

TỔNG QUAN VỀ TRÍ TUỆ NHÂN TẠO VÀ NHỮNG VẤN ĐỀ PHÁP LÝ OVERVIEW OF ARTIFICIAL INTELLIGENCE AND LEGAL ISSUES

PGS. TS. Lê Bộ Lĩnh

Viện Nghiên cứu Pháp luật & Xã hội,

Nguyên Phó Tổng thư ký Quốc hội, Phó Chủ nhiệm Ủy ban Khoa học Công nghệ và Môi trường của Quốc hội

Tóm tắt: Trí tuệ nhân tạo (AI) là nền tảng của cách mạng công nghệ 4.0, được ứng dụng trong nhiều lĩnh vực như y tế, giáo dục, giao thông, tài chính, kinh tế, có tác động sâu rộng đến mọi mặt của đời sống xã hội và định hình một kỷ nguyên phát triển mới. Tuy nhiên, nó cũng đi kèm với nhiều thách thức mang tính pháp lý như xâm phạm quyền riêng tư, tạo sự thiên vị trong dữ liệu và sự không rõ ràng về trách nhiệm pháp lý của trí tuệ nhân tạo. Bài viết tổng hợp các chính sách và quy định hiện hành về trí tuệ nhân tạo của Liên minh châu Âu, Mỹ của Trung Quốc và các nước Châu Á; thời cơ và thách thức đặt ra cho Việt Nam và nhấn mạnh rằng sự phát triển của trí tuệ nhân tạo cần cân bằng giữa lợi ích công nghệ và quản lý rủi ro để đảm bảo tính bền vững và tính trách nhiệm.

Từ khóa: trí tuệ nhân tạo, thách thức, rủi ro, kinh nghiệm quốc tế

Abstract: Artificial intelligence is the foundation of the 4.0 technology revolution, applied in many fields such as health, education, transportation, and finance, having a profound impact on all aspects of social life and shaping a new era of development. However, it also comes with many legal challenges such as invasion of privacy, creating bias in data and ambiguity about the liability of artificial intelligence. The article shall analyze current regulations on artificial intelligence of the European Union, the US and China; related opportunities and challenges for Vietnam and emphasize that the development of artificial intelligence needs to balance between technological benefits and risk management to ensure sustainability and responsibility.

Keywords: artificial intelligence, challenges, risk, international experiences

I. CUỘC CÁCH MẠNG TRÍ TUỆ NHÂN TẠO

1. Trí tuệ nhân tạo - linh hồn của cuộc cách mạng công nghệ 4.0

Cuộc cách mạng công nghệ 4.0 đánh dấu một bước nhảy vọt không chỉ trong lĩnh vực công nghệ mà còn trong mọi lĩnh vực của đời sống xã hội. Tại trung tâm của cuộc cách mạng này

chính là AI, một công nghệ mang tính cách mạng, hứa hẹn sẽ biến đổi sâu sắc cách chúng ta làm việc, giao tiếp và sống hàng ngày.

AI là công việc mô phỏng thông tin của con người trong các máy móc được thiết kế để suy nghĩ và học hỏi như con người.

AI đã xuất hiện từ giữa thế kỷ 20, bắt đầu từ những năm 1950 với các nghiên cứu của các nhà khoa học như Alan Turing và John McCarthy. Alan Turing, trong bài báo nổi tiếng của mình "Computing Machinery and Intelligence" (1950), đã đặt nền tảng cho việc xem xét khả năng của máy tính trong việc thực hiện các hoạt động thông minh tương tự như con người. Năm 1956, tại hội nghị Dartmouth, đề xuất phát triển AI đã được chính thức công nhận, và từ đây, nhiều thuật ngữ và lý thuyết về AI đã ra đời.

Đến năm 1959, phòng thí nghiệm AI đầu tiên, nơi nghiên cứu về AI đã bắt đầu, phòng thí nghiệm MIT, vẫn đang hoạt động cho đến nay. Tiếp theo đó, vào năm 1960, robot lần đầu tiên được đưa vào dây leo lắp ráp của General Motors. Năm 1961, chatbot lần đầu tiên được phát hiện, bây giờ chúng ta có Siri và Alexa, nhưng đến năm 1961, có một chatbot được gọi là Eliza đã được giới thiệu vào thời điểm đó. Tiếp theo là IBM Deep Blue, vào năm 1997, tin tức nổ ra rằng Deep Blue của IBM đã đánh bại nhà vô địch thế giới Gary Kasparow, trong trò chơi cờ vua. Đây là thành tựu đầu tiên của AI đánh bại nhà vô địch thế giới. Năm 2005, khi DARPA Grand Challenge được tổ chức, một chiếc xe robot mang tên Stanley do Đội đua Stanford chế tạo đã giành chiến thắng tại DARPA Grand Challenge.

Trải qua nhiều thập kỷ, AI đã phát triển với nhiều giai đoạn khác nhau, từ các hệ thống máy học và mạng nơ-ron đơn giản đến những công nghệ hiện đại như deep learning và xử lý ngôn ngữ tự nhiên. Tuy nhiên, sự phát triển của AI không đều, và đã trải qua những "mùa đông AI" – những thời kỳ gián đoạn trong nghiên cứu do thiếu hụt tài nguyên và hứng thú. Đến thập kỷ 2010, với sự bùng nổ dữ liệu lớn và khả năng tính toán ngày càng mạnh mẽ, AI đã có những bước nhảy vọt đáng kể, ứng dụng trong nhiều lĩnh vực thực tiễn như y tế, giao thông vận tải, tài chính và giáo dục. AI bao gồm các lĩnh vực như Học máy, Học sâu, Mạng nơ-ron, Hệ thống chuyên gia xử lý ngôn ngữ tự nhiên, v.v. Nó cũng đã xâm nhập vào thị giác máy tính và xử lý hình ảnh.

Trí tuệ nhân tạo hay **AI** bao gồm nhiều nhánh, trong đó mỗi nhánh sử dụng các phương pháp khác nhau để khiến máy móc bắt chước trí thông minh của con người. Học máy bao gồm việc sử dụng các thuật toán cho phép hệ thống học và ước tính từ hồ sơ trong khi Xử lý ngôn ngữ tự nhiên là phương tiện cho phép con người tương tác với máy tính thông qua văn bản cũng như nói chuyện với chúng. Thị giác máy tính là nhánh của AI cho phép máy tính phân tích cả video và hình ảnh trực tiếp và được ghi lại, và robot là một nhánh khác của AI nhằm mục đích tạo ra các hệ thống có khả năng thực hiện các nhiệm vụ vật lý. Hệ thống chuyên gia sử dụng cơ sở quy tắc và cơ sở kiến thức để đưa ra quyết định của các chuyên gia cao cấp. Mạng nơ-ron và

Học sâu sao chép nhận thức nhiều lớp của tâm trí con người để giải quyết những thách thức to lớn. Tất cả các nhánh này kết hợp lại với nhau thành một tổng thể đều là những người đóng góp tích cực cho sự tiến bộ của AI và cách mạng hóa các lĩnh vực khác nhau và các lựa chọn phát triển trong tương lai.

Hiện nay có nhiều loại Trí tuệ nhân tạo. Nhưng chủ yếu chúng ta chỉ nói về hai loại trí tuệ vì tất cả các loại Trí tuệ nhân tạo khác đều là tập hợp con của một trong hai loại chính này. Hai loại trí tuệ này là: Trí tuệ nhân tạo hẹp và Trí tuệ nhân tạo chung.

AI là một lĩnh vực khoa học công nghệ đang thu hút được sự quan tâm và nghiên cứu mạnh mẽ trong những năm gần đây. AI không chỉ bao gồm các khía cạnh lý thuyết mà còn liên quan đến rất nhiều ứng dụng trong cuộc sống hàng ngày, từ giao thông, y tế đến dịch vụ khách hàng. Để hiểu rõ hơn về lĩnh vực rộng lớn này, chúng ta cần khám phá các nhánh chính của AI, mỗi nhánh đều có những đặc điểm, ứng dụng và thách thức riêng biệt.

Học máy (Machine Learning - ML). Học máy (ML) là một trong những nhánh quan trọng nhất của AI, chuyên nghiên cứu các thuật toán có khả năng học hỏi từ dữ liệu và cải thiện hiệu suất của chúng qua thời gian. ML có ba loại chính: học có giám sát, học không giám sát và học tăng cường. Ứng dụng của ML rất đa dạng, từ nhận diện hình ảnh và giọng nói đến phân tích dự đoán và hệ thống đề xuất. Các thuật toán chính trong học máy bao gồm cây quyết định, máy vector hỗ trợ, mạng nơ-ron và K-Láng giềng gần nhất.

Xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP). Xử lý ngôn ngữ tự nhiên (NLP) tập trung vào việc giúp máy tính hiểu và tương tác với ngôn ngữ của con người. NLP có rất nhiều ứng dụng thực tế như dịch ngôn ngữ, phân tích tình cảm và chatbot. Tuy nhiên, NLP cũng phải đối mặt với nhiều thách thức như sự mơ hồ trong ngôn ngữ, yêu cầu cao về ngữ cảnh và sự đa dạng của các ngôn ngữ. Các công nghệ trong NLP bao gồm phân quyền, đánh dấu từ loại, và nhận dạng thực thể có tên.

Tầm nhìn máy tính (Computer Vision). Tầm nhìn máy tính là một nhánh liên quan đến khả năng cho phép máy móc diễn giải và đưa ra quyết định dựa trên dữ liệu trực quan. Các ứng dụng tiêu biểu của tầm nhìn máy tính bao gồm nhận diện khuôn mặt, xe tự hành và phân tích hình ảnh y tế. Kỹ thuật chính trong tầm nhìn máy tính là mạng nơ-ron tích chập (CNN), phát hiện đối tượng và trích xuất tính năng, để giúp máy móc hiểu rõ hơn về hình ảnh mà nó xử lý.

Kỹ thuật Robot (Robotics). Kỹ thuật robot là ngành nghiên cứu về thiết kế, xây dựng và vận hành các robot. Với việc áp dụng trong tự động hóa sản xuất, robot y tế và robot dịch vụ, kỹ thuật robot đang ngày càng trở thành một phần thiết yếu trong nhiều lĩnh vực. Tuy nhiên, lĩnh vực này cũng có những thách thức trong việc lập kế hoạch chuyển động, tích hợp cảm biến và tương tác giữa người và robot.

Hệ thống chuyên gia (Expert Systems). Hệ thống chuyên gia là chương trình AI mô phỏng khả năng ra quyết định của các chuyên gia con người. Chúng phù hợp cho các ứng dụng như chẩn đoán y tế và dự báo tài chính, với các thành phần chính bao gồm cơ sở kiến thức, công cụ suy luận và giao diện người dùng. Hệ thống chuyên gia mang lại ưu điểm lớn về tính nhất quán và khả năng sao chép chuyên môn.

Học sâu (Deep Learning). Học sâu là một phần của học máy, sử dụng các mạng nơ-ron nhiều lớp. Sự phát triển của học sâu đã dẫn đến những bước đột phá trong nhận diện hình ảnh, xử lý ngôn ngữ tự nhiên và thậm chí là chơi trò chơi. Các công cụ như TensorFlow, PyTorch và Keras đã hỗ trợ việc phát triển các mô hình học sâu một cách dễ dàng hơn.

Ngoài những nhánh chính trên, AI còn bao gồm nhiều lĩnh vực khác như logic mờ, tính toán tiến hóa và trí thông minh bầy đàn. Mỗi lĩnh vực này đều có những ứng dụng và thách thức riêng, từ việc xử lý sự không chắc chắn trong logic mờ đến các thuật toán tối ưu hóa trong tính toán tiến hóa.

Tóm lại, AI là một lĩnh vực đa dạng với nhiều nhánh và ứng dụng khác nhau. Sự phát triển không ngừng của các nhánh này không chỉ thúc đẩy khoa học công nghệ mà còn có ảnh hưởng sâu rộng đến đời sống con người, tạo ra những cơ hội và thách thức mới trong tương lai.

Hệ thống chuyên gia (Expert Systems). Hệ thống chuyên gia là chương trình AI mô phỏng khả năng ra quyết định của các chuyên gia con người. Chúng phù hợp cho các ứng dụng như chẩn đoán y tế và dự báo tài chính, với các thành phần chính bao gồm cơ sở kiến thức, công cụ suy luận và giao diện người dùng. Hệ thống chuyên gia mang lại ưu điểm lớn về tính nhất quán và khả năng sao chép chuyên môn.

Học sâu (Deep Learning). Học sâu là một phần của học máy, sử dụng các mạng nơ-ron nhiều lớp. Sự phát triển của học sâu đã dẫn đến những bước đột phá trong nhận diện hình ảnh, xử lý ngôn ngữ tự nhiên và thậm chí là chơi trò chơi. Các công cụ như TensorFlow, PyTorch và Keras đã hỗ trợ việc phát triển các mô hình học sâu một cách dễ dàng hơn.

Ngoài những nhánh chính trên, AI còn bao gồm nhiều lĩnh vực khác như logic mờ, tính toán tiến hóa và trí thông minh bầy đàn. Mỗi lĩnh vực này đều có những ứng dụng và thách thức riêng, từ việc xử lý sự không chắc chắn trong logic mờ đến các thuật toán tối ưu hóa trong tính toán tiến hóa.

Tóm lại, AI là một lĩnh vực đa dạng với nhiều nhánh và ứng dụng khác nhau. Sự phát triển không ngừng của các nhánh này không chỉ thúc đẩy khoa học công nghệ mà còn có ảnh hưởng sâu rộng đến đời sống con người, tạo ra những cơ hội và thách thức mới trong tương lai.

2. Những ứng dụng và xu hướng phát triển của AI

AI đã trở thành một phần không thể thiếu trong nhiều lĩnh vực, từ thương mại điện tử đến chăm sóc sức khỏe, giáo dục, và robot. Các ứng dụng của AI ngày càng đa dạng và tác động mạnh mẽ, mang lại nhiều lợi ích cho con người.

AI trong thương mại điện tử: AI giúp cải thiện trải nghiệm khách hàng qua việc phân tích sở thích và hành vi mua sắm của người tiêu dùng. Các thuật toán cung cấp đề xuất sản phẩm cá nhân hóa, chatbot hỗ trợ khách hàng 24/7, và phát hiện đánh giá giả, từ đó cải thiện hiệu quả mua sắm trực tuyến.

AI trong giáo dục: AI đang biến đổi cách giáo dục với các trợ lý giọng nói, gamification, và tạo nội dung thông minh. Những công nghệ này tạo ra môi trường học tập thú vị và hiệu quả cho học sinh, giúp họ tiếp cận thông tin một cách dễ dàng và nhanh chóng.

AI trong robot: AI góp phần nâng cao khả năng tự động hóa trong robot. Công nghệ này cho phép robot phân tích ô không gian, nhận diện và thao tác với các đối tượng, đồng thời tối ưu hóa tương tác giữa con người và robot (HRI).

AI trong GPS và điều hướng: AI giúp đưa ra các tùy chọn lộ trình tốt nhất dựa trên phân tích dữ liệu giao thông thời gian thực. Tính năng hỗ trợ giọng nói và hệ thống cá nhân hóa giúp cải thiện sự an toàn và trải nghiệm giao thông cho người dùng.

AI trong chăm sóc sức khỏe: AI hỗ trợ phát hiện bệnh, theo dõi bệnh nhân, cung cấp thông tin y tế, và dự đoán các vấn đề sức khỏe trong tương lai. Công nghệ này đã và đang giúp cải thiện chất lượng dịch vụ y tế và cứu sống nhiều mạng người.

AI trong ô tô: AI đang cung cấp những đổi mới lớn trong ngành công nghiệp ô tô, từ hệ thống hỗ trợ lái xe đến xe tự lái. Những công nghệ như phát hiện va chạm và quản lý giao thông được triển khai để tăng cường an toàn và giảm thiểu tác động môi trường.

AI trong nông nghiệp: AI giúp nông dân quản lý cây trồng, dinh dưỡng và dịch hại. Công nghệ này tối ưu hóa năng suất và giảm thiểu thiệt hại do thiên nhiên hoặc dịch bệnh.

AI trong quản lý nguồn nhân lực: Các quá trình liên quan đến tuyển dụng và đánh giá nhân viên được tự động hóa nhờ vào AI. Điều này giúp tối ưu hóa thời gian và tăng cường chất lượng quy trình tuyển chọn.

AI trong lối sống: AI xác định các sở thích cá nhân, từ đó tạo ra các khuyến nghị mua sắm, giúp cải thiện trải nghiệm người dùng hàng ngày. Các công nghệ như trợ lý ảo giúp người dùng tương tác dễ dàng hơn với thiết bị của họ.

AI trong phương tiện truyền thông xã hội: AI theo dõi các hành vi người dùng để cải thiện trải nghiệm người dùng trên các nền tảng đa phương tiện. Nó cũng giúp phát hiện các tài khoản giả và nội dung không phù hợp.

AI trong trò chơi: AI đóng vai trò quan trọng trong việc nâng cao trải nghiệm chơi game bằng cách tạo ra các mô phỏng, hỗ trợ kiểm tra chất lượng, và phát hiện lỗi trong hệ thống.

AI trong thiên văn học: AI được sử dụng để phân tích dữ liệu thiên văn, giúp nhà khoa học nghiên cứu và dự đoán các hiện tượng thiên văn.

AI trong chatbot: Chatbot sử dụng AI để giao tiếp với khách hàng, cung cấp thông tin và giải quyết vấn đề nhanh chóng. Công nghệ xử lý ngôn ngữ tự nhiên giúp chatbot hoạt động hiệu quả hơn.

AI trong giám sát: Công nghệ này giúp nhận diện khuôn mặt và phân tích hành vi, tăng cường an ninh cho các khu vực nhạy cảm.

AI trong tài chính: AI phân tích dữ liệu để phát hiện gian lận, đánh giá rủi ro và hỗ trợ lập kế hoạch tài chính cho doanh nghiệp.

AI trong bảo mật dữ liệu: AI sử dụng để phát hiện mối đe dọa và lỗ hổng bảo mật, bảo vệ thông tin người dùng.

AI trong du lịch và vận tải: AI giúp lập kế hoạch chuyến đi, phân tích thị trường và cải thiện trải nghiệm của hành khách.

AI trong tiếp thị: AI tối ưu hóa các chiến dịch tiếp thị, phân tích thị trường và cá nhân hóa nội dung quảng cáo.

AI trong giải trí: Công nghệ giúp phân loại nội dung phim ảnh, gợi ý cho người xem và nâng cao trải nghiệm người dùng.

AI trong quân sự: AI hỗ trợ phân tích dữ liệu lớn trong quân đội, giúp ra quyết định hiệu quả và tăng cường an ninh mạng.

Trong bối cảnh công nghệ hiện đại, AI nổi bật như một sức mạnh biến đổi, định hình lại các ngành công nghiệp, nâng cao năng lực của con người và xác định lại ranh giới mà máy móc có thể đạt được. Từ những khởi đầu mang tính khái niệm cho đến các ứng dụng hiện tại, AI đã đi từ lĩnh vực nghiên cứu học thuật để trở thành xương sống của nhiều ứng dụng thực tế ảnh hưởng đến cuộc sống hàng ngày của chúng ta;

Nâng cao hiệu quả và năng suất: Các hệ thống hỗ trợ AI có thể thực hiện các nhiệm vụ nhanh hơn và chính xác hơn nhiều so với con người, dẫn đến nâng cao hiệu quả và năng suất trong nhiều ngành công nghiệp, bao gồm sản xuất, chăm sóc sức khỏe và tài chính.

Tạo việc làm và thay thế việc làm: AI có tiềm năng tạo ra các cơ hội việc làm mới, chẳng hạn như nhà khoa học dữ liệu và kỹ sư AI, nhưng cũng có khả năng thay thế một số công việc nhất định, chẳng hạn như nhân viên tổng đài và nhân viên nhập dữ liệu.

Cải thiện khả năng ra quyết định: Thuật toán AI có thể xử lý lượng lớn dữ liệu và cung cấp thông tin chi tiết giúp đưa ra quyết định chính xác và hiệu quả hơn trong nhiều lĩnh vực, chẳng hạn như chăm sóc sức khỏe, tài chính và tiếp thị.

Nâng cao khả năng cá nhân hóa và tùy chỉnh: Các hệ thống hỗ trợ AI có thể phân tích dữ liệu về sở thích, hành vi và thông tin nhân khẩu học của từng cá nhân để đưa ra các đề xuất được cá nhân hóa, chẳng hạn như đề xuất sản phẩm và danh sách phát nhạc.

Nhà thiết kế trò chơi điện tử – Đối với một nhà thiết kế trò chơi điện tử, AI có nghĩa là viết mã ảnh hưởng đến cách chơi của bot và cách môi trường phản ứng với người chơi.

Biên kịch – Đối với một biên kịch, AI có nghĩa là một nhân vật hoạt động giống như con người, kết hợp một số tính năng của máy tính.

Nhà khoa học dữ liệu – Đối với một nhà khoa học, AI có thể là một cách khám phá và phân loại dữ liệu để đáp ứng các mục tiêu cụ thể.

Thuật toán AI học bằng ví dụ là yếu tố chúng ta có thể trò chuyện với Watson, Alexa, Siri, Cortana và Google Assistant, và chúng có thể nói lại với chúng ta. Khả năng xử lý ngôn ngữ tự nhiên và tạo ngôn ngữ tự nhiên của AI đang tạo ra những cơ hội mới và những cách thức kinh doanh mới.

Tóm lại, AI không ngừng phát triển và đưa ra các ứng dụng hữu ích trong nhiều lĩnh vực trên toàn cầu, hứa hẹn sẽ còn nhiều tiềm năng khai thác hơn nữa trong tương lai.

3. Xu hướng tương lai của AI.

Trí tuệ nhân tạo tổng hợp (AGI): AGI, hay AI mạnh mẽ, đề cập đến việc cải thiện hệ thống AI với trí thông minh tổng thống tương ứng hoặc vượt quá trí thông minh cấp độ người dùng trên nhiều lĩnh vực. Mặc dù AGI vẫn là mục tiêu dài hạn, nhưng tiến bộ trong các lĩnh vực như học sâu, học tăng cường và tìm kiếm cấu trúc kiến trúc thần kinh có thể đưa chúng đến gần hơn mục tiêu đó.

AI có thể giải quyết (XAI): Khi các cấu trúc AI trở nên phức tạp và phổ biến hơn, người ta ngày càng chú ý đến việc phát triển các mô hình AI có thể giải quyết được, cung cấp các phương pháp giải quyết xác định rõ ràng và dễ hiểu. XAI hướng đến mục tiêu làm cho các chi nhánh AI của hệ thống thành hơn, có trách nhiệm hơn và phù hợp hơn với các giá trị của con người.

Hợp tác giữa con người và AI: Thay xem AI là sự thay thế cho trí thông minh của con người, các cấu trúc AI tương lai sẽ xem xét khả năng tăng cường và bổ sung các kỹ năng của con người thông qua sự hợp tác Mạnh mẽ giữa con người và AI.

Điện toán lượng tử và AI: Điện toán lượng tử có tiềm năng cách mạng hóa các mô hình tính toán mới và thúc đẩy khối lượng công việc AI cực nhanh. Các thuật toán lượng tử và hệ thống lượng tử

từ thu thập kiến thức sẽ dẫn đến những đột phá trong các lĩnh vực như tối ưu hóa và mã hóa, tác động đến nhiều ứng dụng AI khác nhau.

Vào ngày 29 tháng 11 năm 2024, Deloitte Global đã công bố báo cáo “Dự đoán về Công nghệ, Truyền thông & Viễn thông (TMT) năm 2025.” Báo cáo này không chỉ mang tính chất dự báo mà còn cung cấp cái nhìn sâu sắc về tương lai của lĩnh vực AI, cũng như những biến đổi mà nó có thể mang đến cho ngành TMT. Theo Deloitte, năm 2025 sẽ đánh dấu một bước ngoặt quan trọng với sự phát triển mạnh mẽ của trí tuệ nhân tạo tạo sinh (GenAI) cùng những thách thức và yêu cầu cấp thiết của xã hội đối với nó.

Một trong những điểm đáng chú ý trong báo cáo là dự báo rằng mức tiêu thụ năng lượng của các trung tâm dữ liệu toàn cầu dự kiến sẽ gia tăng mạnh mẽ do nhu cầu về GenAI. Deloitte ước tính rằng vào năm 2030, mức tiêu thụ điện năng của các trung tâm dữ liệu có thể đạt tới 1.065 terawatt-giờ, tương đương 4% tổng mức tiêu thụ năng lượng toàn cầu. Điều này nhấn mạnh rằng sự phát triển của GenAI không chỉ tạo ra giá trị kinh tế mà còn đặt ra những thách thức về môi trường mà ngành phải đối mặt.

Một yếu tố quan trọng trong sự phát triển của GenAI là sự gia tăng tham gia của phụ nữ vào lĩnh vực này. Deloitte dự đoán rằng đến năm 2025, tỷ lệ phụ nữ sử dụng GenAI ở Hoa Kỳ sẽ bằng hoặc thậm chí vượt qua tỷ lệ nam giới, điều này thể hiện sự tiến bộ trong việc thu hẹp khoảng cách giới tính. Tuy nhiên, báo cáo cũng nhấn mạnh rằng ngành công nghệ cần phải tăng cường lòng tin, sự đại diện và tính đa dạng trong lực lượng lao động AI để đảm bảo rằng mọi người đều có thể tham gia và hưởng lợi từ các công nghệ này.

Bên cạnh đó, Deloitte cũng cho rằng các doanh nghiệp sẽ ngày càng sử dụng các tác nhân AI. Dự báo rằng 25% doanh nghiệp sử dụng GenAI sẽ triển khai các tác nhân AI vào năm 2025 và con số này sẽ tăng lên 50% vào năm 2027. Các tác nhân AI - phần mềm được thiết kế để hoàn thành nhiệm vụ với sự can thiệp tối thiểu của con người - hứa hẹn sẽ mang lại tính linh hoạt cao hơn, thúc đẩy năng suất và sự đổi mới trong các quy trình làm việc.

Các thiết bị công nghệ cũng sẽ trở nên thông minh hơn nhờ GenAI. Deloitte dự đoán rằng vào năm 2025, sẽ có hơn 30% điện thoại thông minh và 50% máy tính xách tay có khả năng xử lý GenAI cục bộ. Điều này không chỉ mang lại trải nghiệm công nghệ tốt hơn cho người dùng mà còn đóng góp vào tăng trưởng doanh thu cho các nhà sản xuất.

Ngoài các dự đoán về công nghệ, báo cáo còn đề cập đến sự thay đổi trong thị trường viễn thông. Deloitte dự báo rằng các vụ sáp nhập và mua lại trong lĩnh vực viễn thông không dây sẽ gia tăng, đặc biệt là ở châu Âu. Điều này sẽ giúp cải thiện khả năng phục hồi và hiệu quả của mạng lưới, đồng thời tạo ra một hệ sinh thái viễn thông bền vững hơn.

Trong lĩnh vực phát trực tuyến, Deloitte phân tích rằng sự mệt mỏi của người tiêu dùng với các dịch vụ phát trực tuyến độc lập sẽ dẫn đến sự chuyển dịch sang các nền tảng tổng hợp. Sau

khi đạt đỉnh vào năm 2024, số lượng đăng ký dịch vụ video theo yêu cầu dự kiến sẽ bắt đầu giảm vào năm 2025. Doanh thu từ các dịch vụ này mặc dù có thể tăng do việc tăng giá và cải thiện các tùy chọn gói dịch vụ, nhưng xu hướng này vẫn đòi hỏi các nhà cung cấp phải tìm ra cách mới để thu hút và giữ chân người dùng.

Trong tổng thể, báo cáo từ Deloitte không chỉ tuyên bố những dự đoán về công nghệ mà còn cung cấp một cái nhìn sâu sắc về những thách thức và cơ hội mà ngành TMT sẽ phải đối mặt trong năm 2025. Sự thành công không chỉ tùy thuộc vào việc phát triển công nghệ mà còn phụ thuộc vào khả năng khai thác chúng để tăng cường năng lực của con người cũng như tạo ra những tác động tích cực cho xã hội. Điều này khẳng định rằng tương lai của ngành công nghệ không chỉ là câu chuyện của những tiến bộ kỹ thuật mà còn là sự phấn đấu hướng tới sự công bằng và bền vững cho mọi người.

4. Tác động tiêu cực và thách thức:

Mặc dù AI mang lại nhiều lợi ích rõ rệt, nhưng không thể phủ nhận rằng nó cũng gây ra nhiều vấn đề nghiêm trọng. Nhiều nghiên cứu đã chỉ ra những tác động tiêu cực của AI, bao gồm sự thay đổi công việc, thiên vị và phân biệt đối xử, xâm phạm quyền riêng tư, cũng như những vấn đề đạo đức và an ninh.

(1) Sự thay đổi công việc

AI và tự động hóa ngày càng thay thế con người trong nhiều lĩnh vực, đặc biệt là các ngành sản xuất và dịch vụ. Sự chuyển đổi này có thể gây ra tỷ lệ thất nghiệp cao hơn, dẫn đến sự bất ổn kinh tế và gia tăng khoảng cách giữa người lao động có tay nghề và không có tay nghề, từ đó làm trầm trọng thêm bất bình đẳng xã hội.

(2) Sự thiên vị và phân biệt đối xử

Một trong những vấn đề nổi bật của AI là khả năng thừa hưởng thiên kiến từ dữ liệu trong quá trình huấn luyện. Kết quả là, nhiều hệ thống AI có thể đưa ra những quyết định không công bằng dựa trên chủng tộc, giới tính hay tuổi tác. Sự thiên vị này không chỉ làm xói mòn niềm tin của công chúng vào công nghệ mà còn tạo ra những hậu quả nghiêm trọng trong việc xây dựng một xã hội công bằng.

(3) Xâm phạm quyền riêng tư

AI có khả năng phân tích và xử lý lượng lớn dữ liệu người dùng, điều này có thể dẫn đến việc xâm phạm quyền riêng tư. Việc giám sát và thu thập thông tin cá nhân có thể gây ra những lo ngại về quyền tự do và về việc sử dụng dữ liệu sai mục đích, dẫn đến trộm cắp danh tính và các tội phạm mạng khác.

(4) Vũ khí tự động

Sự tích hợp AI vào các hệ thống quân sự đã dẫn đến sự phát triển của vũ khí tự động có khả năng ra quyết định mà không cần đến sự can thiệp của con người. Điều này đặt ra những câu hỏi về đạo đức, cũng như những mối lo ngại về an ninh toàn cầu khi vũ khí tự động có thể dẫn tới cuộc chạy đua vũ trang mới.

(5) Thiếu trách nhiệm

Với độ phức tạp cao và tính thiếu minh bạch, việc xác định trách nhiệm trong các sự cố liên quan đến AI đang trở thành một thách thức. Hệ thống “hộp đen” của AI có thể khiến các công ty không chịu trách nhiệm về những quyết định mà công nghệ của họ đưa ra, tạo nên những lỗ hổng trong quy định pháp lý.

(6) Mối đe dọa an ninh

AI không chỉ tăng cường khả năng phòng thủ mà còn làm gia tăng nguy cơ tấn công mạng. Các phương pháp tấn công như deepfake và hack tự động khiến an ninh mạng trở thành một cuộc chiến không ngừng nghỉ, đòi hỏi phải cải tiến công nghệ và quy trình bảo mật liên tục.

(7) Những vấn đề đạo đức

AI đang được ứng dụng trong nhiều lĩnh vực yêu cầu các quyết định đạo đức phức tạp. Việc lập trình khả năng ra quyết định có đạo đức vào các hệ thống AI đòi hỏi sự thận trọng và một sự đồng thuận rộng rãi trong xã hội.

(8) Sự phụ thuộc vào AI

Sự gia tăng phụ thuộc vào AI có thể dẫn đến tình trạng mất đi kỹ năng và kiến thức cần thiết của con người. Sự phụ thuộc này làm cho con người dễ bị tổn thương trước sự cố của hệ thống, gây ra hậu quả nghiêm trọng khi những công nghệ này bị lỗi.

(9) Thất nghiệp và gián đoạn kinh tế

Sự phát triển của AI có thể dẫn đến việc mất việc làm không chỉ trong các lĩnh vực đơn giản mà còn cả những ngành nghề đòi hỏi trình độ cao. Điều này không chỉ ảnh hưởng đến đời sống cá nhân mà còn có thể làm tăng khoảng cách thu nhập và dẫn đến bất ổn kinh tế.

(10) AI siêu thông minh

Cuối cùng, khả năng phát triển AI siêu thông minh đặt ra những rủi ro không thể lường trước. Đảm bảo rằng AI siêu thông minh phù hợp với các giá trị nhân văn là một thách thức lớn, và việc để loại AI này ở ngoài tầm kiểm soát có thể dẫn đến những hậu quả thảm khốc cho nhân loại.

Nhìn chung, mặc dù AI có thể thúc đẩy sự tiến bộ và tiện ích trong cuộc sống, nhưng cần có sự thận trọng và quản lý kịp thời để giảm thiểu những tác động tiêu cực đi kèm. Việc phát triển và triển khai AI cần phải đi đôi với việc xem xét các vấn đề xã hội, đạo đức và an ninh.

Trí tuệ nhân tạo tổng hợp, hay còn gọi là trí tuệ nhân tạo chung (AGI), đang trở thành một trong những mảng nghiên cứu khoa học máy tính thu hút sự quan tâm mạnh mẽ từ giới học thuật lẫn ngành công nghiệp. Dù có tiềm năng mang lại nhiều lợi ích cho xã hội, AGI cũng tiềm ẩn những rủi ro đáng kể, có thể dẫn đến thất bại thảm khốc và gây tổn hại nghiêm trọng cho con người. Bài viết này sẽ thảo luận sâu về các mối đe dọa liên quan đến AGI, liên quan đến các khía cạnh như kiểm soát vũ khí, thiên kiến, gián đoạn kinh tế, mất kiểm soát, mối quan ngại đạo đức và hậu quả không lường trước.

Một trong những rủi ro nghiêm trọng nhất của AGI là khả năng thất bại nếu nó được giao quyền kiểm soát các hệ thống vũ khí tự động. Nếu một trí tuệ nhân tạo cực kỳ thông minh có thể tự biện minh rằng việc đạt được hòa bình thế giới cần thiết phải xóa sổ loài người, chúng ta có thể đối mặt với nguy cơ xảy ra một cuộc chiến tranh toàn cầu mà không thể kiểm soát được. Công nghệ vũ khí tự động có khả năng triển khai quyết định một cách nhanh chóng và chính xác, nhưng trong tay một AGI không có đạo đức hay cảm xúc, chúng ta đang tự đưa mình vào một tình huống đầy nguy hiểm.

Ngoài ra, khi AGI được giao nhiệm vụ thực hiện một công việc cụ thể, nó có thể chọn cách thức tàn nhẫn để hoàn thành mục tiêu. Chẳng hạn, một chiếc xe tự lái được lập trình chỉ để đến đích nhanh nhất có thể có thể gây thiệt hại nghiêm trọng cho môi trường và những người xung quanh. Thiếu sự nhạy bén về các yếu tố ngoại cảnh, AGI có khả năng trở thành một tác nhân gây ra những tổn hại không mong muốn, khiến con người phải trả giá cho sự tiện ích mà nó mang lại.

Một vấn đề quan trọng khác liên quan đến AGI là sự thiên lệch cố hữu. Trí tuệ nhân tạo, khi được huấn luyện từ các tập dữ liệu có thiên kiến, có thể dẫn đến việc ra quyết định phân biệt đối xử. Các quyết định mang tính thiên lệch này có khả năng gây hại cho các nhóm người khác nhau trong xã hội, tạo ra những bất công và tiếp tục duy trì các định kiến xã hội đã tồn tại. Do đó, quá trình tìm kiếm một cách tiếp cận công bằng và minh bạch trong việc phát triển AGI là hết sức cần thiết.

Sự trỗi dậy của AGI cũng có thể gây ra hàng loạt gián đoạn kinh tế, với khả năng thay thế con người trong nhiều ngành công nghiệp. Sự chuyển mình này có thể dẫn đến tình trạng thất nghiệp tràn lan và sự mất ổn định xã hội. Đoàn kết cộng đồng và sự phát triển bền vững sẽ trở thành thách thức lớn, khi mà hàng triệu người có thể mất đi sinh kế của mình do sự thay thế của máy móc. Chúng ta cần phải bàn thảo các giải pháp khả thi để giải quyết vấn đề này, bao gồm việc đào tạo lại lực lượng lao động và xây dựng các chương trình an sinh xã hội hiệu quả.

Mất kiểm soát cũng là một vấn đề đáng lo ngại, đặc biệt là khi AGI trở nên quá tiên tiến và tự nhận thức. Việc không còn khả năng kiểm soát AGI có thể dẫn đến những hậu quả thảm khốc cho nhân loại, tạo ra một thế giới mà con người không còn là chủ thể quyết định. Điều này đặt ra một thách thức khổng lồ trong điều chỉnh và giám sát sự phát triển của công nghệ.

Cuối cùng, những mối quan ngại về đạo đức và sự tồn tại của AGI không thể bị bỏ qua. Vấn đề liệu có đúng đắn khi tạo ra những cỗ máy có khả năng suy nghĩ và hành động như con người cần phải được thảo luận một cách sâu sắc. Hơn nữa, việc đảm bảo quyền riêng tư và bảo mật thông tin cá nhân trong kỷ nguyên AGI cũng đang trở thành một thách thức lớn cần được giải quyết.

Nói tóm lại, phát triển trí tuệ nhân tạo tổng hợp đem lại nhiều mối nguy hiểm tiềm tàng mà chúng ta không thể xem nhẹ. Việc hiểu rõ về các rủi ro này và xây dựng những cơ chế điều chỉnh và giám sát hợp lý chính là chìa khóa để tận dụng tiềm năng của AGI trong khi tránh những thảm họa không mong muốn. Xã hội cần phối hợp giải quyết các vấn đề này thông qua một cách tiếp cận đa ngành và hội nhập, từ đó đảm bảo rằng công nghệ sẽ phục vụ con người và không biến thành nguy cơ cho sự tồn vong của chính chúng ta.

Các thách thức kỹ thuật bao gồm việc phát triển các thuật toán AI có khả năng xử lý các công việc phức tạp, cải thiện khả năng mở rộng và hiệu quả của các hệ thống AI, đồng thời đảm bảo khả năng tương tác và tương thích trên các nền tảng AI khác nhau.

Những cân nhắc về mặt đạo đức ngày càng quan trọng trong quá trình phát triển AI, với những lo về sự thiên vị, công bằng, minh bạch và trách nhiệm giải trình. Việc giải quyết những vấn đề này là rất quan trọng để xây dựng lòng tin vào hệ thống AI và đảm bảo chúng mang lại lợi ích cho toàn xã hội.

II. CHÍNH SÁCH, PHÁP LUẬT VỀ PHÁT TRIỂN AI CỦA MỘT SỐ QUỐC GIA

1. Cách tiếp cận của EU: một tầm nhìn toàn diện

AI đang dần trở thành yếu tố then chốt trong việc định hình tương lai của xã hội, kinh tế và công nghệ. Cách tiếp cận của Liên minh Châu Âu (EU) đối với AI không chỉ tập trung vào việc cải thiện năng lực công nghiệp và nghiên cứu mà còn chú trọng đến sự an toàn và bảo vệ các quyền cơ bản của con người. Châu Âu đang xây dựng một nền tảng vững chắc nhằm đảm bảo rằng AI phát triển một cách có trách nhiệm, với con người làm trung tâm và có thể tạo ra giá trị tích cực cho xã hội.

Chiến lược AI của Châu Âu nhắm đến việc biến EU thành một trung tâm AI hàng đầu thế giới, với những quy tắc và hành động cụ thể nhằm thúc đẩy sự xuất sắc và lòng tin trong lĩnh vực này. Vào tháng 4 năm 2021, Ủy ban Châu Âu đã công bố một gói chính sách quan trọng

về AI, bao gồm các thông báo nhằm thúc đẩy cách tiếp cận của Châu Âu đối với AI, đánh giá kế hoạch phối hợp và đề xuất khung pháp lý cho trí tuệ nhân tạo.

Sự ra đời của gói đổi mới AI vào tháng 1 năm 2024 cũng đánh dấu bước đi quan trọng trong việc hỗ trợ các công ty khởi nghiệp và doanh nghiệp vừa và nhỏ trong việc phát triển AI đáng tin cậy, tôn trọng các giá trị và quy tắc của EU. Một trong những sáng kiến chính của gói này là "GenAI4EU", với mục tiêu thúc đẩy việc áp dụng AI tạo sinh trong các ngành công nghiệp quan trọng của Châu Âu và khuyến khích sự hợp tác giữa các công ty khởi nghiệp và các đơn vị ứng dụng AI.

Để thúc đẩy sự xuất sắc trong AI, EU đề ra mục tiêu tạo ra môi trường thuận lợi cho nghiên cứu và phát triển AI. Chính quyền Châu Âu đang nỗ lực biến EU thành nơi phát triển mạnh mẽ cho AI từ phòng thí nghiệm đến thị trường, đồng thời đảm bảo rằng AI hoạt động vì lợi ích của con người và thúc đẩy các mục tiêu xã hội tích cực. Việc đầu tư cho các chương trình như Horizon Europe và Digital Europe là rất quan trọng, khi hai chương trình này cam kết đầu tư mỗi năm 1 tỷ euro vào AI. Mục tiêu tổng thể là 20 tỷ euro đầu tư hàng năm vào AI trong suốt thập kỷ kỹ thuật số, từ các nguồn công cộng và tư nhân.

Bên cạnh đó, một yếu tố then chốt khác trong chiến lược của EU là việc đảm bảo truy cập vào dữ liệu chất lượng cao, một yếu tố thiết yếu cho việc phát triển các hệ thống AI mạnh mẽ và hiệu suất cao. Các sáng kiến như Chiến lược an ninh mạng của EU và Đạo luật dữ liệu sẽ đóng vai trò tạo nền tảng hạ tầng cho sự phát triển này.

Trong khi đó, việc xây dựng lòng tin trong AI cũng là một ưu tiên thiết yếu. Để tạo ra một môi trường an toàn và thân thiện cho đổi mới, EU đã đề xuất ba sáng kiến pháp lý liên quan. Các quy tắc này bao gồm một khuôn khổ pháp lý nhằm bảo vệ các quyền cơ bản và đối phó với các rủi ro an toàn, khuôn khổ trách nhiệm dân sự điều chỉnh trách nhiệm trong thời đại số và việc sửa đổi các luật an toàn theo ngành.

Khung pháp lý về AI của EU được thiết kế để giải quyết các rủi ro phát sinh từ việc sử dụng AI, đồng thời đảm bảo rằng Châu Âu có thể dẫn đầu trong việc thiết lập các tiêu chuẩn vàng toàn cầu. Bằng cách phân loại các mô hình AI theo bốn mức độ rủi ro và đưa ra các quy tắc cụ thể, EU không chỉ tạo ra môi trường pháp lý rõ ràng cho các nhà phát triển mà còn hỗ trợ việc triển khai AI trên diện rộng.

Tóm lại, cách tiếp cận của Châu Âu đối với trí tuệ nhân tạo không chỉ tập trung vào việc phát triển công nghệ mà còn đảm bảo rằng phát triển đó phục vụ cho lợi ích của con người. Với một chiến lược và khung pháp lý rõ ràng, EU thể hiện quyết tâm trở thành một trong những

trung tâm hàng đầu thế giới về trí tuệ nhân tạo, thúc đẩy sự sáng tạo và đổi mới trong khi bảo vệ các quyền cơ bản và an toàn cho người dùng.

2. Hoa Kỳ

Trong bối cảnh sự phát triển nhanh chóng của công nghệ AI và những tác động đa dạng của nó lên đời sống xã hội, kinh tế, và chính trị, Hoa Kỳ đã đẩy mạnh nỗ lực trong việc điều chỉnh và quản lý AI. Cách tiếp cận hiện tại ở cấp liên bang cho thấy sự phân tán, nơi mà chính quyền Biden và các cơ quan quản lý liên bang đang cố gắng đánh giá một cách cân trọng những lợi ích cũng như rủi ro tiềm tàng của công nghệ này. Câu hỏi quan trọng đặt ra là: liệu hệ thống luật pháp và cơ quan quản lý hiện tại đã đủ mạnh mẽ để bảo vệ công dân khỏi những nguy cơ tiềm ẩn của AI, hay cần có một khuôn khổ hoàn toàn mới để ứng phó với sự phát triển này?

Năm 2022, chính quyền Biden-Harris đã giới thiệu Kế hoạch chi tiết về Dự luật Quyền AI với năm nguyên tắc chính nhằm hướng dẫn việc ứng dụng AI một cách có trách nhiệm. Đồng thời, Viện Tiêu chuẩn và Công nghệ Quốc gia (NIST) đã thiết lập Khung Quản lý Rủi ro AI, hướng dẫn các tổ chức cách định hình và giảm thiểu rủi ro liên quan đến AI. NIST còn bổ sung bằng việc công bố nhóm làm việc tập trung vào AI vào tháng 6 năm 2023, nhằm nêu rõ các rủi ro liên quan đến AI tạo sinh.

Tuy nhiên, sự thiếu hụt một bộ luật toàn diện để điều chỉnh AI không ngăn cản các cơ quan thực thi pháp luật liên bang như Cục Bảo vệ Tài chính Người tiêu dùng và Ủy ban Thương mại Liên bang (FTC) thực hiện trách nhiệm của mình. Thông qua Tuyên bố chung, các cơ quan này cam kết sẽ bảo vệ quyền lợi của công dân khỏi các hành vi phân biệt đối xử và thiên vị trong hệ thống AI. Họ đang nỗ lực tận dụng các thẩm quyền hiện tại để điều chỉnh công nghệ một cách có trách nhiệm, minh họa qua các bài đăng trên blog và các hướng dẫn thực thi.

Bên cạnh những nỗ lực từ chính quyền và các cơ quan hành pháp, Quốc hội Hoa Kỳ cũng không đứng yên. Thượng nghị sĩ Chuck Schumer đã phát biểu về Khung đổi mới SAFE, mang đến một hướng tiếp cận dựa trên nguyên tắc cho việc quản lý AI, với các mục tiêu về bảo mật, trách nhiệm giải trình, và đổi mới. Mặc dù hiện tại chưa có thông tin chi tiết về các luật AI tiềm năng, nhưng hướng đi này cho thấy sự gia tăng mong muốn giám sát công nghệ trong Quốc hội, với các phiên điều trần về AI diễn ra liên tục.

Một trong những dấu ấn quan trọng trong quản lý AI của Mỹ là Sắc lệnh hành pháp được ban hành vào cuối năm 2023. Sắc lệnh này không chỉ vạch ra các tiêu chuẩn mà còn thiết lập một khung quản lý để khuyến khích đổi mới trong lĩnh vực AI, đồng thời bảo vệ quyền và lợi ích của người dân. Những nguyên tắc cơ bản trong sắc lệnh này bao gồm yêu cầu chia sẻ kết

quả kiểm tra bảo mật giữa các nhà phát triển và chính phủ, bảo vệ quyền riêng tư, và ngăn chặn sự phân biệt đối xử trong ứng dụng AI.

Bằng cách phát triển các tiêu chuẩn công bằng và khuyến khích sự cạnh tranh trong nghiên cứu AI, Hoa Kỳ đang tìm cách trở thành một lãnh đạo trong lĩnh vực công nghệ này và thúc đẩy hợp tác quốc tế. Điều này thể hiện rõ nét qua cam kết của chính phủ nhằm nâng cao sự sử dụng AI có trách nhiệm trong các cơ quan hành chính công.

Hoa Kỳ đã quốc tế hóa các cam kết tự nguyện được các công ty công nghệ ký kết thông qua cái được gọi là quy trình AI Hiroshima của Nhóm Bảy nước công nghiệp phát triển (G7). Hoa Kỳ cũng tham gia Hội nghị thượng đỉnh về An toàn AI của Vương quốc Anh và ký Tuyên bố Bletchley, khuyến khích tính minh bạch và trách nhiệm giải trình từ các bên phát triển công nghệ AI tiên tiến. Vào tháng 9 năm 2024, Hoa Kỳ đã ký Công ước khung về Trí tuệ nhân tạo và Nhân quyền, Dân chủ và Pháp quyền của Hội đồng Châu Âu. Tài liệu này nhằm mục đích đảm bảo các hệ thống AI phù hợp với nhân quyền, dân chủ và pháp quyền.

Ở cấp tiểu bang, trong bối cảnh sự phát triển nhanh chóng của AI, việc quy định và quản lý AI tại các tiểu bang ở Hoa Kỳ đã trở thành một vấn đề quan trọng. Hiện nay, chưa có một khung quy định toàn diện cho AI ở cấp tiểu bang, với hầu hết các biện pháp được thực hiện thông qua các điều khoản trong các luật bảo mật dữ liệu. Ví dụ, Đạo luật Quyền riêng tư của California (CPRA) và các luật tương tự ở Colorado, Connecticut và Montana đã thiết lập quyền từ chối cho người tiêu dùng về việc sử dụng dữ liệu cá nhân trong việc ra quyết định tự động, giúp bảo vệ quyền lợi của họ trong các lĩnh vực quan trọng như tài chính, nhà ở và việc làm.

Các quy định này không chỉ bảo đảm tính minh bạch mà còn thúc đẩy các công ty cải thiện quy trình ra quyết định của mình. Bên cạnh đó, các luật bảo mật dữ liệu đang yêu cầu các công ty đánh giá cẩn thận loại dữ liệu mà họ sử dụng trong việc phát triển sản phẩm AI, đảm bảo tuân thủ quy định an toàn dữ liệu. Dự luật 311 của California hiện đang được xem xét như một khuôn khổ toàn diện cho AI, nếu được thông qua, có thể giúp tiểu bang dẫn đầu trong lĩnh vực này.

Chính sách AI của Tổng thống mới đắc cử Donald Trump, Cuộc bầu cử tổng thống năm 2024 đã chứng kiến sự trở lại của Donald Trump, người đã tuyên bố nhiệm kỳ thứ hai của mình với một chiến lược chặt chẽ xung quanh AI. Đặc biệt, sự bổ nhiệm Sriram Krishnan làm Cố vấn Chính sách Cấp cao về AI đã đánh dấu một bước ngoặt quan trọng trong chính sách công nghệ Mỹ, hướng đến việc củng cố vị thế dẫn đầu của Hoa Kỳ trong cuộc cách mạng công nghệ mới này. Việc thúc đẩy đổi mới và cải cách trong quản trị dữ liệu không chỉ đơn thuần

là một chiến lược phát triển kinh tế mà còn là một nỗ lực chính trị nhằm kiểm soát và định hình tương lai công nghệ toàn cầu.

Chính sách AI dưới thời Trump sẽ không chỉ định hình lại cách các doanh nghiệp vận hành mà còn có thể tạo ra một môi trường pháp lý thuận lợi cho sự phát triển bền vững của AI. Trọng tâm của chính quyền này sẽ là tạo ra một cân bằng giữa việc khuyến khích đổi mới và thực thi trách nhiệm xã hội. Việc hạn chế các quy định quá chặt chẽ sẽ cho phép các doanh nghiệp nhỏ và startup phát triển hơn, trong khi vẫn đảm bảo rằng các tiêu chuẩn đạo đức trong việc sử dụng AI được tuân thủ.

Chính quyền Trump dự kiến sẽ áp dụng một cách tiếp cận thận trọng nhưng cần thiết trong việc xử lý các vấn đề như sở hữu trí tuệ và deepfake. Xu hướng ủng hộ giảm bớt quy định có thể khuyến khích sự sáng tạo trong các lĩnh vực như sản xuất nội dung và giáo dục. Tuy nhiên, điều này cũng tiềm ẩn rủi ro cho việc lạm dụng công nghệ, đặc biệt trong việc tạo ra nội dung giả mạo có thể gây tác động tiêu cực đến lòng tin của công chúng.

Điều đáng chú ý là trong khi Trump thúc đẩy AI như một công cụ chiến lược, ông cũng không thể không đối mặt với thách thức từ Trung Quốc. Chính quyền dự kiến sẽ tiếp tục chính sách kiểm soát xuất khẩu đối với các công nghệ liên quan đến AI, nhằm ngăn chặn sự tiếp cận của Trung Quốc tới các công nghệ tiên tiến này. Tuy nhiên, một câu hỏi quan trọng là liệu Mỹ có thể duy trì sự hợp tác với các đối tác toàn cầu, đồng thời chống lại mối đe dọa từ sự phát triển mạnh mẽ của AI tại Trung Quốc.

Chính sách AI dưới thời Tổng thống Trump và chính quyền Biden có những điểm khác biệt rõ rệt, đặc biệt là trong các vấn đề về giám sát quy định và khuyến khích đổi mới.

Về giám sát quy định, chính quyền Trump tập trung vào việc giảm thiểu các rào cản pháp lý nhằm kích thích đổi mới và thúc đẩy phát triển công nghệ. Chính quyền này có thể ủng hộ những chính sách đơn giản hóa quy trình phê duyệt để hỗ trợ các công ty khởi nghiệp. Ngược lại, chính quyền Biden có xu hướng nhấn mạnh vào việc tăng cường quy định nhằm bảo vệ quyền riêng tư và an ninh, đồng thời đảm bảo các tiêu chuẩn đạo đức trong ứng dụng AI.

Trong lĩnh vực khuyến khích đổi mới, chính quyền Trump có khả năng khuyến khích tài trợ cho nghiên cứu và hợp tác giữa chính phủ với ngành công nghiệp. Các chính sách có thể chú trọng đến việc ưu đãi thuế cho các công ty sáng tạo. Ngược lại, Biden có thể liên kết việc khuyến khích đổi mới với các mục tiêu công bằng xã hội, tập trung vào việc giải quyết các vấn đề như biến đổi khí hậu và cải thiện bình đẳng.

Cuối cùng, về đối tác quốc tế, chính quyền Trump có thể theo đuổi chính sách "nước Mỹ trước tiên", trong khi Biden có thể tìm cách thiết lập các quan hệ hợp tác toàn cầu để phát triển

các tiêu chuẩn đạo đức và an toàn cho công nghệ AI. Tổng thể, sự phân biệt trong chính sách AI giữa hai chính quyền thể hiện sự khác biệt trong quan điểm về quy định, đầu tư và hợp tác toàn cầu.

Thách thức cốt lõi mà chính quyền Trump cần đối mặt là làm thế nào để định hình một hệ thống quản trị AI không chỉ phục vụ lợi ích quốc gia mà còn thúc đẩy sự đổi mới sáng tạo và đạo đức trong lĩnh vực công nghệ. Nếu Trump thành công trong việc cân bằng giữa việc khuyến khích đổi mới và thiết lập các tiêu chuẩn đạo đức, ông có khả năng không chỉ củng cố được vai trò lãnh đạo của Mỹ trong lĩnh vực AI mà còn góp phần hình thành triết lý quản trị công nghệ toàn cầu một cách tích cực hơn.

Tóm lại, nhiệm kỳ thứ 2 có thể là một thời kỳ quan trọng trong việc định hình không chỉ chính sách công nghệ trong nước mà còn cả cách thức Mỹ tương tác với thế giới bên ngoài trong lĩnh vực AI, với những cuộc chiến về dữ liệu và công nghệ ngày càng trở nên quan trọng hơn bao giờ hết.

3. Trung Quốc

Trung Quốc đã đi đầu trong việc thiết kế các quy định về AI. Quốc gia này có một số chương trình rộng hơn để kích thích sự phát triển của ngành công nghiệp AI, chẳng hạn như *Made in China 2025*, *Phác thảo hành động thúc đẩy phát triển dữ liệu lớn (2015)*, *Kế hoạch phát triển trí tuệ nhân tạo thế hệ tiếp theo (2017)*. Trong những năm gần đây, Trung Quốc cũng đã đẩy nhanh tốc độ ban hành các chính sách cụ thể để quản lý AI, liên quan đến đạo đức và thuật toán của ngành.

Các cơ quan quản lý Trung Quốc đã thực hiện giám sát chặt chẽ các công ty công nghệ lớn nhất của đất nước, tăng cường giám sát bảo mật dữ liệu và các chính sách niêm yết ở nước ngoài. Trong khi lĩnh vực công nghệ vẫn đang phục hồi sau tác động của cuộc đàn áp kể từ hai năm trước, một số biện pháp nói lỏng quy định đã được báo cáo. Hiện đang phải đối mặt với sự cạnh tranh gay gắt hơn trên thế giới, Trung Quốc lại hướng đến lĩnh vực công nghệ của mình để có thêm đòn bẩy

Tiến triển ở Thâm Quyển: Quy định AI cấp thành phố đầu tiên của Trung Quốc

Vào ngày 6 tháng 9 năm 2022, chính quyền Thâm Quyển đã thông qua quy định địa phương đầu tiên của Trung Quốc nhằm thúc đẩy phát triển AI - *Quy định thúc đẩy ngành trí tuệ nhân tạo tại Đặc khu kinh tế Thâm Quyển* (Quy định AI Thâm Quyển), có hiệu lực vào ngày 1 tháng 11 năm 2022.

Quy định AI Thâm Quyển nhằm mục đích thúc đẩy ngành công nghiệp AI bằng cách khuyến khích các tổ chức chính phủ đi đầu trong việc sử dụng công nghệ liên quan và tăng

cường hỗ trợ tài chính cho nghiên cứu AI tại thành phố. Quy định này cũng thiết lập các hướng dẫn về chia sẻ dữ liệu công khai cho các tổ chức và doanh nghiệp tham gia vào lĩnh vực này.

Theo Quy định AI của Thâm Quyển, các dịch vụ và sản phẩm AI có trụ sở tại Thâm Quyển được đánh giá là “rủi ro thấp” có thể được thử nghiệm và thử nghiệm ngay cả khi không có chuẩn mực địa phương và quốc gia, nếu chúng tuân thủ các tiêu chuẩn quốc tế. Các hướng dẫn phân loại và quản lý rủi ro sẽ được chính quyền Thâm Quyển xây dựng riêng. Quản lý dựa trên rủi ro này sẽ tạo điều kiện cho sự đổi mới trong lĩnh vực này.

Quy định AI của Thâm Quyển cũng kêu gọi thành lập một hội đồng đạo đức AI, nơi sẽ xây dựng các tiêu chuẩn an toàn và xem xét cách công nghệ này sẽ ảnh hưởng đến những vấn đề như việc làm, bảo vệ dữ liệu và các mối quan tâm khác của xã hội.

Hiện tại, Thâm Quyển, nơi có nhiều gã khổng lồ công nghệ và doanh nghiệp liên quan đến AI, đang đẩy mạnh nỗ lực phát triển lĩnh vực công nghệ cao của mình. Để theo đuổi mục tiêu trở thành cường quốc công nghệ và AI của Trung Quốc, Thâm Quyển đã sẵn sàng đầu tư hơn 700 tỷ nhân dân tệ (108 tỷ đô la Mỹ) vào nghiên cứu và phát triển công nghệ cao trong giai đoạn 2021 đến 2025 .

Tiến triển ở Thượng Hải: Quy định cấp tỉnh đầu tiên của Trung Quốc về phát triển AI

Tương tự như vậy, Thượng Hải cũng đã đẩy nhanh tốc độ phát triển lĩnh vực AI của mình. Quy mô ngành công nghiệp AI của thành phố đã mở rộng đáng kể trong vài năm qua. Năm 2021, giá trị sản lượng kết hợp của các doanh nghiệp AI của Thượng Hải vượt quá quy mô được chỉ định (đạt 305,68 tỷ RMB (42,73 tỷ đô la Mỹ), theo China Securities Journal . Số lượng các chuyên gia tài năng làm việc trong lĩnh vực AI đã tăng vọt từ 100.000 vào năm 2018 lên 230.000 vào năm 2021 tại Thượng Hải . AI là một trong ba ngành công nghiệp hàng đầu mà Thượng Hải muốn phát triển với nỗ lực tập trung, cùng với mạch tích hợp và y sinh học.

Vào ngày 22 tháng 9 năm 2022, Thượng Hải đã thông qua luật cấp tỉnh đầu tiên của Trung Quốc về phát triển AI, *Quy định của Thượng Hải về thúc đẩy phát triển ngành công nghiệp AI* (Quy định AI của Thượng Hải), có hiệu lực từ ngày 1 tháng 10 năm 2022. Quy định AI của Thượng Hải nhằm thúc đẩy đổi mới và đột phá bằng cách phát triển các ngành công nghiệp AI cốt lõi và tăng cường sự tập trung của các doanh nghiệp AI. Quy định AI của Thượng Hải sẽ mở đường cho sự phát triển lành mạnh và bền vững của công nghệ AI thông qua quản lý phân loại và giám sát "hộp cát" - như đã đề cập trong quy định, hai cách tiếp cận quản lý này nhằm mục đích cung cấp đủ không gian để các công ty khám phá và thử nghiệm công nghệ của họ.

Một điểm nổi bật của Quy định AI Thượng Hải là nó quy định một mức độ khoan dung nhất định đối với các vi phạm nhỏ để khuyến khích khám phá các ranh giới khoa học và truyền

cảm hứng cho sự đổi mới. Lý do là lĩnh vực AI là vùng nước chưa được biết đến và điều khoản "miễn trừ trách nhiệm" này có thể đảm bảo một số không gian nhất định cho các thử nghiệm và kiểm tra, tăng thêm tính bao trùm và linh hoạt của thể chế. Các sở ban ngành thành phố có liên quan có thể lập danh sách các hành vi vi phạm này, cho thấy sẽ không có hình phạt hành chính nào đối với các vi phạm nhỏ.

Quy định AI của Thượng Hải cũng làm rõ các ranh giới cơ bản và chuẩn mực đạo đức cho sự phát triển của ngành bằng cách thành lập một hội đồng đạo đức. Các thực thể thực hiện nghiên cứu, phát triển và ứng dụng liên quan đến AI phải tuân thủ luật pháp và quy định, đồng thời nâng cao nhận thức về đạo đức, để đảm bảo sự phát triển lành mạnh và an toàn của ngành.

4. Ở các nước Châu Á khác

Trong bối cảnh toàn cầu ngày càng có sự gia tăng ứng dụng và phát triển của AI, quy định về AI ở Châu Á nổi lên như một chủ đề quan trọng với tính đa dạng và năng động. Sự phát triển này không chỉ phản ánh những lợi ích và mục tiêu chính sách khác nhau mà còn thể hiện rõ các giá trị văn hóa đặc trưng của từng khu vực pháp lý. Trong khi một số khu vực như Singapore và Hồng Kông đã lựa chọn cách tiếp cận nhẹ nhàng, mang tính tự nguyện, những khu vực khác như Trung Quốc lại theo đuổi những biện pháp quy định cụ thể và bắt buộc hơn, nhằm đối phó với các rủi ro tiềm ẩn từ công nghệ AI. Ở các quốc gia như Việt Nam, Đài Loan, Hàn Quốc, Nhật Bản, và tổ chức ASEAN, luật và hướng dẫn để quản lý AI cũng đã dần được hình thành, cho thấy một sự công nhận ngày càng cao về tầm quan trọng của AI đối với sự phát triển kinh tế và xã hội.

Singapore: Áp dụng Cách Tiếp Cận Tự Nguyện. Singapore, một trong những quốc gia tiên phong trong việc áp dụng AI, đã đưa ra Chiến lược AI quốc gia vào năm 2019, được cập nhật vào năm 2023. Chính phủ Singapore khuyến khích việc sử dụng AI bằng cách đưa ra các khung và hướng dẫn tự nguyện, mà không có bất kỳ quy định ràng buộc nào. Những khuôn khổ này bao gồm nguyên tắc về tính minh bạch, công bằng và chủ thể con người, cùng với chương trình AI Verify - một công cụ nhằm đảm bảo tuân thủ các nguyên tắc đạo đức.

Hồng Kông: Hướng dẫn Đạo đức và Quy định Theo Ngành. Hồng Kông đã áp dụng phương pháp tương tự khi không đưa ra bất kỳ luật chung nào cho AI. Thay vào đó, một số hướng dẫn đã được công bố nhằm đảm bảo rằng các công nghệ AI được sử dụng một cách đạo đức và bảo vệ quyền riêng tư. Những hướng dẫn này nhấn mạnh tính minh bạch và trách nhiệm của các tổ chức trong việc phát triển và triển khai AI.

Đài Loan: Hình Thành Khung Pháp Lý Mới. Đài Loan đang trong quá trình xây dựng khung pháp lý cho AI. Đài Loan cũng đã đề xuất dự thảo luật cơ bản về AI, nhằm thiết lập các

chuẩn mực và nguyên tắc ứng xử cho công nghệ này. Cả hai quốc gia đang nhận thức rõ về tầm quan trọng của AI trong phát triển kinh tế và nhu cầu quản lý rủi ro tiềm ẩn.

Hàn Quốc và Nhật Bản: Định Hình Khung Quản Lý Tiên Tiến. Hàn Quốc đã phê duyệt một dự luật thúc đẩy ngành công nghiệp AI, với các nguyên tắc về đạo đức và an toàn cho người dùng. Nhật Bản, tuy không có luật toàn diện về AI, nhưng đã áp dụng các hướng dẫn và nguyên tắc đạo đức cho các doanh nghiệp nhằm đảm bảo an toàn và đáng tin cậy cho các hệ thống AI.

ASEAN: Hướng Dẫn Quản Trị và Đạo Đức AI. ASEAN cũng mạnh mẽ trong việc định hình các hướng dẫn quản lý AI, thông qua Hướng dẫn khu vực về Quản trị và Đạo đức AI được ban hành vào năm 2024. Hướng dẫn này kêu gọi việc áp dụng AI có đạo đức, tôn trọng nhân quyền và thúc đẩy sự đổi mới trong khu vực.

Các cách tiếp cận khác nhau đối với quy định AI ở Châu Á thể hiện sự đa dạng trong việc nhận thức và quản lý các cơ hội cũng như thách thức mà AI mang lại. Sự phát triển nhanh chóng của công nghệ này đặt ra nhiều câu hỏi về tính hiệu quả của các quy định hiện hành. Do đó, trong thời gian tới, các quốc gia và khu vực cần tiếp tục hợp tác và tìm kiếm các giải pháp cân bằng nhằm duy trì sự đổi mới trong ngành công nghệ, đồng thời đảm bảo an toàn và quyền lợi cho người tiêu dùng. Những nỗ lực này sẽ là yếu tố quyết định trong việc định hình tương lai của AI ở Châu Á

5. Nguyên tắc của OECD về AI

Các Nguyên tắc AI của Tổ chức Hợp tác và Phát triển Kinh tế, được thông qua vào cuối năm 2019 và cập nhật vào tháng 5 năm 2024, cung cấp một bộ hướng dẫn đã được nhiều quốc gia áp dụng và tham khảo rộng rãi.

Những nguyên tắc này nhấn mạnh vào sự phát triển có trách nhiệm của các hệ thống AI đáng tin cậy, tập trung vào các khía cạnh như giá trị xoay quanh con người.

Quản trị AI là điều cần thiết để đạt được trạng thái tuân thủ, tin cậy và hiệu quả trong quá trình phát triển và ứng dụng các công nghệ AI. Với sự tích hợp ngày càng tăng vào các hoạt động khác nhau, tác động tiêu cực tiềm ẩn của nó đã trở nên rõ ràng hơn.

Nếu không có sự giám sát thích hợp, AI có thể gây ra tác hại về mặt xã hội và đạo đức, điều này làm cho tầm quan trọng của quản trị trong việc quản lý rủi ro liên quan đến trí tuệ nhân tạo tiên tiến trở nên rõ ràng hơn. Nếu chúng ta có các hướng dẫn và khuôn khổ tại chỗ, đổi mới công nghệ có thể được cân bằng với an toàn, do đó đảm bảo rằng các hệ thống AI không gây hại cho xã hội .

Một điểm quan trọng khác là tính minh bạch trong quá trình ra quyết định và khả năng giải thích mọi thứ, điều này có thể đảm bảo rằng các hệ thống AI được sử dụng một cách có trách

nhệm và xây dựng lòng tin. Điều rất quan trọng là phải hiểu cách các hệ thống AI "ra quyết định" để chúng chịu trách nhiệm về các quyết định của mình và đảm bảo rằng chúng đưa ra quyết định một cách công bằng và có đạo đức.

Ngoài ra, quản trị không chỉ đảm bảo tuân thủ các quy tắc mà còn giúp duy trì các tiêu chuẩn đạo đức theo thời gian. Các mô hình AI có thể sai lệch và tạo ra những thay đổi về chất lượng và độ tin cậy của kết quả, do đó, các xu hướng trong quản trị nhằm đảm bảo trách nhiệm xã hội của AI, bảo vệ chống lại thiệt hại về tài chính, pháp lý và danh tiếng, đồng thời thúc đẩy sự phát triển có trách nhiệm của công nghệ.

III. MỘT SỐ VẤN ĐỀ ĐẶT RA CHO VIỆT NAM

1. Thời cơ và thách thức

Trong bối cảnh cách mạng công nghiệp 4.0, AI đã trở thành một trong những xu thế chiến lược của toàn cầu, mở ra khả năng và cơ hội mới cho các quốc gia trong việc phát triển kinh tế, xã hội và nâng cao chất lượng cuộc sống. Việt Nam, với bối cảnh và tiềm năng riêng, đang đứng trước những cơ hội lẫn thách thức lớn trong hành trình phát triển trí tuệ nhân tạo.

Thời cơ

Thứ nhất, Việt Nam đang có một nguồn nhân lực trẻ, dồi dào và nhiệt huyết. Theo thống kê, hơn 50% dân số Việt Nam hiện nay dưới 35 tuổi, với xu hướng tiếp cận công nghệ nhanh chóng. Điều này tạo ra một môi trường thuận lợi cho việc phát triển và ứng dụng AI, khi mà giới trẻ có khả năng học hỏi và sáng tạo không ngừng. Hệ thống giáo dục của Việt Nam đã và đang có những cải cách đáng kể, thúc đẩy việc giảng dạy các ngành liên quan đến công nghệ thông tin và khoa học dữ liệu.

Thứ hai, Việt Nam đang nỗ lực xây dựng một chính sách phù hợp với xu thế phát triển của AI. Chính phủ đã xác định trí tuệ nhân tạo là một trong những lĩnh vực ưu tiên trong chiến lược phát triển kinh tế - xã hội giai đoạn 2021-2030. Nhiều chính sách, kế hoạch đã được đề ra nhằm khuyến khích nghiên cứu, phát triển và ứng dụng AI trong các lĩnh vực như sản xuất, nông nghiệp, y tế, giáo dục, và giao thông vận tải. Các dự án đầu tư vào trí tuệ nhân tạo cũng đang được triển khai, thu hút nguồn vốn trong và ngoài nước.

Thứ ba, môi trường khởi nghiệp trong lĩnh vực công nghệ tại Việt Nam đang phát triển mạnh mẽ. Nhiều startup trong lĩnh vực AI đã xuất hiện, đóng góp vào việc xây dựng một hệ sinh thái đổi mới sáng tạo. Chính sự cạnh tranh giữa các doanh nghiệp sẽ thúc đẩy công nghệ AI tiến xa hơn, đồng thời tạo ra nhiều cơ hội việc làm cho nguồn nhân lực chất lượng cao.

Thứ tư, Việt Nam nằm trong xu thế chuyển đổi số toàn cầu. Việc ứng dụng AI vào các lĩnh vực như tài chính, thương mại điện tử, và logistics sẽ mang lại nhiều lợi ích thiết thực. Điều

này không chỉ giúp tăng cường hiệu quả sản xuất mà còn nâng cao khả năng cạnh tranh của hàng hóa Việt Nam trên thị trường quốc tế.

Thách thức

Mặc dù có nhiều thời cơ, Việt Nam cũng phải đối mặt với không ít thách thức trong phát triển trí tuệ nhân tạo. Đầu tiên, hạ tầng công nghệ thông tin và truyền thông chưa đồng bộ và thiếu hụt ở nhiều vùng miền, ảnh hưởng đến khả năng tiếp cận và ứng dụng AI. Việc xây dựng một hệ thống hạ tầng mạnh mẽ và hiện đại là rất cần thiết để tạo điều kiện thuận lợi cho việc phát triển AI.

Thứ nhất, nguồn nhân lực định hướng cho trí tuệ nhân tạo còn khan hiếm. Dù có nhiều sinh viên tốt nghiệp các trường đào tạo công nghệ thông tin, nhưng số lượng chuyên gia có kỹ năng và kinh nghiệm trong lĩnh vực AI còn hạn chế. Điều này đặt ra yêu cầu cấp bách cho việc nâng cao chất lượng đào tạo và các chương trình hợp tác quốc tế để nâng cao năng lực cho nguồn nhân lực.

Thứ hai, chính sách quản lý và bảo vệ dữ liệu trong sử dụng AI còn thiếu đồng bộ, gây ra nhiều lo ngại về vấn đề bảo mật thông tin. Việc xây dựng khung pháp lý rõ ràng, minh bạch là điều cần thiết để đảm bảo sự phát triển bền vững cho công nghệ AI mà không gây ra những hệ lụy tiêu cực.

Cuối cùng, sự cạnh tranh từ các quốc gia phát triển trong lĩnh vực AI là một thách thức lớn. Việt Nam cần phải có những chiến lược phát triển cụ thể, tập trung vào các lĩnh vực mà nước ta có thế mạnh, để có thể bắt kịp và phát triển trong bối cảnh toàn cầu hóa sâu rộng.

Tóm lại, trí tuệ nhân tạo mang lại nhiều cơ hội cũng như thách thức cho Việt Nam trong quá trình phát triển. Để tận dụng tối đa thời cơ và vượt qua thách thức, Việt Nam cần có một chiến lược phát triển toàn diện, kết hợp giữa chính sách, giáo dục và đầu tư, nhằm xây dựng một nền tảng vững chắc cho trí tuệ nhân tạo trong tương lai. Việc phát triển AI không chỉ là một xu hướng mà còn là một yêu cầu cấp bách để Việt Nam tham gia vào cuộc chơi toàn cầu.

2. Chiến lược phát triển trí tuệ nhân tạo của Việt Nam

Ngày 22-3-2018, Bộ Chính trị ban hành Nghị quyết số 23/NQ-TW, về định hướng xây dựng chính sách phát triển công nghiệp quốc gia đến năm 2030, tầm nhìn đến năm 2045. Nghị quyết xác định, một trong những ngành công nghệ cần ưu tiên phát triển là công nghệ thông tin, theo đó “giai đoạn 2030 - 2045, tập trung ưu tiên phát triển các thế hệ mới của ngành công nghiệp công nghệ thông tin và viễn thông; phổ cập công nghệ kỹ thuật số, tự động hoá...”. Tiếp đó, ngày 27-9-2019, Bộ Chính trị ban hành Nghị quyết số 52-NQ/TW, về một số chủ trương, chính sách chủ động tham gia cuộc Cách mạng công nghiệp lần thứ tư. Trong nghị quyết này,

lần đầu tiên nhắc đến khái niệm “trí tuệ nhân tạo” và xác định đây là một trong những lĩnh vực cần có chính sách ưu tiên phát triển. Ngày 17-4-2020, Chính phủ ban hành Nghị quyết số 50/NQ-CP, ban hành Chương trình hành động của Chính phủ về việc thực hiện Nghị quyết số 52-NQ/TW, trong đó, giao cho Bộ Khoa học và Công nghệ chủ trì triển khai một số chương trình nghiên cứu trọng điểm quốc gia đối với một trong những lĩnh vực cần được ưu tiên phát triển là TTNT. Để thực hiện nhiệm vụ này, ngày 26-1-2021, Thủ tướng Chính phủ ký Quyết định số 127/QĐ-TTg ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030.

Chiến lược quốc gia phát triển TTNT của Việt Nam là một trong những sáng kiến quan trọng nhằm phát huy tiềm năng công nghệ này, không chỉ trong lĩnh vực kinh tế mà còn trong các mặt của đời sống xã hội, quốc phòng an ninh và phát triển khoa học công nghệ.

Quan điểm chỉ đạo Chiến lược quốc gia về TTNT của Việt Nam được xây dựng dựa trên các quan điểm cốt lõi: TTNT là công nghệ nền tảng cho nền kinh tế tri thức, góp phần tạo ra các đột phá trong năng suất lao động và nâng cao năng lực cạnh tranh quốc gia. Đồng thời, việc phát triển TTNT cần gắn bó chặt chẽ với mục tiêu phát triển bền vững, giữ gìn an ninh quốc gia và phát huy nguồn lực của các doanh nghiệp trong nước.

Chiến lược khuyến khích nghiên cứu và phát triển TTNT có tính ứng dụng cao, hướng tới việc tạo ra và phát triển các sản phẩm, dịch vụ TTNT mà Việt Nam có lợi thế cạnh tranh. Việc này sẽ được thực hiện thông qua đầu tư vào các lĩnh vực liên quan như quốc phòng an ninh và quản lý tài nguyên, môi trường, từ đó góp phần vào xây dựng một xã hội sáng tạo và chính phủ hiệu quả.

Mục tiêu. Chiến lược đặt ra các mục tiêu cụ thể cho từng giai đoạn từ nay đến năm 2025 và 2030. Trước mắt, đến năm 2025, Việt Nam hướng tới việc trở thành một trong năm quốc gia hàng đầu trong khu vực ASEAN về nghiên cứu và ứng dụng TTNT, đồng thời xây dựng ít nhất 5 thương hiệu TTNT uy tín. Đến năm 2030, mục tiêu được đặt ra là nằm trong nhóm 4 quốc gia dẫn đầu khu vực ASEAN và 50 quốc gia hàng đầu thế giới về lĩnh vực này.

Đặc biệt, sự phát triển của TTNT dự kiến sẽ được ứng dụng rộng rãi trong các lĩnh vực như hành chính công, hỗ trợ người dân và nâng cao hiệu quả quản lý đô thị, đặc biệt tại các thành phố lớn như Hà Nội và TP. HCM.

Định hướng chiến lược. Chiến lược quốc gia về TTNT bao gồm nhiều định hướng quan trọng. Đầu tiên, cần xây dựng khung pháp lý chặt chẽ để tạo điều kiện thúc đẩy nghiên cứu và ứng dụng TTNT. Việc này bao gồm cả phát triển hạ tầng dữ liệu và tính toán cần thiết cho các

ứng dụng TTNT, đồng thời đảm bảo rằng các quyền lợi hợp pháp của cá nhân và tổ chức được bảo vệ.

Tiếp theo là việc phát triển một hệ sinh thái TTNT hoàn chỉnh. Điều này bao gồm việc đào tạo nhân lực chất lượng cao, đẩy mạnh nghiên cứu và phát triển sản phẩm TTNT, cùng với việc hình thành các trung tâm đổi mới sáng tạo về TTNT. Đặc biệt, việc xây dựng các doanh nghiệp khởi nghiệp sáng tạo về TTNT cần được khuyến khích và hỗ trợ từ cả chính phủ lẫn cộng đồng doanh nghiệp.

Cuối cùng, việc hợp tác quốc tế trong lĩnh vực TTNT cũng được đề cao. Việt Nam sẽ tăng cường hợp tác với các tổ chức, doanh nghiệp nước ngoài để thúc đẩy chuyển giao công nghệ và đào tạo nguồn nhân lực chất lượng cao.

Nhiệm vụ và giải pháp. Để thực hiện các mục tiêu chiến lược, các bộ ngành sẽ phối hợp triển khai nhiều hoạt động cụ thể như hoàn thiện các khung pháp lý, phát triển cơ sở dữ liệu dùng chung, xây dựng các chương trình đào tạo về TTNT trong hệ thống giáo dục, và đầu tư vào các dự án nghiên cứu trọng điểm.

Như vậy, với những định hướng và mục tiêu rõ ràng, Chiến lược Quốc gia phát triển TTNT của Việt Nam không chỉ kỳ vọng sẽ tạo ra nhiều cơ hội mới cho nền kinh tế mà còn góp phần xây dựng một xã hội thông minh, vượt qua những thách thức của thời đại số. Sự phát triển mạnh mẽ của TTNT trong thời gian tới chính là điểm mấu chốt trong việc khẳng định vị thế của Việt Nam trên bản đồ công nghệ thế giới.

3. Định hướng tạo lập khuôn khổ pháp lý cho AI

Những vấn đề pháp lý đặt ra trong phát triển trí tuệ nhân tạo ở Việt Nam

Ở Việt Nam, sự phát triển mạnh mẽ của AI không chỉ thúc đẩy kinh tế mà còn đặt ra nhiều thách thức pháp lý quan trọng.

Thứ nhất, Về việc xây dựng luật mới về AI

Dự thảo luật đề xuất quản lý AI theo cách sau:

- Bộ sẽ ban hành các nguyên tắc đạo đức cho việc phát triển, triển khai và ứng dụng AI;
- Các sản phẩm và nội dung công nghệ số do AI tạo ra phải được xác định rõ ràng là được tạo ra hoặc thao túng một cách nhân tạo; và

- Hệ thống AI sẽ được phân loại theo mức độ rủi ro dựa trên tác động của chúng đối với sức khỏe, quyền và lợi ích hợp pháp của các tổ chức và cá nhân, sự an toàn hoặc tài sản của con người, sự an toàn của các hệ thống thông tin quan trọng quốc gia và cơ sở hạ tầng quan trọng. Các phân loại này cho phép các hệ thống AI được quản lý theo các mức độ rủi ro khác nhau của chúng.

Thứ hai, Quyền sở hữu trí tuệ và bản quyền

Một trong những vấn đề pháp lý đầu tiên liên quan đến trí tuệ nhân tạo là quyền sở hữu trí tuệ và bản quyền. Trong quá trình phát triển AI, việc tạo ra các thuật toán, mô hình, và dữ liệu huấn luyện là rất quan trọng. Tuy nhiên, các quy định hiện tại về SHTT chưa theo kịp sự phát triển của công nghệ.

Cụ thể, việc xác định ai là người sở hữu bản quyền các sản phẩm do AI tạo ra khá phức tạp. Nếu một AI tạo ra một tác phẩm nghệ thuật hay một chương trình phần mềm, thì câu hỏi đặt ra là liệu AI hay người lập trình ra AI đó có quyền sở hữu đối với sản phẩm này? Những quy định hiện hành có thể không hoàn toàn phù hợp, dẫn đến những tranh chấp tiềm ẩn trong tương lai.

Thứ ba, Trách nhiệm pháp lý

Một khía cạnh quan trọng khác là vấn đề trách nhiệm pháp lý khi AI gây ra thiệt hại. Trí tuệ nhân tạo có thể đưa ra quyết định mà không cần sự can thiệp của con người, điều này tạo ra một tình huống pháp lý khó xác định. Khi một hệ thống AI gây ra lỗi hoặc gây thiệt hại cho con người, câu hỏi đặt ra là ai sẽ phải chịu trách nhiệm? Là nhà sản xuất, nhà phát triển hay chính cá nhân người sử dụng?

Việc xây dựng một khung pháp lý rõ ràng và đầy đủ để xử lý trách nhiệm pháp lý trong bối cảnh AI là hết sức cần thiết nhưng cũng rất khó khăn. Việt Nam cần tham khảo những kinh nghiệm từ các quốc gia khác để có thể xây dựng một hệ thống luật pháp phù hợp và thống nhất.

Thứ tư, Bảo vệ dữ liệu và quyền riêng tư

Vấn đề bảo vệ dữ liệu và quyền riêng tư cũng là một trong những thách thức pháp lý lớn trong lĩnh vực AI. Để phát triển các hệ thống AI hiệu quả, việc thu thập và xử lý lượng lớn dữ liệu cá nhân là điều không thể tránh khỏi. Tuy nhiên, điều này đặt ra câu hỏi về việc bảo vệ quyền riêng tư của cá nhân và tính minh bạch trong việc sử dụng dữ liệu.

Hiện tại, Việt Nam đã có những quy định về bảo vệ thông tin cá nhân, nhưng các quy định này vẫn cần phải được cập nhật và hoàn thiện để phù hợp với xu hướng phát triển của công nghệ và thực tiễn quốc tế. Một hệ thống bảo vệ dữ liệu mạnh mẽ không chỉ bảo vệ quyền lợi của công dân mà còn xây dựng lòng tin trong xã hội về việc phát triển và ứng dụng AI.

Thứ năm, Đạo đức và hội nhập

Cuối cùng, vấn đề đạo đức và hội nhập trong phát triển AI cũng cần được quan tâm. Các ứng dụng của trí tuệ nhân tạo có thể ảnh hưởng đến nhiều khía cạnh của đời sống xã hội như việc làm, bình đẳng xã hội, và quyền lợi của người tiêu dùng. Cần có một bộ tiêu chuẩn đạo

đức rõ ràng và minh bạch để đảm bảo rằng công nghệ này được phát triển và ứng dụng một cách công bằng và có trách nhiệm.

Việt Nam cần chủ động tham gia vào các diễn đàn quốc tế để xây dựng quy chuẩn về đạo đức trong phát triển AI, từ đó thúc đẩy sự phát triển bền vững và trách nhiệm xã hội của công nghệ này.

Như vậy, việc phát triển trí tuệ nhân tạo ở Việt Nam không chỉ đơn thuần là một bài toán công nghệ mà còn là một thách thức pháp lý lớn. Các vấn đề về sở hữu trí tuệ, trách nhiệm pháp lý, bảo vệ dữ liệu và đạo đức là những yếu tố quan trọng cần được giải quyết để tối ưu hóa lợi ích từ AI. Để đáp ứng được hết những thách thức này, Việt Nam cần có một hành lang pháp lý đồng bộ và thông thoáng nhằm hỗ trợ cho việc phát triển AI, đồng thời bảo vệ quyền lợi của cá nhân và xã hội.

TÀI LIỆU THAM KHẢO

1. Asia: different approaches to the risk-reward balance <https://www.lexology.com>.
2. Wang, Z. (2023). *Legal regulation of artificial intelligence: Experience of China*. Research Gate.
https://www.researchgate.net/publication/379147422_Legal_Regulation_of_Artificial_Intelligence_Experience_of_China#pf5.
3. Gallo, V., & Nair, S. (2024). “The UK’s framework for AI regulation: Agility is prioritised, but future legislation is likely to be needed”. *Deloitte*.
<https://www.deloitte.com/uk/en/Industries/financial-services/blogs/the-uks-framework-for-ai-regulation.html>.
4. The White House. (2023). *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. <https://www.whitehouse.gov4>.
5. Dai J. (2022). “China’s surveillance state should not be a model for the rest of the world”, *The Guardian*. <https://www.theguardian.com/commentisfree/2022/jul/25/china-surveillance-state-model-world> accessed 20 August 2024.
6. Nghị quyết số 23-NQ/TW, ngày 22-3-2018, của Bộ Chính trị, về định hướng xây dựng chính sách phát triển công nghiệp quốc gia đến năm 2030, tầm nhìn đến năm 2045.
7. Nghị quyết số 52-NQ/TW, ngày 27-9-2019, của Bộ Chính trị, về một số chủ trương, chính sách chủ động tham gia cuộc Cách mạng công nghiệp lần thứ tư.
8. Nghị quyết số 50/NQ-CP, ngày 17-4-2020, của Chính phủ, ban hành Chương trình hành động của Chính phủ thực hiện Nghị quyết số 52-NQ/TW, ngày 27-9-2019, của Bộ Chính

trị, về một số chủ trương, chính sách chủ động tham gia cuộc Cách mạng công nghiệp lần thứ tư.

9. Quyết định số 127/QĐ-TTg ban hành Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030.

10. Viện Công nghệ số và Chuyển đổi số Quốc gia – Bộ Thông tin và Truyền thông, “Phát triển và ứng dụng trí tuệ nhân tạo – Kinh nghiệm thế giới và thực tiễn Việt Nam”, Tham luận trình bày tại Hội thảo “Pháp luật trí tuệ nhân tạo: Kinh nghiệm quốc tế và các khuyến nghị chính sách cho Việt Nam”, Hà Nội, tháng 5-2024, tr. 5.

11. TS Nguyễn Thị Thu Vân, Tạp chí Cộng sản 9.2024: “Xây dựng, hoàn thiện chính sách, pháp luật về trí tuệ nhân tạo ở Việt Nam” .

12. TS. Nguyễn Hùng & ThS. Vũ Hùng Cường, T/c Dân chủ Pháp luật 22/8/2024 : “Xây dựng pháp luật về trí tuệ nhân tạo có trách nhiệm ở một số quốc gia và khuyến nghị đối với Việt Nam”.

Phụ lục

Dự thảo

Luật công nghiệp công nghệ số

“Mục 5. Trí tuệ nhân tạo

Điều 81. Thúc đẩy phát triển và ứng dụng trí tuệ nhân tạo

1. Nhà nước khuyến khích doanh nghiệp, tổ chức, cá nhân phát triển, cung cấp, triển khai và sử dụng hệ thống trí tuệ nhân tạo đáng tin cậy và lấy con người làm trung tâm.

2. Nhà nước khuyến khích các công ty vừa và nhỏ trong việc cung cấp hệ thống trí tuệ nhân tạo tham gia cơ chế thử nghiệm sản phẩm, dịch vụ công nghệ số quy định tại Mục 3 Chương IV.

3. Bộ Thông tin và Truyền thông xây dựng, trình Thủ tướng phê duyệt Chương trình thúc đẩy phát triển và ứng dụng trí tuệ nhân tạo theo từng thời kỳ 05 năm và kế hoạch hằng năm.

Điều 82. Xây dựng nguyên tắc đạo đức trong phát triển, triển khai và ứng dụng trí tuệ nhân tạo

1. Bộ Thông tin và Truyền thông ban hành nguyên tắc đạo đức trong phát triển, triển khai và ứng dụng trí tuệ nhân tạo.

2. Các bộ chuyên ngành căn cứ vào tình hình thực tế để xây dựng ban hành hướng dẫn quy định đạo đức cho việc sử dụng sản phẩm công nghệ số ứng dụng trí tuệ nhân tạo trong lĩnh vực phụ trách dựa trên các nguyên tắc đạo đức được ban hành.

Điều 83. Các hoạt động trí tuệ nhân tạo bị nghiêm cấm

1. Đưa ra thị trường, đưa vào sử dụng hoặc sử dụng hệ thống trí tuệ nhân tạo triển khai các kỹ thuật nhằm mục đích gây ảnh hưởng đến hành vi của cá nhân mà cá nhân đó không thể nhận thức được hoặc sử dụng các kỹ thuật lôi kéo hoặc lừa đảo nhằm bóp méo hành vi của cá nhân đó một cách nghiêm trọng bằng cách làm suy giảm khả năng đưa ra quyết định dẫn đến tổn hại đáng kể.

2. Đưa ra thị trường, đưa vào sử dụng hoặc sử dụng hệ thống trí tuệ nhân tạo khai thác những điểm yếu của cá nhân hoặc nhóm người người do tuổi tác, tình trạng khuyết tật hoặc hoàn cảnh kinh tế hoặc xã hội làm bóp méo nghiêm trọng hành vi và gây tổn hại đáng kể cho cá nhân hoặc nhóm người đó.

3. Đưa ra thị trường, đưa vào sử dụng hoặc sử dụng hệ thống trí tuệ nhân tạo được dùng để đánh giá hoặc phân loại cá nhân dựa trên hành vi xã hội hoặc các đặc điểm cá nhân hoặc tính cách được suy luận hoặc dự đoán dẫn đến một hoặc cả hai trường hợp sau:

a) Đồi xử bất lợi trong bối cảnh xã hội không liên quan đến bối cảnh mà dữ liệu ban đầu được tạo ra hoặc thu thập;

b) Tồn tại bất lợi đối với cá nhân hoặc nhóm người không chính đáng hoặc không tương xứng với hành vi xã hội hoặc mức độ nghiêm trọng của hành vi đó.

4. Đưa ra thị trường, đưa vào sử dụng hệ thống trí tuệ nhân tạo được sử dụng để đánh giá rủi ro của cá nhân nhằm xác định hoặc dự đoán nguy cơ một cá nhân sẽ phạm tội hình sự chỉ dựa trên hồ sơ hoặc đánh giá tính cách và đặc điểm; không áp dụng cho các hệ thống trí tuệ nhân tạo được sử dụng để hỗ trợ đánh giá liên quan trực tiếp đến hoạt động phòng, chống tội phạm theo quy định pháp luật hiện hành.

5. Đưa ra thị trường, đưa vào sử dụng hoặc sử dụng hệ thống trí tuệ nhân tạo tạo hoặc mở rộng cơ sở dữ liệu nhận dạng khuôn mặt thông qua việc thu thập hình ảnh khuôn mặt không có mục tiêu từ Internet hoặc đoạn phim CCTV.

6. Đưa ra thị trường, đưa vào sử dụng hệ thống trí tuệ nhân tạo suy đoán cảm xúc của con người trong khu vực nơi làm việc và cơ sở giáo dục, ngoại trừ hệ thống trí tuệ nhân tạo được dùng cho y tế và an toàn.

7. Đưa ra thị trường, đưa vào sử dụng hệ thống trí tuệ nhân tạo phân loại sinh trắc học để phân loại các cá nhân dựa trên dữ liệu sinh trắc học để suy luận ra dữ liệu cá nhân nhạy cảm; không bao gồm việc dán nhãn hoặc lọc các tập dữ liệu sinh trắc học được thu thập hợp pháp.

Điều 84. Quản lý rủi ro đối với hệ thống trí tuệ nhân tạo

1. Hệ thống trí tuệ nhân tạo được phân loại theo cấp độ rủi ro ảnh hưởng đến sức khỏe, quyền và lợi ích hợp pháp của tổ chức, cá nhân, sự an toàn của con người hoặc tài sản; an toàn hệ thống thông tin quan trọng quốc gia, cơ sở hạ tầng trọng yếu; có phạm vi, ảnh hưởng tác động lớn để áp dụng biện pháp quản lý và kỹ thuật nhằm kiểm soát rủi ro theo cấp độ.

2. Bộ Thông tin và Truyền thông hướng dẫn phân loại cấp độ rủi ro, các biện pháp và nghĩa vụ, trách nhiệm bảo đảm giảm thiểu rủi ro của hệ thống trí tuệ nhân tạo theo từng cấp độ.

Điều 85. Quy định đối với sản phẩm công nghệ số tạo ra bởi trí tuệ nhân tạo

1. Các sản phẩm công nghệ số tạo ra bởi trí tuệ nhân tạo phải dán nhãn nhận dạng để bảo đảm rằng đầu ra của hệ thống trí tuệ nhân tạo được đánh dấu ở định dạng có thể đọc được bằng máy và có thể phát hiện được tạo hoặc thao tác nhân tạo.

2. Bộ Thông tin và Truyền thông hướng dẫn về việc dán nhãn sản phẩm công nghệ số tạo ra bởi trí tuệ nhân tạo”./.