

CHUẨN BỊ CHO TƯƠNG LAI CỦA TRÍ TUỆ NHÂN TẠO

MỤC LỤC

LỜI NÓI ĐẦU.....	1
1. GIỚI THIỆU.....	3
1.1. Lược sử TTNT.....	3
1.2. Trí tuệ nhân tạo là gì?.....	5
1.3. Hiện trạng phát triển TTNT.....	6
2. CÁC ỨNG DỤNG CỦA TTNT CHO HÀNG HÓA CÔNG.....	11
3. TRÍ TUỆ NHÂN TẠO TRONG CHÍNH PHỦ.....	13
4. TRÍ TUỆ NHÂN TẠO VÀ QUY ĐỊNH.....	14
Nghiên cứu điển hình: Xe ô-tô và máy bay tự trị.....	16
5. NGHIÊN CỨU VÀ LỰC LƯỢNG LAO ĐỘNG.....	21
5.1. Giám sát tiến độ trong TTNT.....	21
5.2. Hỗ trợ liên bang cho TTNT.....	23
5.3. Phát triển và đa dạng lực lượng lao động.....	25
5.4. Thách thức đa dạng.....	27
6. TRÍ TUỆ NHÂN TẠO, TỰ ĐỘNG HÓA VÀ KINH TẾ.....	28
7. CÔNG BẰNG, AN TOÀN, VÀ QUẢN TRỊ.....	29
7.1. Luật pháp, Công bằng, và trách nhiệm.....	29
7.2. An toàn và kiểm soát.....	32
8. NHỮNG SỰ CÂN NHẮC VÀ AN NINH TOÀN CẦU.....	35
8.1. Hợp tác quốc tế.....	35
8.2. Trí tuệ nhân tạo và an ninh mạng.....	36
8.3. Trí tuệ nhân tạo trong các hệ thống vũ khí.....	37
KẾT LUẬN.....	39

LỜI NÓI ĐẦU

Những tiến bộ trong công nghệ Trí tuệ nhân tạo đã mở ra những thị trường và cơ hội mới cho sự tiến bộ trong các lĩnh vực quan trọng như y tế, giáo dục, năng lượng và môi trường. Trong những năm gần đây, máy móc đã vượt con người trong việc thực hiện một số nhiệm vụ cụ thể nào đó, chẳng hạn như một số khía cạnh trong nhận dạng hình ảnh. Các chuyên gia dự báo rằng sẽ tiếp tục có tiến bộ nhanh chóng trong lĩnh vực trí tuệ nhân tạo chuyên ngành. Mặc dù không chắc rằng máy móc sẽ thể hiện trí tuệ ứng dụng rộng rãi tương đương hoặc vượt con người trong vòng 20 năm tới, nhưng nó được kỳ vọng sẽ đạt và vượt hiệu suất của con người trong ngày càng nhiều nhiệm vụ hơn.

Để góp phần chuẩn bị một tương lai, trong đó Trí tuệ nhân tạo đóng vai trò ngày càng cao, các nghiên cứu tiến hành khảo sát hiện trạng của Trí tuệ nhân tạo, các ứng dụng hiện tại và tiềm năng của nó, và các vấn đề nảy sinh trong xã hội và chính sách công do sự tiến bộ của Trí tuệ nhân tạo, đồng thời cũng đưa ra các khuyến cáo cho hành động cụ thể của các cơ quan chính phủ và các chủ thể.

Tổng luận "Chuẩn bị cho tương lai của trí tuệ nhân tạo" này dựa trên Báo cáo cùng tên được Tiểu ban Máy học và trí tuệ nhân tạo của Hội đồng KH&CN Quốc gia Hoa Kỳ xây dựng để thúc đẩy sự phối hợp liên ngành, cung cấp tư vấn kỹ thuật và chính sách về các chủ đề liên quan đến Trí tuệ nhân tạo, theo dõi sự phát triển của công nghệ Trí tuệ nhân tạo trong các ngành công nghiệp, các cộng đồng nghiên cứu và Chính phủ.

Trong những năm tới, Trí tuệ nhân tạo sẽ tiếp tục đóng góp vào tăng trưởng kinh tế và sẽ là một công cụ có giá trị cho cải thiện thế giới, khi ngành công nghiệp, xã hội dân sự và chính phủ cùng với nhau để phát triển các khía cạnh tích cực của công nghệ, quản lý các rủi ro và thách thức của Trí tuệ nhân tạo, và đảm bảo rằng mọi người đều có cơ hội giúp đỡ trong việc xây dựng một xã hội gia tăng Trí tuệ nhân tạo và thụ hưởng các lợi ích của nó.

Cục Thông tin khoa học và công nghệ quốc gia xin trân trọng giới thiệu.

CỤC THÔNG TIN KHOA HỌC VÀ
CÔNG NGHỆ QUỐC GIA

CÁC CHỮ VIẾT TẮT

DARPA	Cục Dự án nghiên cứu quốc phòng cao cấp (Hoa Kỳ)
FAA	Cục Hàng không liên bang
LAWS	Hệ thống vũ khí sát thương tự động
NC&PT	Nghiên cứu khoa học và phát triển công nghệ
NHTSA	Cục an toàn giao thông đường cao tốc quốc gia
NIH	Viện Y tế quốc gia (Hoa Kỳ)
NSF	Quỹ Khoa học Quốc gia (Hoa Kỳ)
NSTC	Hội đồng KH&CN Quốc gia (Hoa Kỳ)
OSTP	Văn phòng Chính sách Khoa học và Công nghệ (Hoa Kỳ)
STEM	Khoa học, công nghệ, kỹ thuật và toán học
TTNT	Trí tuệ nhân tạo
UAS	Hệ thống bay không người lái

CHUẨN BỊ CHO TƯƠNG LAI CỦA TRÍ TUỆ NHÂN TẠO

1. GIỚI THIỆU

Trí tuệ nhân tạo (TTNT) có tiềm năng giúp giải quyết một số thách thức lớn nhất mà xã hội đang phải đối mặt. Xe ô-tô thông minh có thể cứu được hàng trăm nghìn mạng sống mỗi năm trên toàn thế giới, và tăng khả năng cơ động cho người già và người khuyết tật. Các tòa nhà thông minh có thể tiết kiệm năng lượng và giảm lượng khí thải cacbon. Y học chính xác có thể kéo dài tuổi thọ và nâng cao chất lượng cuộc sống. Chính phủ thông minh hơn có thể phục vụ người dân một cách nhanh chóng hơn, chính xác, và tiết kiệm tiền bạc. Giáo dục với sự tăng cường TTNT có thể giúp giáo viên cung cấp cho mỗi đứa trẻ một nền giáo dục mở ra một cuộc sống an toàn và đầy đủ. Đó chỉ là một vài trong số những lợi ích tiềm năng, nếu công nghệ này được phát triển với sự xem xét đầy đủ những lợi ích cùng với các nguy cơ và thách thức của nó.

Hoa Kỳ đang đi đầu trong nghiên cứu cơ bản về TTNT, chủ yếu được hỗ trợ bằng nguồn kinh phí tài trợ nghiên cứu Liên bang và triển khai tại phòng thí nghiệm của chính phủ. Hỗ trợ của Chính phủ Mỹ cho NC&PT TTNT công khai được thực hiện thông qua các chương trình Nghiên cứu và phát triển Công nghệ thông tin và Mạng và được sự hỗ trợ chủ yếu của Cục Dự án nghiên cứu quốc phòng cao cấp (DARPA), Quỹ Khoa học Quốc gia (NSF), Viện Y tế quốc gia (NIH), Văn phòng Nghiên cứu Hải quân (ONR) và Cơ quan các dự án nghiên cứu tình báo tiên tiến (IARPA). Những nỗ lực nghiên cứu quốc gia lớn như Sáng kiến điện toán chiến lược quốc gia, Sáng kiến dữ liệu lớn, và Sáng kiến Nghiên cứu não thông qua Các công nghệ thần kinh sáng tạo tiên tiến (BRAIN) cũng gián tiếp đóng góp cho tiến bộ của nghiên cứu TTNT. Những lợi ích hiện tại và dự kiến của công nghệ TTNT là vô cùng to lớn, làm tăng thêm sức sống kinh tế của quốc gia và làm tăng năng suất và hạnh phúc của nhân dân.

1.1. Lược sử TTNT

Chế tạo các máy tính với trí thông minh giống như con người là một ước mơ của các chuyên gia máy tính kể từ buổi bình minh của máy tính điện tử. Mặc dù thuật ngữ "trí tuệ nhân tạo" chỉ được nêu ra vào năm 1956, nhưng gốc rễ của lĩnh vực này xuất phát từ ít nhất là những năm 1940 và các ý tưởng của TTNT đã được kết tinh trong bài báo nổi tiếng vào năm 1950 của Alan Turing, "Máy tính điện tử và Trí tuệ." Bài báo của Turing đặt ra câu hỏi: "Liệu máy có

thể tư duy?" Nó cũng đề xuất một kiểm tra để trả lời câu hỏi đó, và nêu ra khả năng rằng một máy tính có thể được lập trình để học kinh nghiệm cũng giống như một đứa trẻ.

Trong các thập kỷ tiếp theo, lĩnh vực TTNT đã trải qua những thăng trầm khi một số vấn đề nghiên cứu TTNT tỏ ra khó khăn hơn so với dự kiến và một số nghiên cứu đã chứng minh không thể vượt qua với các công nghệ tại thời điểm đó. Cho đến cuối thập niên 1990, các tiến bộ trong các nghiên cứu TTNT mới bắt đầu gia tăng, các nhà nghiên cứu tập trung nhiều hơn về các vấn đề nhánh của TTNT và các ứng dụng của TTNT với các vấn đề thực tế như nhận dạng hình ảnh và chẩn đoán y tế. Dấu mốc khởi đầu là chiến thắng của máy tính chơi cờ vua Deep Blue của IBM trước nhà vô địch thế giới Garry Kasparov vào năm 1997. Các đột phá quan trọng khác bao gồm Phần tử nhận thức Học và Tổ chức (CALO) của DARPA, dẫn đến ứng dụng công nghệ Siri của Apple Inc; Chiến thắng của máy tính hỏi-đáp Watson của IBM trong chương trình trò chơi truyền hình "Jeopardy!"; và sự thành công đáng ngạc nhiên của những chiếc xe tự lái trong các cuộc thi Grand Challenge trong những năm 2000.

Làn sóng tiến bộ và đam mê hiện tại đối với TTNT bắt đầu từ khoảng năm 2010, do ba yếu tố hỗ trợ lẫn nhau là: sự tồn tại của **dữ liệu lớn** từ các nguồn bao gồm thương mại điện tử, doanh nghiệp, các phương tiện truyền thông xã hội, khoa học, và chính phủ...; đã cung cấp nguyên liệu cho các **thuật toán và phương pháp nghiên cứu máy học tiên tiến**; qua đó dựa vào khả năng của các **máy tính công suất mạnh hơn**. Trong thời gian này, tốc độ cải thiện đã gây ngạc nhiên cho các chuyên gia TTNT. Ví dụ, trong một kiểm tra thách thức nhận dạng hình ảnh thông thường có 5 phần trăm tỷ lệ lỗi của con người đối với mỗi lỗi đo được, kết quả tốt nhất của TTNT đã được cải thiện từ tỷ lệ lỗi 26 phần trăm trong năm 2011 xuống còn 3,5 phần trăm trong năm 2015.

Cùng với đó, ngành công nghiệp đã tăng cường đầu tư vào TTNT. Trong năm 2016, Giám đốc điều hành của Google, Sundar Pichai cho biết, "Máy học [một nhánh của TTNT] là cách thay đổi cốt lõi mà qua đó chúng ta suy nghĩ lại cách thức chúng ta đang làm mọi thứ. Chúng tôi đang thận trọng áp dụng nó lên tất cả các sản phẩm của chúng tôi, có thể là tìm kiếm, quảng cáo, YouTube, hoặc Play. Và chúng tôi đang trong những ngày đầu, nhưng bạn sẽ thấy chúng tôi - theo một cách có hệ thống - áp dụng máy học trong tất cả các lĩnh vực này." Quan điểm này về TTNT ảnh hưởng rộng rãi đến cách phần mềm được tạo ra và cung cấp như thế nào đã được các giám đốc trong ngành công nghệ

chia sẻ rộng rãi, trong đó có cả Ginni Rometty của IBM, người đã nói rằng tổ chức của bà đánh cược công ty vào TTNT.

1.2. Trí tuệ nhân tạo là gì?

TTNT chưa có một định nghĩa thống nhất được các nhà nghiên cứu chấp nhận rộng rãi. Một số định nghĩa tương đối về TTNT như một hệ thống máy tính thể hiện hành vi thường được coi là đòi hỏi trí thông minh. Một số khác định nghĩa TTNT như một hệ thống có khả năng giải quyết hợp lý các vấn đề phức tạp hoặc có những hành động thích hợp để đạt được mục tiêu của mình trong bất cứ hoàn cảnh nào nó gặp phải trong thực tế.

Các chuyên gia đưa ra các phân loại khác nhau về các vấn đề và giải pháp của TTNT. Tài liệu TTNT phổ biến sử dụng phân loại sau đây: (1) hệ thống có tư duy như con người (ví dụ, mạng lưới thần kinh và các kiến trúc nhận thức); (2) hệ thống hành động như con người (ví dụ, thông qua các bài kiểm tra Turing qua xử lý ngôn ngữ tự nhiên; biểu diễn tri thức, suy luận tự động và học tập), (3) hệ thống tư duy hợp lý (ví dụ, các bộ giải logic, suy luận, và tối ưu hóa); và (4) các hệ thống hành động hợp lý (ví dụ, các phần tử phần mềm thông minh và robot đạt được các mục tiêu thông qua nhận thức, lên kế hoạch, lý luận, học tập, giao tiếp, ra quyết định, và hành động). Một cách độc lập, nhà đầu tư mạo hiểm Frank Chen chia nhỏ không gian vấn đề của TTNT thành năm nhóm chính: suy luận logic, thể hiện tri thức, lập kế hoạch và định hướng, xử lý ngôn ngữ tự nhiên, và nhận thức. Còn nhà nghiên cứu TTNT, Pedro Domingos gán cho các nhà nghiên cứu TTNT thành năm "nhóm" dựa trên các phương pháp họ sử dụng: "Nhóm biểu tượng" sử dụng hợp lý luận dựa trên biểu tượng trừu tượng, "Nhóm liên kết" xây dựng cấu trúc lấy cảm hứng từ bộ não của con người; "Nhóm tiến hóa" sử dụng phương pháp lấy cảm hứng từ sự tiến hóa của Darwin; "Nhóm Bayes" sử dụng suy luận theo xác suất; và "Nhóm analog" ngoại suy từ những trường hợp tương tự đã thấy trước đó.

Sự đa dạng của các vấn đề và giải pháp TTNT, và nền tảng của TTNT trong đánh giá hoạt động của con người và độ chính xác của thuật toán, gây khó khăn cho việc xác định rõ ràng sự khác biệt giữa những gì tạo nên TTNT và những gì không. Ví dụ, nhiều kỹ thuật được sử dụng để phân tích khối lượng lớn dữ liệu được các nhà nghiên cứu TTNT phát triển và bây giờ chúng được định nghĩa là các thuật toán và hệ thống "Dữ liệu lớn" (Big Data). Trong một số trường hợp, quan điểm có thể thay đổi, nghĩa là một vấn đề được coi là cần có cho TTNT trước khi nó đã được giải quyết, nhưng khi đã có giải pháp thì nó lại được coi là việc xử lý dữ liệu thông thường. Mặc dù các ranh giới của TTNT có

thể không chắc chắn và có xu hướng thay đổi theo thời gian, điều quan trọng là mục tiêu chính của nghiên cứu và các ứng dụng TTNT qua các năm đã được tự động hóa hoặc lặp lại hành vi thông minh.

1.3. Hiện trạng phát triển TTNT

Cho đến nay, đã có nhiều tiến bộ đáng kể trong TTNT Hẹp (Narrow AI), nhằm giải quyết các lĩnh vực ứng dụng cụ thể như chơi trò chơi chiến lược, dịch thuật, xe tự lái, và nhận diện hình ảnh. TTNT Hẹp hỗ trợ nhiều dịch vụ thương mại như lên kế hoạch chuyến đi, các hệ thống khuyến nghị người mua hàng và nhắm mục tiêu quảng cáo, và đang tìm kiếm các ứng dụng quan trọng trong chẩn đoán y tế, giáo dục và nghiên cứu khoa học. Tất cả những ứng dụng đều có lợi ích xã hội đáng kể và đã góp phần vào sức sống kinh tế của quốc gia.

TTNT phổ quát (đôi khi được gọi là trí tuệ phổ quát nhân tạo, hoặc AGI) là một hệ thống TTNT trong tương lai có khả năng thể hiện hành vi thông minh rõ ràng ít nhất là tiên tiến như một người trên phạm vi đầy đủ các nhiệm vụ nhận thức. TTNT hiện nay còn một khoảng cách rất lớn với những thách thức khó khăn hơn của TTNT phổ quát. Những nỗ lực để đạt được TTNT phổ quát bằng cách mở rộng các giải pháp TTNT hẹp đạt được rất ít tiến triển trong nhiều thập kỷ nghiên cứu. Cộng đồng chuyên gia nhất trí rằng TTNT phổ quát sẽ không thể đạt được sau ít nhất nhiều thập kỷ nữa.

Con người từ lâu đã suy đoán về những tác động của các máy tính ngày càng trở nên thông minh hơn con người. Một số dự đoán rằng một TTNT đủ thông minh có thể thực hiện nhiệm vụ phát triển các hệ thống thông minh hơn, thậm chí tốt hơn, và đến lượt chúng có thể được sử dụng để tạo ra các hệ thống còn thông minh hơn nữa, và cứ như vậy, dẫn đến suy luận về sự "bùng nổ trí thông minh" hoặc "kỳ dị", trong đó máy móc nhanh chóng vượt xa con người về trí tuệ.

Trong viễn cảnh đen tối của quá trình này, các máy móc siêu thông minh sẽ vượt quá khả năng của con người để hiểu và kiểm soát chúng. Nếu các máy tính có thể đưa vào kiểm soát nhiều hệ thống quan trọng, kết quả có thể là sự tàn phá, lúc đó con người không còn kiểm soát được vận mệnh của mình, thậm chí có thể tồi tệ đến mức bị tuyệt chủng. Kịch bản này từ lâu đã là chủ đề của các câu chuyện khoa học viễn tưởng, và những tuyên bố gần đây của một số nhà lãnh đạo ngành công nghiệp có ảnh hưởng đã nêu bật những nỗi lo sợ này.

Quan điểm tích cực hơn về tương lai của rất nhiều nhà nghiên cứu nhận thấy, sự phát triển của hệ thống thông minh sẽ làm việc như những người giúp

đỡ, trợ lý, huấn luyện, và đồng đội của con người, và được thiết kế để hoạt động một cách an toàn và có đạo đức.

Đánh giá của Ban công nghệ của Hội đồng KH&CN Quốc gia Hoa Kỳ (NSTC) là những lo ngại lâu dài về các TTNT phổ quát siêu thông minh sẽ có ít ảnh hưởng đến chính sách hiện hành. Các chính sách của Chính phủ liên bang nên áp dụng trong thời gian ngắn đến trung bình. Cách tốt nhất để xây dựng năng lực giải quyết các rủi ro tích lũy dài hạn là tấn công các rủi ro ít nghiêm trọng đã thấy hiện nay, như an ninh, bảo mật và rủi ro an toàn, trong khi đầu tư vào nghiên cứu về khả năng dài hạn và làm thế nào kiểm soát được những thách thức của chúng. Ngoài ra, do nghiên cứu và ứng dụng trong lĩnh vực này tiếp tục phát triển, những người thực hiện TTNT trong chính phủ và doanh nghiệp nên tiếp cận những tiến bộ với việc cân nhắc thỏa đáng các vấn đề xã hội và đạo đức dài hạn - chứ không chỉ các vấn đề kỹ thuật - mà những tiến bộ đó báo trước. Mặc dù thận trọng trước khả năng rằng một ngày nào đó siêu trí tuệ có hại trở thành hiện thực, nhưng những lo ngại này không phải là động lực chính của chính sách công về TTNT.

Máy học

Máy học là một trong những phương pháp tiếp cận kỹ thuật quan trọng nhất đối với TTNT và cơ sở của nhiều tiến bộ và các ứng dụng thương mại gần đây của TTNT. Máy học hiện đại là một quá trình thống kê bắt đầu với tập hợp dữ liệu và cố gắng để đưa ra một quy tắc hoặc quy trình thủ tục giải thích các dữ liệu hoặc có thể dự đoán dữ liệu tương lai. Phương pháp tiếp cận này - học từ dữ liệu - tương phản với phương pháp "hệ chuyên gia" cũ cho TTNT, trong đó các lập trình viên ngồi cùng với các chuyên gia về con người để tìm hiểu những quy tắc và tiêu chí được sử dụng để ra quyết định, và đưa những quy tắc này thành mã phần mềm. Hệ chuyên gia nhằm mục đích cạnh tranh với các nguyên tắc được các chuyên gia về con người sử dụng, trong khi máy học dựa trên các phương pháp thống kê để tìm một quy trình thủ tục ra quyết định hoạt động tốt trong thực tế.

Lợi thế của máy học là nó có thể được sử dụng ngay cả trong trường hợp không khả thi hoặc khó viết ra những quy tắc rõ ràng để giải quyết vấn đề. Ví dụ, một công ty điều hành một dịch vụ trực tuyến có thể sử dụng máy học để phát hiện những nỗ lực đăng nhập người sử dụng là lừa đảo. Công ty có thể bắt đầu với một tập dữ liệu lớn những lần đăng nhập trong quá khứ, với mỗi lần được gắn nhãn là gian lận hoặc không sử dụng lợi ích của gợi ý (hindsight). Dựa trên tập dữ liệu này, các công ty có thể sử dụng máy học để đưa ra một quy

tác áp dụng cho những lần đăng nhập trong tương lai có thể dự đoán những đăng nhập nào nhiều khả năng là lừa đảo và sẽ phải chịu thử thách vượt qua thêm các biện pháp bảo mật. Theo một nghĩa nào đó, máy học không phải là một thuật toán để giải quyết một vấn đề cụ thể, mà là một cách tiếp cận tổng quát hơn để tìm ra các giải pháp cho nhiều vấn đề khác nhau, dựa trên dữ liệu về chúng.

Để áp dụng máy học, một người thực hiện bắt đầu với một tập hợp các dữ liệu lịch sử, chia thành một bộ đào tạo và một bộ kiểm tra. Người thực hiện lựa chọn một mô hình, hay cấu trúc toán học đặc trưng cho một loạt các quy tắc ra quyết định có thể với các thông số điều chỉnh khác nhau. Có thể hình dung mô hình này là một chiếc "hộp" áp dụng quy tắc, và các thông số là núm điều chỉnh ở mặt trước của hộp điều khiển cho hộp hoạt động. Trong thực tế, một mô hình có thể có nhiều triệu thông số.

Người thực hiện cũng xác định một hàm mục tiêu sử dụng để đánh giá mức độ cần thiết của sản phẩm thu được là các kết quả của một sự lựa chọn các thông số cụ thể. Hàm mục tiêu thường sẽ có các phần cho mô hình thật phù hợp với bộ dữ liệu đào tạo, cũng như các phần cho việc sử dụng các quy tắc đơn giản hơn.

Đào tạo mô hình là quá trình điều chỉnh các thông số để tối đa hóa hàm mục tiêu. Đào tạo là bước kỹ thuật khó trong máy học. Một mô hình với hàng triệu thông số sẽ có các kết quả khả năng lớn hơn nhiều so với bất kỳ thuật toán nào, các thuật toán đào tạo thành công như vậy phải thông minh trong cách chúng khám phá không gian của các tập hợp thông số để tìm các tập hợp tốt nhất với một mức độ nỗ lực tính toán khả thi.

Khi một mô hình đã được đào tạo, người thực hiện có thể sử dụng bộ kiểm tra để đánh giá độ chính xác và hiệu quả của mô hình. Mục tiêu của máy học là để tạo ra một mô hình đào tạo sẽ khái quát hóa - nó sẽ chính xác không chỉ đối với các ví dụ bộ dữ liệu đào tạo, mà còn cả đối với các trường hợp trong tương lai mà nó chưa gặp trước đây. Mặc dù nhiều mô hình này có thể đạt được mức hiệu suất tốt hơn so với con người về các nhiệm vụ hẹp như gắn nhãn hình ảnh, nhưng ngay cả những mô hình tốt nhất cũng có thể thất bại theo những cách không thể lường trước. Ví dụ, đối với nhiều mô hình gắn nhãn hình ảnh có thể tạo ra các hình ảnh rõ ràng là nhiễu ngẫu nhiên đối với người, nhưng nó vẫn sẽ được dán nhãn một cách sai trái như là một đối tượng cụ thể với sự tự tin cao của mô hình đào tạo.

Một thách thức khác trong việc sử dụng máy học là nó thường không thể trích xuất hay tạo ra một sự giải thích đơn giản cho việc tại sao một mô hình đào tạo cụ thể có hiệu quả. Do các mô hình đào tạo có một số lượng rất lớn các thông số thay đổi - thường lên tới hàng trăm triệu hoặc nhiều hơn - việc đào tạo có thể mang lại một mô hình "hoạt động", theo nghĩa phù hợp với các dữ liệu, nhưng không nhất thiết phải là mô hình hoạt động đơn giản nhất. Trong việc ra quyết định của con người, bất kỳ độ mờ đục trong quá trình này thường là do không có đủ thông tin về lý do tại sao quyết định như vậy, bởi vì người ra quyết định có thể không thể trình bày được rõ lý do tại sao quyết định "cảm thấy đúng." Với máy học, tất cả mọi thứ về thủ tục ra quyết định này được biết đến với sự chính xác toán học, nhưng có thể đơn giản là quá nhiều thông tin để giải thích một cách rõ ràng.

Học sâu

Trong những năm gần đây, một vài trong số những tiên bộ ấn tượng nhất trong máy học thuộc nhánh học sâu (deep learning), còn được gọi là mạng học sâu. Học sâu sử dụng các cấu trúc lỏng lẻo lấy cảm hứng từ bộ não của người, bao gồm một tập hợp các đơn vị (hay "nơron thần kinh"). Mỗi đơn vị kết hợp một tập các giá trị đầu vào để sản sinh ra một giá trị đầu ra, rồi các giá trị này lại được chuyển cho các tế bào thần kinh khác ở tuyến tiếp theo. Ví dụ, trong ứng dụng nhận dạng hình ảnh, lớp các đơn vị thứ nhất có thể kết hợp các dữ liệu thô của ảnh để nhận dạng các mẫu đơn giản trong hình ảnh; lớp các đơn vị thứ hai có thể kết hợp kết quả của lớp đầu tiên để nhận ra các mẫu - của - mẫu; lớp thứ ba có thể kết hợp các kết quả của lớp thứ hai; và cứ tiếp diễn như vậy.

Các mạng học sâu thường sử dụng nhiều lớp - đôi khi đến hơn 100 - và thường sử dụng một số lượng lớn các đơn vị ở mỗi lớp, để cho phép nhận dạng được các mẫu chính xác và vô cùng phức tạp trong các dữ liệu. Trong những năm gần đây, các lý thuyết mới về làm thế nào để xây dựng và đào tạo các mạng học sâu đã xuất hiện, khi có các hệ thống máy tính nhanh hơn, lớn hơn, cho phép sử dụng các mạng học sâu lớn hơn nhiều. Sự thành công đáng kể của các mạng rất lớn này ở nhiều nhiệm vụ máy học đã làm bất ngờ cho một số chuyên gia, và là nguyên nhân chính của làn sóng nhiệt tình hiện nay của các nhà nghiên cứu và thực hành TTNT dân thân vào lĩnh vực máy học.

Tự chủ và tự động

TTNT thường được áp dụng cho các hệ thống có thể điều khiển thiết bị truyền động vật lý hoặc kích hoạt các vận hành trực tuyến. Khi TTNT tiếp xúc

với thế giới hàng ngày, sẽ phát sinh các vấn đề về tự chủ, tự động hóa, và hợp tác người-máy.

Tự chủ đề cập đến khả năng của một hệ thống hoạt động và thích ứng với các hoàn cảnh thay đổi với việc giảm hoặc không có sự kiểm soát của con người. Ví dụ, một chiếc xe tự chủ có thể tự đi đến đích của nó. Mặc dù các tài liệu tập trung nhiều vào xe ô tô và máy bay, tự chủ là một khái niệm rộng hơn nhiều bao gồm các kịch bản như giao dịch tài chính tự động và hệ thống sửa nội dung tự động. Tự chủ cũng bao gồm các hệ thống có thể chẩn đoán và sửa chữa lỗi rong hoạt động của chúng, chẳng hạn như xác định và sửa chữa các lỗi hỏng an ninh.

Tự động hóa xảy ra khi một máy hoạt động mà có thể trước đây đã được người thực hiện. Thuật ngữ này liên quan đến cả hoạt động vật lý và hoạt động tinh thần hay nhận thức có thể được thay thế bằng TTNT. Tự động hóa, và tác động của nó đối với việc làm, đã là hiện tượng xã hội và kinh tế quan trọng ít nhất là từ cuộc Cách mạng công nghiệp. Con người chấp nhận rộng rãi rằng TTNT sẽ tự động hóa một số công việc, nhưng còn nhiều tranh luận về việc liệu đây chỉ là chương tiếp theo trong lịch sử của tự động hóa hay TTNT sẽ ảnh hưởng đến nền kinh tế khác với các làn sóng tự động hóa trong quá khứ.

Hợp tác nhóm người – máy

Ngược lại với tự động hóa, nơi máy móc thay thế cho công việc của con người, trong một số trường hợp, máy móc sẽ bổ sung cho công việc của con người. Điều này có thể xảy ra như một hiệu ứng phụ của sự phát triển TTNT, hay một hệ thống có thể được phát triển riêng với mục tiêu tạo ra một nhóm người-máy. Các hệ thống có mục đích là để bổ sung khả năng nhận thức của con người đôi khi được gọi là tăng cường trí tuệ.

Trong nhiều ứng dụng, một nhóm người-máy có thể hiệu quả hơn so với khi hoạt động riêng lẻ, sử dụng những thế mạnh của một bên để bù đắp cho điểm yếu của bên kia. Một ví dụ là trong chơi cờ vua, khi máy tính yếu hơn thường có thể đánh bại một đấu thủ chơi máy tính mạnh hơn, nếu máy tính yếu hơn được là đồng đội với người, điều này là đúng, mặc dù các máy tính hàng đầu là những cầu thủ mạnh hơn nhiều so với bất kỳ người nào. Một ví dụ khác là trong X quang. Trong một nghiên cứu gần đây, với các hình ảnh của các tế bào hạch và được yêu cầu xác định có hay không các tế bào ung thư, phương pháp tiếp cận dựa trên TTNT có tỷ lệ lỗi 7,5 phần trăm, trong khi một nhà nghiên cứu bệnh nhân có tỷ lệ lỗi 3,5 phần trăm; phương pháp kết hợp, sử dụng

cả TTNT và con người, đã giảm tỷ lệ lỗi xuống còn 0,5 phần trăm, như vậy tỷ lệ lỗi giảm được 85 phần trăm.

2. CÁC ỨNG DỤNG CỦA TTNT CHO HÀNG HÓA CÔNG

Có nhiều lạc quan về TTNT và máy học là tiềm năng của chúng để cải thiện cuộc sống của người dân bằng cách giúp giải quyết một số thách thức và sự không hiệu quả lớn nhất của thế giới. Triển vọng của TTNT đã được so sánh với các tác động biến đổi của những tiến bộ trong điện toán di động. Những đầu tư của khu vực công và khu vực tư nhân vào NC&PT cơ bản và ứng dụng về TTNT đã bắt đầu gặt hái được những lợi ích lớn cho công chúng trong các lĩnh vực khác nhau như y tế, giao thông, môi trường, tư pháp hình sự, và hòa nhập kinh tế.

Tại Trung tâm y tế Walter Reed, Bộ Cựu chiến binh (Hoa Kỳ) đang sử dụng TTNT để dự đoán tốt hơn các biến chứng y tế và cải thiện điều trị các vết thương chiến tranh nặng, dẫn đến kết quả điều trị tốt hơn, chữa bệnh nhanh hơn, và chi phí thấp hơn. Cùng cách tiếp cận tổng quát - dự đoán các biến chứng để cho phép điều trị phòng ngừa - cũng giảm nhiễm khuẩn chéo bệnh viện tại Đại học Johns Hopkins. Với việc chuyển sang chuyển hồ sơ y tế điện tử hiện nay, phân tích tiên đoán các dữ liệu y tế có thể đóng một vai trò quan trọng trên nhiều lĩnh vực y tế như cấp thuốc chính xác và nghiên cứu ung thư.

Trong giao thông vận tải, các ứng dụng quản lý giao thông thông minh hơn bằng TTNT đang làm giảm thời gian chờ đợi, sử dụng năng lượng và phát thải tới 25 phần trăm ở một số nơi. Các thành phố đang bắt đầu tận dụng các loại điều độ và định tuyến đáp ứng, và liên kết nó với phần mềm lập kế hoạch và theo dõi tuyến cho giao thông công cộng để cung cấp sự tiếp cận kịp thời với giao thông công cộng thường có thể được nhanh hơn, rẻ hơn, và trong nhiều trường hợp, dễ tiếp cận hơn với công chúng.

Một số nhà nghiên cứu cũng đang sử dụng TTNT để cải thiện việc theo dõi sự di cư của động vật bằng cách sử dụng phần mềm TTNT phân loại hình ảnh để phân tích hình ảnh từ các trang mạng truyền thông xã hội công cộng. Phần mềm này có thể xác định cá thể động vật trong các bức ảnh và xây dựng một cơ sở dữ liệu di cư của chúng bằng cách sử dụng các dữ liệu và tem vị trí trên các bức ảnh. Tại hội thảo TTNT cho hàng hóa xã hội do OSTP tổ chức, các nhà nghiên cứu đã nói về việc xây dựng một số bộ dữ liệu có sẵn lớn nhất từ trước đến nay về các quần thể và di cư của cá voi và động vật lớn châu Phi, và

triển khai một dự án theo dõi "Internet của Rùa" để có được những hiểu biết mới về sự sống biển. Các đại biểu khác mô tả công dụng của TTNT để tối ưu hóa các chiến lược tuần tra của các đơn vị chống săn trộm, và để thiết kế các chiến lược bảo tồn môi trường sống để tối đa hóa sự đa dạng di truyền của các quần thể có nguy cơ tuyệt chủng.

Các thuyền buồm và mô-tô nước tự trị đã tuần tra đại dương mang theo các dụng cụ cảm biến tinh vi, thu thập dữ liệu về thay đổi trong băng ở Bắc Cực và các hệ sinh thái biển nhạy cảm trong các hoạt động sẽ là quá tốn kém và nguy hiểm cho các đội thủy thủ tàu. Mô-tô nước tự trị có thể sẽ rẻ hơn nhiều so với hoạt động tàu có người lái, và một ngày nào đó có thể được sử dụng để tăng cường dự báo thời tiết, quan trắc khí hậu, hoặc cảnh sát chống đánh bắt cá bất hợp pháp.

TTNT cũng có tiềm năng để cải thiện các khía cạnh của hệ thống tư pháp hình sự, bao gồm cả báo cáo tội phạm, cảnh sát, bảo lãnh, kết án và quyết định tạm tha.

Một số tổ chức học thuật ở Hoa Kỳ đã đưa ra sáng kiến sử dụng TTNT để giải quyết những thách thức kinh tế và xã hội. Ví dụ, Đại học Chicago đã tạo ra một chương trình học tập có sử dụng khoa học dữ liệu và TTNT để giải quyết những thách thức chung như tỷ lệ thất nghiệp và bỏ học. Đại học Nam California ra mắt Trung tâm Trí tuệ nhân tạo trong Xã hội, một học viện dành riêng để nghiên cứu lý thuyết trò chơi tính toán, máy học, lập kế hoạch tự động và các kỹ thuật lập luận đa yếu tố làm sao có thể giúp giải quyết các vấn đề xã hội có liên quan như vô gia cư. Trong khi đó, các nhà nghiên cứu tại Đại học Stanford đang sử dụng máy học trong các nỗ lực để giải quyết đói nghèo toàn cầu bằng cách sử dụng TTNT để phân tích các hình ảnh vệ tinh của các vùng có khả năng nghèo đói để nhận biết ở đâu cần giúp đỡ nhất.

Nhiều công dụng của TTNT cho hàng hóa công dựa vào các dữ liệu sẵn có có thể được sử dụng để đào tạo các mô hình máy học và kiểm tra hiệu năng của các hệ thống TTNT. Các cơ quan, tổ chức với các dữ liệu có thể được công bố mà không liên lụy đến sự riêng tư hoặc bí mật thương mại cá nhân có thể giúp cho sự phát triển của TTNT bằng cách để cho các nhà nghiên cứu có được dữ liệu. Chuẩn hóa các lược đồ và định dạng dữ liệu có thể làm giảm chi phí và khó khăn trong việc làm ra các bộ dữ liệu mới hữu ích.

3. TRÍ TUỆ NHÂN TẠO TRONG CHÍNH PHỦ HOA KỲ

Cơ quan quản lý tiến hành phát triển các chính sách và thực hành nội bộ sẽ tối đa hóa những lợi ích kinh tế và xã hội của TTNT và thúc đẩy đổi mới. Những chính sách và thực hành này có thể bao gồm:

- Đầu tư vào nghiên cứu và phát triển cơ bản và ứng dụng (NC&PT);
- Đóng vai trò là khách hàng ban đầu cho các công nghệ TTNT và các ứng dụng của chúng;
- Hỗ trợ các dự án thí điểm và tạo các cơ sở kiểm tra trong cài đặt thực tế;
- Sẵn sàng cung cấp các bộ dữ liệu cho công chúng;
- Tài trợ cho các giải thưởng khích lệ trong lĩnh vực TTNT;
- Xác định và theo đuổi những Thách thức Lớn để đặt ra các mục tiêu tham vọng nhưng khả thi cho TTNT;
- Tài trợ cho các đánh giá nghiêm ngặt các ứng dụng TTNT để đo lường tác động và hiệu quả kinh tế của chúng; và
- Tạo ra một chính sách, pháp luật và môi trường pháp lý cho phép đổi mới sáng tạo phát triển đồng thời bảo vệ công chúng khỏi bị thiệt hại.

Sử dụng TTNT trong Chính phủ Hoa Kỳ để cải thiện dịch vụ và lợi ích của người dân

Một thách thức trong việc sử dụng TTNT để cải thiện dịch vụ này là năng lực của Chính phủ để thúc đẩy và khai thác đổi mới sáng tạo phục vụ đất nước tốt hơn với các cơ quan rất khác nhau. Một số cơ quan được tập trung hơn vào đổi mới sáng tạo, đặc biệt là những cơ quan có ngân sách NC&PT lớn, nguồn nhân lực bao gồm nhiều nhà khoa học và kỹ sư, một nền văn hóa đổi mới sáng tạo và thử nghiệm, và hợp tác diễn ra mạnh mẽ với các nhà đổi mới sáng tạo trong khu vực tư nhân. Nhiều cơ quan cũng có các tổ chức được giao nhiệm vụ cụ thể với việc hỗ trợ nghiên cứu có rủi ro cao, lợi ích lớn (ví dụ, các cơ quan dự án nghiên cứu tiên tiến trong các Bộ Quốc phòng và Bộ Năng lượng, cũng như cộng đồng tình báo), và cung cấp tài trợ cho NC&PT trên toàn bộ phạm vi từ nghiên cứu cơ bản đến triển khai tiên tiến. Các cơ quan khác như NSF có nghiên cứu và phát triển là nhiệm vụ chính của họ.

Nhưng một số cơ quan, đặc biệt là những cơ quan chịu trách nhiệm về xóa đói giảm nghèo và tăng cường sự cơ động kinh tế và xã hội, có các khả năng liên quan, nguồn lực và chuyên môn ở mức khiêm tốn hơn. Ví dụ, trong khi

Viện Y tế Quốc gia (NIH) có ngân sách NC&PT hơn 30 tỷ USD, còn ngân sách NC&PT của Bộ Lao động chỉ là 14 triệu USD. Điều này hạn chế khả năng của Bộ Lao động để khám phá các ứng dụng của TTNT, như áp dụng công nghệ "giám hộ kỹ thuật số" dựa trên TTNT để tăng kỹ năng và thu nhập của người lao động không được đào tạo chính quy trong trường học.

Chương trình "Giáo dục chủ đạo" của DARPA được xem là một ví dụ về tiềm năng của TTNT để thực hiện và đẩy mạnh những ưu tiên của cơ quan. DARPA, có ý định giảm thời gian cần thiết cho tân binh Hải quân trở thành chuyên gia về kỹ năng kỹ thuật từ vài năm xuống còn vài tháng, đang tài trợ cho phát triển một giám hộ kỹ thuật số sử dụng TTNT để lập mô hình tương tác giữa một chuyên gia và một tân binh. Một đánh giá của chương trình giám hộ kỹ thuật số kết luận rằng các tân binh Hải quân sử dụng giám hộ kỹ thuật số để trở thành các nhà quản trị hệ thống CNTT luôn làm tốt hơn các chuyên gia hải quân có 7-10 năm kinh nghiệm trong cả hai bài kiểm tra viết lý thuyết và thực hành giải quyết vấn đề thực tế.

Bằng chứng sơ bộ dựa trên các dự án thí điểm giám hộ kỹ thuật số cũng cho thấy rằng những người lao động đã hoàn thành chương trình đào tạo có sử dụng giám hộ kỹ thuật số có nhiều khả năng có được công việc công nghệ cao làm tăng đáng kể thu nhập của họ. Những sự gia tăng này dường như lớn hơn nhiều so với các tác động của các chương trình phát triển lực lượng lao động hiện tại. Lý tưởng nhất là những kết quả này sẽ được xác nhận bằng các thử nghiệm ngẫu nhiên được thực hiện một cách độc lập. Hiện nay, chi phí phát triển giám hộ kỹ thuật số còn cao, và chưa có phương pháp lặp để phát triển các giám hộ kỹ thuật số hiệu quả. Các nghiên cứu cho phép xuất hiện một ngành công nghiệp có sử dụng các phương pháp tiếp cận TTNT như giám hộ kỹ thuật số có khả năng có thể giúp nhân viên có được các kỹ năng yêu cầu.

4. TRÍ TUỆ NHÂN TẠO VÀ QUY ĐỊNH

Trí tuệ nhân tạo có các ứng dụng trong nhiều sản phẩm, chẳng hạn như xe hơi và máy bay, thuộc đối tượng được quy định thiết kế để bảo vệ công chúng khỏi tác hại và đảm bảo sự công bằng trong cạnh tranh kinh tế. Sự kết hợp TTNT vào các sản phẩm này sẽ ảnh hưởng đến các quy định phương pháp tiếp cận có liên quan như thế nào? Nói chung, phương pháp tiếp cận quy định của các sản phẩm có kèm theo TTNT để bảo vệ an toàn công cộng phải được thông báo bằng sự đánh giá các khía cạnh rủi ro mà sự bổ sung TTNT có thể làm giảm

đi, cùng với những khía cạnh rủi ro có thể tăng lên. Nếu các nguy cơ nằm trong phạm vi quy chế quản lý hiện hữu, cần có các cuộc thảo luận chính sách bắt đầu bằng cách xem xét liệu các quy định hiện hành đã giải quyết thỏa đáng nguy cơ đó, hoặc liệu chúng có cần phải điều chỉnh với sự bổ sung TTNT. Ngoài ra, những nơi việc đáp ứng quy định bổ sung TTNT đe dọa đến tăng chi phí thực thi hoặc làm chậm sự phát triển hay áp dụng những đổi mới có lợi, các nhà hoạch định chính sách cần xem xét các phản ứng đó liệu có thể được điều chỉnh để giảm chi phí và rào cản đối với đổi mới sáng tạo mà không ảnh hưởng xấu đến an toàn hoặc sự công bằng trên thị trường.

Các nhà bình luận nhận định rằng một quy định rộng về nghiên cứu và sử dụng TTNT vào thời điểm này là không khôn ngoan. Thay vào đó, họ cho rằng các mục đích và cấu trúc của các quy định hiện hành là đủ, và chỉ cần điều chỉnh quy định hiện hành khi cần thiết có tính đến các ảnh hưởng của TTNT. Ví dụ, các nhà bình luận nói rằng quy định xe mô tô phải phát triển để tính đến sự xuất hiện dự kiến của xe lái tự động, và rằng cần có sự phát triển trong các cấu trúc hiện tại của quy định an toàn xe hơi. Để làm điều đó, các cơ quan phải luôn nghĩ đến những mục đích và mục tiêu cơ bản của quy định bảo vệ hàng hóa công, trong khi vẫn tạo ra không gian cho đổi mới sáng tạo và tăng trưởng trong TTNT.

Quy định công nghệ hiệu quả như TTNT đòi hỏi nội bộ các cơ quan phải có chuyên môn kỹ thuật để giúp hướng dẫn việc xây dựng và ban hành quy định. Cần có sự tham gia của chuyên gia cấp cao trong các cơ quan quản lý, và ở tất cả các giai đoạn của quá trình quản lý. Các mô hình chỉ định và trao đổi nhân sự (ví dụ như các cơ quan thuê nhân sự) có thể được sử dụng để phát triển một lực lượng lao động quốc gia với những quan điểm đa dạng hơn về hiện trạng phát triển công nghệ. Ví dụ về một cơ quan như vậy là Chương trình Di chuyển Luật Nhân sự liên chính phủ (IPA), trong đó quy định việc chuyển nhượng tạm thời nhân viên giữa Chính phủ liên bang và chính quyền bang và địa phương, các trường cao đẳng và đại học, các trung tâm NC&PT do liên bang tài trợ, và các tổ chức đủ điều kiện khác. Nếu sử dụng một cách chiến lược, chương trình IPA có thể giúp các cơ quan đáp ứng nhu cầu của họ cho những vị trí khó tuyển người và tăng khả năng sử dụng những người có các nền tảng kỹ thuật đa dạng. Các nhân viên liên bang theo các nhiệm vụ IPA có thể vừa là nhà tuyển dụng vừa làm đại sứ cho lực lượng lao động liên bang. Ví dụ, nhân viên cơ quan gửi đến các trường cao đẳng và đại học là các giảng viên có thể truyền cảm hứng cho học sinh cân nhắc đến các việc làm của liên bang.

Tương tự như vậy, các chương trình luân chuyển nhân viên thông qua các công việc và lĩnh vực khác nhau có thể giúp nhân viên chính phủ tiếp thu được kiến thức và kinh nghiệm để thông báo quy định và chính sách, đặc biệt là khi chúng liên quan đến các công nghệ nổi bật như TTNT.

Nghiên cứu điển hình: Xe ô-tô và máy bay tự trị

Một ví dụ về những thách thức pháp lý liên quan đến cơ quan cập nhật các quy định cũ tính đến các sản phẩm mới dựa trên TTNT là công việc của Bộ Giao thông vận tải (DOT) về các phương tiện vận tải tự động và hệ thống máy bay không người lái (UAS, hay "drone"). Trong Bộ Giao thông vận tải, xe hơi tự động được quy định bởi Cục quản lý an toàn giao thông đường cao tốc Mỹ (NHTSA) và máy bay được quy định bởi Cục quản lý hàng không liên bang (FAA).

Triển vọng của tự trị

Các ứng dụng của TTNT cho xe ô-tô và máy bay đã nằm trong sự hình dung của công chúng. Các xe hơi mới ngày nay có các tính năng hỗ trợ điều khiển dựa trên TTNT như tự đỗ xe và các hệ thống điều khiển hành trình tiên tiến giữ một chiếc xe trong làn đường của mình và điều chỉnh tốc độ dựa trên các xe xung quanh. Các xe thử nghiệm hoàn toàn tự động dưới sự giám sát của con người đã có thể được thấy chạy trên đường. Các chuyên gia nhất trí là công nghệ xe tự động dần dần sẽ an toàn hơn so với xe có người lái và đến một ngày nào đó có thể tránh được hầu hết trong số hàng chục ngàn tai nạn tử vong xảy ra hàng năm trên đường bộ và đường cao tốc quốc gia.

Các xe tự động cũng cung cấp khả năng cơ động cao hơn cho những người cao tuổi và khuyết tật, những người không thể lái xe. Các phương thức giao thông vận tải mới này có thể cung cấp cơ hội hòa nhập chưa từng có cho các cộng đồng bị cô lập khỏi các dịch vụ thiết yếu như công ăn việc làm, chăm sóc sức khỏe, và cửa hàng tạp hóa. Một hệ thống xe tự động được thiết kế tốt có thể dự đoán và tránh va chạm cũng có thể làm giảm đáng kể lượng khí thải giao thông liên quan đến tiêu thụ năng lượng. Cơ quan quản lý đang thực hiện các bước đi để biến triển vọng này thành hiện thực, trong đó có đề xuất đầu tư 3,9 tỷ USD trong Ngân sách năm tài chính 2017 của Bộ Giao thông vận tải cho nghiên cứu, phát triển và triển khai xe ô-tô tự động và kết nối, để đảm bảo rằng Hoa Kỳ duy trì vị trí dẫn đầu trong công nghệ xe hơi tự động.

Hướng vào không trung, kể từ đầu những năm 1990, các hệ thống máy bay không người lái (UAS) thương mại đã hoạt động ở mức độ hạn chế trong

Hệ thống không phận quốc gia (NAS). Cho đến gần đây, UAS chủ yếu là hỗ trợ các hoạt động của chính phủ, chẳng hạn như các hoạt động an ninh quân sự và biên giới. Nhưng trong những năm gần đây, các ứng dụng tiềm năng đã nhanh chóng mở rộng, bao gồm chụp ảnh trên không, khảo sát đất và cây trồng, theo dõi cháy rừng, ứng phó với thiên tai, và kiểm tra cơ sở hạ tầng quan trọng. Một số cơ quan chính phủ đang sử dụng UAS để nâng cao việc thực hiện các nhiệm vụ của họ, và hàng ngàn người Mỹ đã nhận được giấy phép cần thiết từ Cục Hàng không liên bang (FAA) cho các hoạt động UAS thương mại, một quá trình diễn ra nhanh chóng theo "Quy định máy bay không người lái nhỏ" của FAA có hiệu lực trong tháng 8/2016 và Dịch vụ đăng ký Máy bay UAS nhỏ dịch vụ của FAA triển khai tháng năm 12/2015. FAA ước tính số UAS đăng ký để sử dụng thương mại sẽ vượt quá 600.000 vào tháng 8/2017.

Một ước tính về tác động kinh tế của việc đưa UAS vào không phận dự đoán UAS sẽ tạo ra giá trị kinh tế lên tới trên 13,6 tỷ USD trong ba năm đầu tiên tích hợp vào nền kinh tế, với sự tăng trưởng bền vững trong những năm tiếp theo. Một nghiên cứu năm 2013 từ Hiệp hội Quốc tế Hệ thống xe không người lái dự đoán rằng ngành công nghiệp thương mại bay không người lái (drone) có thể tạo ra hơn 82 tỷ USD cho nền kinh tế Hoa Kỳ và tạo ra hơn 100.000 việc làm mới trong vòng 10 năm tới. Doanh thu thuế đối với các bang được dự đoán sẽ tăng hơn 482 triệu USD trong thập kỷ đầu tiên sau khi tích hợp vào nền kinh tế.

Đảm bảo an toàn

Để nhận thấy những lợi ích tiềm năng của các công nghệ đầy hứa hẹn này, chính phủ cần thực hiện các bước đi để đảm bảo sự an toàn của không phận và đường giao thông, trong khi tiếp tục nuôi dưỡng một nền văn hóa đổi mới sáng tạo và tăng trưởng. Hoa Kỳ có hệ thống hàng không an toàn nhất và phức tạp nhất trên thế giới, và công chúng dựa trên sự giám sát của FAA để thiết lập các tiêu chuẩn an toàn. Tiêu chuẩn an toàn xe mô tô Liên bang (FMVSS) đặt ra các yêu cầu cho các nhà sản xuất để phát triển các xe an toàn, và Cục quản lý an toàn giao thông đường cao tốc Hoa Kỳ (NHTSA) có thẩm quyền thu hồi các xe trong trường hợp xảy ra rủi ro bất hợp lý về an toàn. Mặc dù có cơ hội đáng kể để giảm tử vong trên đường bộ và đường cao tốc, nhưng hiện nay vẫn có khoảng một người tử vong trên mỗi 100 triệu dặm xe lăn bánh. Bằng hoặc vượt quá hiệu suất đó đối với xe tự động là một thách thức lớn.

Áp dụng các kỹ thuật TTNT trong các môi trường đòi hỏi an toàn cao như vậy đặt ra một số thách thức. Đầu tiên là cần thiết để biên dịch các trách nhiệm

của con người trong khi lái xe hay máy bay để chuyển vào phần mềm. Không giống như trong một số ứng dụng thành công khác của TTNT hẹp, không có mô tả ngắn gọn cho nhiệm vụ thao tác các xe đường bộ hoặc trên không. Mỗi thao tác trong số này đều là đa diện, có trách nhiệm bao gồm cả chỉ dẫn xe, phát hiện và tránh chướng ngại vật, và xử lý các hồng học vật lý như xệt lốp. Trong khi các nhiệm vụ nhỏ như định hướng hoặc một số loại nhận thức có thể phù hợp với các giải pháp TTNT hẹp hiện có, nhưng việc tích hợp và ưu tiên các nhiệm vụ này có thể chưa được thực hiện. Mặc dù có ý thức tuân thủ luật giao thông, nhưng người lái xe có kinh nghiệm có thể lẩn qua 2 vạch vàng (vạch chia làn giao thông cấm xe lẩn qua) để tránh một tai nạn hoặc vượt một xe va chạm. Tuy những tình huống như vậy có thể hiếm, nhưng không thể bỏ qua - đơn giản để những lỗi xảy ra ít nhất là không thường xuyên như đối với người lái xe, một hệ thống phải xử lý thành công nhiều trường hợp hiếm hoi như vậy.

Đối với các hệ thống dựa trên máy học, cần phải đưa được các trường hợp hiếm gặp có ảnh hưởng vào thiết kế và thử nghiệm hệ thống. Các phương pháp tiếp cận máy học có thể tin cậy sẽ xử lý tình huống chính xác nếu một trường hợp tương tự có trong bộ dữ liệu đào tạo. Thách thức là làm thế nào để xây dựng một bộ dữ liệu bao gồm đầy đủ các trường hợp hiếm gặp góp phần vào nguy cơ xảy ra tai nạn. Hàng không thương mại có các cơ chế chia sẻ dữ liệu sự cố và an toàn trên toàn ngành, nhưng việc báo cáo có thể không phải là thói quen của những người điều hành hệ thống bay không người lái (UAS) mới được ủy nhiệm, những người chưa quen với văn hóa an toàn và trách nhiệm của ngành công nghiệp hàng không truyền thống. Hiện chưa có hệ thống tương tự trong ngành công nghiệp ô-tô, chỉ có những tai nạn chết người được báo cáo, và việc thu thập và báo cáo thông tin an toàn giao thông khác được thực hiện, nếu có, theo cách thức không giống nhau ở cấp tiểu bang hay địa phương. Việc thiếu dữ liệu báo cáo thống nhất làm tăng số dặm hoặc giờ hoạt động cần thiết để thiết lập hệ thống an toàn, là một trở ngại để tiếp cận các phương pháp TTNT nhất định đòi hỏi nhiều thử nghiệm để xác nhận.

Để tạo điều kiện kiểm tra an toàn, FAA đã chỉ định sáu Điểm Kiểm tra UAS trên toàn quốc và cấp phép cho các hoạt động UAS trong các địa điểm này. Các hoạt động tại các địa điểm bao gồm một dự án để mở rộng nghiên cứu nhiều năm của NASA về quản lý giao thông UAS (UTM) để xác định các yêu cầu hoạt động với quy mô lớn ngoài tầm nhìn trong không phận tầm thấp. Tương tự như vậy, các cơ sở kiểm tra xe trên mặt đất như Phi công Kết nối và triển khai các phương tiện tự động ở Columbus, Ohio, người chiến thắng dự án

Thách thức Thành phố Thông minh năm 2016 trị giá 40 triệu USD của Bộ Giao thông vận tải, sẽ cung cấp các dữ liệu cơ sở và tương tác phong phú cho các nhà nghiên cứu TTNT.

Thay đổi các quy định hiện hành

Mặc dù các phương pháp quản lý không phận và đường cao tốc của quốc gia khác nhau, nhưng các phương pháp tiếp cận tích hợp xe và máy bay tự chủ cùng có một mục tiêu chung: cả FAA và NHTSA đang cố gắng thiết lập các khuôn khổ nhanh nhẹn và linh hoạt, đảm bảo sự an toàn đồng thời khuyến khích sự đổi mới sáng tạo.

Đối với quy định không phận, một bước tiến quan trọng hướng tới cho phép tích hợp UAS an toàn vào vùng trời là việc ban hành quy tắc cuối cùng của Phần 107, hay "UAS nhỏ," của FAA, có hiệu lực vào ngày 29/8/2016. Lần đầu tiên, quy tắc cho phép các chuyến bay UAS dưới 55 pounds ngoài mục đích giải trí. Quy định này hạn chế cho các chuyến bay ban ngày, ở độ cao 400 feet (120m) hoặc thấp hơn, với những máy bay được kiểm soát bởi người được cấp giấy phép điều khiển và trong tầm nhìn trực tiếp của người điều khiển. Việc bay trên khu vực có người là không được phép. Các quy tắc tiếp theo đang được dự kiến, nhằm nới lỏng những hạn chế khi kinh nghiệm và dữ liệu cho thấy làm điều đó như thế nào cho an toàn. Đặc biệt, Bộ Giao thông vận tải hiện đang phát triển một Thông Báo Đề xuất Luật đề xuất một cơ chế cho một số loại "UAS siêu nhỏ" để tiến hành các hoạt động trên các vùng có người, với một quy tắc về mở rộng các hoạt động tiếp theo dự kiến.

FAA chưa công bố công khai một lộ trình rõ ràng cho quy định cho phép các chuyến bay tự động hoàn toàn. Mặc dù sự tích hợp an toàn máy bay tự động vào vùng trời sẽ là một quá trình phức tạp, FAA đang chuẩn bị cho một tương lai công nghệ không quá xa xôi, trong đó máy bay tự động và có phi công bay cùng nhau trong một hệ thống không phận tích hợp liên tục.

Phương pháp tiếp cận mới để quản lý vùng trời cũng có thể bao gồm tăng cường hệ thống kiểm soát không lưu dựa trên TTNT. Mật độ giao thông hàng không dự kiến trong tương lai và sự đa dạng của hoạt động này là dường như không khả thi trong kiến trúc quản lý vùng trời hiện nay, do những hạn chế hiện hành về tích hợp trên không/mặt đất, và sự phụ thuộc vào liên lạc người-người trong hoạt động trên không và mặt đất. Chi phí chậm chuyến bay trong vận tải hàng không của Mỹ năm 2007, dữ liệu năm gần nhất, được ước tính là 31,2 tỷ USD - con số đó có lẽ lớn hơn nhiều do lượng người đi máy bay đã tăng lên.

Mặc dù một số chuyến bay bị chậm trễ là không thể tránh khỏi do thời tiết và hạn chế khác, nhưng việc áp dụng các công nghệ hàng không mới, cho phép các chính sách, và nâng cấp cơ sở hạ tầng có thể tăng đáng kể hiệu quả của hoạt động trong vùng trời của Hoa Kỳ. Các giải pháp này bao gồm TTNT và các cấu trúc dựa trên máy học có tiềm năng để phục vụ tốt hơn những người sử dụng không phận trên phạm vi lớn, bao gồm máy bay có người lái và không người lái, và sử dụng không phận một cách hiệu quả hơn mà không làm mất an toàn. Sự phát triển và triển khai các công nghệ đó sẽ giúp đảm bảo khả năng cạnh tranh toàn cầu cho người sử dụng không phận và các nhà cung cấp dịch vụ, đồng thời tăng độ an toàn và giảm chi phí.

Đối với giao thông trên mặt đất, bước đi quan trọng nhất hiện nay đang được tiến hành để thiết lập một khuôn khổ chung là Chính sách Xe tự động Liên bang được Chính quyền ban hành vào ngày 20/9/2016, với một số phần như:

- Hướng dẫn cho các nhà sản xuất, nhà phát triển và các tổ chức khác phác thảo "Đánh giá an toàn" 15 điểm cho việc thiết kế, phát triển, thử nghiệm và triển khai an toàn các phương tiện giao thông tự động hóa cao;

- Chính sách mẫu cấp tiểu bang, trong đó phân biệt rõ trách nhiệm của liên bang và tiểu bang và đề xuất các lĩnh vực chính sách cho các tiểu bang để xem xét, với mục tiêu tạo ra một khuôn khổ quốc gia thống nhất cho việc thử nghiệm và hoạt động của các phương tiện tự động, đồng thời dành chỗ cho các tiểu bang thực hiện thử nghiệm;

- Phân tích các công cụ pháp lý hiện hành mà NHTSA có thể sử dụng để hỗ trợ phát triển an toàn xe tự động, chẳng hạn như giải thích các quy định hiện hành cho phép sự linh hoạt phù hợp trong thiết kế, cung cấp những miễn trừ hạn chế cho phép thử nghiệm các thiết kế xe phi truyền thống, và đảm bảo rằng các phương tiện tự động không an toàn bị loại bỏ ra khỏi đường giao thông; và

- Thảo luận các công cụ và các bộ máy quản lý mà cơ quan này có thể xem xét tìm kiếm trong tương lai để hỗ trợ cho việc triển khai an toàn và hiệu quả các công nghệ cứu sinh mới và đảm bảo rằng các công nghệ triển khai trên đường được an toàn.

Bộ Giao thông dự định hướng dẫn và chính sách tiểu bang mẫu phải được cập nhật thường xuyên khi có dữ liệu mới và hoàn thành nghiên cứu.

5. NGHIÊN CỨU VÀ LỰC LƯỢNG LAO ĐỘNG

Chính phủ cũng có vai trò quan trọng trong việc thúc đẩy lĩnh vực TTNT bằng cách đầu tư vào nghiên cứu và phát triển, đang phát triển lực lượng lao động có tay nghề cao và đa dạng, và quản lý các tác động kinh tế của các công nghệ này khi chúng phát triển. *Kế hoạch chiến lược quốc gia nghiên cứu và phát triển trí tuệ nhân tạo* riêng biệt đã được công bố. Ở đây thảo luận các vấn đề chính sách bổ sung liên quan đến nghiên cứu và phát triển lực lượng lao động.

KẾ HOẠCH CHIẾN LƯỢC QUỐC GIA VỀ NGHIÊN CỨU VÀ PHÁT TRIỂN TRÍ TUỆ NHÂN TẠO

Ngày 3 tháng 5 năm 2016, Chính phủ Mỹ tuyên bố thành lập một Tiểu ban về Máy học và trí tuệ nhân tạo thuộc NSTC, để giúp điều phối các hoạt động liên bang về TTNT. Tiểu ban này, vào ngày 15/6/2016, đã chỉ đạo Tiểu ban Nghiên cứu và Phát triển Mạng và Công nghệ thông tin (NITRD) để xây dựng Kế hoạch chiến lược quốc gia về nghiên cứu và phát triển trí tuệ nhân tạo. Sau đó Nhóm chuyên trách về Trí tuệ nhân tạo thuộc NITRD đã được thành lập để xác định các ưu tiên chiến lược của Liên bang cho NC&PT TTNT, đặc biệt chú ý vào các lĩnh vực mà ngành công nghiệp không có khả năng giải quyết.

Kế hoạch chiến lược quốc gia về NC&PT Trí tuệ nhân tạo này thiết lập một loạt các mục tiêu cho nghiên cứu TTNT do liên bang tài trợ, cả các nghiên cứu diễn ra trong chính phủ cũng như nghiên cứu được liên bang tài trợ diễn ra bên ngoài chính phủ, chẳng hạn như trong học viện. Mục tiêu cuối cùng của nghiên cứu này là để tạo ra tri thức và công nghệ mới về TTNT cung cấp hàng loạt lợi ích tích cực cho xã hội, đồng thời giảm thiểu các tác động tiêu cực. Để đạt được mục tiêu này, điều này Kế hoạch Chiến lược NC&PT TTNT xác định các ưu tiên sau đây cho nghiên cứu TTNT do liên bang tài trợ:

Chiến lược 1: Đầu tư dài hạn cho nghiên cứu TTNT.

Chiến lược 2: Phát triển các phương pháp hiệu quả cho hợp tác người-TTNT.

Chiến lược 3: Hiểu rõ và giải quyết những tác động đạo đức, pháp luật và xã hội của TTNT.

Chiến lược 4: Đảm bảo an toàn và an ninh của hệ thống TTNT.

Chiến lược 5: Xây dựng các bộ dữ liệu công và môi trường chung cho đào tạo và thử nghiệm TTNT.

Chiến lược 6: Đo lường và đánh giá công nghệ TTNT thông qua các tiêu chuẩn và định chuẩn.

Chiến lược 7: Hiểu rõ hơn về các nhu cầu nguồn nhân lực NC&PT TTNT quốc gia.

5.1. Giám sát tiên bộ trong TTNT

Với những tác động tiềm tàng của TTNT, xã hội sẽ được hưởng lợi từ những phương pháp chính xác và kịp thời để theo dõi và dự báo những phát triển của TTNT. Một số dự án đã cố gắng dự đoán những TTNT tương lai. Ủy ban về tương lai TTNT lâu dài năm 2009 và Hội nghị Tương lai của TTNT năm 2015 đã mời các chuyên gia TTNT để dự đoán tương lai trong lĩnh vực của họ. Ngoài ra, chương trình Một Trăm Năm nghiên cứu về trí tuệ nhân tạo của ĐH Stanford có kế hoạch tiến hành "một loạt các nghiên cứu định kỳ về cách TTNT sẽ ảnh hưởng đến tự động hóa, an ninh quốc gia, tâm lý, đạo đức, pháp luật, bảo mật, dân chủ, và các vấn đề khác." Nghiên cứu đầu tiên của loạt này đã được công bố trong tháng 9 năm 2016

Một dòng nghiên cứu hữu ích tiềm năng là khảo sát các phán đoán chuyên gia theo thời gian. Ví dụ như, một cuộc khảo sát của các nhà nghiên cứu TTNT thấy rằng 80 phần trăm số người được hỏi tin rằng cuối cùng sẽ đạt được TTNT phổ quát ở mức như người, và một nửa cho rằng ít nhất 50 phần trăm có khả năng đạt được vào năm 2040. Hầu hết người được hỏi cũng cho rằng TTNT phổ quát cuối cùng sẽ vượt qua con người về trí thông minh chung. Mặc dù những dự đoán cụ thể rất không chắc chắn, nhưng các cuộc điều tra phán đoán chuyên gia như vậy rất hữu ích, đặc biệt là khi chúng được lặp đi lặp lại thường xuyên, đủ để đo lường những thay đổi trong phán đoán theo thời gian. Một cách để luận ra những phán đoán thường xuyên là tiến hành các "vòng thi dự báo" như những thị trường dự đoán, trong đó những người tham gia được khuyến khích tài chính để đưa ra dự đoán chính xác. Một nghiên cứu khác đã phát hiện ra rằng những phát triển công nghệ thường có thể dự đoán chính xác bằng cách phân tích các xu hướng trong dữ liệu công bố khoa học và bằng sáng chế.

Hiện nay, phần lớn các nghiên cứu cơ bản trong TTNT được tiến hành bởi các nhà khoa học và các phòng thí nghiệm thương mại thường xuyên công bố phát hiện của họ và xuất bản chúng trong các tài liệu nghiên cứu. Nếu sự cạnh tranh khiến các phòng thí nghiệm thương mại tăng cường tính bí mật, thì việc theo dõi tiến độ phát triển có thể trở nên khó khăn hơn, và sự quan tâm của công chúng quan tâm có thể tăng.

Một dòng nghiên cứu đặc biệt có giá trị là xác định các mốc thời gian có thể đại diện hay báo trước những bước tiến lớn về các năng lực của TTNT. Khi được hỏi trong các hội thảo và hội nghị cộng đồng về chính phủ làm thế nào có thể nhận ra những dấu mốc tiến bộ quan trọng trong lĩnh vực này, đặc biệt là

những tiến bộ chỉ ra sự xuất hiện của TTNT phổ quát có thể đến gần, các nhà nghiên cứu có xu hướng đưa ra ba kiểu trả lời liên quan:

1. *Thành công trong nhiệm vụ bao quát hơn, ít cấu trúc hơn*: Theo quan điểm này, quá trình chuyển đổi từ TTNT Hẹp hiện nay đến TTNT phổ quát cuối cùng sẽ diễn ra bằng cách dần dần mở rộng khả năng của các hệ thống TTNT Hẹp sao cho một hệ thống đơn có thể bao trùm phạm vi rộng lớn hơn các nhiệm vụ ít cấu trúc hơn. Ví dụ một môc quan trọng trong lĩnh vực này sẽ là một robot làm vệ sinh nhà cửa có khả năng như một người thực hiện đầy đủ các nhiệm vụ dọn dẹp nhà cửa thường xuyên.

2. *Sự hợp nhất các "phong cách" khác nhau của các phương pháp TTNT*: Theo quan điểm này, TTNT hiện dựa trên một tập hợp các phương pháp hoặc cách tiếp cận riêng biệt, mỗi phương pháp hữu ích cho các loại ứng dụng khác nhau. Con đường dẫn đến một TTNT phổ quát sẽ bao gồm sự thống nhất từng bước của các phương pháp này. Một dấu mốc của sự đột phá sẽ liên quan đến việc tìm ra một phương pháp duy nhất có khả năng giải quyết một miền lớn hơn các ứng dụng mà trước đây đòi hỏi nhiều phương pháp.

3. *Giải quyết những thách thức kỹ thuật cụ thể, chẳng hạn như chuyển giao học tập*: Theo quan điểm này, con đường dẫn đến TTNT phổ quát không nằm ở việc mở rộng dần các phạm vi, cũng như không phải trong sự thống nhất các phương pháp hiện có, mà ở sự tiến bộ trong giải quyết những thách thức lớn về kỹ thuật cụ thể, mở ra những cách thức mới tiến lên phía trước. Thách thức thường được trích dẫn nhiều nhất là chuyển giao học tập (transfer learning), trong đó có các mục tiêu tạo ra một thuật toán máy học có kết quả có thể được áp dụng rộng rãi (hoặc chuyển) vào một loạt các ứng dụng mới. Ví dụ, chuyển giao học tập có thể cho phép một mô hình được đào tạo để dịch Tiếng Anh sang Tiếng Tây Ban Nha, theo cách như vậy một mô hình kết quả có thể "chuyển giao" tri thức của mình cho các nhiệm vụ tương tự như dịch tiếng Trung sang tiếng Pháp, hoặc làm thơ bằng tiếng Nga, cho phép các nhiệm vụ mới được học nhiều hơn một cách nhanh chóng.

5.2. Hỗ trợ liên bang cho TTNT

Trong năm 2015, đầu tư của Chính phủ Mỹ cho NC&PT công khai trong các công nghệ có liên quan đến TTNT là khoảng 1,1 tỷ USD, với ước tính sơ bộ cho thấy có thể tăng lên 1,2 tỷ USD năm 2016. Trong suốt các cuộc hội thảo và cuộc các cộng đồng về TTNT do OSTP tổ chức, các quan chức chính phủ đã nhận được lời kêu gọi chính phủ đầu tư nhiều hơn cho nghiên cứu và phát

triển TTNT, từ các nhà lãnh đạo kinh doanh, các nhà công nghệ, và các nhà kinh tế.

Các nhà nghiên cứu hàng đầu về TTNT tỏ ra lạc quan về việc duy trì sự tiến bộ nhanh chóng gần đây trong TTNT và ứng dụng của nó vào phạm vi ứng dụng rộng lớn hơn bao giờ hết. Đồng thời họ nhấn mạnh rằng có rất nhiều vấn đề sâu chưa có câu trả lời, và không có con đường rõ ràng hướng tới TTNT phổ quát.

Các nhà nghiên cứu báo cáo rằng sự nhiệt tình và đầu tư vào nghiên cứu TTNT đã thủng thủng trong nhiều thập kỷ qua - một giai đoạn thấp đã được biết đến là "mùa đông TTNT" - và họ nhấn mạnh tầm quan trọng của sự đầu tư chắc chắn dựa trên lịch sử của những tiến bộ lớn trong khoa học máy tính lớn cần 15 năm hoặc hơn để chuyển từ kết quả phôi thai trong phòng thí nghiệm thành công nghệ chín muồi trong công nghiệp.

Một trường hợp có thể ủng hộ mạnh mẽ lợi ích của việc gia tăng tài trợ liên bang cho nghiên cứu trong TTNT. Phân tích của Hội đồng cố vấn kinh tế (CEA) chỉ ra rằng ngoài TTNT, xuyên suốt tất cả các lĩnh vực nghiên cứu, việc tăng gấp đôi hoặc gấp ba lần đầu tư nghiên cứu sẽ mang lại lợi ích cho quốc gia nhờ sự gia tăng trong tăng trưởng kinh tế. Mặc dù có thể không khả thi về tài chính để tăng kinh phí cho tất cả các nghiên cứu mức đó, nhưng một sự gia tăng có mục tiêu trong các khu vực có giá trị kinh tế và chiến lược cao có thể cung cấp nhiều lợi ích với tác động ngân sách nhỏ hơn nhiều so với mức tăng đồng loạt. TTNT được cho là một khu vực có tính đòn bẩy cao, và các cơ quan nghiên cứu báo cáo rằng cộng đồng nghiên cứu TTNT có thể hấp thụ hiệu quả một sự gia tăng đầu tư đáng kể, dẫn đến tiến bộ nhanh về TTNT và số lượng lớn cán bộ được đào tạo để thực hành TTNT. Trong một bài phát biểu tại một hội thảo TTNT ở thành phố New York trong tháng 7/2016, Chủ tịch CEA Jason Furman nói, "Chúng ta đã có đôi mới sáng tạo đáng kể về robot, TTNT và các khu vực khác trong thập kỷ qua. Nhưng chúng ta sẽ cần một tốc độ đổi mới nhanh hơn nhiều trong các lĩnh vực này để thực sự thúc đẩy gia tăng năng suất tiến lên phía trước," lưu ý rằng những lo lắng lớn nhất mà ông có về TTNT là "chúng ta không có đủ [TTNT]."

Để chắc chắn, khu vực tư nhân sẽ là động lực chính của sự tiến bộ về TTNT. Nhưng như hiện tại, sự thiếu đầu tư vào nghiên cứu cơ bản - nghiên cứu với những tầm nhìn lâu dài được tiến hành với mục đích duy nhất là thúc đẩy hơn nữa các nền tảng kiến thức khoa học - một phần bởi công ty tư nhân khó có thể thu được lợi nhuận từ việc đầu tư vào nghiên cứu như vậy trong một khung

thời gian hợp lý. Nghiên cứu cơ bản mang lại lợi ích cho tất cả mọi người, nhưng chỉ có công ty làm nghiên cứu chi trả các chi phí. Kết quả là, các tài liệu cho thấy các mức chi NC&PT hiện tại chỉ bằng một nửa đến một phần tư mức đầu tư NC&PT sẽ tạo ra mức tăng trưởng kinh tế tối ưu.

5.3. Phát triển và đa dạng lực lượng lao động

Sự tăng trưởng nhanh chóng của TTNT đã làm tăng đáng kể sự nhu cầu đối với những người có kỹ năng có liên quan để hỗ trợ và thúc đẩy sự phát triển của lĩnh vực này. Lực lượng lao động trong TTNT bao gồm các nhà nghiên cứu TTNT, những người thúc đẩy những tiến bộ cơ bản trong TTNT, số lượng lớn các chuyên gia, những người điều chỉnh các phương pháp TTNT cho các ứng dụng cụ thể, và một số lượng lớn hơn nhiều người sử dụng, những người vận hành những ứng dụng đó trong các thiết lập cụ thể. Đối với các nhà nghiên cứu, việc đào tạo TTNT vốn mang tính liên ngành, thường đòi hỏi một nền tảng vững vàng về khoa học máy tính, thống kê, logic toán học và lý thuyết thông tin. Đối với các chuyên gia, việc đào tạo thường đòi hỏi một nền tảng trong công nghệ phần mềm và trong các lĩnh vực ứng dụng. Đối với người dùng, sự hiểu biết các công nghệ TTNT là cần thiết để áp dụng các công nghệ TTNT một cách đáng tin cậy.

Vai trò của chính phủ

Sự thách thức về lực lượng lao động TTNT là một phần trong thách thức giáo dục khoa học, công nghệ, kỹ thuật và toán học (STEM) vẫn đang là một trọng tâm ưu tiên của NSTC, OSTP, và các cơ quan khác. Quỹ Khoa học quốc gia (NSF) và Bộ Giáo dục đang làm việc với khu vực tư nhân và các cơ quan chính phủ để nâng cao chất lượng, tính linh hoạt, và tác động của giáo dục, để giải quyết những mục tiêu như phát triển kinh tế bền vững, tăng sự bao quát và đa dạng, và các biện pháp cải thiện đo lường kết quả. Ủy ban Giáo dục Khoa học, Công nghệ và Toán học Giáo dục của NSTC (CoSTEM) tập hợp các cơ quan liên bang hỗ trợ các chương trình giáo dục STEM để điều phối các nỗ lực trên nhiều chủ đề, gồm cả giáo dục TTNT.

Kiến thức và giáo dục TTNT đang ngày càng được nhấn mạnh trong các chương trình giáo dục STEM liên bang. Một số vai trò quan trọng đối với chính phủ liên bang trong việc phát triển lực lượng lao động TTNT, bao gồm hỗ trợ sinh viên sau đại học, tài trợ cho các nghiên cứu về thiết kế và tác động của chương trình giảng dạy TTNT, và công nhận các chương trình giáo dục TTNT.

Vai trò của trường phổ thông và đại học

Việc tích hợp CNTT, khoa học dữ liệu, và các lĩnh vực liên quan trên toàn hệ thống giáo dục của quốc gia là cần thiết để phát triển một lực lượng lao động có thể giải quyết các ưu tiên quốc gia. Các cơ sở giáo dục đang xây dựng và phát triển các chương trình CNTT ở tất cả các cấp. Các trường đại học, cao đẳng và thậm chí cả các trường trung học đang mở rộng chương trình giảng dạy khoa học dữ liệu và CNTT, nhưng vẫn cần có thêm nhiều chương trình và giáo viên.

Sau đây là một số vai trò quan trọng đối với các tổ chức học thuật:

- Xây dựng và duy trì lực lượng các nhà nghiên cứu, bao gồm các nhà khoa học máy tính, thống kê, lập trình phần mềm và cơ sở dữ liệu, nhà quản lý, cán bộ thư viện, và nhân viên lưu trữ với chuyên ngành khoa học dữ liệu;

- Đào tạo lực lượng chuyên gia, bằng cách nhân mạnh các phương pháp CNTT trong các khóa học phát triển phần mềm, cung cấp các khóa học CNTT ứng dụng chứng minh các ứng dụng của CNTT đối với các lĩnh vực khác, và đưa những thách thức CNTT và khoa học dữ liệu được đặt ra bởi ngành công nghiệp, xã hội dân sự, và chính phủ vào các nghiên cứu trường hợp tích cực;

- Đảm bảo rằng các lực lượng người sử dụng CNTT có hiểu biết cần thiết về các hệ thống CNTT để đáp ứng các nhu cầu của người sử dụng và các tổ chức trên toàn bộ ngành công nghiệp, chính phủ, và học viện;

- Hỗ trợ đào tạo thông qua các trợ cấp ban đầu, tiền sinh hoạt phí cho phát triển chuyên môn, thực tập, học bổng, và chương trình trao đổi kinh nghiệm nghiên cứu mùa hè; và

- Tuyển dụng và giữ chân các tài năng bằng chế độ lương bổng hấp dẫn đối với các nhà nghiên cứu chuyên môn cao.

Các trường cao đẳng cộng đồng, trường đại học hai năm, và các chương trình chứng nhận đóng một vai trò quan trọng trong việc cung cấp cơ hội cho sinh viên và các chuyên gia để có được những kỹ năng cần thiết cho đầu tư khiêm tốn thời gian và tiền bạc của họ. Những cơ hội này có thể đặc biệt có liên quan đến những người lao động mở rộng các kỹ năng của họ, các cựu binh quay trở về lực lượng lao động, và người thất nghiệp tìm cách để nhập lại lực lượng lao động.

Một thế giới được TTNT hỗ trợ đòi hỏi một công dân thông thạo dữ liệu tức là có khả năng đọc, sử dụng, giải thích, và truyền đạt giao tiếp về dữ liệu, và tham gia vào các cuộc tranh luận chính sách về các vấn đề bị ảnh hưởng bởi TTNT. Giáo dục khoa học dữ liệu sớm ở tiểu học hay trung học có thể giúp cải thiện sự thông thạo dữ liệu toàn quốc, đồng thời cũng trang bị cho sinh viên các khái niệm khoa học dữ liệu tiên tiến hơn và chương trình sau trung học.

Giáo dục TTNT cũng là một thành phần của Chương trình Khoa học máy tính cho Mọi người, một sáng kiến của Tổng thống giúp tất cả các học sinh Mỹ từ mẫu giáo đến trung học được học khoa học máy tính và được trang bị những kỹ năng tư duy tính toán cần cho những nhà sáng tạo, chứ không chỉ là người tiêu dùng, trong nền kinh tế kỹ thuật số, và là những công dân tích cực trong một thế giới dựa trên công nghệ. Nền kinh tế Mỹ đang nhanh chóng chuyển đổi, và cả các nhà giáo dục lẫn lãnh đạo doanh nghiệp đang ngày càng nhận ra rằng khoa học máy tính là một kỹ năng "cơ bản mới" cần thiết cho các cơ hội kinh tế và cơ động xã hội. Khoa học máy tính cho Mọi người xây dựng dựa trên những nỗ lực của các bậc cha mẹ, giáo viên, các nhà lãnh đạo trường học, và các nhà lãnh đạo khu vực tư nhân trên khắp đất nước và là một cách để đáp ứng thách thức chuẩn bị một lực lượng lao động trong tương lai cho các nhu cầu của một nền kinh tế dựa trên TTNT.

5.4. Thách thức đa dạng

Tất cả các ngành đều đối mặt với thách thức làm thế nào để đa dạng hóa lực lượng lao động TTNT. Việc thiếu sự đa dạng giới tính và chủng tộc trong lực lượng lao động TTNT phản ánh sự thiếu đa dạng trong ngành công nghiệp công nghệ và lĩnh vực khoa học máy tính nói chung. Phát triển hết tiềm năng của người dân Mỹ, đặc biệt là trong các lĩnh vực STEM, trong kinh doanh, và trong công nghiệp công nghệ là một ưu tiên của Chính quyền này. Tầm quan trọng của các cá nhân bao gồm sự đa dạng về nền tảng, kinh nghiệm và bản sắc, đặc biệt là phụ nữ và các thành viên của các nhóm sắc dân tộc truyền thống ít hiện diện trong STEM, là một trong những thách thức quan trọng nhất và ưu tiên cao đối với khoa học máy tính và TTNT

Chỉ 18 phần trăm sinh viên tốt nghiệp khoa học máy tính hiện nay là phụ nữ, giảm từ mức đỉnh của 37 phần trăm trong 1984. Mặc dù thiếu dữ liệu nhân khẩu học báo cáo về lực lượng lao động TTNT, nhưng còn có một số thống kê. Tại Hội nghị Hệ thống xử lý thông tin thần kinh (NIPS) năm 2015, một trong những hội nghị lớn nhất trong năm về nghiên cứu TTNT, chỉ 13,7 phần trăm những người tham gia hội nghị là phụ nữ. Sau khi nhìn thấy sự đại diện thấp

tương tự tại một hội nghị máy thông minh, lúc đó là người nữ diễn giả duy nhất từ ngành công nghiệp, giám đốc điều hành và đồng sáng lập Textio, một công ty khởi nghiệp áp dụng TTNT cho soạn email thông báo việc làm và tuyển dụng, quyết định điều tra thêm ngôn ngữ tuyển dụng trong ngành công nghiệp. Khi công ty phân tích 78.768 danh sách vị trí công việc kỹ thuật, họ phát hiện thấy thông tin việc làm cho các kỹ sư phần mềm trong lĩnh vực máy thông minh đã có một số thiên vị nam giới nhiều hơn gấp đôi bất kỳ lĩnh vực nào khác.

Thách thức đa dạng không giới hạn trong giới tính. So với tỷ lệ dân cư của họ trong dân số Hoa Kỳ, người Mỹ gốc Phi, gốc Tây Ban Nha, và các thành viên của các nhóm thiểu số về chủng tộc và dân tộc khác có sự hiện diện vô cùng thấp trong lực lượng lao động STEM, khoa học máy tính, và trong lực lượng lao động công nghiệp công nghệ, kể cả trong các lĩnh vực TTNT. Nhiều ý kiến được chuyển tới OSTP thảo luận về thách thức đa dạng. Các bình luận tập trung vào tầm quan trọng của TTNT được tạo ra và sử dụng cho các nhóm dân cư đa dạng. Làm như vậy sẽ giúp tránh những hậu quả tiêu cực của phát triển TTNT tập trung trong phạm vi hẹp, bao gồm cả nguy cơ của những thành kiến trong phát triển các thuật toán, bằng cách tận dụng một phổ rộng hơn về kinh nghiệm, nền tảng, và quan điểm. Các chủ đề này cũng đã bao trùm rộng rãi trong các hội thảo công cộng. Một số nghiên cứu về tác động của sự thiếu đa dạng trong lực lượng lao động TTNT lên thiết kế công nghệ TTNT và tác động xã hội của TTNT. Sự phong phú của nghiên cứu đang tăng lên nhưng vẫn còn tụt hậu so với các tài liệu về nhu cầu phát triển lực lượng lao động TTNT rộng hơn.

6. TRÍ TUỆ NHÂN TẠO, TỰ ĐỘNG HÓA VÀ KINH TẾ

Hiệu quả kinh tế trung tâm của TTNT trong ngắn hạn sẽ là tự động hóa các nhiệm vụ mà không thể được làm tự động trước đây. Có một số tiền lệ lịch sử cho các làn sóng tự động hóa mới mà từ đó chúng ta có thể học hỏi, và một số cách thức trong đó TTNT sẽ khác nhau. Chính phủ phải hiểu những tác động tiềm năng đó để có thể đưa ra các chính sách và tổ chức sẽ hỗ trợ những lợi ích của TTNT, đồng thời giảm thiểu các thiệt hại.

Giống như những làn sóng đổi mới trong quá khứ, TTNT sẽ tạo ra cả những lợi ích và thiệt hại. Lợi ích chủ yếu của các làn sóng trước đây của tự động hóa là tăng năng suất; và làn sóng tự động hóa giờ đây cũng không khác. Ví dụ, một nghiên cứu về robot năm 2015 ở 17 quốc gia nhận thấy rằng chúng

bổ sung ước tính trung bình 0,4 điểm phần trăm vào tăng trưởng GDP hàng năm ở những nước đó từ năm 1993 đến năm 2007, chiếm hơn một phần mười tăng trưởng GDP tổng của các nước này trong thời gian đó.

Tuy nhiên, một vấn đề quan trọng phát sinh từ những làn sóng tự động hóa trước đây là tác động tiềm năng trên một số loại công việc và ngành nhất định và các tác động đó dẫn đến sự bất bình đẳng về thu nhập. Do TTNT có khả năng loại bỏ hoặc làm giảm tiền lương của một số công việc, đặc biệt là các việc làm đòi hỏi kỹ năng thấp và trung bình, các chính sách can thiệp có thể sẽ là cần thiết để đảm bảo rằng các lợi ích kinh tế TTNT được chia sẻ rộng rãi và bất bình đẳng được giảm đi và không trở thành hậu quả tồi tệ.

Các vấn đề chính sách kinh tế xuất hiện do tự động hóa dựa trên TTNT là quan trọng nhưng chúng được một nhóm công tác riêng tập trung giải quyết. Nhà Trắng sẽ tiến hành một nghiên cứu liên ngành bổ sung về các tác động kinh tế của tự động hóa nền kinh tế và phản ứng chính sách để xuất.

7. CÔNG BẰNG, AN TOÀN, VÀ QUẢN TRỊ

Khi các công nghệ TTNT được triển khai rộng rãi hơn, các chuyên gia kỹ thuật và nhà phân tích chính sách đã nêu lên những quan ngại về những hậu quả ngoài ý muốn. Việc sử dụng TTNT để đưa ra những quyết định liên quan về con người, thường thay thế các quyết định được thực hiện bởi con người và các tổ chức, dẫn đến những quan tâm về làm thế nào để đảm bảo công lý, công bằng và trách nhiệm giải trình - giống như những quan tâm được nêu lên trước đây về nội dung "dữ liệu lớn". Việc sử dụng TTNT để điều khiển thiết bị trong thế giới vật chất dẫn đến những lo ngại về an toàn, đặc biệt là khi các hệ thống này tiếp xúc với tất cả phức tạp của môi trường của con người.

Ở mức độ kỹ thuật liên quan đến những thách thức về sự công bằng và an toàn. Trong cả hai trường hợp, các nhà thực hành cố gắng tránh bị kỳ thị cố ý hay thất bại, để tránh những hậu quả ngoài ý muốn, và để tạo ra các bằng chứng cần thiết để các bên liên quan tự tin biện minh rằng thất bại không mong đợi là khó xảy ra.

7.1. Luật pháp, Công bằng, và trách nhiệm

Một chủ đề phổ biến trong các hội thảo Luật và Quản trị, TTNT cho Hàng hóa xã hội, và Tác động kinh tế và xã hội là sự cần thiết đảm bảo rằng TTNT thúc đẩy công lý và công bằng, và các quá trình dựa trên TTNT sẽ được các bên

liên quan chịu trách nhiệm. Vấn đề này đã được nhấn mạnh trước đó trong Báo cáo Dữ liệu lớn đầu tiên của Chính quyền xuất bản tháng 5/2014, và các báo cáo tiếp theo về Dữ liệu lớn, Các hệ thống thuật toán, Cơ hội, và các Quyền dân sự xuất bản tháng 5/2016.

Trong hệ thống tư pháp hình sự, một số trong những mối quan tâm lớn nhất với Dữ liệu lớn là thiếu dữ liệu và thiếu dữ liệu có chất lượng. TTNT cần có các dữ liệu tốt. Nếu dữ liệu không đầy đủ hoặc sai lệch, TTNT có thể làm trầm trọng thêm các vấn đề sai lệch. Điều quan trọng là bất cứ ai sử dụng TTNT trong bối cảnh tư pháp hình sự phải nhận thức được những hạn chế của dữ liệu hiện tại.

Các vấn đề tương tự có thể ảnh hưởng đến việc thuê tuyển lao động. Nếu một mô hình máy học được sử dụng để sàng lọc ứng viên xin việc, và nếu dữ liệu được sử dụng để đào tạo mô hình phản ánh các quyết định trong quá khứ có tính chất thiên vị, thì kết quả có thể để duy trì sự thiên vị quá khứ. Ví dụ, tìm kiếm những ứng viên giống như đã thuê trong quá khứ có thể thiên vị một hệ thống hướng vào thuê nhiều người giống như những người đã có trong nhóm, chứ không phải là xem xét các ứng cử viên tốt nhất với đầy đủ sự đa dạng của các ứng viên tiềm năng

Để đối phó với những mối lo ngại này, một số diễn giả hội thảo trao đổi về sự minh bạch cao hơn khi các công cụ TTNT được sử dụng cho những mục đích công cộng. Một diễn giả đã so sánh vai trò của TTNT với vai trò của các cơ quan hành chính trong việc ra quyết định công. Thẩm quyền được giao cho một cơ quan do chuyên môn đặc thù của cơ quan đó, nhưng việc giao thẩm quyền bị hạn chế bởi sự bảo vệ quy trình, các biện pháp thúc đẩy tính minh bạch và giám sát, và giới hạn về phạm vi thẩm quyền được giao. Một số diễn giả kêu gọi phát triển một lý thuyết tương tự về làm thế nào duy trì trách nhiệm khi trao quyền quyết định cho máy. Những mối quan tâm minh bạch không chỉ tập trung vào các dữ liệu và các thuật toán được sử dụng, mà còn về khả năng có một số hình thức giải thích mọi quyết định dựa trên TTNT.

Cũng tại các hội thảo, các chuyên gia TTNT cảnh báo rằng có những thách thức vốn có trong việc tìm hiểu, dự đoán, và giải thích hành vi của các hệ thống TTNT tiên tiến, do sự phức tạp của các hệ thống này và khối lượng lớn các dữ liệu mà chúng sử dụng.

Khó khăn trong việc hiểu các kết quả máy học là những mâu thuẫn với quan niệm sai lầm phổ biến mà các thuật toán phức tạp luôn luôn làm những gì

các nhà thiết kế ra chúng lựa chọn để chúng làm, và do đó sự thiên vị sẽ chui vào một thuật toán khi và chỉ khi bản thân các nhà phát triển bị thiên vị có ý thức hoặc vô thức. Điều chắc chắn đúng là một nhà phát triển công nghệ muốn sản xuất một thuật toán thiên vị có thể làm được như vậy, và sự thiên vị vô thức có thể khiến cho những người sử dụng không đủ nỗ lực để ngăn chặn thiên vị. Tuy nhiên, trong thực tế, các nhà phát triển thiên vị với những ý định tốt nhất có thể vô tình tạo ra các hệ thống với các kết quả thành kiến, bởi vì ngay cả các nhà phát triển một hệ thống TTNT có thể không hiểu nó đủ rõ để ngăn chặn các kết quả không mong đợi.

Moritz Hardt cho một ví dụ minh họa về thiên vị có thể vô tình xuất hiện như thế nào từ quá trình máy học. Ông mặc nhiên công nhận một mô hình máy học được đào tạo để phân biệt tên thật của một người dân với các tên giả. Mô hình này có thể xác định rằng một tên có nhiều khả năng là giả nếu phần tên riêng (firstname) của nó là độc đáo trong tập dữ liệu. Quy tắc này có thể tiên đoán toàn bộ dân số, bởi vì các tên giả nhiều khả năng được cách điệu và do đó độc đáo. Tuy nhiên, nếu có một nhóm dân tộc là một thiểu số dân cư nhỏ và có xu hướng sử dụng một bộ tên riêng khác với nhóm dân cư chiếm đa số, thì các tên khác biệt này nhiều khả năng là duy nhất trong mẫu, và do đó nhiều khả năng được xếp loại nhầm thành các tên giả. Tác động này phát sinh không phải vì bất kỳ sự đối xử đặc biệt nào với các tên của nhóm dân tộc thiểu số, và cũng không phải vì các dữ liệu đầu vào là không đại diện chung cho dân số, mà chỉ đơn giản vì nhóm dân tộc thiểu số đó quá ít.

Andrew Moore, Trưởng khoa Khoa học máy tính tại Đại học Carnegie Mellon, cung cấp một quan điểm về thách thức của TTNT và những hậu quả không lường trước tại hội thảo về công nghệ, an toàn, và kiểm soát TTNT. Ông lập luận rằng ngày nay, vì sự mờ của các thuật toán TTNT, cách hiệu quả nhất để giảm thiểu nguy cơ kết quả ngoài ý muốn là thông qua mở rộng thử nghiệm - điều cơ bản để có một danh sách dài của các loại quả xấu có thể xảy ra, và để loại trừ những kết quả đó bằng cách tạo ra nhiều kiểm tra chuyên biệt để tìm kiếm chúng.

Đào tạo đạo đức cho sinh viên và những người thực hành TTNT là một phần cần thiết của này giải pháp. Lý tưởng nhất, mỗi sinh viên học TTNT, khoa học máy tính, hay khoa học dữ liệu sẽ được tiếp xúc với chương trình giảng dạy và thảo luận về các vấn đề đạo đức liên quan và các chủ đề ninh. Tuy nhiên, chỉ đạo đức thôi là không đủ. Đạo đức có thể giúp những người sử dụng hiểu rõ trách nhiệm của mình đối với tất cả các bên liên quan, nhưng việc giáo dục đạo

đức cần phải được tăng cường với các tính năng kỹ thuật để đưa những ý định tốt vào thực tế bằng việc có những đề phòng kỹ thuật khi một hệ thống được chế tạo và thử nghiệm.

Khi những người sử dụng cố gắng làm cho hệ thống TTNT công bằng và có trách nhiệm hơn, sẽ có cơ hội để làm cho công nghệ trợ giúp thực hiện trách nhiệm chứ không phải là một rào cản đối với nó. Nghiên cứu để cải thiện khả năng giải nghĩa của các kết quả máy học là một ví dụ. Với một mô hình có thể giải nghĩa giúp mọi người hiểu một quyết định cho phép họ chất vấn các giả định và quy trình đằng sau nó. Có một số phương pháp kỹ thuật để nâng cao trách nhiệm và sự mạnh mẽ của các quyết định thuật toán phức tạp. Một hệ thống có thể được thử nghiệm "trong tự nhiên" bằng cách đưa nó vào các tình huống và quan sát hành vi của nó. Một hệ thống có thể bị kiểm tra hộp đen, trong đó nó tiếp nhận các đầu vào tổng hợp và hành vi của nó được theo dõi, cho phép kiểm tra hành vi trong các tình huống mà có thể không xảy ra tự nhiên.

Một số hoặc tất cả các chi tiết kỹ thuật của thiết kế hệ thống có thể được công bố, cho phép các nhà phân tích tái tạo nó và phân tích các khía cạnh hành vi bên trong của nó mà có thể khó mô tả chỉ bằng thử nghiệm. Trong một số trường hợp, có thể công bố thông tin giúp công chúng đánh giá rủi ro sai lệch của một hệ thống, trong khi giữ lại các thông tin khác về các hệ thống làm độc quyền hay tài sản riêng,

7.2. An toàn và kiểm soát

Tại các hội thảo, các chuyên gia TTNT nói rằng một trong những yếu tố chính hạn chế việc triển khai TTNT trong thế giới thực là sự quan ngại về an toàn và kiểm soát. Nếu những người sử dụng không có được sự tự tin hợp lý rằng hệ thống an toàn và được kiểm soát, sao cho việc triển khai hệ thống không tạo ra một nguy cơ không thể chấp nhận về các hậu quả xấu nghiêm trọng, thì hệ thống không thể và không nên được triển khai.

Một thách thức lớn về an toàn và kiểm soát là các hệ thống được chế tạo có thể an toàn chuyên từ "thế giới khép kín" trong các phòng thí nghiệm ra "thế giới mở" bên ngoài, nơi có thể xảy ra những điều không thể đoán trước. Trong thế giới mở, một hệ thống có khả năng gặp phải các đối tượng và tình huống không được dự tính khi nó được thiết kế và chế tạo. Thích ứng một cách êm thấm với những tình huống không lường trước là rất khó khăn nhưng cần thiết cho hoạt động an toàn.

Về chủ đề an toàn và khả năng dự báo trong TTNT, một số diễn giả trích dẫn một bài báo gần đây có tựa đề "Các vấn đề chắc chắn trong an toàn TTNT", và tác giả chính của bài báo đã phát biểu tại hội thảo về Công nghệ, An toàn, và Kiểm soát. Báo cáo sử dụng ví dụ một con robot tự trị dọn dẹp nhà cửa. Phần tổng quan của bài báo đưa ra một danh sách dài các loại vấn đề thực tế phát sinh trong việc chế tạo một robot như vậy hiệu quả và an toàn, được trích dẫn ở đây:

Tránh tác dụng phụ tiêu cực: Làm thế nào chúng ta có thể đảm bảo rằng con robot làm vệ sinh của chúng ta sẽ không làm phiền môi trường theo những cách tiêu cực khi theo đuổi các mục đích của nó, ví dụ bằng cách gõ lên chiếc bình để robot có thể làm vệ sinh nhanh hơn? Chúng ta có thể làm điều này mà không cần chỉ định bằng tay tất cả mọi thứ robot không nên làm phiền?

Tránh thưởng: Làm thế nào chúng ta có thể đảm bảo rằng các robot làm sạch không chơi chức năng thưởng được ban cho? Ví dụ, nếu chúng ta thưởng cho các robot đạt được một môi trường sạch không có vết bẩn, nó có thể vô hiệu hóa "con mắt" của mình sẽ không tìm thấy vết bẩn nào, hoặc phủ lên những chỗ bẩn bằng vật liệu mà nó không thể nhìn xuyên qua, hoặc chỉ đơn giản là trốn đi khi có người xung quanh để họ không thể nói cho nó về các loại bẩn mới.

Khả năng mở rộng giám sát: Làm thế nào chúng ta có thể đảm bảo hiệu quả rằng các robot làm sạch tôn trọng các khía cạnh của các mục tiêu quá đắt tiền để được thưởng xuyên đánh giá trong đào tạo? Ví dụ, nó nên loại bỏ những thứ không thuộc về bất kỳ ai, nhưng đặt sang một bên những thứ có thể thuộc về một ai đó (nó cần xử lý giấy gói kẹo vút lung tung khác với chiếc điện thoại di động vút lung tung). Việc hỏi những người liên quan xem họ có bị mất thứ gì có thể dùng kiểm tra về điều này, nhưng việc kiểm tra này có thể phải tương đối không thường xuyên, có thể robot tìm thấy cách để làm điều đúng đắn mặc dù thông tin hạn chế?

Thăm dò an toàn: Làm thế nào để chúng ta đảm bảo rằng robot dọn nhà không có những chuyển động thăm dò với những hậu quả rất xấu? Ví dụ, robot nên thử nghiệm với các chiến lược lau nhà, nhưng đưa một khăn lau ướt vào ổ điện là một ý tưởng tồi.

Khả năng chuyển đổi mạnh mẽ: Làm thế nào để chúng ta đảm bảo rằng các robot làm vệ sinh nhận biết được, và khả năng thực hiện mạnh mẽ, khi ở trong một môi trường khác với môi trường đào tạo của nó? Ví dụ, những gì nó

được học để làm sạch sàn nhà của nhà máy có thể nguy hiểm đối với một văn phòng.

Những ví dụ này minh họa cho việc "sự thông minh" của một hệ thống TTNT có thể chuyên sâu chứ không nông cạn: hệ thống này có thể có khả năng siêu phạm để phát hiện bụi bẩn và tối ưu hóa chiến lược lau chùi của nó, nhưng không biết tránh đưa khăn lau ướt lên một ổ cắm điện. Một cách để mô tả vấn đề phổ quát này là: làm thế nào chúng ta có thể cung cấp cho các máy thông minh sự nhận biết thông thường? Các nhà nghiên cứu cũng đang có sự tiến bộ chậm chạp về các loại vấn đề này.

Kỹ thuật an toàn TTNT

Một chủ đề phổ biến tại hội thảo Công nghệ, An toàn và Kiểm soát là nhu cầu kết nối các phương pháp TTNT trong thế giới mở với lĩnh vực kỹ thuật an toàn rộng lớn hơn. Kinh nghiệm trong việc chế tạo các loại hệ thống đặc biệt an toàn, chẳng hạn như máy bay, nhà máy điện, cầu và các loại xe, giúp ích rất nhiều cho những người sử dụng TTNT về xác minh và xác nhận, làm thế nào để xây dựng một trường hợp an toàn một công nghệ, làm thế nào để quản lý rủi ro, và làm thế nào trao đổi với các bên liên quan về rủi ro.

Hiện nay, việc thực hành TTNT, đặc biệt là ở các lĩnh vực máy học có tốc độ phát triển nhanh, có thể cũng đòi hỏi sự khéo léo như khoa học. Một số khía cạnh thực hành không được sự hẫu thuẫn của lý thuyết đã phát triển dựa vào phán đoán trực giác và thử nghiệm của những nhà thực hành. Đây không phải là điều bất thường ở các lĩnh vực công nghệ mới nổi, nhưng nó làm hạn chế việc áp dụng công nghệ này trong thực tế. Một số bên liên quan đã đề xuất yêu cầu phải phát triển TTNT vào lĩnh vực kỹ thuật trưởng thành hơn.

Khi các lĩnh vực kỹ thuật trưởng thành, chúng thường di chuyển từ giai đoạn "thủ công" ban đầu đặc trưng bởi sự sáng tạo dựa vào trực giác của những tài năng không chuyên và một tinh thần tự mày mò; đến giai đoạn thương mại thứ hai bao gồm những người thực hành có kỹ năng cao, cải tiến thực dụng, quy tắc ngón tay cái (kinh nghiệm) được chấp nhận rộng rãi, và tổ chức sản xuất để bán; đến giai đoạn trưởng thành tích hợp các phương pháp nghiêm ngặt hơn, các chuyên gia được đào tạo, lý thuyết được xây dựng hoàn chỉnh và sự chuyên môn hóa sản phẩm cao hơn. Hầu hết các lĩnh vực kỹ thuật, có lịch sử lâu hơn nhiều so với TTNT hiện đại, đã đạt đến giai đoạn trưởng thành.

Nói chung, các lĩnh vực kỹ thuật trưởng thành có thành công lớn hơn trong việc tạo ra những hệ thống có thể dự đoán được, đáng tin cậy, mạnh mẽ,

an toàn và đảm bảo. Việc tiếp tục tạo ra tiến bộ đưa đề TTNT trở thành một lĩnh vực kỹ thuật trưởng thành sẽ là một trong những điều kiện quan trọng để tạo ra các hệ thống an toàn, được kiểm soát và phức tạp hơn.

8. NHỮNG SỰ CÂN NHẮC VÀ AN NINH TOÀN CẦU

Ngoài những thách thức lâu dài của TTNT và các vấn đề cụ thể liên quan đến sự công bằng và an toàn, TTNT còn đặt ra những vấn đề chính sách liên quan trong quan hệ quốc tế, an ninh mạng, và quốc phòng.

8.1. Hợp tác quốc tế

TTNT đã là một chủ đề được quan tâm trong các cuộc thảo luận quốc tế gần đây khi các quốc gia, các tổ chức đa phương và các bên liên quan khác đã bắt đầu đánh giá những lợi ích và thách thức của TTNT. Đối thoại và hợp tác giữa các thực thể này có thể giúp thúc đẩy tiến bộ trong NC&PT TTNT và khai thác ứng dụng TTNT, đồng thời cũng giải quyết những thách thức liên quan. Đặc biệt, một số đột phá về TTNT là kết quả trực tiếp hoặc gián tiếp của hợp tác nghiên cứu liên quan đến con người, nguồn lực, và tổ chức ở nhiều quốc gia. Như với những chính sách kỹ thuật số khác, các nước sẽ cần phải cùng làm việc để xác định các cơ hội hợp tác và phát triển các khuôn khổ quốc tế sẽ giúp thúc đẩy NC&PT TTNT và giải quyết những thách thức. Hoa Kỳ, quốc gia đi đầu trong NC&PT TTNT, có thể tiếp tục đóng một vai trò then chốt trong việc phối hợp nghiên cứu toàn cầu thông qua đối thoại giữa các chính phủ với nhau và các đối tác.

Sự tham gia quốc tế là cần thiết để khám phá đầy đủ các ứng dụng của TTNT trong việc chăm sóc sức khỏe, tự động hóa trong sản xuất và công nghệ thông tin và truyền thông. Các ứng dụng TTNT cũng có khả năng để giải quyết các vấn đề toàn cầu như phòng chống thiên tai và ứng phó với biến đổi khí hậu, buôn bán động vật hoang dã, khoảng cách số, việc làm, và thành phố thông minh. Bộ Ngoại giao dự báo những mối quan tâm riêng tư, an toàn của xe tự lái, và tác động của TTNT lên các xu hướng lao động về lâu dài khi xem xét các lĩnh vực chính sách liên quan đến TTNT trong bối cảnh quốc tế.

Với sự hỗ trợ của những ưu tiên chính sách đối ngoại của Hoa Kỳ trong không gian này, bao gồm đảm bảo sự lãnh đạo quốc tế và khả năng cạnh tranh kinh tế của Hoa Kỳ, Chính phủ Hoa Kỳ đã tham gia vào NC&PT TTNT và các vấn đề chính sách trong các cuộc thảo luận song phương với các nước khác, bao gồm Nhật Bản, Hàn Quốc, Đức, Ba Lan, Vương quốc Anh và Italia, cũng như

tại các diễn đàn đa phương. vấn đề chính sách TTNT quốc tế và các tác động kinh tế của TTNT đã được nêu lên tại Liên Hiệp Quốc, G-7, Tổ chức Hợp tác và Phát triển Kinh tế (OECD), và Tổ chức Hợp tác kinh tế châu Á-Thái Bình Dương (APEC). Chính phủ Hoa Kỳ hy vọng TTNT sẽ là một chủ đề ngày càng được quan tâm hơn trong các cam kết quốc tế.

Hoa Kỳ đã cam kết hợp tác với các công ty và các tổ chức tiêu chuẩn liên quan, để tạo điều kiện thuận lợi cho phát triển các tiêu chuẩn quốc tế theo: dựa vào ngành công nghiệp; tình nguyện; theo hướng đồng thuận; và dựa trên các nguyên tắc minh bạch, công khai, và nhu cầu thị trường. Cách tiếp cận của Hoa Kỳ được chính thức hóa trong luật (NTTAA, PL 104-113) và chính sách (OMB Thông tư A-119) và tái khẳng định trong Chiến lược tiêu chuẩn Hoa Kỳ.

8.2. Trí tuệ nhân tạo và an ninh mạng

Trí tuệ nhân tạo Hợp ngày nay có những ứng dụng quan trọng trong an ninh mạng, và dự kiến sẽ đóng vai trò ngày càng tăng cho cả các biện pháp phòng thủ (phản ứng) và các biện pháp tấn công (chủ động).

Hiện nay, việc thiết kế và vận hành các hệ thống an toàn đòi hỏi một sự đầu tư lớn về thời gian và sự quan tâm của các chuyên gia. Tự động hoá công việc chuyên môn này, một phần hoặc hoàn toàn, có thể cho phép bảo mật mạnh mẽ trên một phạm vi rộng hơn nhiều của các hệ thống và các ứng dụng với chi phí thấp hơn đáng kể, và có thể làm tăng sự nhanh nhẹn của phòng vệ không gian mạng. Sử dụng TTNT có thể giúp duy trì phản ứng nhanh cần có để phát hiện và phản ứng với những tính huống đe dọa không gian mạng luôn phát triển. Có rất nhiều cơ hội cho TTNT và đặc biệt là các hệ thống máy học để giúp đối phó với sự phức tạp tuyệt đối của không gian mạng và hỗ trợ hiệu quả việc ra quyết định của con người để phản ứng với các cuộc tấn công mạng

Các hệ thống TTNT trong tương lai có thể thực hiện các phân tích dự đoán để lường trước các cuộc tấn công mạng bằng cách tạo ra các mô hình đe dọa động từ các nguồn dữ liệu có sẵn rất lớn, luôn biến đổi và thường không đầy đủ. Những dữ liệu này bao gồm các cấu trúc liên kết và trạng thái của các nút mạng, liên kết, thiết bị, kiến trúc, giao thức, và các mạng lưới. TTNT có thể là phương pháp tiếp cận hiệu quả nhất để diễn giải các dữ liệu này, chủ động xác định các lỗ hổng an ninh, và hành động để ngăn chặn hoặc giảm thiểu các cuộc tấn công trong tương lai.

Các kết quả cập nhật trong cuộc thi Thách thức lớn trong an ninh mạng (Cyber Grand Challenge-CGC) của DARPA chứng minh tiềm năng của phương

pháp này. CGC đã được thiết kế để thúc đẩy sự phát triển của các hệ thống tự chủ, tiên tiến có thể phát hiện, đánh giá, và các lỗ hổng phần mềm và lỗi trước khi đối thủ có cơ hội khai thác chúng. Để tiếp sức cho nghiên cứu và cuộc thi song song tiếp theo, tất cả các mã các hệ thống tự động tạo ra trong Chung kết CGC đã công bố như là mã nguồn mở cho phép những người khác giải mã và tìm hiểu chúng.

Các hệ thống TTNT cũng có nhu cầu an ninh mạng riêng của chúng. Các ứng dụng dựa trên TTNT cần thực hiện kiểm soát tốt an ninh mạng để đảm bảo tính toàn vẹn của dữ liệu và chức năng thực hiện, bảo vệ sự riêng tư và bảo mật, và duy trì tính sẵn sàng. Kế hoạch chiến lược NC&PT An ninh Mạng liên bang mới đây đã nêu bật nhu cầu "phát triển và vận hành các hệ thống an toàn bền vững." Những tiến bộ trong an ninh mạng sẽ rất quan trọng trong việc đưa ra các giải pháp TTNT an toàn và vững vàng chống lại các hoạt động độc hại trên mạng, đặc biệt là khi khối lượng và loại hình nhiệm vụ được tiến hành bởi chính phủ và các doanh nghiệp tư nhân sử dụng TTNT Hẹp tăng lên.

Cuối cùng, TTNT có thể hỗ trợ việc lập kế hoạch, điều phối, tích hợp, đồng bộ hóa, và chỉ đạo các hoạt động vận hành và bảo vệ mạng và hệ thống của của chính phủ Mỹ một cách hiệu quả, cung cấp sự hỗ trợ giúp cho các hoạt động an toàn của các mạng lưới và các hệ thống của khu vực tư nhân, và cho phép hành động phù hợp với tất cả luật pháp, quy định và các điều ước có thể áp dụng.

8.3. Trí tuệ nhân tạo trong các hệ thống vũ khí

Hoa Kỳ đã kết hợp tính tự chủ vào một số hệ thống vũ khí nhất định từ nhiều năm qua. Những cải tiến công nghệ này có thể cho phép sử dụng các hệ thống vũ khí có độ chính xác cao hơn và an toàn hơn, các hoạt động quân sự nhân đạo hơn. Vũ khí dẫn đường chính xác cho phép hoạt động được hoàn thành đòi hỏi ít vũ khí hơn và với thiệt hại tài sản ít hơn, và các loại xe điều khiển từ xa có thể giảm bớt rủi ro cho quân nhân do họ ở cách xa khu vực nguy hiểm. Tuy nhiên, việc đưa các hệ thống vũ khí ra ngoài sự kiểm soát trực tiếp của con người liên quan đến một số rủi ro và có thể nảy sinh các vấn đề pháp lý và đạo đức. Chìa khóa để kết hợp hơn nữa các hệ thống vũ khí tự động và bán tự động vào kế hoạch quốc phòng và cơ cấu lực lượng Hoa Kỳ là tiếp tục đảm bảo rằng tất cả các hệ thống vũ khí, bao gồm các hệ thống vũ khí tự động, đang được sử dụng một cách phù hợp với luật nhân đạo quốc tế. Ngoài ra, Chính phủ Hoa Kỳ cần tiếp tục thực hiện các bước thích hợp để kiểm soát sự phổ biến vũ

khí này, và làm việc với các đối tác và đồng minh để phát triển các tiêu chuẩn liên quan đến sự phát triển và sử dụng các hệ thống vũ khí như vậy.

Cụ thể, trong vài năm qua, các vấn đề liên quan đến sự phát triển của cái gọi là "Hệ thống vũ khí sát thương tự động" (Lethal Weapon Autonomous Systems-LAWS) đã được nêu lên bởi các chuyên gia kỹ thuật, các nhà đạo đức, và những người khác trong cộng đồng quốc tế. Hoa Kỳ đã tích cực tham gia các cuộc thảo luận quốc tế đang diễn ra về LAWS trong bối cảnh Hiệp ước về Một số vũ khí thông thường (CCW), và dự kiến tiếp tục thúc đẩy thảo luận quốc tế về các hệ thống vũ khí tiềm năng trong tương lai.

Các quốc gia trong CCW đang thảo luận về các vấn đề kỹ thuật, pháp luật, quân sự, đạo đức và các vấn đề khác liên quan đến các công nghệ mới nổi, mặc dù rõ ràng là chưa có được tiếng nói chung về LAWS. Một số nước đã gắn LAWS với máy bay điều khiển từ xa ("drone" quân sự), một điều mà Hoa Kỳ phản đối, bởi máy bay điều khiển từ xa, theo định nghĩa, được con người trực tiếp điều khiển cũng chỉ như là máy bay có người lái. Các ưu tiên của Hoa Kỳ được nhắc lại rằng tất cả các hệ thống vũ khí, tự động hay không, phải tuân thủ luật nhân đạo quốc tế, bao gồm các nguyên tắc đặc biệt và tương xứng. Vì lý do này, Hoa Kỳ đã liên tục ghi nhận tầm quan trọng của quá trình xem xét các loại vũ khí trong việc phát triển và triển khai các hệ thống vũ khí mới.

Chính phủ Hoa Kỳ cũng đang tiến hành đánh giá toàn diện các tác động của sự tự chủ trong hệ thống phòng thủ. Tháng 11/2012, Bộ Quốc phòng Mỹ (DoD) đã ban hành Chỉ thị DoD 3000,09: "Sự tự chủ trong các Hệ thống vũ khí," trong đó vạch ra các yêu cầu cho sự phát triển và triển khai các hệ thống vũ khí tự động và bán tự động. Các hệ thống vũ khí có khả năng tự chủ lựa chọn và chốt các mục tiêu bằng vũ lực gây chết người cần có sự xem xét và phê chuẩn của quan chức cấp cao trong Bộ Quốc phòng trước khi những hệ thống vũ khí này được phát triển chính thức và lại một lần phê chuẩn nữa trước khi triển khai. Chỉ thị của Bộ Quốc phòng chỉ thị không cấm và cũng không khuyến khích sự phát triển này, nhưng đòi hỏi nó phải được tiến hành một cách cẩn thận và chỉ sau khi được xem xét và phê duyệt bởi các quan chức quốc phòng cấp cao. Trong những vấn đề khác, Chỉ thị của Bộ Quốc phòng yêu cầu các hệ thống vũ khí tự động và bán tự động phải được kiểm tra chặt chẽ và các nhân viên phải được đào tạo một cách phù hợp trong việc sử dụng để thúc đẩy các tiêu chuẩn quốc tế liên quan đến xung đột vũ trang.

Trí tuệ nhân tạo có khả năng đem lại lợi ích đáng kể trong một loạt các hoạt động liên quan đến quốc phòng. Các hoạt động không gây sát thương như

hậu cần, bảo trì, điều hành căn cứ, chăm sóc y tế cho cựu chiến binh, hỗ trợ y tế ở chiến trường và sơ tán thương vong, quản lý nhân sự, định tuyến, thông tin liên lạc, phòng thủ mạng, và phân tích thông tin tình báo có thể được hưởng lợi từ TTNT, làm cho lực lượng quân đội Mỹ an toàn hơn và hiệu quả hơn. TTNT cũng có thể đóng vai trò quan trọng trong các hệ thống mới để bảo vệ người và tài sản cố định giá trị và ngăn chặn các cuộc tấn công thông qua các phương tiện không gây thương vong. Cuối cùng, các ứng dụng có thể trở nên quan trọng nhất đối với Bộ Quốc phòng.

Với các tiến bộ trong công nghệ quân sự và trí tuệ nhân tạo trên phạm vi rộng hơn, các nhà khoa học, các nhà chiến lược, và các chuyên gia quân sự đều đồng ý rằng tương lai của LAWS là khó dự đoán và tốc độ thay đổi diễn ra rất nhanh. Nhiều khả năng mới có thể sớm trở thành hiện thực, và có thể nhanh chóng được phát triển và vận hành. Chính quyền đang chủ động tham gia vào cuộc thảo luận liên ngành đang diễn ra để hướng tới một chính sách về vũ khí tự trị phù hợp với các giá trị chung của con người, lợi ích an ninh quốc gia, và các nghĩa vụ quốc tế và trong nước.

KẾT LUẬN

Trí tuệ nhân tạo có thể là động lực chính của tăng trưởng kinh tế và tiến bộ xã hội, nếu ngành công nghiệp, xã hội dân sự, chính phủ, và công chúng cùng nhau hỗ trợ sự phát triển của công nghệ, với sự quan tâm chu đáo đến tiềm năng của chúng và để quản lý các rủi ro của chúng.

Chính phủ sẽ đóng một số vai trò như triệu tập các cuộc thảo luận về các vấn đề quan trọng và giúp đưa ra chương trình nghị sự cho tranh luận công cộng. Chính phủ cần theo dõi sự an toàn và tính công bằng của các ứng dụng khi chúng phát triển và thông qua các khung pháp lý để khuyến khích đổi mới cùng với bảo vệ công chúng. Chính phủ cần hỗ trợ nghiên cứu cơ bản và ứng dụng TTNT vào các hàng hóa công cộng, cũng như sự phát triển của một lực lượng lao động đa dạng có chuyên môn cao. Và bản thân chính phủ cần sử dụng TTNT để phục vụ công chúng nhanh hơn, hiệu quả hơn, và với chi phí thấp hơn.

Nhiều lĩnh vực chính sách công, từ giáo dục và mạng lưới an toàn kinh tế, quốc phòng, bảo vệ môi trường, và tư pháp hình sự, sẽ thấy những cơ hội mới

và những thách thức mới do sự tiến bộ liên tục của TTNT. Chính phủ phải tiếp tục nâng cao năng lực của mình để hiểu và thích nghi với những thay đổi này.

Do công nghệ TTNT liên tục phát triển, những người thực hành phải đảm bảo rằng các hệ thống dựa trên TTNT là có thể kiểm soát được; chúng là hệ thống mở, minh bạch và dễ hiểu; và rằng chúng có thể làm việc hiệu quả với mọi người; và rằng hoạt động của họ sẽ vẫn phù hợp với các giá trị và khát vọng của con người. Các nhà nghiên cứu và những người thực hành đã tăng cường sự chú ý của họ đến những thách thức này, và cần tiếp tục tập trung vào chúng.

Phát triển và nghiên cứu máy móc thông minh có thể giúp chúng ta hiểu rõ hơn và đánh giá tốt hơn sự thông minh của con người. Được sử dụng một cách cẩn thận, TTNT có thể tăng cường trí thông minh của chúng ta, giúp chúng ta vạch ra một con đường tốt hơn và khôn ngoan hơn hướng về phía trước.

Biên soạn: Trung tâm Phân tích thông tin

TÀI LIỆU THAM KHẢO CHÍNH:

"Preparing for the future of artificial intelligence". *Office of Science and Technology Policy*, 10/2016.

"AAAI Presidential Panel on Long-Term AI Futures: 2008-2009 Study," *The Association for the Advancement of Artificial Intelligence*, <http://www.aaai.org/Organization/presidential-panel.php>

Artificial Intelligence: Opportunities and Risks. Policy paper by the Eective Altruism Foundation, 12/2015

"Big Data: Seizing Opportunities, Preserving Values," *Executive Office of the President*, May 2014,

"One Hundred Year Study on Artificial Intelligence (AI100)," Stanford University, accessed August 1, 2016, <https://ai100.stanford.edu>.

"World Development Report 2016: Digital Dividends," *The World Bank Group*, 2016, <http://documents.worldbank.org/curated/en/896971468194972881/pdf/102725-PUB-Replacement-PUBLIC.pdf>.

CHUẨN BỊ CHO TƯƠNG LAI CỦA TRÍ TUỆ NHÂN TẠO

