BẢN TIN
CHIẾN LƯỢC PHÁT TRIỂN

KHOA HỌC CÔNG NGHỆ KINH TẾ

Số 6
2024

(BẢN TIN CHỌN LỌC PHỤC VỤ LÃNH ĐẠO)

HOA KỲ CẬP NHẬT CHIẾN LƯỢC NGHIÊN CỨU VÀ PHÁT TRIỂN TRÍ TUỆ NHÂN TẠO QUỐC GIA

NASATI
BỘ KHOA HỌC VÀ CÔNG NGHỆ
CỤC THÔNG TIN KHOA HỌC VÀ CÔNG NGHỆ QUỐC GIA

CỤC THÔNG TIN KHOA HỌC VÀ CÔNG NGHỆ QUỐC GIA

Địa chỉ: 24, Lý Thường Kiệt, Hoàn Kiếm, Hà Nội.

Tel: (024)38262718, Fax: (024)39349127

BAN BIÊN TẬP

TS. Trần Đắc Hiến (*Trưởng ban*);

ThS. Nguyễn Lê Hằng; ThS. Phùng Anh Tiến.

MỤC LỤC

HOA KỲ CẬP NHẬT CHIẾN LƯỢC
NGHIÊN CỨU VÀ PHÁT TRIỂN TRÍ TUỆ NHÂN TẠO QUỐC GIA

HOA KỲ CẬP NHẬT CHIẾN LƯỢC

NGHIÊN CỨU VÀ PHÁT TRIỂN TRÍ TUỆ NHÂN TẠO QUỐC GIA

Giới thiệu

Chiến lược nghiên cứu và phát triển trí tuệ nhân tạo (AI) quốc gia của Hoa Kỳ, công bố lần đầu năm 2016 và được cập nhật năm 2019 và cuối năm 2023, sau khi tiếp thu các đánh giá của cơ quan quản lý và phản hồi từ cộng đồng chuyên gia. Bản cập nhật 2023 bao gồm 9 nội dung chiến lược, bổ sung thêm một số nội dung mới, như Chiến lược 9 về hợp tác quốc tế trong nghiên cứu AI.

Trong Chiến lược mới cập nhật, Hoa Kỳ sẽ tăng cường đầu tư nghiên cứu lớn và dài hạn vào AI, nghiên cứu AI cơ bản, quản lý rủi ro, an ninh mạng, và phát triển phần cứng AI, tuân thủ nguyên tắc khoa học, để tối đa hóa lợi ích xã hội và kinh tế của AI phù hợp với các giá trị của Hoa Kỳ. Chiến lược cũng xác định những thách thức nghiên cứu chính trong AI để phối hợp và tập trung các khoản đầu tư cho nghiên cứu và phát triển (R&D) của liên bang. Chính phủ liên bang luôn đặt con người và cộng đồng vào vị trí trung tâm bằng cách đầu tư vào R&D có trách nhiệm phục vụ lợi ích công cộng, bảo vệ quyền lợi, sự an toàn của mọi người và thúc đẩy các giá trị dân chủ. Chiến lược kêu gọi thiết lập và mở rộng quan hệ đối tác công - tư, phát triển phương pháp hợp tác hiệu quả giữa con người và AI; tăng thêm kinh phí cho một số cơ quan chủ chốt trong lĩnh vực khoa học và AI, trong đó có Quỹ Khoa học Quốc gia, Viện Tiêu chuẩn và Công nghệ Quốc gia.

Ngoài ra, Chiến lược mới cập nhật còn đề cập đến việc chuẩn bị lực lượng lao động hiện tại và tương lai của Hoa Kỳ để tích hợp hệ thống AI vào tất cả các lĩnh vực và phối hợp các hoạt động AI đang diễn ra trong tất cả các cơ quan liên bang. Đề cập đến tương lai của công việc, Chiến lược cũng thừa nhận tầm quan trọng của việc hiểu được nhu cầu về lực lượng lao động AI để lấp đầy khoảng cách chênh lệch, phát triển các chương trình đào tạo và giáo dục nhằm đáp ứng nhu cầu ngày càng tăng đối với các chuyên gia AI. Việc tăng cường năng lực AI tại Hoa Kỳ sẽ thúc đẩy khả năng cạnh tranh quốc gia, giúp Hoa Kỳ thu hút và giữ chân các doanh nghiệp và nhân tài trên nhiều cấp độ của chuỗi cung ứng AI. Cuối cùng, để tăng cường vai trò lãnh đạo toàn cầu của Hoa Kỳ về AI, Chiến lược nhấn mạnh tầm quan trọng của hợp tác quốc tế trong lĩnh vực AI và cam kết Hoa Kỳ sẽ hợp tác toàn cầu rộng rãi hơn.

Chiến lược 1: Đầu tư dài hạn vào nghiên cứu cơ bản trí tuệ nhân tạo và phát triển trí tuệ nhân tạo có trách nhiệm

Hoa Kỳ duy trì vị thế dẫn đầu trong lĩnh vực AI nhờ đầu tư liên tục và nhất quán vào nghiên cứu AI cơ bản lâu dài. Nhiều sản phẩm và dịch vụ AI ngày nay bắt nguồn từ nghiên cứu cơ bản được tài trợ bởi liên bang từ nhiều thập kỷ trước. Kể từ khi ban hành Chiến lược năm 2019, đầu tư vào R&D AI của Chính phủ liên bang đã tăng đáng kể, dẫn đến các phát hiện đột phá trong nhiều lĩnh vực. Hoa Kỳ cần tiếp tục thúc đẩy nghiên cứu AI cơ bản lâu dài và có trách nhiệm để đạt được những khám phá và đổi mới mang lại lợi ích lâu dài. Đầu tư vào R&D AI bao gồm từ nghiên cứu cơ bản đến nghiên cứu ứng dụng, với mục tiêu cải thiện toàn xã hội và tôn trọng quyền tự do cá nhân.

Đặc biệt quan trọng là đầu tư vào việc phát triển các hệ thống AI bảo đảm tin cậy, cần thiết cho các ứng dụng liên quan đến an toàn và quyết định của AI ảnh hưởng đến cá nhân, nhóm, cộng đồng và môi trường. Nghiên cứu AI chủ yếu tập trung vào các nhiệm vụ cá nhân, nhưng cũng cần giải quyết các thách thức khoa học và công nghệ đa lĩnh vực, hướng tới AI đa năng. Chiến lược này ưu tiên phát triển các hệ thống AI có khả năng mở rộng và đa năng, cải thiện khả năng nhận thức và hành động của AI, và phát triển hệ thống làm việc trong môi trường thực và ảo. Cuối cùng, việc hiểu lý thuyết về khả năng và giới hạn của AI là quan trọng để định hướng R&D và bảo đảm sử dụng AI an toàn.

Nội dung Chiến lược 1 gồm 10 định hướng sau đây:

(1) Phát triển các phương pháp luận tiếp cận tập trung vào dữ liệu để khám phá tri thức

Chiến lược nhấn mạnh nhu cầu phát triển các công cụ và công nghệ mới để hiểu dữ liệu và khám phá tri thức từ dữ liệu lớn, như đã đề cập trong Chiến lược Nghiên cứu và Phát triển Dữ liệu Lớn Liên bang 2016. Nó cũng nhấn mạnh việc cải tiến các hệ thống AI tiên tiến hơn để phát hiện thông tin hữu ích, giải quyết các vấn đề về tính xác thực và sự phù hợp của dữ liệu, cải thiện hiệu quả của các kỹ thuật làm sạch và gán nhãn dữ liệu, và xử lý các vấn đề về quyền riêng tư. Chiến lược cũng đề cập đến nhu cầu nghiên cứu thêm về khai thác dữ liệu và siêu dữ liệu, giải quyết vấn đề thiếu dữ liệu đại diện trong các lĩnh vực nhạy cảm như y tế, và phát triển học máy (ML) đa phương thức để xử lý dữ liệu không đồng nhất; tầm quan trọng của việc có một cơ sở hạ tầng chuẩn để mã hóa tri thức cho các hệ thống AI.

(2) Thúc đẩy các cách tiếp cận học máy liên kết (Federated ML)

Các phương pháp học liên kết mới sẽ quan trọng trong thế giới ngày càng kết nối và lo ngại về quyền riêng tư, bảo mật dữ liệu. Học liên kết cho phép nhiều thiết bị cộng tác xây dựng mô hình ML toàn cầu dựa trên dữ liệu cục bộ. Quá trình này bao gồm các thiết bị huấn luyện mô hình cục bộ và chia sẻ cập nhật mô hình để cải thiện mô hình toàn cầu, không chia sẻ dữ liệu thô. Mô hình toàn cầu sau đó được phân phối lại cho các thiết bị để tiếp tục huấn luyện cho đến khi đạt độ chính xác yêu cầu. Học liên kết

cải thiện độ chính xác và công bằng của mô hình ML bằng cách bao gồm dữ liệu bảo vệ cục bộ từ nhiều nguồn khác nhau, quan trọng đối với các ngành như y tế, tài chính, và viễn thông. Tuy nhiên, có nhiều thách thức trong việc xử lý đặc điểm không đồng nhất của thiết bị và dữ liệu, cải thiện hiệu quả giao tiếp và cập nhật mô hình ML, cũng như tăng cường bảo vệ và bảo mật dữ liệu.

(3) Hiểu về khả năng và giới hạn lý thuyết của AI

Việc hiểu rõ khả năng và giới hạn lý thuyết của AI là cần thiết để phát triển các hệ thống đa năng và bảo đảm sử dụng AI an toàn và có trách nhiệm. Hiện tại, vẫn chưa có sự hiểu biết rõ ràng về khả năng và giới hạn lý thuyết của AI, đặc biệt là với các kỹ thuật AI tạo sinh. Cần nghiên cứu lý thuyết để hiểu cách các kỹ thuật này hoạt động và các tính chất mới xuất hiện. Dù nhiều ngành (toán học, khoa học điều khiển, khoa học máy tính) đang nghiên cứu vấn đề này, nhưng hiện tại vẫn thiếu các mô hình lý thuyết thống nhất để hiểu hiệu suất của hệ thống AI. Cần thêm nghiên cứu về khả năng giải quyết tính toán, để xác định các lớp vấn đề mà AI có thể hoặc không thể giải quyết. Việc hiểu những vấn đề không thể giải quyết về lý thuyết có thể dẫn đến phát triển các giải pháp hoặc mở ra nghiên cứu mới về phần cứng cho hệ thống AI trong tương lai.

(4) Nghiên cứu về các hệ thống AI đa năng có thể mở rộng quy mô

Phát triển của các mô hình cơ sở trong lĩnh vực AI, được huấn luyện trên dữ liệu lớn không được gán nhãn và có thể được áp dụng trong nhiều lĩnh vực khác nhau như pháp luật, chăm sóc sức khỏe và khoa học. Các ví dụ về các mô hình ngôn ngữ được huấn luyện bao gồm BERT, GPT-4, và các hệ thống AI khác. Tuy nhiên, các mô hình này phải đối mặt với các thách thức như sự sản sinh không mong muốn và thiên vị từ dữ liệu huấn luyện, cũng như cần nghiên cứu thêm về tính hợp lệ, đáng tin cậy, bảo mật của chúng. Cần tiếp tục nghiên cứu để phát triển các kỹ thuật giải thích và hiểu kết quả của mô hình, đồng thời cần xem xét và thiết kế các biện pháp bảo vệ phù hợp cho các hệ thống này.

(5) Phát triển hệ thống AI và mô phỏng trên môi trường thực và ảo

Một xu hướng mới nổi trong mô hình hóa và mô phỏng là sự phát triển của "bản sao kỹ thuật số" (digital twins). Bản sao kỹ thuật số là một biểu diễn hoặc mô hình ảo đóng vai trò là đối tác kỹ thuật số thời gian thực của một đối tượng hoặc quy trình vật lý. Các ứng dụng trong thế giới thực bao gồm bảo trì dự đoán động cơ máy bay, quy hoạch đô thị và quản lý thành phố thông minh, và sản xuất 3D. Một yêu cầu chính là hệ thống vật lý phải được trang bị để dữ liệu thu thập được có thể tương tác với mô hình kỹ thuật số hoặc mô hình tính toán của chính nó. Phương pháp bản sao kỹ thuật số cho phép tự động hóa thông minh các hệ thống vật lý trên môi trường thực và ảo. Những thách thức cụ thể đối với nhiều ứng dụng khác nhau, chẳng hạn như tính chất lượng dữ liệu, độ trễ và quyền riêng tư, cũng như độ chính xác khác nhau mà các hiện tượng khác nhau có thể được mô hình hóa.

(6) Cải thiện khả năng nhận thức của hệ thống AI

Nhận thức là khả năng của hệ thống thông minh nhìn thấy và hiểu thế giới xung quanh. Nó bắt đầu từ dữ liệu cảm biến và kết hợp thông tin để đưa ra nhận thức về môi trường và tình hình. Để cải thiện, cần phát triển cả phần cứng và thuật toán để cảm biến có thể thu thập dữ liệu ở xa và trong thời gian thực. Hệ thống nhận thức cần tích hợp dữ liệu từ nhiều nguồn và cần giải quyết các thách thức như nhận diện đối tượng trong điều kiện khó khăn và xử lý quyền riêng tư. Đồng thời, nhận thức về con người cũng cần được cải thiện để tăng cường tương tác hiệu quả giữa con người và hệ thống AI.

(7) Phát triển các robot có năng lực và đáng tin cậy hơn

Robot, một lĩnh vực chủ chốt trong AI, tập trung vào nhận thức, thao tác vật lý và điều hướng. Các tiến bộ gần đây trong công nghệ robot đã tạo ra ảnh hưởng đáng kể trên nhiều lĩnh vực bao gồm sản xuất, logistics, chăm sóc sức khỏe, quốc phòng, nông nghiệp và sản phẩm tiêu dùng. Đáng chú ý, các robot được điều khiển bởi AI đang biến đổi nghiên cứu với sự xuất hiện của các phòng thí nghiệm tự động, đổi mới trong thiết kế thuốc và tổng hợp vật liệu. Hợp tác giữa robot và con người ngày càng phổ biến. Tuy nhiên, vẫn còn thách thức trong việc làm cho các hệ thống robot trở nên mạnh mẽ, đáng tin cậy, dễ sử dụng và an toàn. Cải tiến cần được thực hiện trong nhận thức, tri thức, tính thích ứng, di động, thao tác, hợp tác và an toàn. Nghiên cứu cũng cần được thực hiện để đối phó với các hệ thống đối đầu và bảo đảm sự tương tác mượt mà với con người trong môi trường đa dạng. Tóm lại, tạo tiến bộ cho các hệ thống robot đòi hỏi cải tiến về khả năng, tính khả dụng và an toàn.

(8) Phần cứng tiên tiến để cải thiện AI

Nghiên cứu AI đang phụ thuộc nhiều vào tiến bộ phần cứng, như các chip GPU và bộ gia tốc, cùng với các công nghệ cải tiến như bộ nhớ, tốc độ xử lý, và tiết kiệm năng lượng. Phát triển phần cứng tối ưu hóa cho AI có thể cung cấp hiệu suất cao hơn, như các chip được thiết kế để hoạt động giống như các mạng nơ -ron trong não bộ của con người (chip neuromorphic). Tiến bộ trong thiết bị lưu trữ cũng hỗ trợ triển khai AI. Nghiên cứu tiếp tục cần được thực hiện để cải thiện khả năng học từ dữ liệu tốc độ cao và phát triển phương pháp phản hồi thông minh hơn.

(9) Tạo AI để cải thiện phần cứng

Cải thiện phần cứng có thể làm cho AI mạnh mẽ hơn và ngược lại, AI cũng có thể cải thiện hiệu suất và sử dụng tài nguyên của phần cứng. Sự tương tác này sẽ dẫn đến các tiến bộ tiếp theo trong hiệu suất phần cứng. AI có thể được sử dụng để dự đoán hiệu suất và tài nguyên, tối ưu hóa trực tuyến và tạo ra các hệ thống tự điều chỉnh. Cải tiến trong thuật toán AI có thể giảm di chuyển dữ liệu giữa các bộ xử lý và bộ nhớ, cũng như tăng cường hiệu suất hệ thống.

(10) Áp dụng AI và Hệ thống máy tính bền vững

Cần chú ý đến chi phí tính toán ngày càng tăng của việc phát triển và vận hành các hệ thống AI tiên tiến. Sự phổ biến của AI dựa vào dữ liệu sẽ tăng đáng kể nhu cầu tính toán và ảnh hưởng môi trường. Do vậy cần phải thiết kế các thuật toán, hệ thống và ứng

dụng AI nhận thức tài nguyên và coi trọng các khái niệm bền vững hơn chỉ là tiêu thụ năng lượng.

Chiến lược 2: Phát triển các phương pháp hiệu quả cho sự hợp tác giữa con người và trí tuệ nhân tạo

Việc phát triển phương pháp hiệu quả cho sự hợp tác giữa con người và AI ngày càng trở nên quan trọng, đặc biệt khi AI ngày càng phổ biến trong xã hội. Trong khi các hệ thống hoàn toàn tự động vẫn quan trọng trong nhiều bối cảnh, sự kết hợp giữa con người và AI có thể hiệu quả nhất trong các ứng dụng khác nhau, từ khắc phục thiên tai đến khám phá khoa học. Chiến lược này nhấn mạnh sự cần thiết của nghiên cứu đa ngành để tạo ra một môi trường hợp tác hiệu quả giữa con người và AI.

(1) Phát triển khoa học về sự hợp tác giữa con người và AI

Cần nghiên cứu để hiểu về mặt con người trong tương tác giữa con người và máy. Nghiên cứu này sẽ giúp hiểu các yếu tố và yêu cầu của các nhóm con người-máy hiệu quả để thực hiện nhiệm vụ một cách hiệu quả. Cần xem xét các khả năng bổ sung mà máy cần có để trở thành một đồng đội hiệu quả cho các nhiệm vụ và môi trường tương ứng, bao gồm mô hình hóa tương tác con người. Các vai trò chức năng của hệ thống AI trong bối cảnh làm việc nhóm đã được xác định: thực hiện các chức năng cùng với con người, thực hiện chức năng khi con người gặp quá tải nhận thức và thực hiện chức năng thay thế cho con người. Để trở thành đồng đội thực sự, máy cần linh hoạt và thích ứng với trạng thái của đồng đội con người, cũng như môi trường, để đoán trước thông minh khả năng và ý định của đồng đội con người, và tổng hợp các trải nghiệm học hỏi cụ thể vào các tình huống hoàn toàn mới. Đây đều là thách thức nghiên cứu. Các câu hỏi khác còn mở liên quan đến hợp tác giữa con người và AI bao gồm việc xác định thành phần nhóm, quản lý nhận thức tình hình và các mô hình tương tác điều khiển hệ thống AI.

(2) Tìm kiếm các mô hình tiên tiến và số liệu về hiệu suất

Một phương pháp truyền thống để xây dựng các đội ngũ con người-AI hiệu quả là xem xét các khả năng của cả con người và hệ thống AI một cách riêng lẻ, sau đó nghiên cứu cách kết hợp chúng một cách tối ưu. Tuy nhiên, việc mô hình hóa sự hợp tác này đầy thách thức, đặc biệt là khi cần tính đến những sự kiện không mong đợi và các vấn đề như niềm tin và ảnh hưởng của con người và AI. Để hiểu rõ hơn về hiệu quả của các đội ngũ con người-AI, cần phải có nhiều nghiên cứu về lý thuyết, mô hình, dữ liệu và công cụ tính toán.

(3) Nuôi dưỡng lòng tin trong tương tác giữa con người và AI

Tính không rõ ràng của quá trình lập trình và ra quyết định trong các hệ thống AI có thể làm giảm sự tin cậy cần thiết cho việc hợp tác hiệu quả giữa con người và AI. Một thách thức quan trọng là con người thường mong đợi các hệ thống tự động hoạt động theo cách xác định, nhưng các hệ thống AI có thể hoạt động không xác định, đặc biệt khi phản ứng với thông tin thế giới thực không hoàn hảo. Cần nghiên cứu để thiết

lập và duy trì mức độ tin cậy phù hợp giữa con người và AI trong các điều kiện không chắc chắn.

(4) Theo đuổi sự hiểu biết sâu sắc hơn về hệ thống AI của con người

Để tăng sự tin cậy và thành công của hợp tác giữa con người và AI, việc học từ những thất bại và phát triển phương pháp kiểm thử hiệu quả là rất quan trọng. Cần có các "bộ ghi" để ghi lại dữ liệu và chẩn đoán các thất bại, đặc biệt là trong việc làm việc nhóm giữa con người và AI. Tiếp tục nghiên cứu trong môi trường ảo và phát triển các phương pháp kiểm thử là các bước tiếp theo quan trọng.

(5) Phát triển các mô hình mới cho tương tác và cộng tác AI

Nghiên cứu về tính sử dụng và thiết kế tập trung vào con người đã chứng minh vai trò quan trọng của cơ chế tương tác trong việc cải thiện hiệu suất người dùng. Tương tự, cần có nghiên cứu để hiểu về tính sử dụng và ảnh hưởng của thiết kế tương tác trong hợp tác con người-AI. Nghiên cứu này sẽ tập trung vào ảnh hưởng của thiết kế tương tác đối với quyết định, sự giữ kỹ năng, đào tạo, sự hài lòng với công việc, và tổng thể hiệu suất và sự đàn hồi của đội ngũ con người-AI. Ngoài ra, cần phát triển các mô hình mới cho tương tác con người-AI để tạo điều kiện cho sự hợp tác, quyết định, và sự kiểm soát của con người. Điều thách thức là truyền đủ thông tin cho người dùng mà không gây quá tải tư duy, và giúp máy và người dùng hiểu khi nào nên chuyển quyền kiểm soát qua lại. Cuối cùng, cần có thêm nghiên cứu với người dùng cuối để bảo đảm sự hiểu biết và áp dụng hiệu quả của các ứng dụng hợp tác con người-AI.

Chiến lược 3: Hiểu và giải quyết các vấn đề về đạo đức, pháp lý và xã hội của trí tuệ nhân tạo

AI mang lại cơ hội lớn nhưng cũng đặt ra rủi ro có thể ảnh hưởng tiêu cực. Rủi ro này có thể là dài hạn hoặc ngắn hạn, xác suất cao hoặc thấp, và có ảnh hưởng hệ thống hoặc cục bộ. Nếu không có biện pháp kiểm soát, các hệ thống AI có thể làm tăng cường, duy trì hoặc làm trầm trọng hóa kết quả không công bằng hoặc không mong muốn. Chính phủ Hoa Kỳ đã đưa ra một bản kế hoạch bảo vệ quyền lợi khi sử dụng AI. Cần có nghiên cứu để hiểu và giảm thiểu các rủi ro đạo đức và xã hội của AI, cũng như sử dụng AI để giải quyết các vấn đề đạo đức, pháp lý và xã hội.

(1) Đầu tư vào nghiên cứu cơ bản để thúc đẩy các giá trị cốt lõi thông qua thiết kế hệ thống xã hội kỹ thuật và về các hàm ý đạo đức, pháp lý và xã hội của AI

Cần nghiên cứu cơ bản để thiết kế các hệ thống AI phù hợp với giá trị và hiểu rõ các hậu quả đạo đức, pháp lý và xã hội của AI. Cần nghiên cứu về giao tiếp, tâm lý học và các phương pháp đo lường và giảm thiểu các rủi ro liên quan đến AI để xây dựng các hệ thống an toàn, công bằng và đáng tin cậy.

(2) Hiểu và giảm thiểu rủi ro xã hội và đạo đức của AI

Cần nghiên cứu ngay lập tức để xác định cấu trúc quản trị AI hiệu quả có thể giảm thiểu các rủi ro, xây dựng hệ thống và triển khai AI đáng tin cậy và tạo niềm tin công

cộng phù hợp thông qua sự tương tác hiệu quả. Một phương pháp có thể là nghiên cứu và thích nghi các phương tiện từ các lĩnh vực khác, như y học, có hệ thống quản trị và quy định mạnh mẽ. Cần nghiên cứu về các công cụ để xác định và giảm thiểu các thiên vị có hại trong dữ liệu, đặc biệt là trong dữ liệu huấn luyện mới. Cơ chế để phát triển, đánh giá và duy trì các hệ thống AI giảm thiểu rủi ro và tối đa hóa lợi ích là cần thiết.

(3) Sử dụng AI để giải quyết các vấn đề đạo đức, pháp lý và xã hội

Phát triển hệ thống AI một cách hợp lý, giảm thiểu thiên vị và hại đến quyền dân sự, tự do dân chủ, và lợi ích của những người bị ảnh hưởng bởi hệ thống có thể giúp giải quyết các thách thức xã hội phức tạp. AI có thể cung cấp dữ liệu để giải quyết các vấn đề liên quan đến công bằng, thích ứng và giảm nhẹ biến đổi khí hậu, việc làm và chăm sóc sức khỏe, đặc biệt là cho những người bị bỏ lại truyền thống. Đồng thời, cần phát triển và điều chỉnh các công cụ AI khác nhau để đối mặt với các thách thức trong các lĩnh vực khác nhau. Tuy nhiên, cần cẩn trọng với việc giải quyết các vấn đề này, vì giải pháp công nghệ có thể không phù hợp hoặc không hiệu quả trong một số trường hợp.

(4) Hiểu được tác động rộng hơn của AI

AI hứa hẹn mang lại những thay đổi to lớn cho xã hội. Mặc dù nhiều thay đổi này sẽ tích cực, nhưng có khả năng sẽ có những hậu quả tiêu cực, và những ảnh hưởng này cũng có thể phân phối không đồng đều. Nghiên cứu và phát triển trong các vấn đề đạo đức, pháp lý và xã hội của AI cần thiết để hiểu, dự đoán và giảm thiểu tổn thất cũng như hiểu rõ sự phân phối của những lợi ích có thể xảy ra.

Chiến lược 4: Bảo đảm tính an toàn và bảo mật của hệ thống trí tuệ nhân tạo

Mặc dù hệ thống AI hứa hẹn cải thiện hiệu suất trong nhiều ứng dụng khác nhau, nhưng độ phức tạp ngày càng tăng, công nghệ nền tảng phát triển nhanh chóng và nhu cầu dữ liệu đáng kể của chúng có thể dẫn đến rủi ro gia tăng khi triển khai. Do đó, việc nhấn mạnh đến tính an toàn và bảo mật của hệ thống AI ngày càng quan trọng, đòi hỏi một cách tiếp cận đa ngành.

Chiến lược này định nghĩa "an toàn" là giảm thiểu việc hệ thống gây ra tác hại mới và "bảo mật" là giám sát tính toàn vẹn của hệ thống. An toàn và bảo mật của AI cần được quan tâm xuyên suốt vòng đời của hệ thống, từ thiết kế ban đầu, xây dựng dữ liệu/mô hình đến xác minh, triển khai, vận hành và giám sát.

Các lĩnh vực nghiên cứu quan trọng bao gồm:

- Phát triển các phương pháp thử nghiệm phù hợp với nhu cầu ngày càng tăng của các hệ thống AI hiện đại và các hệ thống phức tạp.

- Cải thiện các phương pháp để bảo đảm tính bảo mật của hệ thống AI chống lại thao túng dữ liệu đầu vào, đảo ngược mô hình và các dạng tấn công đối nghịch khác.

- Đầu tư thêm vào phát triển tiêu chuẩn, hệ thống và nghiên cứu để xây dựng lòng tin vào hiệu suất của các hệ thống AI được triển khai.

Chiến lược này chia chương trình nghiên cứu và phát triển an toàn và bảo mật theo hai hướng chính: xây dựng AI an toàn và bảo mật AI.

(1) Xây dựng AI an toàn

Với sự phổ biến của AI, cần một phương pháp tiếp cận quốc gia về nghiên cứu an toàn AI. Điều này bao gồm phát triển các phương pháp tạo ra, đánh giá, triển khai và giám sát AI an toàn. Cần các giải pháp mở rộng cho các hệ thống AI lớn và phức tạp, và một hệ sinh thái đổi mới quốc gia để mọi người có thể tiếp cận và phân tích các mô hình AI quy mô lớn.

Nghiên cứu cần tập trung vào tương tác an toàn giữa con người và máy móc, sử dụng các phương pháp mới để xác định hành vi AI và phát triển các kỹ thuật như ngôn ngữ lập trình mới, xác minh chính thức và lập trình neurosymbolic (sử dụng các biểu tượng để đại diện cho kiến thức và thông tin, đồng thời sử dụng các mạng nơ-ron để học hỏi từ dữ liệu).

Cần giải quyết thách thức trong hệ thống phức hợp AI và phát triển phương pháp xác minh độc lập cho các hệ thống con để bảo đảm an toàn và bảo mật. Các cơ sở thử nghiệm mới có thể hỗ trợ nghiên cứu này.

Rủi ro lớn và dài hạn như AI đa dụng và các vấn đề xã hội, môi trường cần được nghiên cứu thêm để dự đoán và quản lý hiệu quả.

(2) Bảo mật AI

Nhu cầu bảo mật AI ngày càng tăng khi phần mềm và hệ thống trở nên phức tạp hơn, đồng thời tăng cường tính dễ tổn thương trước các mối đe dọa an ninh mạng. Điều này thể hiện qua nhu cầu đào tạo thêm cho các chuyên gia trong các cơ quan chính phủ và nhận thức về bảo mật AI như một lĩnh vực nghiên cứu độc lập liên quan đến an ninh mạng và AI.

Các vấn đề bảo mật AI bao gồm "data poisoning" (nhiễm độc dữ liệu), các cuộc tấn công đối kháng nhắm vào hệ thống AI, và việc thao túng dữ liệu âm thanh hoặc hình ảnh mà con người không thể nhận biết nhưng AI có thể xử lý sai lệch. Một số rủi ro này có thể được xác định qua "red teaming" (mô phỏng tấn công) và mô hình toán học. Cần nghiên cứu để cải thiện cả hai phương pháp này.

Chuỗi cung ứng phát triển AI hiện tại cũng là một mối đe dọa, do chỉ có một số công cụ được sử dụng cho việc phát triển và triển khai hệ thống AI, có nguy cơ bị thao túng. Nỗ lực bảo vệ những công cụ này và phát triển bộ công cụ mạnh mẽ hơn để bảo vệ chuỗi cung ứng phát triển AI là cần thiết.

Cần nghiên cứu các phương pháp cải thiện bảo mật hệ thống AI, bao gồm việc chống lại thao túng dữ liệu đầu vào, đảo ngược mô hình và các hình thức tấn công đối kháng khác.

Chiến lược 5: Phát triển dữ liệu và tạo môi trường chung cho đào tạo và kiểm tra trí tuệ nhân tạo

Sự tiến bộ trong AI phụ thuộc nhiều vào dữ liệu và tính toán. Cần có dữ liệu hợp pháp, có đạo đức và cơ sở hạ tầng mạng để hỗ trợ liên kết, quản lý dữ liệu và tái sản xuất. Truy cập vào các hệ thống tính toán tiên tiến như tính toán hiệu suất cao và tài nguyên đám mây giúp thúc đẩy đổi mới AI.

Tuy nhiên, việc tiếp cận dữ liệu và tài nguyên tính toán quy mô lớn vẫn là trở ngại. Nhiều nhà nghiên cứu chuyển sang công nghiệp do thiếu tài nguyên. Sự tập trung tài nguyên vào các công ty công nghệ lớn và đại học có nguồn lực tốt có thể làm lệch hướng nghiên cứu AI.

Lực lượng Đặc nhiệm về Nguồn tài nguyên nghiên cứu AI quốc gia (NAIRR) đã công bố lộ trình và kế hoạch triển khai cơ sở hạ tầng mạng quốc gia để nhà nghiên cứu tiếp cận với dữ liệu, tính toán và môi trường thử nghiệm. Các nỗ lực này dựa trên Nguyên tắc FAIR (tìm thấy được, truy cập được, tương thích được, và tái sử dụng được) và Đạo luật Dữ liệu Chính phủ Mở.

Chiến lược này có bốn định hướng chính: Phát triển và cung cấp dữ liệu cho ứng dụng AI đa dạng; Chia sẻ tài nguyên tính toán và phần cứng quy mô lớn; Đáp ứng nhu cầu thử nghiệm cho lợi ích công và thương mại; và Phát triển thư viện và bộ công cụ phần mềm mã nguồn mở.

(1) Phát triển và cung cấp dữ liệu đáp ứng nhu cầu của các ứng dụng AI đa dạng

Bảo đảm truy cập vào các bộ dữ liệu chất lượng là rất quan trọng để đạt được kết quả khoa học đáng tin cậy, có giá trị đạo đức và công bằng. Các bộ dữ liệu phải đủ đại diện để giải quyết các vấn đề thực tế và phải có tài liệu chi tiết về nguồn gốc dữ liệu và công việc liên quan trước đây. Hạ tầng kỹ thuật cần thiết để hỗ trợ nghiên cứu có thể đưa vào sản xuất là một thách thức quan trọng.

Nhiều ứng dụng học máy cần dữ liệu được tích hợp, làm sạch và tinh chỉnh để sử dụng. Hạ tầng dữ liệu cần thiết kế để đáp ứng yêu cầu cụ thể của các ứng dụng AI và cần được cập nhật liên tục dựa trên tiến bộ công nghệ và thay đổi trong nghiên cứu. Chính phủ cần cải thiện truy cập dữ liệu và tạo điều kiện sử dụng các phương pháp học máy và phân tích dữ liệu.

Các tổ chức chính phủ và phi chính phủ cần hợp tác để cung cấp nhiều dữ liệu hơn, bảo đảm tính bảo mật và quyền riêng tư. Việc tạo ra các tập dữ liệu tổng hợp có thể giúp khi dữ liệu thực không thể chia sẻ vì lý do quyền riêng tư. Đầu tư vào các phương pháp thu thập dữ liệu đại diện và hợp tác công-tư là cần thiết.

Cần nghiên cứu các phương pháp liên kết dữ liệu hợp lý, tạo điều kiện cho việc phát hiện và sử dụng dữ liệu. Việc bảo đảm sử dụng dữ liệu một cách phản ánh các giá trị của Hoa Kỳ là quan trọng, bao gồm bảo vệ thông tin cá nhân khi dữ liệu chính phủ được công khai và nghiên cứu về quản trị dữ liệu hiệu quả.

(2) Phát triển và chia sẻ tài nguyên tính toán và phần cứng quy mô lớn và chuyên biệt

Để thúc đẩy đổi mới trong lĩnh vực AI, việc cung cấp truy cập dễ dàng đến các tài nguyên tính toán tiên tiến là cần thiết. Điều này đặc biệt quan trọng đối với các nhà nghiên cứu và sinh viên tại các cơ sở giáo dục nhỏ và các doanh nghiệp khởi nghiệp. Lực lượng Đặc nhiệm NAIRR đã đề xuất một kế hoạch triển khai để phát triển các tài nguyên tính toán mới, bao gồm các hệ thống đám mây và trên điện toán hiệu suất cao, để phục vụ cộng đồng nghiên cứu và phát triển AI.

(3) Phát triển các nguồn tài nguyên phần cứng và máy tính tiên tiến chuyên biệt và quy mô lớn có thể được chia sẻ

Để thúc đẩy đổi mới AI, cần tiếp cận tài nguyên máy tính mạnh trong nghiên cứu AI. Các tổ chức lớn thường có sẵn tài nguyên này, nhưng nhiều trường học nhỏ, doanh nghiệp khởi nghiệp lại gặp khó khăn. Để khuyến khích nghiên cứu AI, cần cải thiện việc tiếp cận tài nguyên máy tính cho những nhóm này. Có thể tận dụng tài nguyên sẵn có, nâng cấp cơ sở hạ tầng nghiên cứu và hỗ trợ tài chính để giảm thiểu rào cản gia nhập lĩnh vực AI.

(4) Làm cho các nguồn tài nguyên thử nghiệm đáp ứng được lợi ích thương mại và công cộng

Để đáp ứng sự phức tạp ngày càng tăng của các hệ thống AI, cần phát triển các tài nguyên kiểm thử mạnh mẽ. Các tài nguyên kiểm thử liên bang, như các bộ kiểm thử và khung làm việc, có thể giải quyết những hạn chế của các phương pháp kiểm thử hiện có. Ví dụ, Chương trình Kiểm thử Nhận dạng Khuôn mặt của NIST cung cấp thông tin về độ chính xác của các thuật toán nhận dạng khuôn mặt, và chương trình của Cơ quan Dự án Nghiên cứu Cao cấp Quốc phòng hỗ trợ các cơ chế kiểm thử mới trong bảo mật học máy. Mở rộng phạm vi của các tài nguyên kiểm thử liên bang là rất quan trọng để áp dụng lành mạnh các hệ thống AI mới.

(5) Phát triển các thư viện và bộ công cụ phần mềm nguồn mở

Chính phủ cần đầu tư vào việc cung cấp truy cập và hỗ trợ cho các thư viện và bộ công cụ AI mã nguồn mở. Việc truy cập và hỗ trợ liên tục cho các thư viện và bộ công cụ này có thể tăng tốc quá trình nghiên cứu và phát triển công nghệ, từ nghiên cứu cơ bản đến hỗ trợ chuyển giao công nghệ. Sự phát triển của các thư viện và bộ công cụ mã nguồn mở đã tạo điều kiện cho sự phát triển tương ứng của các ứng dụng và kỹ năng AI. Nhiều cơ quan và các nhà nghiên cứu được tài trợ bởi cơ quan cũng đưa mã nguồn mã nguồn mở của họ lên các nền tảng thương mại phổ biến như GitHub. Chính phủ cũng có thể cần khuyến khích việc phát triển, bảo trì và chăm sóc tiếp tục cho phần mềm và công cụ để ngăn chúng trở nên lỗi thời. Ví dụ, chương trình Pathways to Enable Open-Source Ecosystems của NSF nhằm mục đích tận dụng sức mạnh của việc phát triển mã nguồn mở cho việc tạo ra các giải pháp công nghệ mới cho các vấn đề quan trọng của quốc gia và xã hội.

Chiến lược 6: Đo lường và đánh giá hệ thống AI thông qua các tiêu chuẩn và điểm chuẩn

Tiêu chuẩn, điểm chuẩn, và các bộ kiểm thử, cũng như sự áp dụng của chúng bởi cộng đồng AI, là rất quan trọng để hướng dẫn và thúc đẩy nghiên cứu và phát triển về các hệ thống AI. Sự nhận thức về vai trò này ngày càng tăng ở Hoa Kỳ và trên toàn thế giới. Các văn kiện quan trọng đều đặc biệt nhấn mạnh tầm quan trọng của tiêu chuẩn. Ngoài ra, Tổ chức Tiêu chuẩn Hóa Quốc tế (ISO) và Ủy ban Điện kỹ thuật Quốc tế (IEC) đã tổ chức một nhóm chuyên môn liên quan đến AI để phát triển tiêu chuẩn và các yếu tố liên quan cho các hệ thống AI.

Sự tăng đột biến trong các hoạt động tiêu chuẩn liên quan đến AI đã vượt xa sự khởi đầu của các điểm chuẩn (benchmarks) mới và các đánh giá AI, đặc biệt là liên quan đến tính đáng tin cậy của các hệ thống AI. Các xem xét về công bằng và thiên vị trong các tập dữ liệu đánh giá đã trở nên ngày càng quan trọng. Cần phải mở rộng các nỗ lực này đến các đánh giá kỹ thuật xã hội của các hệ thống AI và đánh giá ảnh hưởng toàn diện của các công nghệ AI. Đánh giá, thúc đẩy, và cung cấp bảo đảm về mọi khía cạnh của tính đáng tin cậy của AI đòi hỏi đánh giá và đo lường hiệu suất công nghệ AI thông qua các điểm chuẩn và tiêu chuẩn.

(1) Phát triển một loạt các tiêu chuẩn AI

Việc phát triển tiêu chuẩn cho AI cần được đẩy nhanh để theo kịp sự phát triển và mở rộng của các ứng dụng AI. Các tiêu chuẩn này bảo đảm hệ thống AI hoạt động an toàn, đáng tin cậy, và có khả năng tương tác, đồng thời cung cấp các định nghĩa và thuật ngữ kỹ thuật nhất quán. Việc áp dụng tiêu chuẩn tạo sự tin cậy cho công nghệ và mở rộng thị trường.

Một số tiêu chuẩn AI hiện có gồm P1872-2015 của IEEE và ISO/IEC 22989:2022, nhưng cần thêm tiêu chuẩn cho các tiểu lĩnh vực AI. NIST đang phát triển khung quản lý rủi ro AI, cần thêm hỗ trợ nghiên cứu để đánh giá rủi ro, cấp chứng nhận và bảo hiểm cho AI.

Các lĩnh vực cần tiêu chuẩn bao gồm kỹ thuật phần mềm, tính năng và độ tin cậy, chỉ số hiệu suất, an toàn, khả năng tương tác, bảo mật, quyền riêng tư, công bằng và khả năng giải thích, tính linh hoạt, hợp tác quốc tế, truy xuất nguồn gốc, và các lĩnh vực cụ thể như y tế và sản xuất. Tiêu chuẩn cũng giúp quản lý hiệu suất môi trường của AI, giảm dấu chân carbon và thúc đẩy sử dụng AI có trách nhiệm.

(2) Thiết lập điểm chuẩn công nghệ AI

Việc thiết lập các điểm chuẩn công nghệ AI (AI Technology Benchmarks) thông qua các bài kiểm tra và đánh giá giúp thúc đẩy đổi mới và cung cấp dữ liệu khách quan để theo dõi sự phát triển của khoa học và công nghệ AI. Cần phát triển và chuẩn hóa các phương pháp và chỉ số đánh giá hiệu quả. Các chỉ số chuẩn cần đo lường tính chính xác, độ phức tạp, độ tin cậy, khả năng giải thích, độ lệch không mong muốn, so sánh với hiệu suất con người và tác động kinh tế.

Các chỉ số thường dùng như độ chính xác, độ chính xác tiên đoán không cung cấp đủ thông tin toàn diện. Đánh giá AI nên sử dụng các chỉ số liên quan đến ngữ cảnh sử dụng, và dữ liệu nên được cải thiện liên tục và liên kết với các vấn đề thực tế. Các tiêu chuẩn và điểm chuẩn bổ sung cần được phát triển cho nhiều lĩnh vực khác nhau để bảo đảm ứng dụng rộng rãi và hiệu quả của AI.

Chính phủ liên bang nên xác thực và tổng hợp các đánh giá từ các nhà nghiên cứu độc lập để tạo ra danh mục các bài kiểm tra được phê duyệt cho các mô hình đang triển khai và phát triển. Việc nhấn mạnh đặc điểm hiệu suất trong các điều kiện sử dụng là quan trọng để tránh tình trạng AI hoạt động kém hiệu quả hoặc gây hại.

(3) Tăng cường tính khả dụng của các nền tảng thử nghiệm AI

Các nền tảng thử nghiệm (testbeds) rất quan trọng để các nhà nghiên cứu mô phỏng và thí nghiệm trên dữ liệu thực tế. Hiện tại cần nhiều testbeds hơn cho mọi lĩnh vực AI. Chính phủ có lượng dữ liệu lớn nhưng không thể chia sẻ rộng rãi, do đó cần chương trình để các nhà nghiên cứu làm việc trong môi trường thử nghiệm an toàn. Các khung kiểm tra và tiêu chuẩn hóa cần được tạo ra để đánh giá hệ thống AI, bảo đảm chúng hoạt động công bằng, an toàn, bảo mật và đáng tin cậy. NAIRR sẽ hỗ trợ mục tiêu này.

(4) Thu hút cộng đồng AI tham gia vào các tiêu chuẩn và điểm chuẩn

Chính phủ cần lãnh đạo và phối hợp để hỗ trợ tiêu chuẩn hóa AI và khuyến khích việc sử dụng rộng rãi trong các cơ quan, khu vực hàn lâm và công nghiệp. Cộng đồng AI, bao gồm chính phủ, khu vực hàn lâm, công nghiệp và xã hội dân sự, cần tham gia vào việc phát triển tiêu chuẩn và chương trình đánh giá.

Yêu cầu từ người dùng định hình mục tiêu và thiết kế, còn tiêu chuẩn từ nhà phát triển thúc đẩy tiến bộ công nghệ. Công cụ mô phỏng và phân tích sẽ tăng tốc phát triển AI. Kết quả từ các tiêu chuẩn này giúp xác định công nghệ phù hợp và tạo ra tiêu chí khách quan cho việc tuân thủ.

Khuyến khích sự tham gia của các chuyên gia vào hoạt động tiêu chuẩn hóa và đánh giá là rất quan trọng. Việc cập nhật quy trình mua sắm để bao gồm yêu cầu tiêu chuẩn AI sẽ thúc đẩy sự tham gia này. Các điểm chuẩn cộng đồng thúc đẩy cạnh tranh lành mạnh và cung cấp chỉ số hiệu suất so sánh khách quan. Cần cải thiện phương pháp và tài nguyên thử nghiệm để các cơ quan đánh giá khả năng AI.

(5) Phát triển các tiêu chuẩn để kiểm toán và giám sát các hệ thống AI

Hệ thống AI cần được kiểm toán đúng cách và giám sát thường xuyên để xác định và giảm thiểu các rủi ro kỹ thuật (như độ chính xác, độ tin cậy) và rủi ro do con người (như sự thiên lệch và quyền riêng tư). Việc kiểm toán và giám sát AI vẫn còn nhiều câu hỏi chưa được giải đáp và khả năng mở rộng kiểm toán là một thách thức lớn. Khi AI ngày càng phổ biến, cần phát triển các kỹ thuật kiểm toán có thể mở rộng, tạo ra công cụ phân tích định tính mới, đào tạo đủ nhân lực, nhận phản hồi từ con người và xây dựng năng lực tổ chức trong chính phủ và công nghiệp để thực hiện, giám sát và phản hồi các kiểm toán.

Chiến lược 7: Hiểu rõ hơn nhu cầu về lực lượng lao động nghiên cứu và phát triển trí tuệ nhân tạo quốc gia

Sự phát triển nhanh chóng của AI làm tăng nhu cầu về chuyên gia khoa học máy tính và thông tin cũng như kỹ năng mới trong lực lượng lao động. Tại Hoa Kỳ, các vị trí việc làm liên quan đến khoa học máy tính và thông tin dự kiến tăng 22% trong giai đoạn 2020 - 2030. Ngành công nghiệp khu vực tư nhân sẽ dẫn đầu nhu cầu này với sự hỗ trợ tài chính và cơ sở hạ tầng tiên tiến. Nghiên cứu AI có thể đóng góp tới 11,5 nghìn tỷ USD vào tăng trưởng kinh tế của các nước G20 trong cùng giai đoạn trên. Tuy nhiên, các tổ chức giáo dục Hoa Kỳ đang gặp khó khăn trong việc đáp ứng sự bùng nổ về số lượng sinh viên theo học AI. Đặc biệt, tuyển sinh tiến sĩ của công dân Hoa Kỳ và người thường trú giảm, ảnh hưởng đến lực lượng lao động AI, nhất là ở các vị trí cần an ninh. Chính phủ cần hiểu rõ nhu cầu lực lượng lao động và hỗ trợ phát triển tài năng AI để tạo ra lực lượng lao động AI bền vững. Chiến lược này chia gồm 10 định hướng dưới đây:

(1) Mô tả và đánh giá lực lượng lao động AI

Nhấn mạnh tính đa ngành của lực lượng lao động AI và nghiên cứu thêm về nhu cầu lực lượng lao động quốc gia hiện tại và tương lai cho nghiên cứu và phát triển AI. Cần làm rõ thêm về thành phần và nhu cầu của lực lượng lao động AI, bao gồm các mặt dân số học, để hiểu rõ hơn về khả năng và những khoảng trống cần điều chỉnh. Việc này sẽ giúp tập trung nỗ lực và đầu tư hiệu quả trong nhiều lĩnh vực khác nhau, đồng thời giải quyết các chênh lệch dân tộc và tăng cường công bằng và đa dạng.

(2) Phát triển chiến lược cho tài liệu hướng dẫn AI ở mọi cấp độ

Hoa Kỳ sẽ có lợi khi làm cho nghiên cứu AI dễ tiếp cận hơn với mọi người. Đồng thời, việc giới thiệu AI và khoa học dữ liệu cho học sinh từ cấp tiểu học và trung học chuẩn bị họ cho thế giới ngày càng áp dụng AI. Cần có bài học chất lượng cao để phát triển kỹ năng tư duy phản biện và đánh giá các hệ thống AI. Nghiên cứu cần tập trung vào việc biên soạn nội dung phù hợp và phương pháp giảng dạy hiệu quả, đồng thời bảo đảm tính công bằng dân tộc và văn hóa trong việc tiếp cận với tài nguyên và chương trình đào tạo AI.

(3) Hỗ trợ nhân viên giáo dục đại học AI

Các nỗ lực của lực lượng lao động cũng nên nghiên cứu các cơ hội để bảo đảm lực lượng lao động đại học đủ để đào tạo các thế hệ tương lai của lực lượng lao động AI tại các trường cao đẳng và đại học, bao gồm các chương trình cấp bằng liên kết, cử nhân, thạc sĩ và tiến sĩ. Những nỗ lực này cho phép giảng viên tham gia vào nhiều lĩnh vực.

(4) Khám phá tác động của sự đa dạng về chuyên môn và đa ngành

Phát triển và triển khai AI an toàn và công bằng đòi hỏi hiểu biết rộng về những người và nơi bị ảnh hưởng, cùng với kỹ thuật sâu về AI. Giáo dục đa ngành sẽ bảo đảm tiếp cận công bằng đến thông tin và cơ hội, phát triển thị trường ý tưởng đa dạng về

công nghệ. Việc tuyển dụng đa dạng từ các góc độ học thuật, chuyên môn và kinh nghiệm cần được ưu tiên. Nghiên cứu viên nên khai thác vị trí và quan điểm để nghiên cứu tác động của các lĩnh vực khác nhau đối với AI, bảo đảm tính đại diện và đa văn hóa trong lực lượng lao động AI liên bang.

(5) Xác định và thu hút những nhân tài giỏi nhất thế giới

Hoa Kỳ sở hữu nhiều tài năng trong nhiều lĩnh vực nhưng lịch sử đã dựa vào tài năng nước ngoài để tăng cường lực lượng lao động công nghệ, đặc biệt là trong nghiên cứu và phát triển các công nghệ mới nổi. Một nửa số chuyên gia AI hiện tại trong khu vực hàn lâm và công nghiệp ở Hoa Kỳ đều là người nước ngoài. Tài nguyên liên bang có thể hỗ trợ các nỗ lực của khu vực đại học, công nghiệp và xã hội dân sự để tiếp đón sinh viên và học giả nước ngoài, cung cấp lộ trình đạt được quốc tịch Hoa Kỳ. Xây dựng các đối tác quốc tế với các chính phủ và đại học nước ngoài cũng hỗ trợ chiến lược này.

(6) Phát triển chuyên môn AI ở địa phương

Hoa Kỳ với quy mô và đa dạng địa lý là nguồn lực quan trọng để tổng hợp và tận dụng đào tạo AI và cơ hội kinh tế một cách công bằng và rộng rãi. Phối hợp các nguồn lực địa phương như cơ sở hạ tầng dữ liệu, máy tính, và ngành công nghiệp bổ trợ sẽ thúc đẩy sự phát triển của nền kinh tế AI ở địa phương và gia tăng tiến bộ nghiên cứu AI trên toàn quốc. Các nỗ lực liên bang cần hướng tới việc mở rộng cơ hội tiếp cận nền kinh tế AI cho các vùng miền ít được quan tâm.

(7) Nghiên cứu các lựa chọn để tăng cường lực lượng lao động AI của liên bang

Chính phủ liên bang nên tài trợ và thực hiện nghiên cứu để tăng cường nguồn nhân lực AI, bằng cách đẩy nhanh và tận dụng các chương trình giáo dục AI và phát triển nguồn nhân lực. Các nỗ lực này sẽ xây dựng đối tác giữa chính phủ, khu vực hàn lâm và ngành công nghiệp, giúp tuyển dụng và đào tạo chuyên gia mới và sinh viên tham gia các cơ quan liên bang trong các lĩnh vực chuyển đổi số, quản lý dữ liệu, phân tích dữ liệu và AI. Các đối tác có thể bao gồm việc làm việc tại các tổ chức chính phủ, đẩy nhanh việc triển khai AI. Đạo luật Đào tạo AI yêu cầu phát triển chương trình đào tạo AI cho lực lượng lao động liên bang. Đạo luật CHIPS và Khoa học năm 2022 ủy quyền cho NSF nghiên cứu và thiết lập chương trình học bổng phục vụ AI liên bang và mở rộng chương trình Học bổng CyberCorps cho người học các chủ đề AI, nhằm thu hút và đào tạo chuyên gia AI cho chính phủ.

(8) Đưa các nội dung về đạo đức, pháp lý và xã hội vào giáo dục và đào tạo AI

Những tác động đạo đức, pháp lý và xã hội của AI ngày càng trở nên quan trọng. Do đó, những người phát triển, sử dụng và giám sát hệ thống AI cần hiểu rõ và cam kết tuân thủ các giá trị này. Cần có chuyên gia thông thạo các vấn đề này và các hệ thống dữ liệu và AI để giúp giáo dục lực lượng lao động và xây dựng chương trình đào tạo. Cũng cần các chuyên gia về chính sách, luật và quản trị hiểu biết về các khía cạnh đạo đức, pháp lý, xã hội và công nghệ của AI. Tuy nhiên, các chương trình học hiện tại gặp khó khăn trong việc đào tạo chuyên gia toàn diện. Để giải quyết vấn đề này, chính phủ

nên hỗ trợ các chương trình đại học và sau đại học, cũng như cơ hội sau tiến sĩ để xây dựng năng lực liên ngành, và hỗ trợ nghiên cứu và phổ biến tài liệu giáo dục về các khía cạnh đạo đức, pháp lý và xã hội của AI trong chương trình giáo dục và đào tạo AI.

(9) Phổ biến các ưu tiên của lực lượng lao động liên bang cho các bên liên quan bên ngoài

Truyền đạt cho các tổ chức tư nhân, đại học và công chúng về nhu cầu và ưu tiên nhân lực của chính phủ liên bang là rất quan trọng. Việc mô tả, tuyển dụng và phát triển nhân lực phải công bằng, minh bạch và có trách nhiệm, và cần được truyền đạt nhất quán đến các bên liên quan. Các cơ quan liên bang có thể làm điều này qua việc đăng tải câu chuyện thành công, tiếp cận doanh nghiệp nhỏ, tham gia hội chợ thương mại và hội nghị khoa học, và thông báo tài trợ chương trình. Các cơ hội khác bao gồm các chương trình giáo dục và phát triển nhân lực kết hợp nghiên cứu như trong Viện Nghiên cứu AI Quốc gia và các hợp tác hiện có giữa đại học, ngành công nghiệp và chính phủ.

Chiến lược 8: Mở rộng quan hệ đối tác công - tư để thúc đẩy tiến bộ trong trí tuệ nhân tạo

Sự lãnh đạo của Hoa Kỳ trong nghiên cứu và đổi mới khoa học kỹ thuật dựa trên hệ sinh thái R&D của chính phủ, đại học và ngành công nghiệp. Vị thế đổi mới sáng tạo của Hoa Kỳ phụ thuộc vào đối tác nghiên cứu Chính phủ - đại học - ngành công nghiệp mạnh mẽ hơn, nhất là để tạo ra và triển khai các đột phá công nghệ AI. Nghiên cứu công nghệ thông tin tại các trường đại học với tài trợ liên bang và trong ngành công nghiệp đã tạo ra các ngành kinh tế hàng tỷ USD. Các tiến bộ trong đối tác này sẽ củng cố lẫn nhau và dẫn đến một lĩnh vực AI đổi mới và sôi động. Chiến lược này phản ánh tầm quan trọng ngày càng tăng của các đối tác công-tư với ba chủ đề chính sau đây:

(1) Đạt được nhiều hơn từ sự hợp tác công tư

Khu vực tư nhân xem AI là công cụ tiềm năng cho kinh doanh, trong khi tài trợ công tập trung vào tác động dài hạn và lợi ích xã hội. Tích hợp hai quan điểm này sẽ thúc đẩy đổi mới khoa học và kỹ thuật nhanh chóng hơn. Chia sẻ mô hình, dữ liệu và kết quả AI giúp giảm tài nguyên và trùng lặp. Các đối tác R&D giữa chính phủ, đại học và ngành công nghiệp giải quyết các thách thức thực tế và chuyển đổi kết quả nghiên cứu thành sản phẩm thực tế. Sự hợp tác giữa các cơ quan liên bang tối ưu hóa đầu tư ở các lĩnh vực giao thoa nhiệm vụ. Hỗ trợ các nỗ lực liên chính phủ như Viện Nghiên cứu AI Quốc gia là chìa khóa cho tiến bộ dài hạn. Các khoản đầu tư này thúc đẩy nghiên cứu AI có trách nhiệm và ứng dụng, qua hợp tác giữa chính phủ, học viện, ngành công nghiệp và xã hội dân sự. Mở rộng các chương trình cho phép các nhà nghiên cứu làm việc trong các lĩnh vực khác nhau sẽ tăng cường hiệu quả hợp tác. Ngành công nghiệp thương mại hóa AI được hỗ trợ bởi R&D từ các trường đại học và phòng thí nghiệm liên bang.

(2) Mở rộng quan hệ đối tác với nhiều bên liên quan đa dạng hơn

Mở rộng quan hệ đối tác giữa công - tư bao gồm các tổ chức xã hội dân sự giúp tích hợp góc nhìn đa dạng vào phát triển AI. Phát triển R&D tập trung vào trách nhiệm, công bằng và tôn trọng giá trị dân chủ và quyền con người là rất quan trọng. Hợp tác quốc tế có thể thúc đẩy tiến bộ AI cho lợi ích toàn cầu. Sự tham gia của xã hội dân sự trong thảo luận về tiếp cận công bằng và độ tin cậy là cần thiết. Các công ty AI hướng dẫn và giảm rủi ro trong phát triển AI, trong khi tổ chức phi lợi nhuận nhỏ đóng góp lớn cho các nỗ lực "AI vì lợi ích xã hội" với các chương trình tình nguyện. Hợp tác công-tư và xã hội dân sự là cần thiết để đạt được tiếp cận công bằng và giải quyết các vấn đề xã hội toàn cầu. Chính phủ và tổ chức quốc tế đặt ra tiêu chuẩn cho việc sử dụng AI công bằng và có trách nhiệm. Hệ sinh thái AI mở với sự tham gia của các công ty lớn và nhỏ, khả năng tính toán tiên tiến và các nguồn lực từ chính phủ cùng với sự đa dạng của các tổ chức sẽ dẫn đến sử dụng AI đạo đức hơn. Hợp tác đa dạng này hỗ trợ mô hình mới như quan hệ đối tác giữa các tổ chức phục vụ thiểu số và Viện Nghiên cứu AI Quốc gia. Quan hệ đối tác cũng hỗ trợ tính liên ngành của R&D AI, kết hợp nhiều lĩnh vực từ khoa học máy tính đến triết học. Mặc dù còn thách thức, kết quả cuối cùng sẽ thúc đẩy sự phát triển và đánh giá các hệ thống AI công bằng, minh bạch, có trách nhiệm, an toàn và bảo mật.

(3) Cải thiện, mở rộng và tạo ra các cơ chế cho quan hệ đối tác R&D

R&D là nỗ lực tập thể, thường được thực hiện bởi các nhóm hoạt động tại nhiều cơ sở khác nhau. Hợp tác công - tư đòi hỏi các sắp đặt tổ chức để hỗ trợ việc huy động tài nguyên nhằm đạt được hiệu quả và tác động tích cực nhanh chóng, tránh lãng phí nỗ lực. Các hình thức và cơ chế hợp tác công - tư cho các ứng dụng AI đã được phát triển trong vài thập kỷ qua. Mở rộng phạm vi, cải thiện hoạt động và sản phẩm đầu ra cho nhiều người tham gia và không gian ứng dụng, cũng như tạo ra các hình thức mới của hợp tác công tư là những nỗ lực quan trọng và có giá trị. Các ví dụ bao gồm: hợp tác dựa trên dự án cá nhân; chương trình chung để thúc đẩy nghiên cứu cơ bản trước cạnh tranh mở; hợp tác để triển khai và cải thiện cơ sở hạ tầng nghiên cứu; hợp tác để phát triển nguồn nhân lực; cuộc thi giải thưởng liên bang; chia sẻ dữ liệu và mô hình. Việc tận dụng sức mạnh của mỗi đối tác để mang lại lợi ích cho tất cả là rất quan trọng nhằm đạt được tác động lớn nhất.

Chiến lược 9: Thiết lập một cách tiếp cận có nguyên tắc và hợp tác quốc tế trong nghiên cứu trí tuệ nhân tạo

Mặc dù Hoa Kỳ dẫn đầu thế giới về chi tiêu R&D hàng năm, nhưng các đối thủ cạnh tranh đang tìm cách bắt kịp. Không có quốc gia nào dẫn đầu trong tất cả các khía cạnh của khoa học và kỹ thuật trong thế giới ngày nay. Trong lĩnh vực AI, số lượng bài báo được xuất bản hàng năm đã tăng gấp đôi từ năm 2010 đến năm 2020 và phân bố về địa lý đã trở nên rộng rãi hơn. Để bảo đảm rằng Hoa Kỳ vẫn là trung tâm quan trọng trong hệ sinh thái R&D AI, cần tiếp tục tham gia vào các chương trình quốc tế, cơ sở hạ tầng, bộ dữ liệu và cơ chế chia sẻ dữ liệu an toàn; duy trì sự tiếp cận với tài năng toàn cầu; hợp tác quốc tế hiệu quả và làm việc với các cấu trúc quốc tế hiện có về dữ

liệu, cơ sở hạ tầng và tài năng mà hệ sinh thái R&D AI cần. Các đối tác quốc tế đóng vai trò quan trọng trong việc thúc đẩy các nỗ lực trong tất cả các lĩnh vực này.

(1) Xây dựng một nền văn hóa toàn cầu về phát triển và sử dụng AI đáng tin cậy

Nghiên cứu khoa học tiên tiến dựa trên sự hợp tác quốc tế là cốt lõi để phát triển và triển khai AI một cách đáng tin cậy và công bằng. Đối với Hoa Kỳ, "AI đáng tin cậy" phải tuân thủ các chuẩn mực đạo đức, pháp lý và xã hội, bao gồm tính hợp pháp, mục đích rõ ràng, đáng tin cậy, an toàn, minh bạch và thúc đẩy công bằng. Hợp tác quốc tế củng cố nỗ lực nghiên cứu của chính phủ Hoa Kỳ, thúc đẩy các giá trị chung như công bằng và trách nhiệm. Các mối quan hệ như các chương trình liên minh với Úc và các thỏa thuận hành chính với Ủy ban châu Âu là ví dụ điển hình cho các nỗ lực này. Vai trò lãnh đạo của Hoa Kỳ tại các diễn đàn đa phương đã định hướng các hướng dẫn về AI toàn cầu và khởi động các sáng kiến như Đề xuất về AI của OECD và Đối tác Toàn cầu về AI. Tiếp tục tham gia vào các diễn đàn này thể hiện cam kết phát triển AI đạo đức và hợp tác, đồng thời giải quyết các rủi ro từ các đối tác không dân chủ. Các chiến lược hợp tác cũng nhắm đến ngăn chặn sử dụng AI sai mục đích như áp bức chính trị, hoạt động tội phạm hay chi phối xã hội, phù hợp với ưu tiên toàn cầu và Chiến lược 3.

(2) Hỗ trợ phát triển các hệ thống, tiêu chuẩn và khuôn khổ AI toàn cầu

Cần hợp tác nghiên cứu quốc tế để phát triển các phương pháp chia sẻ dữ liệu an toàn, ứng dụng AI trong y tế công cộng và bền vững, cũng như tiêu chuẩn chất lượng, an ninh, và công cụ chuẩn hóa cho thiết kế, phát triển và sử dụng hiệu quả các hệ thống AI đáng tin cậy. Quan trọng là nghiên cứu các cơ chế hợp tác công tư và thỏa thuận quốc tế, như trong Chiến lược 8. Việc này phức tạp và giao thoa, nhưng các ví dụ có thể hướng dẫn các cơ quan trong nghiên cứu này. Ví dụ: Tuyên bố của Hoa Kỳ và Vương quốc Anh về Hợp tác trong R&D AI, nhằm thúc đẩy tầm nhìn chung về AI và hướng tới một hệ sinh thái R&D AI hỗ trợ lẫn nhau; cam kết mới của Nhóm bộ tứ (Hoa Kỳ, Ấn Độ, Úc và Nhật Bản) thành lập các nhóm tiếp xúc về tiêu chuẩn kỹ thuật, bao gồm một nhóm về truyền thông tiên tiến và AI tập trung vào các hoạt động phát triển tiêu chuẩn và nghiên cứu tiền chuẩn hóa cơ bản. Ngoài ra, nghiên cứu hợp tác quốc tế cần tập trung vào quản lý dữ liệu, quản trị và chia sẻ dữ liệu, bảo vệ quyền sở hữu dữ liệu và bảo đảm an ninh quốc gia khi chuyển dữ liệu xuyên biên giới cho hợp tác nghiên cứu AI.

(3) Thúc đẩy trao đổi ý tưởng và chuyên môn quốc tế

Các chuyên gia và khu vực đổi mới sáng tạo hàng đầu về công nghệ mới đang phân bố rải rác trên nhiều quốc gia. Để bảo đảm ý tưởng có thể chia sẻ dễ dàng, hợp tác quốc tế giữa các cơ quan và các thỏa thuận hai bên là rất cần thiết. Việc này có thể thực hiện qua các chương trình hiện có như Chương trình Học giả Khoa học tại Đại sứ quán, Chương trình Đại sứ Khoa học Hoa Kỳ, Fulbright, và các sáng kiến TechCamps, tập trung vào lĩnh vực AI. Các chương trình thực tập và học bổng quốc tế có thể giúp xây dựng lực lượng lao động STEM của Hoa Kỳ. Hợp tác này cũng giúp thu hút và giữ chân các tài năng AI hàng đầu và phát triển các đối tác lâu dài với các nhà nghiên cứu AI của

Hoa Kỳ. Các thử thách lớn là cần một cơ chế hiệu quả để các chính phủ có thể sử dụng các đối tác và công nghệ để giải quyết các thách thức phức tạp của xã hội và công nghiệp, như sức khỏe, thiên tai, và an ninh lương thực.

(4) Khuyến khích phát triển AI vì lợi ích toàn cầu

Một số ứng dụng của AI chống lại các giá trị và lợi ích của Hoa Kỳ, đặc biệt khi AI được sử dụng cho mục đích áp bức chính trị, lạm dụng, vi phạm pháp luật quốc tế và thao túng xã hội. Cần có nghiên cứu bổ sung về cách ngăn chặn việc sử dụng AI cho các mối đe dọa này. Việc hợp tác với cộng đồng quốc tế qua các đối tác cho phép Hoa Kỳ hạn chế các quốc gia đối thủ và đối nghịch khỏi việc tiếp cận các công nghệ AI tiên tiến, bảo vệ an ninh quốc gia. Các liên minh và đối tác có lợi song phương về AI cũng giúp Hoa Kỳ duy trì ổn định toàn cầu và ngăn ngừa các hành động quá khích. Ngoài ra, đầu tư chung với các nước có giá trị phù hợp trong AI cũng mang lại cơ hội giải quyết các thách thức toàn cầu như đại dịch, thiên tai, ô nhiễm, an ninh lương thực và bền vững. Công tác tuyên truyền và tương tác với cộng đồng rộng rãi là rất quan trọng để nâng cao nhận thức về khả năng và hạn chế của AI. Khi vai trò của AI ngày càng quan trọng trên toàn cầu, hợp tác quốc tế trở nên càng cần thiết để thúc đẩy việc áp dụng AI một cách an toàn, hiệu quả và đạo đức, trong đó Hoa Kỳ có thể tiếp tục dẫn đầu trong nghiên cứu, đổi mới sáng tạo và tạo ra các giá trị vì lợi ích toàn cầu.