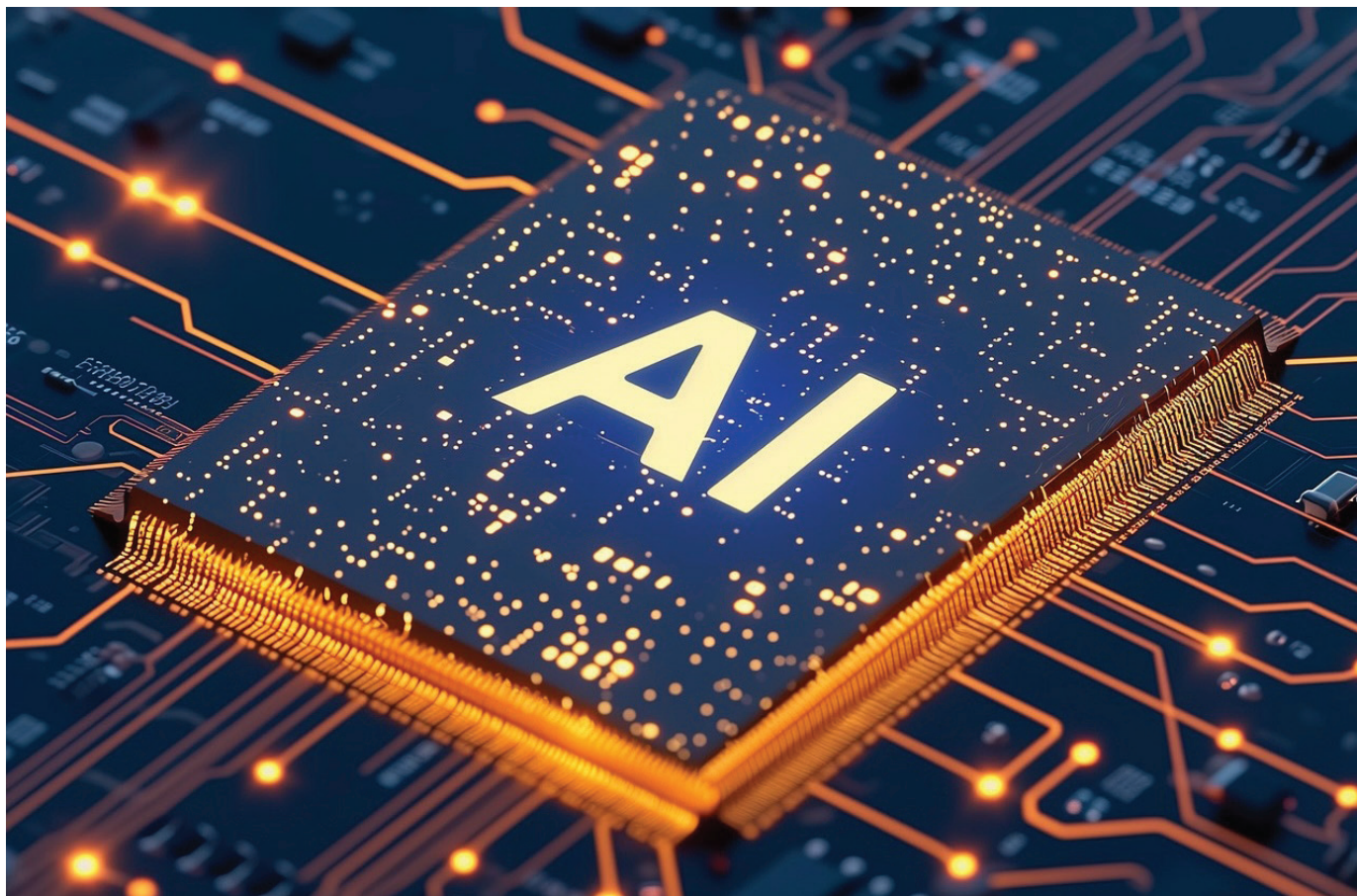




THÚC ĐẨY NGHIÊN CỨU PHÁT TRIỂN VÀ SỬ DỤNG TRÍ TUỆ NHÂN TẠO MỘT CÁCH CÓ TRÁCH NHIỆM Ở VIỆT NAM

Trần Anh Tú, Đinh Hải Đăng

Vụ Công nghệ cao, Bộ Khoa học và Công nghệ



Trí tuệ nhân tạo (Artificial Intelligence - AI) đã và đang ảnh hưởng ngày càng sâu rộng trong nhiều lĩnh vực đời sống xã hội. Chính vì vậy, việc nghiên cứu, phát triển và sử dụng AI như thế nào cho “có trách nhiệm” là một thách thức không nhỏ đối với các cơ quan quản lý và các tổ chức nghiên cứu, ứng dụng công nghệ này.



Thế nào là trí tuệ nhân tạo có trách nhiệm?

Trí tuệ nhân tạo có trách nhiệm (Responsible artificial intelligence - RAI) là một khái niệm đề cập đến một cách tiếp cận để phát triển và khai thác AI nhằm tuân theo các quan điểm về cả đạo đức và pháp lý. Mục tiêu của cách tiếp cận đối với AI có trách nhiệm là hướng đến việc phát triển và sử dụng AI theo cách an toàn, đáng tin cậy và có đạo đức.

Với cách tiếp cận AI có trách nhiệm, các chuyên gia đều hy vọng rằng việc hình thành một khuôn khổ quản trị được áp dụng rộng rãi bao gồm các thông



lệ tốt nhất về AI sẽ giúp các tổ chức, cá nhân trên thế giới dễ dàng hơn trong việc đảm bảo các hệ thống/ứng dụng AI sẽ lấy con người làm trung tâm. Từ đó, sẽ có các ứng dụng/hệ thống AI có trách nhiệm đảm bảo tính công bằng, độ tin cậy và minh bạch. Nói cách khác, việc tiếp cận có trách nhiệm này là để đảm bảo đạo đức AI (hay vấn đề đạo đức trong việc nghiên cứu, thiết kế, phát triển, kinh doanh và sử dụng các ứng dụng, sản phẩm AI). Điều này bao gồm việc xem xét ý nghĩa đạo đức của các quyết định do máy móc (các thuật toán) đưa ra và xác định tình trạng pháp lý của sản phẩm, dịch vụ AI.

Thực tế cho thấy, việc thiết kế, phát triển và sử dụng có trách nhiệm các hệ thống AI liên quan đáng kể đến một số ứng dụng như phương tiện tự lái, rô-bốt chăm sóc sức khỏe... đây là những thứ đã và đang ảnh hưởng đến con người, xã hội hoặc sẽ ảnh hưởng rộng rãi trong một vài năm tới. Trong các ứng dụng kiểu này, AI có thể cần được xem xét dưới các góc độ về giá trị xã hội, đạo đức, các vấn đề về văn hóa/sắc tộc; việc giải trình/giải thích lý do cũng như tính minh bạch của ứng dụng/thuật toán sử dụng...

Việc phát triển và sử dụng trí tuệ nhân tạo có trách nhiệm trên thế giới

Trong vài năm qua, đã nổi lên những cuộc tranh luận gay gắt về việc phát triển và sử dụng AI một cách có trách nhiệm. Các quốc gia và tổ chức lớn trên thế giới đã ban hành nhiều loại hướng dẫn để đảm bảo các ứng dụng/hệ thống AI đáp ứng các yêu cầu về đạo đức, điển hình là Khuyến nghị về đạo đức AI của Tổ chức Giáo dục, Khoa học và Văn hóa của Liên Hợp Quốc (UNESCO), trong đó nêu cao tính minh bạch, khả năng giải thích, an toàn, trách nhiệm và quyền riêng tư... Tuy nhiên, để phát triển và sử dụng các ứng dụng/hệ thống AI đảm bảo được các nguyên tắc nêu trên đòi hỏi phải có những hướng dẫn rõ ràng hơn trong những bối cảnh văn hóa hay những quan niệm khác biệt về giá trị tại những môi trường ứng dụng khác nhau.

Liên minh châu Âu

Ở châu Âu, nhằm bảo vệ quyền con người, khuyến khích sự sáng tạo và giảm thiểu những rủi ro tiềm ẩn, Liên minh châu Âu (EU) đã tiến hành xây dựng một hệ thống các quy định và hướng dẫn, để bảo đảm AI được phát triển và sử dụng một cách an toàn và có đạo đức. Cụ thể, Đạo luật về AI đã được ban hành năm 2021 nhằm đảm bảo AI được sử dụng một cách có trách nhiệm theo hướng phân loại các hệ thống AI theo mức độ rủi ro (từ thấp đến cao) và thiết lập các yêu cầu cụ thể cho từng loại. Đối với các hệ thống có

mức độ rủi ro cao như sinh trắc học, y tế, sẽ áp dụng các biện pháp kiểm soát nghiêm ngặt để đảm bảo an toàn và tôn trọng quyền con người. Ngoài ra, châu Âu cũng xây dựng Hướng dẫn đảm bảo đạo đức về AI đáng tin cậy, nhằm cung cấp một bộ nguyên tắc cơ bản định hướng cho việc phát triển và triển khai AI. Bên cạnh đó, châu Âu cũng tăng cường các nghiên cứu, xây dựng các quy định quản lý, phát triển dữ liệu và đảm bảo an toàn dữ liệu. Theo đó, các hệ thống AI cần tuân thủ quy định này, đặc biệt trong các vấn đề liên quan đến bảo mật và quyền riêng tư.

Vương quốc Anh

Tại Vương quốc Anh, Chiến lược quốc gia về AI được thông qua năm 2021, trong đó đã có định hướng cụ thể về các báo cáo trách nhiệm sử dụng các ứng dụng AI, cách thức thương mại sản phẩm AI và thúc đẩy việc ứng dụng, phát triển các ứng dụng AI. Bên cạnh đó, Sách trắng về AI cũng có những quy định rõ ràng và đưa ra công cụ tư vấn cho chính phủ. Tuy nhiên, thực tế AI đang phát triển quá nhanh do đó vẫn cần tiếp cận ở góc độ con người, kết nối cộng đồng để cùng nhau giải quyết các thách thức. Hiện nay, Vương quốc Anh đã phát triển trung tâm tiêu chuẩn AI nhằm thúc đẩy sử dụng AI có trách nhiệm. Trung tâm có vai trò quan trọng giúp các bên vừa áp dụng các tiêu chuẩn chung nhưng vẫn có thể thực hiện được nhiệm vụ đặc thù. Các quy định được chuẩn hóa giúp khai phá tiềm năng, là công cụ để thực hiện đổi mới sáng tạo, đồng thời trao quyền cho các bên liên quan được tham gia xây dựng tiêu chuẩn.

Hoa Kỳ

Tổng thống Hoa Kỳ đã ban hành Kế hoạch chiến lược nghiên cứu và phát triển AI quốc gia (cập nhật mới nhất năm 2023). Nhiều tập đoàn công nghệ lớn cũng có ý kiến kêu gọi xây dựng các quy định quản lý các ứng dụng AI, nhưng Chính phủ chưa đưa ra quy định cụ thể nào, chủ yếu quản lý dựa trên các quy định đối với các công nghệ mới hiện hành. Hiện nay, Viện Khoa học và Công nghệ quốc gia (NIST) đã công bố khung quản trị các ứng dụng AI về việc sử dụng các ứng dụng AI (bản mới nhất tháng 7/2024) và xem xét việc tích hợp các tiêu chuẩn quốc tế vào khung quản trị này.

Trung Quốc

Từ năm 2017, Trung Quốc đã công bố “Kế hoạch phát triển AI thế hệ mới”, đưa AI trở thành một phần trong chiến lược mũi nhọn của quốc gia này. Từ chiến lược quốc gia, có thể thấy rằng lĩnh vực AI không chỉ liên quan đến đổi mới công nghệ và kinh tế công

nghiệp, mà còn liên quan đến khía cạnh quản trị. Theo đó, Trung Quốc đã tăng cường các cơ chế, biện pháp để quản trị như luật pháp, khuôn khổ đạo đức, tiêu chuẩn, chứng nhận để thúc đẩy AI có trách nhiệm, tin cậy và lấy con người làm trung tâm, thể hiện ở việc thành lập Ủy ban Đạo đức công nghệ quốc gia (2019); ban hành chính sách hỗ trợ quản lý đạo đức công nghệ; xây dựng các quy tắc bảo vệ sở hữu AI; Ủy ban Quản lý Tiêu chuẩn hóa Quốc gia đã thành lập Nhóm tổng hợp chuẩn hóa AI quốc gia; Ủy ban kỹ thuật chuẩn hóa CNTT quốc gia đã thành lập một tiểu ban kỹ thuật về AI và ban hành “Hướng dẫn xây dựng Hệ thống tiêu chuẩn AI quốc gia thế hệ mới” nhằm mục đích thiết lập một hệ thống quản lý thông qua các tiêu chuẩn an toàn/đạo đức và thúc đẩy sự phát triển lành mạnh và bền vững của AI. Tháng 9/2024, Trung Quốc đã ban hành Khung quản lý an toàn AI, theo đó hướng dẫn phân loại, xác định các rủi ro và hướng dẫn đảm bảo an toàn cho các bên liên quan (người phát triển mô hình/thuật toán, nhà cung cấp dịch vụ, các lĩnh vực ứng dụng và người sử dụng).

Quản lý phát triển trí tuệ nhân tạo ở Việt Nam

Ở Việt Nam, AI ngày càng phát triển nhanh chóng và đang có những tác động không nhỏ đối với các ngành kinh tế - xã hội. Bên cạnh những lợi ích to lớn, sự phát triển của AI cũng đã và đang làm dấy lên những quan ngại sâu sắc về các rủi ro tiềm ẩn từ khía cạnh đạo đức, xã hội, pháp lý.

Thứ nhất, về chính sách thúc đẩy nghiên cứu, ứng dụng. Công nghệ nói chung, trong đó có công nghệ AI, cơ bản đã được chỉ ra tại các văn bản của Đảng, Chính phủ hướng đến việc khai thác công nghệ hiện đại/tiên tiến, trong đó có AI. Cụ thể như Nghị quyết số 50/NQ-CP ngày 17/04/2020 của Chính phủ ban hành Chương trình hành động của Chính phủ thực hiện Nghị quyết số 52-NQ/TW ngày 27/09/2019 của Bộ Chính trị về một số chủ trương, chính sách chủ động tham gia cuộc Cách mạng công nghiệp lần thứ tư xác định AI là ngành công nghệ ưu tiên, cần ưu tiên nguồn lực triển khai nghiên cứu, ứng dụng; Quyết định số 749/QĐ-TTg ngày 03/06/2020 của Thủ tướng Chính phủ phê duyệt “Chương trình Chuyển đổi số quốc gia đến năm 2025, định hướng đến năm 2030” xác định AI là công nghệ cốt lõi mà Việt Nam có thể đi tắt đón đầu, có khả năng tạo bứt phá; xây dựng hệ thống điện toán có năng lực đủ mạnh để xử lý, phân tích dữ liệu, huy động được sự tham gia của cộng đồng, cho phép các tổ chức, doanh nghiệp cùng khai thác phục vụ phát triển hệ sinh thái sản phẩm sáng tạo ứng dụng AI; ứng dụng AI trong hành chính công; đào tạo nhân lực AI... Cụ thể hơn, ngày 26/01/2021, Thủ tướng Chính phủ đã ban hành “Chiến lược quốc gia về nghiên cứu, phát

triển và ứng dụng trí tuệ nhân tạo đến năm 2030” tại Quyết định số 127/QĐ-TTg. Trong đó, xác định trụ cột xây dựng hệ thống văn bản quy phạm pháp luật và hành lang pháp lý liên quan đến AI nhằm thúc đẩy nghiên cứu, phát triển và ứng dụng AI vào cuộc sống; đồng thời quản lý, kiểm soát để tránh lạm dụng công nghệ và xâm phạm quyền, lợi ích hợp pháp của tổ chức, cá nhân...

Thứ hai, về việc phát triển và ứng dụng AI một cách có trách nhiệm, trong phạm vi quản lý, Bộ Khoa học và Công nghệ đã ban hành Quyết định số 1290/QĐ-BKHCN ngày 11/06/2024 về việc hướng dẫn một số nguyên tắc về nghiên cứu, phát triển các hệ thống AI có trách nhiệm; trong đó, nêu ra một số nguyên tắc chung cần chú ý trong nghiên cứu, phát triển các hệ thống AI có trách nhiệm và khuyến nghị tự nguyện tham khảo, áp dụng trong quá trình nghiên cứu, thiết kế, phát triển, cung cấp các hệ thống AI. Theo đó, các cơ quan, tổ chức KH&CN, tổ chức, doanh nghiệp, cá nhân có hoạt động nghiên cứu, thiết kế, phát triển, cung cấp các hệ thống AI được khuyến khích áp dụng 9 nguyên tắc nghiên cứu, phát triển các hệ thống AI có trách nhiệm và hướng dẫn thực hiện bao gồm: tinh thần hợp tác, thúc đẩy đổi mới sáng tạo; tính minh bạch; khả năng kiểm soát; an toàn; bảo mật; quyền riêng tư; tôn trọng quyền và phẩm giá con người; hỗ trợ người dùng; trách nhiệm giải trình.

Thứ ba, về mặt chuẩn hóa. Hiện nay, Bộ Khoa học và Công nghệ đã phối hợp với Bộ Thông tin và Truyền thông xây dựng, công bố một số tiêu chuẩn quốc gia về AI như TCVN 13902:2023 dựa trên ISO/IEC 22989:2022 (về khái niệm, thuật ngữ)...; TCVN 13903:2023 dựa trên ISO/IEC TR 24028:2020 (tổng quan về tính tin cậy trong AI)... Ngoài ra, Bộ Thông tin và Truyền thông cũng đã xây dựng một số dự thảo tiêu chuẩn về AI dựa trên các tiêu chuẩn quốc tế. Trong thời gian tới, Bộ Khoa học và Công nghệ sẽ phối hợp với các bộ, ngành liên quan xây dựng và công bố các tiêu chuẩn liên quan khác về AI.

Thứ tư, ở góc độ nghiên cứu khoa học. Trên cơ sở hỗ trợ của Chương trình Aus4Innovation, Trường Đại học Luật (Đại học Quốc gia Hà Nội) cũng triển khai dự án nghiên cứu xây dựng bộ nguyên tắc và một số hướng dẫn cho phát triển AI có trách nhiệm ở Việt Nam, nhằm tăng cường đối thoại chính sách với các cơ quan hoạch định chính sách, cộng đồng doanh nghiệp... Cho đến nay, dự án đã nghiên cứu, khảo sát việc sử dụng và quản lý các ứng dụng/hệ thống AI trong một số ngành như tư pháp, y tế và nông nghiệp... từ đó định hướng khuyến nghị việc quản lý, đảm bảo các yêu cầu về đạo đức đối với các ứng dụng AI được phát triển, sử dụng trong các ngành, lĩnh vực ở Việt Nam.



Định hướng thúc đẩy nghiên cứu phát triển và sử dụng trí tuệ nhân tạo một cách có trách nhiệm

Sau một thời gian triển khai thực hiện Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng AI đến năm 2030 (theo Quyết định số 127/QĐ-TTg ngày 26/01/2021), qua đánh giá sơ bộ có thể thấy, cần phải nhanh chóng nghiên cứu, xây dựng bộ khung pháp lý để bảo đảm phát triển và sử dụng AI một cách có trách nhiệm, an toàn. Trong đó, cần tham khảo cách mà các nước trên thế giới đang triển khai và bảo đảm các quy định mang tính hội nhập quốc tế cao, có thể khai thác tối đa các ưu thế của AI, đồng thời giảm thiểu thấp nhất các mặt trái của nó. Đặc biệt là ngăn ngừa các hành vi lạm dụng công nghệ này cho những mục đích xấu, có hại cho xã hội và vi phạm pháp luật.

Thứ nhất, về mặt chính sách. Qua nghiên cứu kinh nghiệm của một số quốc gia, có thể thấy cần quan tâm xây dựng các định hướng, quy định quản lý theo hướng quản trị rủi ro, có nghĩa là xác định trước một số rủi ro mức độ cao, từ đó có biện pháp quản lý phù hợp. Ở giai đoạn hiện nay, cần quan tâm: i) Nghiên cứu, xem xét các mối quan hệ về mặt pháp lý liên quan đến AI như quan hệ về tài sản, quyền sở hữu, sở hữu trí tuệ, quan hệ lao động, bồi thường thiệt hại... để điều chỉnh về mặt pháp luật cho phù hợp (ví dụ: quyền sở hữu các sản phẩm tạo ra bởi các ứng dụng AI); ii) Đề khai thác, ứng dụng AI phục vụ phát triển kinh tế - xã hội theo xu hướng chung trên thế giới, vẫn cần các biện pháp, chính sách để thúc đẩy nghiên cứu, phát triển; trong đó, rất cần các cơ chế mới, đặc thù để tạo thuận lợi cho các doanh nghiệp nghiên cứu, phát triển và thử nghiệm các sản phẩm, ứng dụng từ công nghệ AI (bao gồm các sandbox); iii) Quan tâm, phát triển các cơ sở phục vụ khai thác, ứng dụng AI ở Việt Nam như xây dựng các cơ sở dữ liệu quốc gia dùng chung, các hạ tầng tính toán, thử nghiệm... làm cơ sở để phát triển các ứng dụng AI nhưng cũng để đảm bảo nguồn dữ liệu an toàn, phù hợp với các quy định quản lý dữ liệu tương ứng; iv) Nghiên cứu, xây dựng các quy định/hướng dẫn để quản lý, phát triển các sản phẩm/ứng dụng AI về nội dung dữ liệu (tính trung thực); các vấn đề kỹ thuật (thuật toán sử dụng, việc đánh giá an toàn thông tin, xác thực); tính minh bạch (đáp ứng các hướng dẫn, tiêu chuẩn liên quan)... nhằm đảm bảo sự an toàn và quyền lợi của người sử dụng... Do tính đặc thù chuyên ngành, các bộ/ngành có thể nghiên cứu, xây dựng các hướng dẫn để phát triển và sử dụng các sản phẩm/dịch vụ AI trong ngành/lĩnh vực quản lý.

Thứ hai, việc nghiên cứu, phát triển các ứng dụng AI có tính chất hệ thống là cần thiết. Do vậy, cần được đầu tư thông qua nhiệm vụ nghiên cứu các cấp. Về

nội dung, có thể bao gồm cả việc nghiên cứu, dự báo các cơ hội, thách thức về công nghệ, pháp lý; lý luận, thực tiễn và đề ra những giải pháp giải quyết thách thức, đặt ra các nghiên cứu xã hội học về những quy chuẩn đạo đức, pháp lý trong xây dựng và ứng dụng AI trong đời sống xã hội; đánh giá tác động của xã hội (đến từng cá nhân) khi triển khai ứng dụng AI trong cuộc sống...

Thứ ba, về chuẩn hóa. Các chuyên gia đều khuyến nghị xây dựng tiêu chuẩn, xem xét mức độ rủi ro ứng dụng công nghệ và hướng tới lợi ích người dùng. Các tiêu chuẩn có mục đích là định hướng ứng dụng công nghệ, nhưng đồng thời cũng là công cụ để quản lý, phát triển và ứng dụng công nghệ một cách hệ thống, đảm bảo trật tự, quyền lợi và an toàn cho các bên liên quan đến sản phẩm, dịch vụ. Mỗi ngành có thể cần các tiêu chuẩn khác nhau do tính đặc thù chuyên ngành. Việc định hướng, xác định đối tượng chuẩn hóa và đặt ra lộ trình xây dựng, áp dụng các tiêu chuẩn/quy chuẩn kỹ thuật vào thời điểm hiện nay là cần thiết và phù hợp; nhất là khi đã có thể tham khảo kinh nghiệm của các tổ chức tiêu chuẩn hóa quốc tế, khu vực cũng như các quốc gia có điều kiện tương đồng với Việt Nam (đặc biệt là Trung Quốc).

Thứ tư, về hướng dẫn, phổ biến thông tin cho các bên liên quan (từ nghiên cứu phát triển, đến cung cấp dịch vụ và khai thác sử dụng) luôn cần thiết nhằm đảm bảo nhận thức về các lợi ích và rủi ro, khai thác và sử dụng các ứng dụng AI một cách có trách nhiệm. Vì vậy, ở góc độ quản lý ngành/lĩnh vực, các bộ/ngành có thể thực hiện tuyên truyền hay triển khai các chương trình đào tạo về sản phẩm/dịch vụ AI trong lĩnh vực quản lý...

*

* *

Để thúc đẩy nghiên cứu phát triển và sử dụng AI một cách có trách nhiệm ở Việt Nam, các cơ quan, đơn vị cần tiếp tục quan tâm triển khai các nhiệm vụ được giao tại Quyết định số 127/QĐ-TTg ngày 26/01/2021 của Thủ tướng Chính phủ. Theo đó, tùy theo chức năng, nhiệm vụ, thực hiện xây dựng và ban hành các quy tắc ứng xử/quy chuẩn đạo đức trong việc phát triển, ứng dụng AI trong ngành/lĩnh vực; nghiên cứu, xây dựng các tiêu chuẩn, quy chuẩn kỹ thuật về công nghệ và sản phẩm AI; tiếp tục thúc đẩy các hoạt động nghiên cứu và ứng dụng công nghệ AI trong các ngành, lĩnh vực.