

QUẢN LÝ TRÍ TUỆ NHÂN TẠO

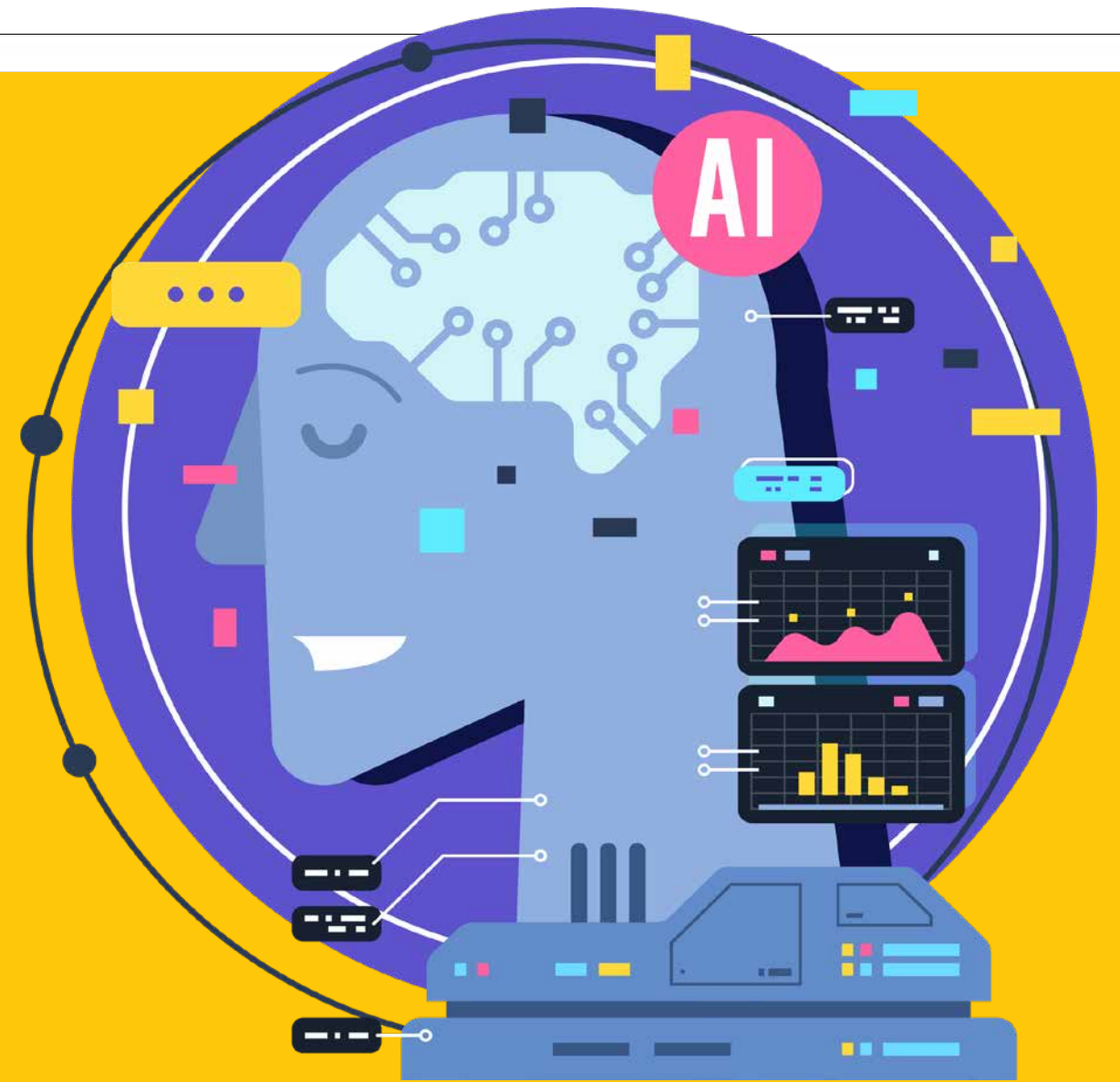
KINH NGHIỆM QUỐC TẾ VÀ GỢI MỞ VỚI VIỆT NAM



⇒ NGUYỄN ĐỨC THỦY

Phó trưởng phòng Nghiên cứu phát triển Công nghệ -
Viện Công nghiệp phần mềm và Nội dung số - Bộ TT&TT

Trí tuệ nhân tạo hay trí thông minh nhân tạo (Artificial Intelligence - AI) được xem là một ngành khoa học máy tính (Computer Science), liên quan đến việc tự động hóa các hành vi thông minh đang ngày càng được ứng dụng trong thực tế. Cũng như việc ứng dụng công nghệ mới, khi AI được ứng dụng vào cuộc sống cần phải có chính sách quản lý, thúc đẩy hiệu quả.



Ứng dụng AI và bài toán quản lý

Có thể hiểu một cách đơn giản, AI là công nghệ mô phỏng các quá trình suy nghĩ và học tập của con người bằng máy móc, đặc biệt là các hệ thống máy tính. Các quá trình này bao gồm việc học tập (thu thập thông tin và các quy tắc sử dụng thông tin), lập luận (sử dụng các quy tắc để đạt được kết luận gần đúng hoặc xác định) và tự sửa lỗi. Trí tuệ này có thể tư duy, suy nghĩ, học hỏi v.v.. như trí tuệ con người. AI cho phép xử lý dữ liệu ở mức rộng lớn hơn, quy mô hơn, mang tính hệ thống khoa học và nhanh hơn so với con người. Các ứng dụng đặc biệt của AI bao gồm các hệ thống chuyên gia, nhận dạng tiếng nói và thị giác máy tính (nhận diện khuôn mặt, vật thể hoặc chữ viết).

Các công nghệ được phát triển và sau đó là ứng dụng vào thực tế đều tác động ảnh hưởng đến đời sống kinh tế xã hội của một quốc gia. Đối với AI cũng vậy.

Đối với các hệ thống AI, người tạo ra hệ thống không thể dự đoán trước được hành động của AI, mà việc dự đoán hành động này rất quan trọng đối với việc tiếp cận pháp lý vì cần để xác định trách nhiệm pháp lý đối với hành động do AI thực hiện. Đặc điểm này đặt ra vấn đề liên quan đến khả năng dự đoán và khả năng hành động độc lập, đồng thời không chịu trách nhiệm của AI. Khi AI được áp dụng vào những lĩnh vực rủi ro cao, thì áp lực đặt ra cần có quy định đối với sản phẩm AI phải có trách nhiệm, công bằng và minh bạch.

Việc tiếp cận xây dựng quy định về AI ban đầu có thể tập trung vào các nhóm nhỏ, cụ thể như sau.

Thứ nhất, quản trị đạo đức: tập trung vào vấn đề đạo đức phát sinh từ AI như là công bằng, minh bạch và riêng tư, việc phân bổ dịch vụ và hàng hóa và dịch chuyển nền kinh tế (mất việc do tự động hóa dựa trên AI).

Thứ hai, khả năng giải thích: đây là cơ chế để tăng tính công bằng, minh bạch của thuật toán. Ví dụ, ý tưởng về “quyền giải thích” quyết định của thuật toán được tranh cãi tại châu Âu. Đây là quyền cho phép cá nhân có quyền được giải thích nếu thuật toán đó quyết định liên quan đến họ, ví dụ từ chối đơn vay tiền. Tuy nhiên, quyền này chưa được đảm bảo. Hơn nữa, việc giải thích thuật toán để mở và cách giải thích được lưu lại trong hệ thống AI.

Thứ ba, kiểm soát đạo đức: đối với hệ thống thuật toán khó hiểu và phức tạp, cơ chế trách nhiệm không thể chỉ dựa vào giải thích. Cơ chế kiểm toán được đề xuất như là một giải pháp khả thi mà kiểm tra đầu ra và đầu vào của thuật toán để tìm ra sự sai lệch, gây hại, thay vì giải đáp cách thức hoạt động của hệ thống. Bên cạnh đó, nhiều quan điểm tranh luận về việc đưa ra các khái niệm xã hội như công bằng, minh bạch, có trách nhiệm là những khái niệm mơ hồ và thiếu cơ chế thực thi, khó có thể áp dụng các khái niệm trừu tượng áp dụng vào những số liệu, thuật toán, mã code. Điều này một phần gây bối rối cho các nhà nghiên cứu, phát triển công nghệ và đồng thời cũng gây cho các nhà hoạch định chính sách hiểu sai, hiểu lầm khi áp dụng các thuật ngữ trừu tượng vào máy học.

Mặc dù còn thiếu vắng khung quy định pháp luật về AI, nhưng các nước trên thế giới, vẫn tiếp tục đầu tư mạnh mẽ vào lĩnh vực này. Điều này, một lần nữa tạo ra áp lực cho các nhà nghiên cứu, nhà khoa học thiết lập các quy định phù hợp với những đặc điểm của AI.

Kinh nghiệm thúc đẩy, quản lý AI của một số nước

1. Kinh nghiệm của Mỹ

Hoa Kỳ là quốc gia đầu tiên xây dựng Kế hoạch chiến lược phát triển AI quốc gia (ban hành tháng 10/2016, dưới thời của Tổng thống Obama) với mục tiêu để Hoa Kỳ là cường quốc phát triển AI thế giới, từ đó, thúc đẩy nền kinh tế và an ninh quốc gia của nước này phát triển. Bản kế hoạch này đã đưa ra bảy chiến lược và hai khuyến nghị.

Bảy chiến lược đó là: (1) Đầu tư dài hạn cho nghiên cứu phát triển AI; (2) Phát triển các phương pháp hiệu quả cho sự hợp tác giữa con người với AI, (hay nói cách khác là con người làm việc cùng đồng nghiệp là rô-bốt); (3) Nắm vững, giải quyết các vấn đề phát sinh về pháp lý, đạo đức và ảnh hưởng xã hội của AI; (4) Bảo đảm an toàn, an ninh của các hệ thống phát triển AI; (5) Phát triển các bộ dữ liệu công chia sẻ (shared public datasets) và môi trường để huấn luyện và kiểm thử AI; (6) Đo lường, đánh giá các công nghệ AI thông qua các tiêu chuẩn và điểm chuẩn (benchmark); (7) Hiểu rõ hơn nhu cầu nguồn nhân lực nghiên cứu phát triển AI quốc gia.

Hai khuyến nghị đó là: (1) Phát triển khung triển khai nghiên cứu AI để xác định các cơ hội khoa học – công nghệ và hỗ trợ hiệu quả việc hợp tác trong đầu tư nghiên cứu phát triển AI phù hợp với chiến lược từ 1–6. (2) Nghiên cứu xây dựng không gian phát triển AI quốc gia, bảo đảm duy trì bền vững nguồn nhân lực nghiên cứu phát triển AI phù hợp chiến lược 7. Hai điểm nổi bật trong bảy chiến lược và hai khuyến nghị là các khu vực tập trung đầu tư của Chính phủ Hoa Kỳ và nguồn nhân lực nghiên cứu phát triển AI cho Hoa Kỳ.

Đến khi Tổng thống Donald J. Trump cầm quyền (2017), ông đã khởi động Sáng kiến Trí tuệ nhân tạo Mỹ - chiến lược của Quốc gia nhằm thúc đẩy sự lãnh đạo của Mỹ trong lĩnh vực AI, bằng cách ký Sắc lệnh hành pháp 13859 vào ngày 11/2/2019. Sáng kiến Trí tuệ nhân tạo Mỹ tập trung nguồn

lực của Chính phủ Liên bang để hỗ trợ đổi mới AI nhằm tăng cường thịnh vượng, an ninh quốc gia và cải thiện chất lượng cuộc sống cho người dân Mỹ. Chiến lược này nhấn mạnh 6 chính sách và thực tiễn chính như sau.

1) **Đầu tư vào nghiên cứu và phát triển AI:** Mỹ phải thúc đẩy đầu tư của Liên bang vào nghiên cứu và phát triển AI phối hợp với các ngành công nghiệp, học viện, các đối tác quốc tế và đồng minh cũng như các tổ chức không thuộc Liên bang khác để tạo ra những đột phá công nghệ trong AI. Tổng thống Trump kêu gọi tăng gấp 2 lần hoạt động nghiên cứu và phát triển AI phi quốc phòng trong năm tài chính 2019 và 2021.

2) **Khai thác tài nguyên AI:** Mỹ phải tăng cường quyền truy cập vào dữ liệu, mô hình và tài nguyên máy tính chất lượng cao của Liên bang để tăng giá trị của chúng cho hoạt động R&D của AI, đồng thời duy trì và mở rộng các biện pháp bảo vệ an toàn, bảo mật, quyền riêng tư và bí mật. Chiến lược kêu gọi các cơ quan Liên bang xác định các cơ hội mới để tăng cường khả năng tiếp cận và sử dụng dữ liệu và mô hình của Liên bang. Vào năm 2019, Văn phòng Quản lý và Ngân sách của Nhà Trắng đã thiết lập Chiến lược Dữ liệu Liên bang như một khuôn khổ cho các nguyên tắc hoạt động và thực tiễn tốt về cách các cơ quan Liên bang sử dụng và quản lý dữ liệu.

3) **Xóa bỏ các rào cản đối với đổi mới AI:** Mỹ phải giảm các rào cản đối với việc phát triển, thử nghiệm, triển khai và áp dụng các công nghệ AI một cách an toàn bằng cách cung cấp hướng dẫn để quản trị AI phù hợp với các giá trị của quốc gia và bằng cách thúc đẩy sự phát triển của kỹ thuật AI phù hợp tiêu chuẩn.

4) **Đào tạo lực lượng lao động sẵn sàng với AI:** Mỹ phải trao quyền cho các thể hệ công nhân Mỹ hiện tại và tương lai thông qua học việc; các chương trình kỹ năng; và giáo dục về khoa học, công nghệ, kỹ thuật và toán học (STEM), chú trọng vào khoa học máy tính, nhằm đảm bảo rằng công

nhân Mỹ, bao gồm cả công nhân Liên bang, có khả năng tận dụng tối đa các cơ hội của AI. Tổng thống Trump chỉ đạo tất cả các cơ quan Liên bang ưu tiên liên quan đến AI, các chương trình và cơ hội học nghề, đào tạo việc làm.

Ngoài trọng tâm R&D, chương trình Viện Nghiên cứu AI Quốc gia mới của Quỹ Khoa học Quốc gia cũng sẽ đóng góp vào sự phát triển của lực lượng lao động, đặc biệt là các nhà nghiên cứu AI.

5) **Thúc đẩy môi trường quốc tế hỗ trợ sự đổi mới AI của Mỹ:** Mỹ phải tham gia quốc tế để thúc đẩy một môi trường toàn cầu hỗ trợ nghiên cứu và đổi mới AI của Mỹ, mở ra thị trường cho các ngành công nghiệp AI của Mỹ đồng thời bảo vệ lợi thế công nghệ của Mỹ trong AI. Năm 2021, Mỹ đã dẫn đầu những nỗ lực lịch sử tại Tổ chức Hợp tác và Phát triển Kinh tế (OECD) nhằm phát triển các thỏa thuận đồng thuận quốc tế đầu tiên về các nguyên tắc cơ bản để quản lý AI đáng tin cậy. Mỹ cũng đã làm việc với các đối tác quốc tế của mình trong G7 và G20 để áp dụng các nguyên tắc AI tương tự.

6) **Sử dụng AI đáng tin cậy cho các dịch vụ và sứ mệnh của chính phủ:** Mỹ phải nắm lấy công nghệ như AI để cải thiện việc cung cấp và hiệu quả các dịch vụ của chính phủ cho người dân Mỹ và đảm bảo ứng dụng của nó thể hiện sự tôn trọng đúng mức đối với các giá trị của quốc gia, bao gồm quyền riêng tư, dân sự quyền và tự do dân sự. Cơ quan Quản lý Dịch vụ Tổng hợp đã thành lập Trung tâm AI xuất sắc để cho phép các cơ quan Liên bang xác định các phương pháp hay nhất để kết hợp AI vào tổ chức của họ.

Như vậy, có thể thấy, chính sách ưu tiên phát triển AI là chính sách xuyên suốt các đời Tổng thống Mỹ.

2. Kinh nghiệm của Trung Quốc

Tháng 7/2017, Hội đồng Nhà nước Trung Quốc công bố Kế hoạch phát triển AI thế hệ mới. Kế hoạch tập trung vào AI như là mục tiêu quốc gia, trong đó chi tiết hóa chiến lược xây dựng ngành

công nghiệp AI quốc gia để đạt trị giá 150 tỷ USD trong tương lai gần nhằm đưa Trung Quốc trở thành siêu cường AI hàng đầu thế giới vào năm 2030. Mục tiêu kế hoạch được chia làm 3 giai đoạn:

Giai đoạn 1 (đến năm 2020): Tổng sản lượng của ngành công nghiệp AI lõi của Trung Quốc và tổng sản lượng của ngành liên quan đến AI sẽ lần lượt vượt 22,5 tỷ USD và 150,8 tỷ USD, đưa Trung Quốc vào liên minh các nước tiên tiến nhất về phát triển AI và công nghệ tiên tiến nói chung.

Giai đoạn 2 (đến năm 2022): Trung Quốc có kế hoạch đạt các sản lượng cốt lõi của ngành công nghiệp liên quan đến AI và thuần AI lần lượt vượt quá 60,3 tỷ USD và 754 tỷ USD, đưa Trung Quốc trở thành công ty hàng đầu thế giới trong nhiều lĩnh vực AI.

Giai đoạn 3 (đến năm 2020): Trung Quốc tìm cách trở thành trung tâm đổi mới AI chính của thế giới, với một sản lượng cốt lõi liên quan đến AI và thuần AI lần lượt vượt quá 150,8 tỷ USD và 1,5 nghìn tỷ USD.

Những đặc điểm chính trong hoạch định chủ trương, chính sách phát triển AI của Trung Quốc là:

Thứ nhất, phương pháp tiếp cận từ trên xuống của Chính phủ Trung Quốc đối với mục tiêu AI quốc gia. Sự hỗ trợ của Chính phủ cho AI ở Trung Quốc là rất lớn. Các chính sách khác nhau đang được chính quyền địa phương và Trung ương thực hiện để thu hút tài năng và dự án AI phù hợp. Họ cũng tìm kiếm các dự án AI và có nhiều chính sách thuế có lợi cho các công ty công nghệ phát triển hoặc áp dụng AI. Các chính sách này tạo ra một hệ sinh thái vĩ mô có lợi cho AI với nhiều tiếp cận vào Chính phủ, tài trợ và hỗ trợ và khuyến khích nhà đầu tư khi phát triển các sản phẩm AI sáng tạo.

Ngoài ra, Bộ Khoa học và Công nghệ Trung Quốc cũng đã chỉ định bốn công ty AI lớn nhất là Baidu, Alibaba, Tencent và iFlyTex để dẫn dắt sự phát triển của các nền tảng đổi mới AI quốc gia trong các lĩnh vực như: ô tô tự lái, thành phố thông minh, tầm nhìn máy tính để chẩn đoán y tế và trí thông minh giọng nói.

Thứ hai, đầu tư và tài trợ mạnh mẽ cho AI. Chính phủ đóng một vai trò lớn hơn trong việc tài trợ cho các dự án AI. Thông qua quỹ hướng dẫn của Chính phủ do chính quyền địa phương và các công ty nhà nước thành lập, Trung Quốc đã đầu tư hơn 1 tỷ USD vào các công ty khởi nghiệp trong nước trong năm 2016.

Cùng với đầu tư, Trung Quốc đang cố gắng thu hút nhân tài AI từ khắp thế giới. Không chỉ giữ chân các tài năng Trung Quốc hiện có, các công ty và tổ chức khu vực tư nhân cũng đang cấp học bổng và các gói lương cạnh tranh cho các chuyên gia AI phương Tây đến và làm việc tại Trung Quốc.

Thứ ba, quy mô dân số là lợi thế làm ngày càng tăng cơ sở dữ liệu người tiêu dùng của Trung Quốc. Hơn 800 triệu người ở Trung Quốc đang hoạt động trên Internet, trong đó 98% là người dùng di động (788 triệu người). Người dùng kết nối nhiều hơn có nghĩa là nhiều dữ liệu hơn, điều này sẽ càng tốt cho sự phát triển của AI, làm AI càng chính xác.

Thứ tư, sự cởi mở để chia sẻ dữ liệu trong nước. Trung Quốc có các biện pháp bảo vệ tương đối lỏng lẻo và việc các đại gia công nghệ của Trung Quốc thu thập các kho dữ liệu khổng lồ và chia sẻ giữa các cơ quan chính phủ và các công ty là phổ biến. Tuy nhiên, Trung Quốc chỉ chọn lọc chia sẻ dữ liệu trong nước chứ không phải quốc tế. Chính phủ Trung Quốc đã kiểm duyệt, chặn Facebook và Google, do đó cho phép sự gia tăng của các nền tảng trong nước như Wechat, Weibo...

Bởi vì các sản phẩm AI liên quan nhiều tới dữ liệu nên Cục An ninh mạng Trung Quốc (CAC), cơ quan giám sát bảo vệ dữ liệu và an ninh mạng của Trung Quốc đã ban hành "Quy định quản lý các thuật toán khuyến nghị cho các dịch vụ thông tin Internet", quy định này có hiệu lực áp dụng từ ngày 01/3/2022. Đây là một bộ quy tắc mở rộng - một trong những quy định về sử dụng các thuật toán được đề xuất dựa trên AI, được ban hành và hiệu lực áp dụng đầu tiên trên thế giới.

Mục tiêu ban hành các quy định hướng tới điều chỉnh việc áp dụng các thuật toán trong hoạt động

cung cấp các dịch vụ thông tin Internet tại Trung Quốc và phù hợp với các quy định hiện hành, chẳng hạn như Luật An ninh mạng, Luật Bảo vệ dữ liệu, Luật Bảo vệ dữ liệu cá nhân, các quy định về cung cấp thông tin, dịch vụ Internet của nước này. Đối tượng áp dụng là các tổ chức, doanh nghiệp, cá nhân áp dụng các thuật toán được đề xuất để cung cấp dịch vụ thông tin Internet.

Các thuật toán được đề xuất trong quy định này được hiểu là các thuật toán được thực hiện phần lớn liên quan đến công nghệ AI, chẳng hạn như tổng hợp nhân tạo, xử lý dữ liệu cá nhân, sắp xếp và lựa chọn thông tin, lọc và truy xuất dữ liệu, lập lịch, ra quyết định v.v... Do đó phạm vi áp dụng của các quy định này cho các thuật toán theo phương pháp liệt kê không đầy đủ.

Do đó các điều khoản quy định về cơ bản không giới hạn đối với các loại thuật toán nêu trên. Nội dung các quy định tập trung vào các vấn đề như:

- Các nhà cung cấp dịch vụ Internet không sử dụng các thuật toán được đề xuất cho các hoạt động vi phạm quy định của luật pháp như: Gây nguy hiểm cho an ninh quốc gia, lợi ích xã hội và công cộng, gây rối trật tự kinh tế và xã hội, xâm phạm quyền và lợi ích hợp pháp của các tổ chức, cá nhân, phổ biến thông tin xấu và độc hại, phổ biến pháp luật và các quy định hành chính.

- Các nhà cung cấp dịch vụ Internet sử dụng các thuật toán được đề xuất phải chịu trách nhiệm chính về bảo mật thuật toán, tạo lập, cải tiến và giám sát cơ chế hoạt động của thuật toán, giám sát về khía cạnh đạo đức khoa học và công nghệ, đăng ký người dùng, giám sát việc phát hành thông tin, bảo mật dữ liệu và bảo vệ thông tin cá nhân, chống gian lận mạng viễn thông, giám sát và đánh giá các chức năng bảo mật người dùng và bảo mật hệ thống, có các biện pháp kỹ thuật ứng phó khẩn cấp với sự cố, xây dựng và xuất bản các quy tắc liên quan cho dịch vụ được cung cấp áp dụng thuật toán được đề xuất, cung cấp các chuyên gia và hỗ

trợ kỹ thuật tương xứng với quy mô của dịch vụ áp dụng thuật toán được đề xuất.

- Các nhà cung cấp dịch vụ Internet sử dụng các thuật toán được đề xuất phải thường xuyên xem xét, đánh giá và xác minh các cơ chế thuật toán, mô hình, dữ liệu và kết quả ứng dụng v.v.. và không được thiết lập các mô hình thuật toán khiến người dùng nghiện hoặc sử dụng quá mức, vi phạm các quy tắc đạo đức.

- Các nhà cung cấp dịch vụ Internet sử dụng các thuật toán được đề xuất phải tăng cường cải thiện cơ sở dữ liệu, tính năng để xác định và xử lý thông tin bất hợp pháp và thông tin xấu, cải thiện các tiêu chuẩn, quy tắc và thủ tục lưu trữ. Các thuật toán được đề xuất phải có các chức năng xử lý thông tin với khả năng diễn giải, giải trình một cách minh bạch theo các quy tắc đề ra.

- Các nhà cung cấp dịch vụ Internet sử dụng các thuật toán được đề xuất không được sử dụng thuật toán để tạo và đăng ký tài khoản, giao dịch bất hợp pháp, thao túng thông tin, can thiệp vào việc trình bày thông tin, thực hiện các hành vi gây ảnh hưởng đến dư luận trực tuyến hoặc trốn tránh sự giám sát và quản lý của các cơ quan có thẩm quyền; không được sử dụng thuật toán để hạn chế một cách bất hợp lý các nhà cung cấp dịch vụ thông tin Internet khác, cản trở, phá hoại hoạt động bình thường của các dịch vụ thông tin Internet do họ cung cấp hợp pháp và thực hiện các hành vi độc quyền và cạnh tranh không lành mạnh.

Các quy định đảm bảo quyền lợi của người dùng, đặc biệt là đối với trẻ em và người già; trách nhiệm quản lý của các cơ quan có thẩm quyền về thủ tục, trình tự cấp phép; các chế tài xử phạt vi phạm các quy định cũng được nêu ra trong văn bản này.

3. Kinh nghiệm của Pháp

Có thể nói châu Âu nói chung và Pháp nói riêng đi sau Bắc Mỹ và Trung Quốc trong xây dựng chiến lược phát triển AI. Do vậy Pháp đã chủ trương tăng

tốc phát triển AI. Ngày 29/3/2018, chính phủ Pháp công bố chiến lược phát triển AI quốc gia với mục tiêu đưa Pháp trở thành “AI hub” của khu vực. Tổng thống Macron nhận định rằng, AI sẽ tạo đột phá trong tất cả các mô hình kinh doanh khác nhau.

Tuyên bố của Tổng thống Pháp về chiến lược phát triển AI là tóm tắt các điểm chính của Báo cáo chiến lược phát triển AI của Pháp và châu Âu do Cédric Villani (nhà toán học Pháp được giải thưởng Fields năm 2010 và là một nghị sỹ Pháp) và cộng sự xây dựng. Những nội dung chính trong báo cáo đó là:

Một là, xây dựng một chính sách dữ liệu tích cực, khuyến khích các công ty cùng tạo lập và chia sẻ dữ liệu, tạo lập dữ liệu được xã hội quan tâm, hỗ trợ quyền sao lưu dữ liệu.

Hai là, bốn khu vực phát triển AI chiến lược trọng tâm là y tế, giao thông, môi trường, quốc phòng và an ninh, tiến hành các chính sách riêng theo từng khu vực chiến lược hướng vào các vấn đề chính, tạo nền tảng đặc thù khu vực, kiểm tra vùng đổi mới cho mỗi khu vực.

Ba là, thúc đẩy lợi thế tiềm năng nghiên cứu phát triển AI của Pháp, thành lập các tổ chức AI liên ngành tại các trường đại học và viện nghiên cứu được chọn lựa, phân bổ nguồn lực phù hợp dành cho nghiên cứu (bao gồm một siêu máy tính được thiết kế riêng cho ứng dụng AI với sự hợp tác của nhà sản xuất); tăng lương cho các nhà nghiên cứu và tăng cường trao đổi hàn lâm – công nghiệp.

Bốn là, lập kế hoạch ứng phó tác động của công nghệ AI tới người lao động, thiết lập phòng thí nghiệm công về chuyển đổi công việc, phát triển nghiên cứu về sự hỗ trợ của máy móc tới con người, đánh giá phương pháp tài trợ mới cho đào tạo nghề.

Năm là, nâng cao tính thân thiện với môi trường của công nghệ AI, xây dựng trung tâm nghiên cứu về AI với hệ sinh thái, giảm thiểu năng lượng cho sử dụng AI, chuyển đổi hệ sinh thái song hành với độ tự do dữ liệu hệ sinh thái.

Sáu là, bảo đảm tính minh bạch của công nghệ phát triển AI, phát triển tính minh bạch và kiểm toán thuật toán, lưu ý trách nhiệm của các tác nhân AI liên quan tới đe dọa đạo đức, thành lập ủy ban đạo đức tư vấn cho các công nghệ số và AI với trách nhiệm tổ chức các tranh luận công khai về đạo đức AI, bảo đảm nguyên tắc trách nhiệm là của con người (đặc biệt khi công cụ AI dùng trong dịch vụ công).

Để thực hiện chiến lược nói trên, Chính phủ Pháp sẽ đẩy mạnh dữ liệu mở của cả khối công và tư để tạo môi trường cho AI phát triển, nhất là các công ty start-up; quan tâm đến vấn đề đạo đức của AI, đảm bảo các bộ dữ liệu huấn luyện AI có ít sự thiên vị nhất. Pháp đầu tư 1,5 tỷ Euro trong 5 năm (2018 – 2022) đã được nhìn nhận như đại diện cho chiến lược phát triển AI của các quốc gia châu Âu.

Bài học kinh nghiệm về quản lý, phát triển AI

Từ những nghiên cứu tổng quan về sự hình thành phát triển và ứng dụng của công nghệ AI, cũng như kinh nghiệm quản lý AI của các quốc gia có thể rút ra một số nhận định, bài học về quản lý thúc đẩy phát triển AI như sau:

Thứ nhất, sản phẩm AI sẽ được ứng dụng mang tính phổ biến ở mọi lĩnh vực của nền kinh tế, xã hội, giúp các quốc gia tăng cường tính cạnh tranh của nền kinh tế. AI là một trong những trụ cột của Cách mạng 4.0, là động lực chính trong chiến lược chuyển đổi số của các nước phát triển.

Thứ hai, xu thế phát triển AI là tất yếu và AI sẽ có tác động to lớn, toàn diện cả mặt tích cực và tiêu cực trong mọi mặt đời sống xã hội. Do vậy hầu hết các quốc gia trên thế giới đều có chủ trương, chính sách và các biện pháp triển khai để thúc đẩy phát triển, ứng dụng sản phẩm AI trong nước.

Thứ ba, đối với các nước lớn, AI có khả năng là chìa khóa để thay đổi cán cân quyền lực kinh tế, quân sự, chính trị của quốc gia ở phạm vi khu vực cũng như toàn cầu. Điều này đặc biệt đáng lưu ý là

cuộc chạy đua phát triển AI giữa hai cường quốc lớn Mỹ và Trung Quốc.

Thứ tư, xây dựng cơ chế chính sách quản lý phát triển AI phải có mục tiêu rõ ràng, trên cơ sở đó xây dựng chiến lược, lộ trình, các nội dung triển khai và mục tiêu cụ thể cho từng giai đoạn.

Thứ năm, các chính sách thúc đẩy phát triển AI chủ yếu theo hướng tiếp cận từ trên xuống. Để xây dựng và phát triển AI thành công cần phải có sự hợp tác giữa chính phủ, khu vực tư nhân, nhà đầu tư và học viện trong quá trình phát triển AI. Chính phủ đóng vai trò đầu tàu, là mắt xích giữa các bên liên quan trong quá trình xây dựng và phát triển AI.

Thứ sáu, chính phủ cần nghiên cứu kỹ lưỡng để hình thành những chiến lược, lộ trình, giải pháp cụ thể để phát triển và ứng dụng AI phù hợp với thực tế và tiềm năng của đất nước.

Thứ bảy, chính phủ cũng cần có những biện pháp hỗ trợ các doanh nghiệp trong việc đầu tư, ứng dụng AI vào hoạt động sản xuất kinh doanh của mình như tạo cơ chế pháp lý thông thoáng, có chính sách thuế có lợi các công ty công nghệ phát triển hoặc áp dụng AI. Đồng thời chính phủ cũng cần có chính sách phối hợp với các trường đại học trong việc đào tạo nguồn nhân lực AI chất lượng cao, có chính sách thu hút nhân tài AI

Thứ tám, chính sách xây dựng kiến tạo các nguồn dữ liệu lớn tạo điều kiện triển khai các ứng dụng AI.

Tuy nhiên các quốc gia hiện nay vẫn đang “thận trọng” trong việc xây dựng và ban hành các luật hoặc quy định nhằm quản lý các sản phẩm, ứng dụng, hệ thống AI.

Hướng tiếp cận chủ yếu hiện nay của các quốc gia là đưa ra các quy định điều chỉnh để việc sử dụng AI không vi phạm các quy định hiện hành (chẳng hạn như luật về dữ liệu riêng tư, an toàn thông tin v.v.). Do tính chất đặc thù của AI nên các quốc gia hiện nay hướng tới nghiên cứu các quy định về đạo đức, lòng tin, tính rủi ro trong sử dụng

AI nhằm hạn chế các tác hại có thể xuất hiện trong việc sử dụng AI. Việc xây dựng và ban hành các tiêu chuẩn, quy chuẩn liên quan đến các vấn đề này hiện đang được các quốc gia xem xét.

Thay lời kết

AI đang được phát triển mạnh mẽ ở mọi lĩnh vực, là một trong số các công nghệ chủ chốt của Cách mạng công nghiệp lần thứ 4. AI đang được các quốc gia chú trọng phát triển để khẳng định vị thế về chính trị, kinh tế; là động lực đổi mới sản xuất, kinh doanh, tăng cường tính cạnh tranh của nền kinh tế.

Các quốc gia đều nhận thức được tầm quan trọng của AI do vậy họ đều có những chủ trương, chiến lược, lộ trình để phát triển và ứng dụng AI trong các lĩnh vực theo đặc điểm của quốc gia. Những vấn đề chủ yếu được quan tâm trong chiến lược phát triển của các quốc gia đó là phát triển AI có ưu tiên vào các lĩnh vực mang tính trọng tâm như sản xuất, kinh doanh, tài chính, ngân hàng, quốc phòng. Xây dựng và ban hành luật, các quy định quản lý các sản phẩm, ứng dụng và hệ thống AI hiện nay là vấn đề được chính phủ quan tâm xúc tiến triển khai nhưng cần tiếp cận một cách thận trọng.

Từ những kinh nghiệm trên, chúng ta cần triển khai nghiên cứu xây dựng và ban hành các chính sách, quy định quản lý AI phù hợp với thực tiễn phát triển ở Việt Nam. ■ THÔNG TIN & THUYẾT TRÌNH

Tài liệu tham khảo:

1. Cục Quản lý không gian mạng Trung Quốc – Các quy định về quản lý các thuật toán sử dụng cho các dịch vụ thông tin Internet, 1/2022
2. WIPO, Artificial Technology Trends 2019
3. Expert Group on Architecture for AI Principles, AI Governance in Japan, 2021
4. Chiến lược hiện thực hóa trí tuệ nhân tạo đáng tin cậy, dự thảo liên bộ, ngành, 5/2021
5. Monetary Authority of Singapore, Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector, 12/2018
6. White House M-21-06, Guidance for Regulation of Artificial Intelligence Applications, 6/2021
7. Asia Pacific Foundation of Canada, Artificial Intelligence Policies in East Asia: An Overview from the Canadian Perspective, 7/2019