồ sơ kỹ thuật số" về bản chất con người, một kho tri thức khổng lồ về các hành vi tiềm ẩn, nằm trong tay các cơ quan quyền lực nhà nước hoặc các tập đoàn công nghệ có liên kết chính trị.

Cơ chế *thứ hai* là cảm giác bị giám sát mọi lúc và mọi nơi. Cơ chế quyền lực của Panopticon không phụ thuộc vào việc tù nhân có bị quan sát hay không, mà phụ thuộc vào việc họ luôn tin rằng mình có thể bị quan sát bất cứ lúc nào. AI khuếch đại cảm giác này lên mức độ cực đại. Sự vô hình của thuật toán, sự không rõ ràng của cơ chế thu thập dữ liệu (hộp đen thuật toán), và các vụ rò rỉ thông tin liên tục tạo ra một nỗi sợ hãi mơ hồ, thường trực về việc bị theo dõi. Công dân không thể biết chính xác dữ liệu nào đang được sử dụng, bởi ai, và nhằm mục đích gì. Cảm giác bị giám sát liên tục này tạo ra một áp lực định hình hành vi mạnh mẽ hơn bất kỳ luật lệ hữu hình nào. Nó không chỉ kiểm soát các hoạt động công cộng mà còn xâm nhập vào các không gian cá nhân, tâm lý và ý thức.

Cơ chế *thứ ba* là tự giám sát và tự kỷ luật dẫn đến sự tuân thủ. Mục tiêu tối thượng của "Panopticon kỹ thuật số" là tạo ra sự tự kỷ luật. Khi con người biết rằng mọi hành vi của mình đều có thể được ghi lại và phân tích bởi AI, họ sẽ tự động điều chỉnh hành vi để phù hợp với những tiêu chuẩn được hệ thống chấp nhận.

Ví dụ điển hình là hệ thống nhận diện khuôn mặt, được triển khai rộng rãi ở Trung Quốc để giám sát công dân thông qua mạng lưới camera công cộng, kết hợp với dữ liệu cá nhân để phát hiện hành vi bất thường. Một ứng dụng khác là dự đoán tội phạm, như hệ thống PredPol ở Mỹ, sử dụng AI để phân tích dữ liệu lịch sử tội phạm và dự báo khu vực có nguy cơ cao, hoặc Strategic Subjects List ở Chicago, liệt kê cá nhân có khả năng phạm tội dựa trên thuật toán. Những công nghệ này không chỉ nâng cao hiệu quả thực thi pháp luật mà còn tạo ra một mạng lưới kiểm soát vô hình, nơi công dân luôn bị theo dõi, dẫn đến hành vi tự kiểm duyệt. Như Zhidas Daskalovski đã mô tả: "Chế độ giám sát đã được tăng cường trên khắp thế giới … Chúng ta được chứng kiến chính phủ tăng cường giám sát, thu thập và lưu trữ dữ liệu của công dân và trì hoãn việc thông báo về những cuộc tìm kiếm này… Họ có sử dụng kho thông tin hay không và sử dụng như thế nào vẫn là một câu hỏi lớn"[13].

AI đã thực sự biến xã hội thành một "*Panopticon kỹ thuật số*", nơi dữ liệu cá nhân được thu thập liên tục qua camera, thiết bị di động và mạng xã hội, thúc đẩy kỷ luật xã hội. Trong khi quyền lực nhà nước truyền thống dựa trên khả năng cưỡng chế thì thông qua AI, quyền lực nhà nước đang chuyển sang mô hình dự báo và phòng ngừa. Hệ thống AI có thể dự đoán hành vi bất ổn hoặc phạm tội trước khi nó xảy ra, cho phép nhà nước can thiệp sớm. Mặc dù có vẻ hiệu quả, mô hình này gây ra thách thức nghiêm trọng đối với các nguyên tắc pháp lý cốt lõi như sự vô tội cho đến khi chứng minh có tội và tự do ý chí, bởi vì cá nhân có thể bị xử lý dựa trên một "khả năng"

phạm tội do thuật toán tính toán.

AI tạo ra sự nghịch lý trong cấu trúc quyền lực. *Một mặt*, quyền lực trở nên phân tán hơn khi các thuật toán và dữ liệu len lỏi vào mọi góc cạnh xã hội (từ giao thông, y tế, đến giáo dục). *Mặt khác*, nó lại tập trung hơn bao giờ hết vào tay những người kiểm soát và lập trình thuật toán. Quyền lực của nhà nước được khuếch đại bởi các thuật toán, khiến việc phản biện và giám sát quyền lực trở nên khó khăn hơn. Người dân không thể chất vấn một thuật toán, hay yêu cầu một "hộp đen" thuật toán giải trình các quyết định của nó.

Từ góc nhìn Foucault, mối quan hệ biện chứng giữa AI và quyền lực chính trị thể hiện ở chỗ: AI củng cố quyền lực giám sát, làm cho nó trở nên lan tỏa và vô hình, nhưng đồng thời cũng tạo ra khả năng kháng cự khi cá nhân nhận thức được về sự thao túng. Nếu không có sự can thiệp dân chủ, AI sẽ duy trì sự thống trị của nhà nước giám sát, dẫn đến tha hóa tự do. AI không chỉ là một công cụ giám sát hiệu quả hơn mà còn là một cơ chế mới của quyền lực. Nó mở rộng khái niệm *Panopticon* của Foucault ra phạm vi toàn cầu, định hình hành vi công dân một cách vô hình. Điều này đặt ra một thách thức lớn cho triết học chính trị, yêu cầu phải xem xét lại các khái niệm về tự do, quyền riêng tư và quyền công dân trong một kỷ nguyên mà sự giám sát đã trở thành một phần không thể thiếu của cuộc sống.

3. Hiệu suất AI và quyền lực ban hành quyết định chính trị theo góc nhìn lý thuyết Hành động giao tiếp của Jürgen Habermas

Khi các chính phủ và chính trị gia sử dụng AI để tham khảo, hình thành hoặc thậm chí đề xuất các quyết định chính trị, nó lập tức tạo ra những tranh cãi xoay quanh hai câu hỏi cốt lõi: Tính hợp pháp của quyết định đó đến từ đâu? Và liệu nó có thực sự thuyết phục được công chúng hay không?

Theo nhà triết học xã hội Jürgen Habermas với lý thuyết Hành động giao tiếp, tính hợp pháp của một quyết định chính trị không thể đến từ hiệu quả kỹ thuật, sự cưỡng chế, hay truyền thống, mà phải đến từ một quá trình giao tiếp hợp lý và không bị chi phối. Habermas nhấn mạnh, quyền lực chính trị chỉ được coi là chính đáng khi nó được hình thành qua giao tiếp công khai và đồng thuận xã hội. Ông cho rằng, hành động giao tiếp là hành động mà các chủ thể tham gia nhằm mục đích đạt được sự hiểu biết lẫn nhau, không nhằm mục đích thao túng hay kiểm soát đối phương. Quyết định chính trị hợp pháp phải dựa trên sức mạnh của lập luận tốt nhất được đưa ra trong quá trình giao tiếp này, chứ không phải dựa trên quyền lực áp đặt.

Với Habermas, quyền lực chính trị hợp pháp được xây dựng trên lĩnh vực công cộng - một không gian lý tưởng nơi các công dân có thể đối thoại và hình thành ý kiến chung. Quyết định được coi là đúng đắn khi nó là kết quả của một sự đồng thuận hợp lý, dựa trên sức mạnh của lập luận tốt nhất. Trong tác phẩm Sự chuyển đổi cấu trúc của khu vực công cộng: Một nghiên cứu về phạm trù xã hội dân sự (*The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*) năm 1962, ông khẳng định, khu vực công cộng là những định chế xã hội trong đó tạo cơ hội cho những cuộc thảo luận mang tính chất lý tính và cởi mở giữa các công dân để hình thành nên ý kiến của công chúng và cuộc thảo luận có thể được thực hiện một cách trực tiếp hoặc thông qua sự trao đổi thư

từ hay thông qua trung gian bởi các tờ báo, tạp chí và các hình thức giao tiếp điện tử khác. Ý tưởng về một không gian công được mở ra cho tất cả và sự đồng thuận được bảo đảm thông qua tính thuyết phục của những lập luận tốt hơn là một sự diễn trò hay là việc sử dụng quyền lực áp đặt. Ông định nghĩa: "Khu vực công cộng dân sự trên hết có thể được quan niệm như là không gian của các cá nhân tụ họp cùng nhau như một cộng đồng; họ sớm tuyên bố khu vực công cộng được điều hành bởi cái gì đó ngược lại với bản thân các quyền lực công, để đi vào trong một cuộc tranh luận về các quy tắc chung điều khiển các mối quan hệ trong không gian về cơ bản được cá nhân hóa nhưng liên quan đến cộng đồng về sự trao đổi hàng hóa và lao động xã hội. Môi trường của sự gặp gỡ chính trị này rất kì lạ và chưa từng có tiền lệ: Người ta sử dụng lý tính của mình một cách công khai"[14].

Trong khu vực công cộng, các chuyên gia hay quy trình kỹ trị có thể hỗ trợ, nhưng không thể thay thế việc ra quyết định của con người. Vấn đề phát sinh khi AI được sử dụng để tối ưu hóa hiệu suất, giảm thiểu sự không chắc chắn và sai sót của con người. Điều này có thể dẫn đến việc các quyết định được ban hành không phải vì chúng là kết quả của một quá trình thảo luận dân chủ, mà vì chúng là kết quả của một thuật toán được chứng minh là hiệu quả nhất.

Sự xuất hiện và ứng dụng rộng rãi của AI, đặc biệt là trong quy trình ra quyết định của chính phủ, đặt ra một thách thức nghiêm trọng đối với nền tảng giao tiếp của Habermas, đó là nguy cơ kỹ trị. Hiệu suất cao của AI, vốn ưu tiên sự tối ưu hóa, độ chính xác, và tốc độ xử lý dữ liệu, có xu hướng làm lu mờ tầm quan trọng của quá trình giao tiếp và đồng thuận xã hội.

Vấn đề phát sinh khi các quyết định chính trị được ban hành không phải vì chúng là kết quả của một quá trình thảo luận dân chủ, mà vì chúng là kết quả của một thuật toán được chứng minh là hiệu quả nhất. Khi một chính phủ áp dụng AI để dự đoán chính sách kinh tế hay hành vi tội phạm, họ hành động dựa trên "độ chính xác" của thuật toán mà không cần phải trải qua một quá trình thảo luận rộng rãi. Điều này làm suy yếu vai trò của các cuộc tranh luận công khai và tính hợp pháp dựa trên sự đồng thuận.

Thêm vào đó, sự phụ thuộc vào hiệu suất AI có nguy cơ biến việc ra quyết định chính trị thành một vấn đề kỹ thuật thay vì một quá trình đạo đức và quyền lực chính trị. Quyết định không còn đến từ lý lẽ của con người mà từ "lý trí" được lập trình sẵn. Điều này dẫn đến sự thao túng, nơi các quyết định được "hợp pháp hóa" bởi tính toán máy móc, thay vì sự tham gia của công dân.

Đặc biệt, nếu AI bị kiểm soát bởi các nhóm quyền lực, nó có thể trở thành công cụ thao túng dư luận, làm méo mó không gian công. Ví dụ điển hình là thuật toán mạng xã hội ưu tiên nội dung gây tranh cãi, làm suy yếu khả năng đối thoại lý tính. Khi thông tin được lọc, ưu tiên, hoặc kiểm duyệt bởi thuật toán, khu vực công cộng (dù là trực tuyến hay ngoại tuyến) không còn là nơi diễn ra các cuộc đối thoại bình đẳng và lý trí, mà trở thành không gian bị định hướng bởi lợi ích kỹ trị hoặc tư bản. Habermas đã cảnh báo, khi lý trí kỹ trị chiếm ưu thế, nó sẽ làm xói mòn tính tự chủ lý tính của công dân, những người thay vì tham gia vào việc hình thành ý chí chung, lại trở thành đối tượng được "tối ưu hóa" bởi các hệ thống thuật toán.

Tuy nhiên, lập luận của Habermas cũng mở

ra một hướng tiếp cận khác. AI không tất yếu dẫn đến sự suy yếu dân chủ nếu nó được sử dụng đúng cách. Thay vì để AI ra quyết định, nó có thể được dùng như một công cụ để tăng cường hiệu quả trong khu vực công. Ở trường hợp này, hiệu suất của AI không thay thế mà lại phục vụ cho quá trình giao tiếp.

AI có thể xử lý dữ liệu lớn, phân tích xu hướng xã hội, hỗ trợ ra quyết định chính sách dựa trên bằng chứng. Nếu được thiết kế minh bạch, AI có thể tăng cường năng lực giao tiếp lý tính bằng cách cung cấp thông tin khách quan cho các cuộc tranh luận chính trị. Ví dụ, AI có thể giúp phân tích các luồng ý kiến công chúng, tóm tắt các lập luận phức tạp, hoặc làm nổi bật các điểm bất đồng để thúc đẩy một cuộc đối thoại hiệu quả hơn. Bên cạnh đó, AI giúp con người đưa ra các quyết định có thông tin tốt hơn, đồng thời vẫn giữ vững nguyên tắc rằng quyền đưa ra quyết định cuối cùng vẫn thuộc về các chủ thể chính trị có lý trí. Sự hỗ trợ này giúp giảm bớt sự không chắc chắn và sai sót của con người, nhưng không vượt qua vai trò của chủ thể chính trị.

Tóm lại, trong mối quan hệ biện chứng giữa AI và quyền lực chính trị, AI vừa là công cụ củng cố quyền lực (giám sát, tối ưu hóa quyết định), vừa là yếu tố phân tán quyền lực (hỗ trợ giao tiếp dân chủ). Ngày nay, tầm quan trọng của mối quan hệ này nổi bật khi các quốc gia như: Trung Quốc (kiểm soát qua tín dụng xã hội), Mỹ (dự đoán tội phạm), và Việt Nam (ứng dụng thí điểm trong xây dựng thành phố thông minh) đang triển khai các mô hình quản trị AI khác nhau. Thực trạng cho thấy việc sử dụng AI mang lại tiềm năng to lớn nhưng cũng tiềm ẩn rủi ro. Cần định hướng việc quản trị AI phù hợp với điều kiện thực tiễn Việt Nam nhằm bảo vệ dân chủ và giá trị con người, thông qua khung pháp lý minh bạch và sự tham gia của công dân trong không gian công cộng số. Do đó, triết học chính trị cần tiếp tục khám phá để AI phục vụ lợi ích xã hội, tránh lạm dụng quyền lực và duy trì bản sắc dân tộc trong kỷ nguyên số ■

[1] Xem: Nguyễn Minh Hải, *Trí tuệ nhân tạo AI là gì? Khám phá lợi ích và thách thức*, https://vnptai.io, ngày 13-2-2025.

[2] James Barrat, *Phát minh cuối cùng - Trí tuệ nhân tạo và sự cáo chung của kỷ nguyên con người*, Nxb. Thế giới, Hà Nội, 2018, tr.25-26.

[3] Văn Toản, *Những cột mốc đánh dấu sự hình thành và phát triển của trí tuệ nhân tạo*, https://nhandan.vn, ngày 13-3-2023.

[4] Đức Anh, *Quốc gia đầu tiên trên thế giới dùng AI để viết luật*, https://vneconomy.vn, ngày 21-4-2025.

[5] Thanh Danh, *Thủ tướng Thụy Điển gây tranh cãi vì tham khảo quyết sách bằng AI*, https://vnexpress.net, ngày 6-8-2025.

[6] Xem: Luis Suarez Villa, *Technocapitalism: A Critical Perspective on Technological Innovation and Corporatism*, Temple University Press, Philadelphia, 2009, p.4.

[7, 8] C.Mác và Ph.Ăngghen, *Toàn tập*, t.46, phần II, Nxb. Chính trị quốc gia, Hà Nội, 2000, tr.372-373, 367.

[9, 10] Shoshana Zuboff, *Kỷ nguyên của chủ nghĩa tư bản giám sát - Cuộc chiến vì tương lai loài người ở biên giới mới của quyền lực*, Nxb. Chính trị quốc gia Sự thật, Hà Nội, 2021, tr.560-561, 561.

[11] Michel Foucault, *Discipline and Punishment: The Birth of the Prison, Vintage Books*, New York, 1995, pp.27-28.

[12] Xem: Michel Foucault, *The Foucault Reader*, Penguin Books, New York, 2020, p.265.

[13] Zhidas Daskalovski, *Số hóa: Lựa chọn giữa an ninh trật tự và dân chủ tự do*, in trong: Fatima Roumate (Chủ biên), *Trí tuệ nhân tạo và ngoại giao số - Thách thức và cơ hội*, Nxb. Chính trị quốc gia Sự thật, Hà Nội, 2022, tr.162-163.

[14] Jürgen Habermas, *The Structural Transformation of the Public Sphere - An Inquiry into a Category of Bourgeois Society*, The MIT Press, Cambridge, 1991, p.27.