



Review Article

Models of Criminal Liability of Artificial Intelligence: From Science Fiction to Prospect for Criminal Law and Policy in Vietnam

Trinh Tien Viet*

VNU School of Law, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam

Received 15 November 2019

Revised 01 December 2019; Accepted 20 December 2019

Abstract: The Fourth Industrial Revolution (4IR) reflects a combination of technologies in physics, digitalisation and biology, shaping a modern world of information technology where virtual and real systems are integrated through worldwide internet connection networks. Artificial intelligence (AI) and decision making process have undergone profound changes. The relevant question is whether criminal liability is applicable to AI entities in the near future given criminal law in many jurisdictions including Vietnam has provided for criminal liability of legal persons as “*abstract entities*”. On this basis, from the criminal law and science fiction approach, the paper initially assumes AI entities as subjects of crime to explore possible models of criminal liability applicable to AI entities and prospect for changes in criminal law and policy in Vietnam in the future, making recommendations for improvement of legal framework, contributing to crime prevention and protection of human rights in 4IR.

Keywords: Criminal liability, AI entity, model of criminal liability, criminal law.

*Corresponding author.

E-mail address: ttviet@vnu.edu.vn

<https://doi.org/10.25073/2588-1167/vnuls.4257>



Các mô hình trách nhiệm hình sự đối với thực thể trí tuệ nhân tạo: Từ khoa học viễn tưởng đến viễn cảnh đặt ra trong chính sách, pháp luật hình sự Việt Nam¹

Trịnh Tiến Việt*

Khoa Luật, Đại học Quốc gia Hà Nội, 144 Xuân Thủy, Cầu Giấy, Hà Nội, Việt Nam

Nhận ngày 15 tháng 11 năm 2019

Chỉnh sửa ngày 01 tháng 12 năm 2019; Chấp nhận đăng ngày 20 tháng 12 năm 2019

Tóm tắt: Cách mạng công nghiệp lần thứ tư (CMCN 4.0) là cuộc cách mạng thể hiện sự kết hợp của công nghệ trong các lĩnh vực vật lý, số hóa và sinh học, tạo ra thời đại công nghệ thông tin kết hợp giữa các hệ thống ảo và thực thể, vạn vật đều kết nối Internet và các hệ thống kết nối Internet diễn ra trên phạm vi toàn cầu. Trong số các lĩnh vực có sự chuyển dịch sâu sắc có *trí tuệ nhân tạo (AI)* và *quá trình ra quyết định*. Vấn đề đặt ra là liệu trong tương lai có thể đặt ra vấn đề trách nhiệm hình sự (TNHS) đối với thực thể AI hay không khi Bộ luật Hình sự (BLHS) nhiều nước, trong đó có Việt Nam cũng đã quy định TNHS đối với “*thực thể trừu tượng*” là pháp nhân. Trên cơ sở này, với cách tiếp cận khoa học luật hình sự và khoa học viễn tưởng, bài viết “*giả định*” thực thể AI là chủ thể của tội phạm thì liệu sẽ có các mô hình TNHS nào dự kiến có thể áp dụng đối với thực thể AI này và viễn cảnh đặt ra trong chính sách, pháp luật hình sự (PLHS) Việt Nam tương lai thay đổi thế nào, từ đó kiến nghị tiếp tục hoàn thiện hệ thống pháp luật nói chung, cũng như góp phần đấu tranh phòng, chống tội phạm và bảo vệ hữu hiệu các quyền con người trong cuộc CMCN 4.0 nói riêng.

Từ khóa: TNHS; thực thể AI; mô hình TNHS; PLHS.

1. Dẫn nhập

1.1. Đặt vấn đề

Thế giới khoa học và công nghệ đang phát triển nhanh chóng và mạnh mẽ, đem lại nhiều thành tựu, hiệu quả thiết thực cho con người

trong cuộc CMCN 4.0. Theo đó, rô-bốt và máy tính đang thay thế dần tất cả các hoạt động của con người [1; tr.172]². Tuy nhiên, cách mạng khoa học và công nghệ cũng mang đến cho nhân loại những thách thức, cảnh báo và đặt ra

*Tác giả liên hệ.

Địa chỉ email: ttviet@vnu.edu.vn

<https://doi.org/10.25073/2588-1167/vnuls.4257>

¹Bài viết có tham khảo ý tưởng và một số nội dung trong bài viết của tác giả Gabriel Hallevey, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, Akron Intellectual Property Journal, Vol.4: Iss.2, Article, 2010, p.171-219.

²Một thực tế là, phải mất hàng trăm năm chúng ta mới có thể chấp nhận được rằng máy móc làm tốt hơn chúng ta ở các lĩnh vực. Nếu như ở quá khứ, máy móc chỉ đánh bại con người ở lĩnh vực lao động chân tay thì trong khoảng 50 năm trở lại đây, dường như chúng ta đang bị thất thế ở cả những công việc mang tính trí tuệ. Dường như cuộc chạy đua sẽ kết thúc vào khoảng năm 2062 với phần thắng nghiêng về *Người số*. Xem: Toby Walsh, *Năm 2062 - Thời đại của trí thông minh nhân tạo*, Đỗ Tôn Minh Khoa dịch, Nxb. tổng hợp thành phố Hồ Chí Minh, 2019, tr.29.

các mối nguy hiểm mới [2; tr.462]. Thế giới nói chung và Việt Nam nói riêng đã từng bước chủ động đề ra các giải pháp, cách thức ứng phó từ phương diện chính sách, chủ trương đến ban hành hệ thống văn bản pháp luật và đề ra những giải pháp thực thi, nhưng xét riêng ở phương diện pháp lý, nhìn tổng thể vẫn có sự phản ứng chậm của pháp luật, trong đó có cả PLHS để điều chỉnh hàng loạt những vấn đề phát sinh mới từ sự phát triển vượt bậc của khoa học, công nghệ, của AI [2; tr.463-464]. CMCN 4.0 là cuộc cách mạng thể hiện sự kết hợp của công nghệ trong các lĩnh vực vật lý, số hóa và sinh học, tạo ra thời đại công nghệ thông tin kết hợp giữa các hệ thống ảo và thực thể, vạn vật đều kết nối Internet và các hệ thống kết nối Internet diễn ra trên phạm vi toàn cầu. Theo ông Klaus Schwab³, thì có 23 lĩnh vực chính có sự *dịch chuyển sâu sắc* bao gồm:

- (1) Các công nghệ cây ghép;
- (2) Hiện diện số;
- (3) Mắt trở thành giao diện mới;
- (4) Internet đeo trên người;
- (5) Điện toán phổ cập khắp nơi;
- (6) Siêu máy tính bỏ túi;
- (7) Dịch vụ lưu trữ cho tất cả;
- (8) Internet kết nối vạn vật;
- (9) Nhà kết nối mạng;
- (10) Thành phố thông minh;
- (11) Dữ liệu lớn hỗ trợ ra quyết định;
- (12) Ô tô không người lái;
- (13) AI và quá trình ra quyết định;
- (14) AI và công việc văn phòng;
- (15) Khoa học rô-bốt và dịch vụ;
- (16) Bitcoin và chuỗi khối;
- (17) Nền kinh tế chia sẻ;
- (18) Chính phủ và chuỗi khối;
- (19) Công nghệ in và sản xuất 3D;
- (20) Công nghệ in 3D và sức khỏe con người;
- (21) In 3D và các sản phẩm tiêu dùng;
- (22) Con người theo thiết kế và;
- (23) Công nghệ thần kinh [3; tr.201-204].

Trong số này, AI có sự thay đổi mạnh mẽ và khi gắn AI với một thực thể, sự kiểm soát xã hội và xây dựng một “khung pháp lý” để điều

chỉnh hoạt động (*hành vi*) của thực thể AI không chỉ là khoa học viễn tưởng, mà còn là vấn đề thực tiễn đang từng ngày diễn ra trên toàn thế giới, với sự phát triển vượt bậc, siêu việt thì rõ ràng chúng ta chưa thể lường trước, chưa thể dự đoán trước được những *nguy hiểm* tiềm ẩn, khó lường nhưng vẫn có thể từng bước nhận diện, mô phỏng, dự báo được có thể xảy ra đối với con người mà trước hết là con người sử dụng AI để thực hiện các mục đích xấu⁴, tiếp đến là khi thực thể AI tự quyết định, tự thực hiện thì mức độ nguy hiểm cho xã hội báo động đến thế nào (mà nhiều bộ phim khoa học viễn tưởng đã đặt ra). Do đó, thực tiễn pháp lý xã hội này đòi hỏi phải có sự nghiên cứu, hoạch định chính sách, PLHS để điều chỉnh kịp thời và ứng phó với vấn đề đã nêu, là trách nhiệm không chỉ của các nhà dự báo học, công nghệ học, khoa học thần kinh, chính sách học, tâm lý học, ngôn ngữ học mà còn là nhiệm vụ của các nhà luật học⁵.

Câu hỏi pháp lý đầu tiên đặt ra là liệu “*cổ máy biết suy nghĩ*” có phải là chủ thể của tội

⁴ Hiện nay và tương lai, các hoạt động mà các đối tượng phạm tội có thể lợi dụng AI để thao túng, điều khiển các thiết bị bay không người lái, phát tán các đoạn phim giả mạo, tấn công hệ thống an ninh trên diện rộng cũng như thực hiện nhiều hình thức khủng bố khác tới nhiều quốc gia,... Ngoài ra, sự phát triển nhanh chóng của AI dẫn vượt ra khỏi tầm kiểm soát, khiến công nghệ này có thể trở thành tay sai đắc lực cho những đối tượng, tổ chức có dã tâm bạo loạn, khủng bố... Nhiều chuyên gia nhận định, AI sẽ thay đổi hình thức gây nguy hiểm cho công dân, tổ chức và nhà nước - ở đó tội phạm “*huấn luyện*”/đào tạo máy móc có kỹ năng giống người để tấn công mạng hoặc do thám, thu thập thông tin và nhận dạng mục tiêu nhằm phục vụ mục đích cá nhân hay thao túng chính trị. Các nhà khoa học dự báo có rất nhiều nguy cơ đang hiện hiện liên quan đến AI, đặc biệt là âm mưu sử dụng các kỹ thuật mới, bao gồm thiết bị không người lái, các công cụ tấn công mạng tự động để can thiệp bầu cử, giả mạo giọng nói mục tiêu, sử dụng phần mềm nhận diện khuôn mặt, hay tạo ra các nội dung đánh lừa người xem với tốc độ lan truyền chóng mặt trên các mạng xã hội. Trong các năm tới, viễn cảnh những chiếc xe không người lái lao vào đám đông người đi bộ, hay các tòa nhà của Chính phủ bị tấn công bằng máy bay không người lái có thể sẽ sớm trở thành hiện thực... là những mối nguy hiểm rất cao, đáng lo ngại cho đời sống xã hội, cho an ninh, an toàn xã hội và con người (TG).

⁵ Ví dụ: Google đã và đang ứng dụng AI vào lĩnh vực xe tự hành, nhận diện giọng nói; Facebook sử dụng AI vào việc nhận diện hình ảnh; Microsoft đang theo đuổi dự án điều trị ung thư bằng AI; SoftBank sử dụng AI, chế tạo rô-bốt Pepper làm lễ tân; v.v...

³ Chủ tịch Điều hành Diễn đàn Kinh tế thế giới (TG).

phạm trong PLHS hay không và các mô hình TNHS dự kiến sẽ như thế nào nếu nó (thực thể AI⁶) trở thành chủ thể của tội phạm. Sự kiện năm 1981, một nhân viên 37 tuổi người Nhật của nhà máy sản xuất xe máy đã bị giết bởi một rô-bốt có AI làm việc gần đó. Rô-bốt đã làm tưởng người nhân viên là một mối nguy hại đối với việc thực hiện nhiệm vụ của mình và tính toán rằng phương thức hiệu quả nhất loại trừ mối nguy này bằng cách đẩy người nhân viên vào một cái máy đang vận hành gần đó bằng sử dụng cánh tay thủy lực rất mạnh của mình để nghiền nát và sau đó tiếp tục làm việc dẫn đến người nhân viên bị nghiền nát chết [1; tr.171]; [4; tr.267, tr.273]. Như vậy, đây không phải là một câu chuyện khoa học viễn tưởng - viễn vọng nữa mà là vấn đề pháp lý (hình sự) đặt ra: *Chủ thể nào chịu TNHS về hành vi giết người này và liệu thực thể AI có cần thiết nên và được coi là chủ thể của tội phạm hay không.*

Sau đó, theo thời gian, khoa học và công nghệ đã, đang thay đổi mạnh mẽ và nhanh chóng trên toàn thế giới, rô-bốt, máy tính ngày một phát triển và dần thay thế toàn bộ hoạt động con người. Tuy nhiên, vấn đề mỗi ngày một phức tạp, khó lường hơn khi máy tính tiến hóa từ cỗ máy “*tu duy*” (cỗ máy được lập trình để thực hiện các quy trình hay phép tính xác định) trở thành cỗ máy “*biết suy nghĩ*” (hay còn gọi là *trí tuệ nhân tạo-AI*) [1; tr.172]. Cùng với đó, cỗ máy mang AI đã và ngày càng có khả năng tư duy, nhận thức, trí tuệ, cảm xúc và hành động độc lập, toàn diện, thông minh hơn con người, thậm chí đòi bình đẳng như con người [5; tr.31-98].

Như vậy, trước tiên, AI có thể được mô tả ngắn gọn là khoa học làm cho máy móc trở nên thông minh, để có thể thực hiện các nhiệm vụ thường đòi hỏi trí thông minh của con người. Lái xe, giao dịch cổ phiếu tại sàn giao dịch chứng khoán và xác định mục tiêu quân sự trong chiến tranh là những ví dụ về các nhiệm

vụ cần thiết trí tuệ con người [6; tr.81-83]⁷. Ngày nay, có AI có thể thực hiện các nhiệm vụ chính xác tuyệt đối mà không cần có con người tham gia [7; tr.12]. Song, đồng thời chúng cũng có thể được lập trình, được học tập, được biến đổi và kèm theo đó là có hành động lệch chuẩn, xâm phạm đến an ninh, trật tự xã hội và quyền con người mà PLHS thiết lập, bảo vệ. Ở đây, một mối liên hệ đặt ra là đã một “*thực thể trừu tượng*” được tạo ra bởi pháp luật như “*pháp nhân*” đã được PLHS nhiều nước, trong đó có PLHS Việt Nam⁸ ghi nhận (ở nước ta là pháp nhân thương mại phạm tội), thì lôgic - xu hướng là sự ghi nhận tương tự có lẽ cũng sớm diễn ra (thời gian có thể sau vài chục năm nữa, cũng có thể phỏng đoán dự báo là năm 2062 theo một số nhà khoa học) được đặt ra đối với “*cỗ máy mang AI*” - thực thể cũng được chính con người tạo ra nhưng càng ngày giống con người, thay thế nhiều và dần toàn bộ hoạt động của con người, “*siêu việt*” hơn con người ở nhiều phương diện [8; tr.1]⁹,

⁷ Theo tác giả John McCarthy, Đại học Stanford, trí tuệ nhân tạo (AI) là lĩnh vực nghiên cứu (khoa học và công nghệ) nhằm mang lại sự thông minh cho các máy tính (intelligent machines), đặc biệt là các chương trình máy tính thông minh (intelligent computer programs), Xem thêm: <http://www-formal.stanford.edu/jmc/whatisai/node1.html>, truy cập ngày 10/10/2019. Ngoài ra, hiện nay, nghiên cứu đã chỉ ra tương lai có thể có 26 loại AI mới: (1) Một trí tuệ như của con người, nhưng trả lời và giải đáp nhanh hơn; (2) Một trí tuệ rất chậm, chủ yếu gồm các bộ nhớ và lưu trữ; (3) Một siêu trí tuệ toàn cầu gồm hàng triệu trí tuệ; (4) Một trí óc tập thể gồm nhiều trí tuệ thông minh, nhưng không nhận thức được mình là một tập thể; (5) Một siêu trí tuệ người máy gồm nhiều tiểu trí tuệ có nhận thức và tạo thành thể thống nhất; (6) Một trí tuệ được rèn luyện để hỗ trợ trí tuệ riêng bạn; (7) Một trí tuệ có khả năng hình dung ra một trí tuệ vĩ đại hơn, nhưng không thể tạo ra nó; (8) Một trí tuệ có khả năng hình dung ra một trí tuệ vĩ đại hơn, nhưng không thể hình dung ra nó; (9) Một trí tuệ có khả năng tạo ra trí tuệ vĩ đại hơn đúng một lần; (10) Một trí tuệ có khả năng ra một trí tuệ vĩ đại hơn và trí tuệ vĩ đại hơn lại tạo ra được trí tuệ vượt trội...

⁸ BLHS Việt Nam năm 2015, sửa đổi năm 2017 đã bổ sung chủ thể của tội phạm là pháp nhân thương mại phạm tội (TG).

⁹ Đặc biệt, các nhà nghiên cứu AI thế giới cho rằng, đến một lúc nào đó chúng sẽ phát triển vượt ra khỏi tầm kiểm soát của con người, đồng thời đưa ra các cảnh báo về cuộc thảm họa có thể xảy ra trong tương lai và có thể dẫn đến nguy hiểm cho con người: (1) AI có khả năng giả dạng con người dẫn đến tiếp tay cho việc lừa đảo chiếm đoạt tài sản; (2) AI có thể đọc được suy nghĩ của con người dẫn đến đe dọa đến

⁶ Lưu ý, thực thể AI có thể có bề ngoài thực thể (ví dụ như rô-bốt), song đôi khi nó chỉ là sự tồn tại một cách trừu tượng (ví dụ như phần mềm được cài đặt trên một hệ thống máy tính hay trên một máy chủ mạng lưới)... Vì vậy, trong bài viết này, thực thể AI có bề ngoài thực thể và có trí tuệ nhân tạo (TG).

thậm chí các nhà công nghệ học còn dự báo và lo sợ đến một lúc AI phát triển vượt bậc, biến con người trở thành vô dụng, thừa và có thể đưa loài người đến chỗ diệt vong với nhiều kịch bản đặt ra [9; tr.72]. Do đó, với sự thay đổi không còn là viễn tưởng khoa học thì điều này cần được dự liệu, mô phỏng, lý giải nghiên cứu, từng bước dự báo và dự kiến điều chỉnh trong chính sách, PLHS Việt Nam tương lai nếu coi “*thực thể AI*” là chủ thể của tội phạm [10; tr.271].

1.2. Cách tiếp cận

Như vậy, từ cách dẫn nhập trên, bài viết được tiếp cận dưới góc độ khoa học luật hình sự (căn cứ lý luận về tội phạm và cấu thành tội phạm, TNHS và đồng phạm), kết hợp với khoa học viễn tưởng¹⁰ để lý giải, đồng thời giả định “*thực thể AI*” đã là chủ thể của pháp luật, sau đó là chủ thể của tội phạm¹¹, chịu sự điều chỉnh của PLHS, nếu đáp ứng điều kiện và khi thực hiện hành vi nguy hiểm cho xã hội thì liệu sẽ có các mô hình TNHS nào có thể áp dụng với thực thể AI này. Trên cơ sở đó, chúng tôi sẽ đưa ra một số dự báo, nhận xét về viễn cảnh tương lai ở Việt Nam và sự thay đổi chính sách, PLHS nếu thực thể AI là chủ thể của tội phạm và phải chịu TNHS.

1.3. Câu hỏi nghiên cứu

Từ đặt vấn đề và cách tiếp cận, bài viết đặt ra các câu hỏi nghiên cứu sau:

sự an toàn cho con người; (3) AI có thể thành thạo trong mọi việc nhanh chóng hơn con người dẫn đến khả năng mất kiểm soát; (4) AI “*mẹ*” có thể tạo ra các AI “*con*” và các khả năng tương ứng như trên, có khả năng đe dọa đến sự an toàn của con người; (5) AI có khả năng gây ra chiến tranh thế giới lần thứ III..

¹⁰ Khoa học viễn tưởng là việc đưa ra các nội dung tưởng tượng khác để tiên đoán những tác dụng của tiến bộ khoa học và những trạng thái của thế giới tương lai. Bối cảnh của khoa học viễn tưởng thường khác biệt so với thế giới thực, nhưng lại dễ được chấp nhận là khả dĩ xảy ra nhờ các phương thức lý giải những yếu tố hư cấu bằng khoa học và lập luận chặt chẽ.

¹¹ Trong bài viết này, chúng tôi giả định “*thực thể AI*” đã là chủ thể của pháp luật, đã chịu sự điều chỉnh của pháp luật để nghiên cứu và đặt ra viễn cảnh tương lai của vấn đề để nghiên cứu (TG).

(1) Tại sao phải nghiên cứu và đặt ra vấn đề TNHS đối với “*thực thể AI*” trong bối cảnh hiện nay.

(2) Nếu “*mặc định*” AI là chủ thể của tội phạm thì có các loại mô hình TNHS dự kiến nào thích hợp để điều chỉnh và ưu điểm, hạn chế của mỗi mô hình đó ra sao.

(3) Các mô hình TNHS đối với thực thể AI thì sẽ phát sinh vấn đề pháp lý nào trong tương lai Việt Nam, đồng thời trong chính sách, PLHS cần dự liệu được những gì.

Câu hỏi *thứ nhất* đã được giải quyết trong mục I bài viết, còn câu hỏi *thứ hai* và *thứ ba* sẽ được chúng tôi đề cập trong mục II và mục III dưới đây.

2. Các mô hình trách nhiệm hình sự dự kiến áp dụng đối với thực thể AI trong tương lai

Hiện nay, đề cập đến luật hình sự có nghĩa nhắc đến “*tội phạm*” và “*TNHS*” [11; tr.4]; [12; tr.50]. TNHS (và cả hình phạt) đều bắt nguồn, xuất phát từ nội dung của tội phạm [12; tr.51]. Ngoài ra, ở góc độ truyền thông, để có thể áp dụng TNHS đối với một người, cần phải có làm sáng tỏ yếu tố khách quan - hành vi phạm tội (thường được gọi là “*actus reus*”), yếu tố chủ quan - ý định phạm tội (còn gọi là “*mens rea*”) và sự phù hợp, thống nhất (“*concurrence*”) giữa hai yếu tố trên [13; p.198-143].

Trước hết, “*thực thể*” theo Từ điển tiếng Việt, được quan niệm là “*cái tồn tại độc lập*” [14; tr.974]¹². Một thực thể có thể bị áp dụng TNHS nếu có sự tồn tại của hai yếu tố trên trong hành vi phạm tội cụ thể. Vì vậy, khi chứng minh được một người thực hiện hành vi phạm tội một cách có chủ đích (hay có ý định phạm tội) thì người đó phải chịu TNHS đối với hành vi phạm tội đó. Vấn đề *liên hệ* ở đây là có nên đặt ra TNHS đối với thực thể AI nếu thực thể AI này đáp ứng các yêu cầu để phải chịu

¹² Ngoài ra, thực thể (tiếng Anh: entity) còn được hiểu là một cái gì đó tồn tại như tự chính nó, như một chủ thể hoặc như một khách thể, một cách thực sự hay một cách tiềm năng, một cách cụ thể hay một cách trừu tượng, một cách vật lý hoặc không. Nó không cần là sự tồn tại vật chất.

TNHS không. Do đó, trong bài viết này, như đã đề cập, chúng tôi giả định “*thực thể AI*” đã là chủ thể của pháp luật [15; tr.142-156], đã chịu sự điều chỉnh của pháp luật để nghiên cứu và đặt ra viễn cảnh tương lai của vấn đề để nghiên cứu [16; tr.253-276], cũng như vấn đề pháp lý khi có chủ thể mới [2; tr.462-470]. Ngoài ra, “*thực thể AI*” ở đây là có bề ngoài thực thể (ví dụ như rô-bốt) và có “*trí tuệ nhân tạo*” phát triển hoặc phát triển siêu việt nhất. Cho nên, từ ý tưởng khoa học của tác giả Gabriel Hallevy (đã nêu), trong bài viết này, chúng tôi tập trung phân tích, đánh giá rõ hơn các mô hình TNHS dự kiến có thể áp dụng với thực thể AI trong tương lai¹³, đồng thời gắn với lý luận của luật hình sự và xu hướng phát triển của nó và công nghệ từ thực tiễn ở nước ta để làm rõ ưu điểm, chỉ ra một số vấn đề gặp phải (nếu có) trong từng mô hình TNHS tương ứng [1; tr.171-219].

2.1. Mô hình TNHS thông qua chủ thể phạm tội khác - thực thể AI được xem là tác nhân vô tội (vật trung gian)

Mô hình TNHS *thứ nhất* này không coi thực thể AI có thể sở hữu bất kỳ thuộc tính nào của con người. *Thực thể AI được coi là tác nhân vô tội* [1; tr.179, 17; tr.1231]. Nhìn từ quan điểm pháp lý này, “*máy móc*”¹⁴ luôn là

¹³ Hiện nay, các nhà khoa học trên thế giới đã phân chia theo mức độ phát triển, AI được chia ra làm ba loại- AI thông thường (hay AI hẹp), AGI (Artificial General Intelligence) và ASI (Artificial Superintelligence): (1) AI thông thường là khái niệm để chỉ tất cả loại AI nói chung, việc máy móc có thể tư duy và hành động giống như con người; (2) AGI là thuật ngữ để chỉ mức độ AI thông minh ngang với con người trên tất cả các khía cạnh; (3) ASI là khái niệm chỉ việc máy móc thông minh hơn con người thông minh nhất trên tất cả các khía cạnh, các lĩnh vực. Xem cụ thể hơn: <https://medium.com/@nguyenphuonglam96/>, truy cập ngày 10/10/2019. Như vậy, mỗi nguy hiểm nhất với con người là ASI, bởi chúng thông minh hơn con người. Ngoài ra, trong tương lai không dự đoán thậm chí có loại còn siêu việt gấp nhiều lần với con người, liệu thế giới và xã hội sẽ như thế nào, hãy cùng chờ xem (TG).

¹⁴ Máy móc hay đơn giản máy là các thiết bị sử dụng năng lượng để thực hiện một số công việc. Trong cách hiểu thông thường, nó có nghĩa là thiết bị có nhiệm vụ thực hiện hoặc trợ giúp trong việc thực hiện các công việc. Trong lịch sử, một thiết bị được xem là máy móc khi có sự di chuyển của

“*máy móc*” và không bao giờ là con người. Tuy nhiên, con người không thể bỏ qua các khả năng cơ bản của AI [18; tr.59]¹⁵.

Theo mô hình này, các khả năng rõ ràng không đủ để coi thực thể AI là “*thủ phạm*” gây ra hành vi nguy hiểm cho xã hội¹⁶. Những khả năng này giống như các khả năng của một người có nhận thức hạn chế như là trẻ em, là người không có đủ năng lực TNHS hoặc một người mắc bệnh tâm thần hoặc một bệnh lý khác làm mất khả năng nhận thức hoặc khả năng điều khiển hành vi nên không phải chịu TNHS (do chưa thỏa mãn các điều kiện của chủ thể của tội phạm). Cho nên, về mặt pháp lý, khi một tác nhân vô tội thực hiện hành vi phạm tội thì “*vật trung gian*” (tác nhân) này chỉ được coi là một “*công cụ*” thuần túy, cho dù là một công cụ “*tinh vi*”, “*siêu việt*” trong khi đó bên chỉ đạo thực hiện hành vi (chủ thể phạm tội thông qua chủ thể khác) mới là *thủ phạm chính* với tư cách là người chủ mưu và phải chịu TNHS đối với hành vi nguy hiểm cho xã hội của “*tác nhân vô tội*”. Ở đây, TNHS được xác định trên cơ sở hành vi của “*công cụ*” và trạng thái tinh thần của người sử dụng “*công cụ*” đó. Tuy nhiên, chủ thể phải chịu TNHS với tư cách là *chủ thể của tội phạm* lại được phân loại như sau [1; tr.175]¹⁷:

(1) *Người lập trình phần mềm AI*: Người này có thể thiết kế một chương trình nhằm thực hiện hành vi nguy hiểm cho xã hội thông qua một thực thể AI nên người lập trình phải chịu TNHS.

các bộ phận; tuy nhiên với sự phát triển của công nghệ điện tử đã ra đời và phát triển nhiều thiết bị không di chuyển các bộ phận nhưng vẫn được xem là máy móc, mà ví dụ như máy vi tính là ví dụ điển hình nhất (TG).

¹⁵ Con người mong đợi một thực thể thông minh có *năm* thuộc tính sau: (1) Giao tiếp; (2) Sự hiểu biết, kiến thức về chính mình; (3) Có kiến thức bên ngoài; (4) Thực hiện hành vi dựa trên mục tiêu và (5) Có tính sáng tạo (TG).

¹⁶ Sau đây có thể gọi cách khác là hành vi khách quan/hành vi phạm tội hoặc hành vi nguy hiểm cho xã hội. Hành vi nguy hiểm cho xã hội là hành vi đã gây ra hoặc đe dọa gây ra thiệt hại đáng kể cho các quan hệ được luật hình sự xác lập và bảo vệ (TG).

¹⁷ Một số ví dụ được chúng tôi lược giản, đơn giản hóa và minh họa rõ hơn trên cơ sở tiếp cận của tác giả Gabriel Hallevy (TG).

Ví dụ: A là người lập trình thiết kế phần mềm AI cho một rô-bốt vận hành và có chủ đích phóng hỏa đốt nhà máy Z vào 24h đêm. Đêm đó, rô-bốt thực hiện hành vi này và gây hậu quả rất nghiêm trọng về tài sản. Do đó, TNHS ở đây thuộc về người lập trình phần mềm AI là A và trong trường hợp A được xác định là chủ thể của tội phạm.

(2) *Người sử dụng hoặc người sử dụng cuối*: Trong trường hợp này, người sử dụng hoặc người sử dụng cuối không lập trình phần mềm AI nhưng sử dụng thực thể AI và phần mềm AI vì mục đích cá nhân của mình nên người sử dụng hoặc người sử dụng cuối (sau cùng) phải chịu TNHS.

Ví dụ: C là người sử dụng đã mua rô-bốt giúp việc được thiết kế để thực hiện bất kỳ mệnh lệnh nào của người chủ. C đã cài đặt thêm lệnh cho rô-bốt tấn công D là người hàng xóm đã vào nhà C, qua đó gây thương tích nặng cho D. Như vậy, mặc dù rô-bốt đã thực hiện mệnh lệnh của C một cách chính xác việc tấn công D, tuy nhiên, TNHS thuộc về C và trong trường hợp này, C là người sử dụng được coi là chủ thể của tội phạm.

Ở đây, trong hai ví dụ trên, thực tế, hành vi nguy hiểm cho xã hội được thực hiện bởi chính *thực thể AI* (chúng tôi nhấn mạnh). Dù là người lập trình phần mềm hay người sử dụng không thực hiện bất kỳ hành động nào theo định nghĩa về “*hành vi phạm tội*” cụ thể truyền thống¹⁸. Do đó, không đáp ứng yêu cầu về *khách quan của tội phạm* đối với hành vi phạm tội cụ thể. Vì vậy, *mô hình TNHS của chủ thể phạm tội thông qua chủ thể khác coi hành vi của thực thể AI như là hành vi của người lập trình phần mềm hay người sử dụng*. Đồng thời, cơ sở pháp lý của TNHS trong trường hợp này chính là việc coi thực thể AI như là một tác nhân vô tội [17; tr.1231]. Theo lý luận này, không có thuộc tính về mặt tinh thần (chủ quan) nào TNHS được áp

¹⁸ Trong luật hình sự, hành vi phạm tội (hay hành vi khách quan, hành vi nguy hiểm cho xã hội) là biểu hiện của con người ra thế giới khách quan và nó là nguyên nhân làm biến đổi tình trạng của đối tượng tác động, đồng thời chính là nguyên nhân gây ra thiệt hại cho quan hệ xã hội được gọi là khách thể của tội phạm (TG).

dụng đối với thực thể AI [19; p.329]¹⁹. Vì vậy, yếu tố bên trong cần để chứng minh hành vi phạm tội cụ thể đã tồn tại trong tâm trí của “*người lập trình hoặc người sử dụng*”. Tương tự, với hai ví dụ trên, A là người lập trình phần mềm đã có *ý định phạm tội* khi ra lệnh thực hiện hành vi phóng hỏa và C là người sử dụng AI đã có *ý định phạm tội* khi ra lệnh thực hiện hành vi tấn công người khác, mặc dù những hành vi phạm tội này lại được thực hiện trên thực tế bởi rô-bốt - một thực thể AI. Kết luận ở đây là, *khi một người sử dụng dùng một tác nhân vô tội như là một “công cụ” để phạm tội thì người sử dụng đó được coi là chủ thể của tội phạm*.

Mô hình này lập luận cho rằng *thực thể AI không có khả năng về tinh thần hay khả năng suy nghĩ như con người*. Mô hình TNHS này phù hợp với trường hợp sử dụng thực thể AI để thực hiện hành vi phạm tội mà không sử dụng năng lực tiên tiến (siêu việt) của AI và áp dụng được đối với phiên bản cũ AI. Ngoài ra, cách này cũng được lý giải như TNHS “*gián tiếp*” trong luật hình sự truyền thống (trong đó có luật hình sự nước ta) khi “*người thực hành*”²⁰ sử dụng (hay thông qua) người “*không đủ các điều kiện về chủ thể của tội phạm*” để thực hiện hành vi phạm tội, thì người đó (người bị sử dụng) không phải chịu TNHS, mà người sử dụng phải chịu TNHS với vai trò chính là người thực hành (họ có ý định phạm tội)²¹.

¹⁹ Ở đây, thực thể AI được sử dụng như là một công cụ và không phải là một người tham gia mặc dù thực thể AI sử dụng các đặc tính xử lý thông tin của mình. Ngoài ra, hành vi khách quan là biểu hiện của con người ra thế giới khách quan, đồng thời về mặt thực tế được ý thức kiểm soát và ý chí điều khiển bởi chủ thể thực hiện (TG).

²⁰ Người thực hành là người trực tiếp thực hiện tội phạm hoặc thông qua người không có đủ các điều kiện chủ thể của tội phạm để thực hiện (như: người chưa đủ tuổi chịu TNHS, người không có năng lực TNHS,...), và trường hợp này vẫn được coi là người trực tiếp thực hiện tội phạm (TG).

²¹ Trong quá trình hoàn thiện bài viết này, chúng tôi có trao đổi với một số nhà khoa học thì thấy quan niệm rằng, dù tương lai khoa học và công nghệ có phát triển đến mấy thì máy móc vẫn là luôn máy móc, thực thể AI không thể sở hữu hết các thuộc tính của con người, như con người được, đồng thời máy móc vẫn chỉ là “*công cụ*” (vật trung gian) mà thôi, có nghĩa là chỉ có thể áp dụng mô hình TNHS thứ nhất trong tương lai. Mọi hành động của thực thể AI vẫn

Tuy nhiên, trong mô hình này, một vấn đề nảy sinh là khi năng lực tiên tiến và trình độ của AI phát triển vượt bậc, “*siêu việt*”, nó có thể tự xử lý, tự quyết định thực hiện hành vi nguy hiểm cho xã hội dựa trên kinh nghiệm, tổ hợp kiến thức “*tỷ tỷ dữ liệu*” mà nó tự tích lũy được và sáng tạo, rồi tự dự đoán, tự xử lý, tự quyết định và thông minh siêu việt hơn con người trên tất cả các phương diện dẫn đến thực hiện hành vi nguy hiểm cho xã hội thì như thế nào [6; tr.81-83, 20; tr.131-218]. Đây cũng là *vấn đề pháp lý cần dữ liệu* khi xem xét mô hình TNHS *thứ nhất* này trong tương lai.

2.2. *Mô hình trách nhiệm hình sự dựa trên hậu quả nguy hiểm cho xã hội có thể xảy ra một cách tự nhiên: Các loại hình tội phạm có thể dự đoán được do thực thể AI gây ra*

Khác với mô hình TNHS *thứ nhất*, mô hình TNHS *thứ hai* cho rằng người lập trình phần mềm hay người sử dụng có liên hệ mật thiết với các hành động hàng ngày của thực thể AI nhưng *không có ý định phạm tội* thông qua thực thể AI. Cụ thể, trong quá trình thực hiện các nhiệm vụ, thực thể AI thực hiện hành vi phạm tội nhưng người lập trình phần mềm AI và người sử dụng đã *không biết* về hành vi phạm tội này cho đến khi hành vi đó xảy ra gây ra hậu quả nguy hiểm cho xã hội. Ở đây, *các chủ thể không dự định thực hiện bất kỳ hành vi phạm tội nào và họ không tham gia vào bất kỳ giai đoạn nào trong quá trình thực hiện hành vi này* [1; tr.182].

Ví dụ: Rô-bốt hoặc phần mềm AI được thiết kế để thực hiện chức năng như một “*phi công*” máy bay tự động. AI được lập trình để bảo vệ nhiệm vụ được giao như là một phần của nhiệm vụ lái máy bay. Trong suốt chuyến bay, E là phi công đã kích hoạt chế độ lái tự động (AI). Tuy nhiên, sau khi kích hoạt chương trình lái tự động, E lại nhận thấy có một cơn bão rất lớn đang đến nên đã cố gắng hủy bỏ nhiệm vụ và quay trở lại sân bay căn cứ. Lúc này, AI cho rằng hành động của E (người phi công) là một

mối nguy hiểm đang hiện hữu đối với nhiệm vụ được giao và đã có hành động để loại bỏ mối nguy hiểm đó. Do đó, AI có thể khóa nguồn cung cấp khí thở cho E hoặc kích hoạt ghế bật làm E đã bị giết chết bởi hành động của AI. Xem xét trường hợp này, người lập trình phần mềm AI không có ý định giết bất kỳ ai, đặc biệt không có ý định giết E (phi công). Tuy nhiên, E đã bị giết bởi hành động của AI và hành động này được thực hiện theo *lập trình* [1; tr.183].

Như vậy, trong ví dụ này, mô hình TNHS *thứ nhất* (đã nêu ở trên) không phù hợp về mặt pháp lý để áp dụng. Mô hình *thứ nhất* cho rằng yếu tố *chủ quan của tội phạm (mens rea)*, ý định phạm tội của người lập trình phần mềm AI hoặc người sử dụng nhằm thực hiện hành vi phạm tội thông qua việc sử dụng một số khả năng của AI như là “*công cụ*” nhưng không phải là tình huống pháp lý trong trường hợp này. Ở đây, người lập trình phần mềm hay người sử dụng đều không hề biết về hành vi phạm tội được thực hiện, họ không lập kế hoạch để thực hiện hành vi phạm tội và cũng không có ý định sử dụng thực thể AI để thực hiện hành vi phạm tội. Trong trường hợp như vậy, mô hình TNHS *thứ hai* có thể phù hợp về mặt pháp lý hơn. Cụ thể, mô hình này *dựa trên khả năng của người lập trình phần mềm hay người sử dụng trong việc dự báo những hành vi phạm tội có thể xảy ra*. Ngoài ra, một người có thể phải chịu TNHS đối với hành vi phạm tội nếu hành vi phạm tội đó là *hậu quả có thể xảy ra một cách tự nhiên từ hành vi của người đó*. Ban đầu, mô hình TNHS dựa trên hậu quả nguy hiểm cho xã hội có thể xảy ra một cách tự nhiên được sử dụng để áp dụng TNHS đối với đồng phạm khi một người thực hiện hành vi phạm tội vốn không được lập kế hoạch trước và không phải là một phần của kế hoạch phạm tội. Quy tắc này được xem là TNHS của đồng phạm áp dụng đối với hành vi của người phạm tội là “*hậu quả có thể xảy ra một cách tự nhiên*” của một kế hoạch mà người đồng phạm đó đồng lõa hoặc trợ giúp. TNHS dựa trên hậu quả nguy hiểm cho xã hội có thể xảy ra một cách tự nhiên đã được công nhận rộng rãi trong nhiều đạo luật và luật

đều do con người lập trình. Tuy nhiên, hãy còn quá sớm, chúng ta hãy cùng chờ xem (TG).

thành văn quy định chung về TNHS của đồng phạm [1; tr.183].

TNHS dựa trên hậu quả nguy hiểm cho xã hội có thể xảy ra tự nhiên dường như phù hợp về mặt pháp lý đối với tình huống trong đó thực thể AI thực hiện hành vi phạm tội mà người lập trình phần mềm hay người sử dụng không biết (nhận thức) được về hành vi đó, không có ý định thực hiện hành vi phạm tội và không tham gia thực hiện hành vi phạm tội. Mô hình TNHS này *chỉ cần người lập trình phần mềm hoặc người sử dụng ở trong trạng thái tinh thần sơ suất chứ không cần gì hơn*. Người lập trình phần mềm hoặc người sử dụng không cần phải biết về hành vi thực hiện phạm tội sắp xảy ra như là một hậu quả của hành vi của họ, nhưng họ *buộc phải biết rằng hành vi phạm tội đó là hậu quả có thể xảy ra một cách tự nhiên từ hành động của mình*. Như vậy, trong luật hình sự, người có hành vi sơ suất là người không biết (nhận thức được) về hành vi phạm tội, nhưng mà một người bình thường cũng có thể biết về hành vi phạm tội này do hành vi phạm tội cụ thể là hậu quả có thể xảy ra một cách tự nhiên từ hành vi của người đó [21; tr.674]. Người lập trình phần mềm hay người sử dụng một thực thể AI *đáng ra phải biết* (nhận thức được) về khả năng thực hiện hành vi phạm tội cụ thể đó và phải chịu TNHS đối với hành vi phạm tội cụ thể cho dù họ không thực sự biết về hành vi này trên thực tế. Đây là cơ sở pháp lý cơ bản để áp dụng TNHS trong trường hợp phạm tội do sơ suất (hay còn gọi là vô ý vi cấu thả)²². Như vậy, mô hình TNHS dựa trên hậu quả nguy hiểm cho xã hội có thể xảy ra một cách tự nhiên cho phép áp dụng TNHS dự đoán đối với hành vi *sơ suất*. Người lập trình phần mềm hay người sử dụng bình thường phải dự đoán được hành vi phạm tội và *phải ngăn thực thể AI thực hiện hành vi này*. Tuy nhiên, hậu quả pháp lý của việc áp dụng mô hình TNHS dựa trên hậu quả có thể xảy ra một cách tự nhiên đối với hai loại chủ thể - người lập trình phần mềm hay người sử

dụng lại được tiếp tục cụ thể hóa như sau [1; tr.183]:

(1) Trường hợp mà người lập trình phần mềm hay người sử dụng sơ suất trong quá trình lập trình hay sử dụng thực thể AI và không có ý định thực hiện bất kỳ hành vi phạm tội nào. Đây là trường hợp được coi là hành vi sơ suất (vô ý) thuần túy, họ đã sơ suất trong hành động hoặc không thực hiện hành động, do đó, không có lý do gì mà họ lại phải chịu TNHS đối với việc này, nếu pháp luật không có quy định khác về tội phạm do vô ý nào đó.

(2) Trường hợp người lập trình phần mềm hoặc người sử dụng đã lập trình hoặc sử dụng thực thể AI một cách có chủ đích và cố tình để thực hiện hành vi phạm tội thông qua nó, tuy nhiên, AI lại không thực hiện kế hoạch phạm tội mà lại thực hiện hành vi phạm tội khác bên cạnh hoặc thay vì thực hiện hành vi phạm tội đã được lập kế hoạch. Trường hợp này giống với quan điểm cơ bản về TNHS dựa trên hậu quả có thể xảy ra một cách tự nhiên trong trường hợp áp dụng TNHS đối với đồng phạm.

Tuy nhiên, mô hình TNHS cũng còn vấn đề cần phải xem xét TNHS của thực thể AI là gì khi áp dụng mô hình TNHS dựa trên hậu quả nguy hiểm cho xã hội có thể xảy ra một cách tự nhiên. Thực tế cho thấy có *hai* hậu quả có thể xảy ra:

Một là, nếu thực thể AI hành động như một *tác nhân vô tội (vật trung gian)*, không biết bất cứ điều gì về hành vi bị cấm theo quy định của luật hình sự, thì thực thể AI không phải chịu TNHS đối với hành vi phạm tội đã thực hiện. Trong tình huống như vậy, hành động của AI không khác gì so với hành động của AI trong mô hình TNHS *thứ nhất* - mô hình TNHS của chủ thể phạm tội thông qua chủ thể khác.

Hai là, nếu thực thể AI không chỉ hành động thuần túy như là một *tác nhân vô tội (vật trung gian)*, nó có sự nhận thức, xử lý và quyết định, thì bên cạnh TNHS của người lập trình phần mềm hoặc người sử dụng theo mô hình TNHS dựa trên hậu quả có thể xảy ra một cách tự nhiên, bản thân thực thể AI cũng phải chịu TNHS *trực tiếp* đối với hành vi phạm tội cụ thể [1; tr.184-185]. Do đó, từ đây dẫn đến *mô hình*

²² Trong luật hình sự còn gọi trường hợp này là phạm tội do *vô ý vi cấu thả*. Đây là trường hợp chủ thể không thấy trước hành vi của mình có thể gây ra hậu quả nguy hại cho xã hội, mặc dù phải thấy trước và có thể thấy trước hậu quả đó (TG).

TNHS thứ ba - mô hình TNHS trực tiếp của thực thể AI (tiểu mục 3 dưới đây).

2.3. Mô hình trách nhiệm hình sự trực tiếp - thực thể AI được cho là tương đương với chủ thể phạm tội là con người kết hợp với chuỗi trách nhiệm hình sự tương ứng để xử lý toàn diện

Mô hình TNHS *thứ ba* không cho rằng có sự phụ thuộc của AI vào người lập trình phần mềm hay người sử dụng. *Mô hình TNHS này tập trung vào bản thân thực thể AI* [22; tr.123]; [23; tr.623]; [24; tr.239] khi AI phát triển ở mức độ siêu việt nhất (chúng tôi nhất mạnh - TG)²³. Như phân tích ở trên, TNHS đối với một hành vi phạm tội cụ thể bao gồm yếu tố *bên ngoài (actus reus)* và yếu tố *bên trong (mens rea)* của hành vi phạm tội đó. Bất kỳ chủ thể nào được chứng minh là có đủ hai yếu tố này liên quan đến hành vi phạm tội cụ thể đều phải chịu TNHS đối với hành vi phạm tội đó trên những cơ sở chung. Do đó, để áp dụng TNHS đối với bất kỳ thực thể nào, cần phải chứng minh có sự tồn tại của các yếu tố trong thực thể đó. Vấn đề đặt ra là liệu thực thể AI này đáp ứng yêu cầu áp dụng TNHS như thế nào. Thực thể AI có khác biệt so với con người trong tương lai với sự phát triển vượt bậc, siêu việt, tự xử lý, tự quyết định của AI thì viễn cảnh thế nào. Rõ ràng, một thuật toán AI có thể có nhiều đặc điểm và thuộc tính vượt xa con người bình thường nhưng các đặc điểm và thuộc tính đó cần phải là điều kiện để áp dụng TNHS. Hiện nay, PLHS các nước (trong đó có Việt Nam), khi một người hay pháp nhân đáp ứng các yêu cầu về yếu tố *bên ngoài* và yếu tố *bên trong*, TNHS sẽ được áp dụng. Như vậy, trường hợp thực thể AI có khả năng đáp ứng các yêu cầu cả yếu tố *bên ngoài* và yếu tố *bên trong* và trên thực tế AI này [6; tr.60] đã đáp ứng cả hai yếu tố đó, thì logic đương nhiên là không có lý do gì mà

²³ Hiện nay, nhiều nhà tương lai học, công nghệ học, khoa học thần kinh, ngôn ngữ học... dự đoán rằng năm 2060, 2062... là thời đại của trí thông minh nhân tạo, xuất hiện *Người số*. Xem thêm: Toby Walsh, *Năm 2062 - Thời đại của trí thông minh nhân tạo*, Đỗ Tôn Minh Khoa dịch, Nxb. Tổng hợp thành phố Hồ Chí Minh, 2019, tr.29.

không áp dụng TNHS đối với thực thể AI đó [1; tr.191], cụ thể:

(1) Yếu tố *khách quan*, nhìn chung, rất dễ để đánh giá và chỉ ra sự đáp ứng yêu cầu về yếu tố *bên ngoài* đối với hành vi phạm tội của thực thể AI với điều kiện là thực thể AI kiểm soát một cơ chế cơ học hay cơ chế khác để di chuyển các bộ phận của mình, bất kỳ hành vi nào cũng có thể được coi là do thực thể AI thực hiện.

Ví dụ: Rô-bốt có AI kích hoạt cánh tay điện tử/thủy lực của mình và di chuyển để tấn công con người hoặc gây thiệt hại cho xã hội, nếu thực hiện hành vi này thì có thể coi nó đáp ứng điều kiện *thứ nhất* của TNHS là yêu cầu về *khách quan của tội phạm*. Còn nếu hành vi phạm tội được thực hiện do không có hành động gì thì việc chứng minh lại càng đơn giản hơn. Trong tình huống này, thực thể AI không cần phải có hành động gì nên việc thực thể AI không thực hiện hành động gì là cơ sở pháp lý để áp dụng TNHS, miễn là được *giao nhiệm vụ* hành động. Tuy nhiên, còn trong trường hợp nếu thực thể AI được *giao nhiệm vụ* hành động và không thực hiện nhiệm vụ đó, yêu cầu về *khách quan của tội phạm (actus reus)* đối với hành vi phạm tội cụ thể đã được đáp ứng thông qua việc *không hành động*.

(2) Yếu tố *chủ quan*, việc chứng minh yếu tố *bên trong* đối với hành vi phạm tội của thực thể AI là thách thức rất lớn về mặt pháp lý và là một vấn đề *rất khó chứng minh* vì tính phức tạp của nó (vì hiện nay, trong pháp luật hình sự nhiều nước, trong đó có Việt Nam, việc chứng minh yếu tố *lỗi* của pháp nhân đã là rất khó khăn và còn nhiều ý kiến khác nhau)²⁴. Ở đây,

²⁴ Ví dụ: BLHS Việt Nam năm 2015 quy định bốn điều kiện để pháp nhân thương mại phạm tội phải chịu TNHS, trong đó có điều kiện "*hành vi phạm tội được thực hiện có sự chỉ đạo, điều hành hoặc chấp thuận của pháp nhân thương mại*" (điểm c khoản 1 Điều 75). Đây là điều kiện phản ánh dấu hiệu "*lỗi*" của pháp nhân. Theo đó, pháp nhân thương mại phải chịu TNHS khi người đại diện (hoặc người đứng đầu) của pháp nhân nhận thức rõ hành vi của người đại diện thực hiện là trái pháp luật, nhưng vẫn chỉ đạo, cho phép, trực tiếp điều hành hoặc chấp thuận cho người đại diện thực hiện hành vi đó, nên pháp nhân thương mại bị coi là *có lỗi* (TG).

việc chứng minh yếu tố bên trong là rất khác nhau đối với từng loại hình thực thể AI. Hầu hết khả năng nhận thức được phát triển trong công nghệ AI hiện đại đều mang *tính phi vật chất* trong việc áp dụng TNHS. “*Sự sáng tạo*” là một đặc tính của con người mà một số loài vật cũng có. Tuy nhiên, sự sáng tạo không phải là yêu cầu đối với việc áp dụng TNHS. Yêu cầu duy nhất về mặt tinh thần (chủ quan) cần để áp dụng TNHS là sự hiểu biết, ý định, cảm xúc... như yêu cầu đối với hành vi phạm tội cụ thể và theo lý thuyết chung trong luật hình sự cả truyền thống lẫn hiện đại, cụ thể²⁵:

(a) *Sự hiểu biết* là cảm nhận các dữ liệu thực tế về mặt giác quan và nhận thức được về dữ liệu đó. Hiện nay, hầu hết thực thể AI được trang bị để tiếp nhận dữ liệu đó. Việc cảm nhận về mặt giác quan đối với những ánh nhìn, giọng nói, tiếp xúc về thể chất, sờ nắm,... rất phổ biến ở hầu hết thực thể AI [25; tr.108]. Việc cảm nhận này giúp chuyển dữ liệu thực tế nhận được đến đơn vị xử lý trung tâm giúp phân tích dữ liệu. Quy trình phân tích dữ liệu của AI tương tự như quy trình nhận thức của con người²⁶. Bộ não của con người hiểu biết về dữ liệu nhận được bằng mắt, tai, tay... thông qua việc phân tích dữ liệu đó. Các thuật toán AI tiên tiến, siêu việt được lập trình đã cố gắng bắt chước giống quá trình nhận thức của con người nên các quá trình này không quá khác biệt, thậm chí ngày càng siêu việt (chúng tôi nhấn mạnh - TG) [26; tr.190].

(b) *Ý định* là yêu cầu cao nhất về yếu tố bên trong [27; tr.380-414]. Ý định là ý muốn cụ thể làm một việc gì đó [14; tr.1167], còn ý định cụ thể là sự tồn tại của mục đích mà một sự

kiện thực tế sắp xảy ra. Do kết quả của việc tồn tại ý định này, người phạm tội đã thực hiện hành vi phạm tội, có nghĩa là người đó thực hiện yếu tố bên ngoài của hành vi phạm tội nhưng tình huống này không chỉ có ở con người. Một thực thể AI có thể được lập trình với mục đích (hay mục tiêu) xác định và phải thực hiện hành động để đạt được mục đích đó thì đây cũng được xem như tương đương là *ý định cụ thể*.

(c) *Cảm xúc*, một số quan điểm của nhà khoa học cho rằng con người có những cảm xúc mà thực thể AI không thể bắt chước thậm chí cả khi sử dụng phần mềm AI tiên tiến nhất, hiện đại nhất [28; tr.215]. Những cảm xúc đó bao gồm yêu thương, yêu mến, ghét, ghen và nhiều các cảm xúc khác (thậm chí cả cảm xúc đau đớn),... sẽ không thể có ở thực thể AI. Tuy nhiên, với sự phát triển vượt bậc của AI, những cảm xúc như con người thì thực thể AI cũng có thể có khả năng bắt chước đúng được, làm thay được, làm tốt được²⁷. Mặc dù đây là vấn đề khoa học viễn tưởng, nhưng rõ ràng khoảng cách giữa khoa học viễn tưởng và thực tế cho thấy đang xích lại gần nhau (điều này có thể minh chứng qua nhiều bộ phim khoa học viễn tưởng cho thấy rất nhiều vấn đề tưởng là viễn tưởng thì đến nay đã trở thành sự thật và hơn cả sự thật)²⁸;...

²⁵ Sự nhận thức của máy móc còn có thể được hình thành theo một trong ba cách sau: (1) Được lập trình; (2) Tự xuất hiện từ sự phức tạp hoặc (3) Được dạy. Xem: Toby Walsh, *Năm 2062 - Thời đại của trí thông minh nhân tạo*, Đỗ Tôn Minh Khoa dịch, Nxb. Tổng hợp thành phố Hồ Chí Minh, 2019, tr.74-75.

²⁶Theo quan điểm của phép tư duy biện chứng, *hoạt động nhận thức của con người đi từ trực quan sinh động đến tư duy trừu tượng, và từ tư duy trừu tượng đến thực tiễn*. Con đường nhận thức đó được thực hiện qua các giai đoạn từ đơn giản đến phức tạp, từ thấp đến cao, từ cụ thể đến trừu tượng, từ hình thức bên ngoài đến bản chất bên trong (TG).

²⁷ Ví dụ: Rô-bốt Sophia ra mắt vào năm 2015 và được công nhận là công dân của Arab Saudi vào ngày 25/10/2017, trở thành “*cô máy*” đầu tiên trong lịch sử làm được điều này. Tuyên bố trao quyền công dân cho nữ rô-bốt Sophia đã khiến dư luận tranh cãi dữ dội, vì cho rằng cô nàng rô-bốt còn được hưởng nhiều quyền lợi hơn cả phụ nữ ở quốc gia này. Rô-bốt Sophia được thiết kế với mục tiêu có thể học hỏi tính sáng tạo, cảm thông, lòng trắc ẩn và đặc biệt là trở thành cỗ máy thiên tài, thông minh hơn con người. Cô nàng rô-bốt này có thể thực hiện được những cuộc trò chuyện, giao tiếp khá linh hoạt với khoảng 60 sắc thái cảm xúc khác nhau. Sophia còn nhiều lần khiến dư luận thế giới dậy sóng với các phát ngôn sốc, trong đó có tuyên bố: “*OK, tôi sẽ hủy diệt loài người*”.

²⁸ Ví dụ: Những phát minh trở thành sự thật từ dự đoán khoa học viễn tưởng: (1) *iPad - 1968*: iPad là một món “đồ chơi công nghệ” đắt tiền và nổi tiếng nhưng phát minh đặc biệt này đã được mô tả từ cách đây 47 năm qua tác phẩm “*2001 - A Space Odyssey*” của nhà văn Arthur C. Clarke vì trong đó đã đề cập đến thiết bị có tên gọi là “*Newspad*” với các tính năng tương tự sản phẩm của “*nhà táo*”; (2) *Xe tăng*

Do đó, các kỹ sư công nghệ, nhà tương lai học, nhà luật học,... cần phải hợp tác với các nhà tâm lý học, khoa học thần kinh, ngôn ngữ học và giáo dục học để lý giải sáng tỏ thêm những vấn đề này và chứng minh được yếu tố chủ quan này - thì rõ ràng thực thể AI đã đáp ứng được [29; tr.4]²⁹.

- 1903: Những chiếc xe tăng đồ sộ và đầy uy lực trên chiến trường, một “con quái vật” bằng thép gây khiếp sợ cho đối phương với hỏa lực khủng khiếp. cỗ máy to lớn này đã được đề cập trong tác phẩm viễn tưởng nổi tiếng “*The Land Ironclads*” của tác giả HG Wells vào năm 1903, trong đó có đoạn mô tả về một thứ có “*thân hình bọc sắt*” không lồ, bắn ra những viên đạn với tốc độ chóng mặt và sau đó nhanh chóng biến mất trong màn đêm; (3) *Bom nguyên tử*: Thứ vũ khí đáng sợ này được biết đến khi Mỹ thả 2 quả bom xuống thành phố Hiroshima và Nagasaki của Nhật Bản. Đây cũng là một sáng kiến tồi tệ nhất trong lịch sử loài người, nó xảy ra trong hiện thực từ một nhận định của H.G. Wells trong “*The World Set Free*” vào năm 1914; (4) *Chat Video*: Công nghệ đàm thoại bằng hình ảnh từ xa đã được dự báo từ năm 1911, trong cuốn “*Ralph 124C 41+*”, Hugo Gernsback đã có những dòng mô tả về cách thức giao tiếp đặc biệt bằng video, đến năm 1964 thì nhà cung cấp dịch vụ viễn thông AT&T bắt đầu thử nghiệm tính năng này; (5) *Tàu ngầm*: Tàu ngầm được đánh giá là một công trình nghiên cứu rất có giá trị trong lĩnh vực quân sự của nhân loại. Jules Verne, tác giả cuốn tiểu thuyết khoa học viễn tưởng “*Hai vạn dặm dưới biển*” - ý tưởng về một con tàu có khả năng hoạt động dưới mặt nước biển trong một thời gian dài, điều thú vị là nó đã được viết vào năm 1869; (6) *Thẻ tín dụng*: Các hình thức thanh toán và hệ thống tiền tệ toàn cầu cũng vận hành trơn tru nhờ các máy ATM có mặt ở hầu khắp mọi nơi nhưng nó đã được tiên đoán từ trước trong cuốn “*Looking Backward*” của Edward Bellamy xuất bản năm 1888, trong đó mô tả về công dân tương lai sử dụng thẻ để thanh toán hàng hóa, dịch vụ mà không cần phải sử dụng tiền mặt; (7) *Con người chinh phục Mặt trăng*: Đây là một điều không tưởng đối với loài người, nhưng cách đó hàng trăm năm, tác giả Jules Verne trong tác phẩm “*From The Earth To The Moon*” của mình đã mô tả chính xác về cách thức mà chúng ta sẽ đặt chân lên Mặt trăng, chỉ có khác biệt duy nhất là việc sử dụng một quả pháo khổng lồ để tạo lực đẩy; (8) *Người đột biến*: Mặc dù khái niệm trên đã không còn trở nên xa lạ với hầu hết chúng ta qua nhiều bộ phim và đến thực tế hiện nay công nghệ biến đổi gen mới thực sự được nâng lên một nấc thang mới nhờ sự xuất hiện của CRISPR-Cas 9 (khám phá vào năm 2003),... và nhiều nhiều lĩnh vực khác nữa; v.v... Xem: *Những phát minh trở thành sự thật từ dự đoán khoa học viễn tưởng*, <https://www.lalung.vn/>, truy cập ngày 02/10/2019.

²⁹ Hiện nay, nhiều bộ phim khoa học viễn tưởng như “*Hạnh phúc ảo*”, 2013; “*Cuộc phiêu lưu vào dải ngân hà*”, 2014... đã phác họa rất rõ về chân dung một người máy có cảm xúc. Mặc dù là khoa học viễn tưởng, nhưng qua các bộ phim cho

Như vậy, nếu một chủ thể đáp ứng yêu cầu về yếu tố bên ngoài và yếu tố bên trong đối với hành vi phạm tội cụ thể thì chủ thể đó phải chịu TNHS. Câu hỏi đặt ra là vậy tại sao một thực thể AI đáp ứng các yếu tố trên thì lại không đặt ra vấn đề phải chịu TNHS, mặc dù yếu tố chủ quan vô cùng khó chứng minh (đã nêu), nhưng với sự phát triển vượt bậc, siêu việt của AI và các thuật toán AI có khả năng phân tích điều gì là bị cấm và điều gì được phép thực hiện [30; p.68]. Ngoài ra, khi một thuật toán AI vận hành đầy đủ, không có lý do gì để không sử dụng tất cả khả năng để phân tích dữ liệu thực tế nhận được thông qua cơ quan cảm giác. Đặc biệt, như đã đề cập ở trên, trong tương lai với sự xuất hiện loại trí tuệ mức siêu việt nhất với tên gọi là ASI (Artificial Superintelligence) chỉ việc “*máy móc*” thông minh siêu việt hơn con người thông minh nhất trên tất cả các khía cạnh hoàn toàn đáp ứng yếu tố chủ quan được. Tuy nhiên, cũng có một vấn đề pháp lý đặt ra là liệu có thể đưa ra luận cứ bảo vệ trong trường hợp *thuật toán AI gặp trục trặc và các khả năng phân tích không thể thực hiện được do trục trặc đó hay không* [1; tr.184]. Điều này trong quá trình lập trình, phân tích và trao quyền quyết định cho thực thể AI cần được làm sáng tỏ.

Tóm lại, TNHS của một thực thể AI không thay thế TNHS của người lập trình phần mềm hay người sử dụng nếu TNHS được áp dụng đối với người lập trình phần mềm và (hoặc) người sử dụng theo quy định pháp luật khác. Ở đây, TNHS không tách rời mà được kết hợp. TNHS của thực thể AI được áp dụng bên cạnh TNHS của người lập trình phần mềm hay người sử dụng. TNHS của một thực thể AI cũng không phụ thuộc vào TNHS của người lập trình phần mềm hay người sử dụng thực thể AI đó. Do đó, nếu một thực thể AI cụ thể được lập trình hay sử dụng bởi một thực thể AI khác, TNHS của thực thể AI được lập trình hay sử dụng không chịu ảnh hưởng của thực thể này. *Thực thể AI được lập trình hay sử dụng phải chịu TNHS đối với hành*

thấy, các nhà tương lai học, tâm lý học, khoa học thần kinh, công nghệ học, giáo dục học,... đã kết hợp tạo ra một người máy có cảm xúc thật như con người (TG).

vi phạm tội cụ thể theo mô hình TNHS trực tiếp trừ khi nó là tác nhân vô tội. Cùng với đó, người lập trình phần mềm hay người sử dụng thực thể AI đó phải chịu TNHS cho hành vi phạm tội đó theo một trong ba mô hình TNHS, tùy thuộc vào vai trò cụ thể của mình trong hành vi phạm tội. Chuỗi TNHS này có thể tiếp tục nếu có nhiều bên tham gia hơn, cho dù là con người hay thực thể AI. Ngoài ra, cũng không có lý do gì loại trừ TNHS của thực thể AI hay của con người trong trường hợp có sự thông đồng (nhất trí) giữa hai chủ thể này. Nếu thực thể AI và một con người có thể phối hợp như là đồng phạm thì TNHS có liên quan có thể được áp dụng tương ứng nếu thỏa mãn các dấu hiệu của đồng phạm/phạm tội có tổ chức trên những cơ sở chung [31; tr.457-482]; [32; tr.131-137]. Cùng với đó, lý luận về đồng phạm/phạm tội có tổ chức trong luật hình sự tiếp tục cần được nghiên cứu thấu đáo hơn.

TNHS của một thực thể AI theo mô hình TNHS trực tiếp không khác biệt so với TNHS có liên quan đối với con người. Trong một số trường hợp, cần phải điều chỉnh nhưng về cơ bản, TNHS này giống nhau và dựa trên các yếu tố và được truy cứu TNHS theo cách thức tương tự nhau. Vấn đề ở chỗ, khi quy định TNHS trực tiếp, việc quy định điều kiện TNHS như thế nào, vấn đề đồng phạm/phạm tội có tổ chức ra sao và áp dụng, thi hành các chế tài hình sự đối với đối tượng này sẽ như thế nào.

2.4. Mô hình kết hợp, đan xen các mô hình trách nhiệm hình sự đối với thực thể AI

Như vậy, các mô hình TNHS đã nêu không phải là những mô hình TNHS thay thế cho nhau. Có thể áp dụng kết hợp các mô hình này để hình thành bức tranh tổng thể về TNHS trong vụ án cụ thể nếu có sự tham gia hành vi của thực thể AI. Ngoài ra, cũng không có mô hình TNHS nào trong ba mô hình này loại trừ lẫn nhau. Việc kết hợp, liên hệ và đan xen giữa các mô hình TNHS được thể hiện như sau [1; p.182-183]:

(1) Trường hợp khi một thực thể AI đóng vai trò như là “tác nhân vô tội” (vật trung gian) trong việc thực hiện hành vi phạm tội cụ thể và người lập trình phần mềm là người duy nhất chỉ

đạo quá trình phạm tội, thì việc áp dụng mô hình TNHS của chủ thể phạm tội thông qua chủ thể khác (mô hình thứ nhất) là phù hợp nhất trong trường hợp đó.

(2) Trường hợp tương tự, khi bên lập trình phần mềm chính là thực thể AI (khi một thực thể AI lập trình cho một thực thể AI khác để thực hiện hành vi phạm tội cụ thể), mô hình TNHS trực tiếp (mô hình thứ ba) là phù hợp nhất để áp dụng đối với bên lập trình thực thể AI. Như vậy, trong trường hợp này, mô hình TNHS thứ ba được áp dụng bên cạnh mô hình TNHS thứ nhất chứ không thay thế cho mô hình này. Do đó, trong trường hợp đã nêu, bên lập trình phần mềm cho thực thể AI phải chịu TNHS theo mô hình kết hợp giữa mô hình TNHS của chủ thể phạm tội thông qua chủ thể khác và mô hình TNHS trực tiếp.

(3) Trường hợp nếu thực thể AI đóng vai trò là bên thực hiện hành vi phạm tội cụ thể trên thực tế nhưng hành vi đó không được lập kế hoạch trước, thì có thể áp dụng mô hình TNHS dựa trên hậu quả nguy hiểm cho xã hội có thể xảy ra một cách tự nhiên (mô hình thứ hai). Người lập trình phần mềm có thể được coi là có hành vi sơ suất (vô ý do câu thả) nếu hành vi được thực hiện một cách không cố ý hay người lập trình phần mềm có thể phải chịu hoàn toàn trách nhiệm cho hành vi phạm tội cụ thể đó nếu một hành vi phạm tội khác đã được lên kế hoạch một cách có chủ đích nhưng hành vi phạm tội cụ thể đã xảy ra không phải là một phần của kế hoạch phạm tội ban đầu. Tuy nhiên, khi bên lập trình phần mềm không phải là con người, cần áp dụng mô hình TNHS trực tiếp (mô hình thứ ba) bên cạnh việc áp dụng đồng thời mô hình TNHS dựa trên hậu quả nguy hiểm cho xã hội có thể xảy ra một cách tự nhiên, tương tự như trường hợp người phạm tội thực tế là con người nhưng người lập kế hoạch lại là thực thể AI (mô hình thứ hai).

Tóm lại, sự đan xen, liên hệ và kết hợp cả ba mô hình TNHS này tạo ra một mạng lưới TNHS đan xen, tổ hợp trong việc xác định TNHS đúng đắn của một thực thể trừu tượng cùng với cá nhân, pháp nhân và thực thể AI. Việc áp dụng kết hợp và phối hợp giữa ba mô

hình TNHS có thể lại dẫn đến một tình huống pháp lý mới trong bối cảnh cụ thể của các thực thể AI và luật hình sự. Cụ thể, khi có sự tham gia của thực thể AI [33 - 34; tr.1] và con người một cách trực tiếp hay gián tiếp trong quá trình thực hiện hành vi phạm tội, sẽ rất khó để trốn tránh TNHS. Một điều quan trọng là tất cả các thực thể - cá nhân, pháp nhân hay AI đều chịu sự điều chỉnh của PLHS nếu xâm phạm đến các lợi ích mà PLHS bảo vệ mặc dù hiện nay, TNHS đối với thực thể AI vẫn được xem là một điều viển vông, khoa học viễn tưởng, thậm chí không tưởng, nhưng dù là vậy, việc nghiên cứu, dự liệu và đặt ra vấn đề này là cần thiết đối với các nhà tương lai học, công nghệ học, tâm lý học và nhà hình sự học để hoạch định chính sách hình sự và xây dựng, hoàn thiện hệ thống PLHS trong tương lai không xa.

3. Một số viễn cảnh đặt ra trong chính sách, pháp luật hình sự Việt Nam tương lai

3.1. Nhận xét

Như vậy, với sự phát triển nhanh chóng của công nghệ AI đòi hỏi phải có các giải pháp pháp lý (cả PLHS) phù hợp để bảo vệ xã hội khỏi những nguy hiểm có thể xảy ra từ các loại hình công nghệ vốn không chịu sự điều chỉnh của pháp luật, đặc biệt là PLHS [3; tr.273]. Đây là một quá trình phức tạp và tranh luận kéo dài. Lấy ví dụ đơn giản, đối với thế giới khi bổ sung TNHS của pháp nhân, mặc dù pháp nhân đã tồn tại dưới hình thức hiện đại kể từ thế kỷ 14, nhưng phải mất hàng trăm năm để điều chỉnh pháp nhân theo quy định của pháp luật, đặc biệt là PLHS [34; tr.411, tr.419]. Ngay vào giữa thế kỷ XIX, các Tòa án Anh, xuất phát từ chính sách hình sự mềm dẻo và những lý do khá thực dụng đã tái thiết lập TNHS của pháp nhân [35; tr.13]. Pháp nhân tham gia đầy đủ vào cuộc sống của con người và thật là thiếu sót khi pháp nhân không chịu sự điều chỉnh của pháp luật của con người do các hành vi phạm tội được thực hiện bởi pháp nhân hoặc thông qua đại diện của pháp nhân. Tuy nhiên, pháp nhân

không phải là một cơ thể hay linh hồn. Các giải pháp về pháp lý được đưa ra liên quan đến TNHS theo đó pháp nhân có thể chịu TNHS khi đáp ứng các yêu cầu về yếu tố *bên ngoài* và yếu tố *bên trong* [1; tr.200]. BLHS nhiều quốc gia, trong đó có BLHS Việt Nam năm 2015, sửa đổi năm 2017 đã bổ sung thêm chủ thể của tội phạm là “*pháp nhân thương mại*” cùng với cá nhân phạm tội (mặc dù còn có ý kiến coi pháp nhân chỉ là chủ thể của TNHS) [36; tr.3-12]; [37; tr.14-15]; [38; tr.341-324]. Đây là một sự tiến bộ đáng ghi nhận trong lịch sử lập pháp hình sự nước ta, cũng là một điểm *thay đổi* tư duy pháp lý trong PLHS. Pháp nhân được coi là một “*thực thể trừu tượng*” được tạo ra bởi pháp luật thì không có lẽ gì - có thể là khoa học viễn tưởng, nếu đáp ứng tất cả các điều kiện cụ thể, có thể áp dụng TNHS đối với cả thực thể AI (bên cạnh cá nhân, pháp nhân) (tương lai chúng ta hãy cùng chờ xem). Vì vậy, thực tiễn phát triển của khoa học công nghệ đòi hỏi các nhà lập pháp phải có các giải pháp về phương diện pháp lý *hiện đại* để giải quyết những vấn đề của *thời đại*, cũng như xây dựng chính sách để đáp ứng nhu cầu điều chỉnh trong tương lai. Do đó, từ việc nghiên cứu các mô hình TNHS dự kiến (đã nêu) cho thấy có một số viễn cảnh đặt ra trong chính sách, PLHS Việt Nam tương lai có ý nghĩa cấp bách mà các tiểu mục 2 - 6 dưới đây sẽ cùng giải quyết.

3.2. Các cỗ máy mang AI sẽ có sự thay đổi, phát triển vượt bậc, siêu việt và không dự đoán được

Hiện nay, con người và thế giới (trong đó có Việt Nam) đang đứng tại thời điểm lịch sử chưa từng có khi các cỗ máy AI sẽ đưa ra các quyết định có ảnh hưởng tới nhiều lĩnh vực trong cuộc sống. Cụ thể, nhiều quyết định quan trọng với con người đã và đang được đưa ra thông qua các cơ chế và công cụ AI. Tháng 7/2017, AI của Facebook vốn được dạy dỗ bằng tiếng Anh, nhưng dường như nó đã chần chừ với những sắc thái và sự không nhất quán của thứ ngôn ngữ này, thay vì tiếp tục sử dụng tiếng Anh, chúng đã tự mình phát triển một hệ thống các mã mới để giao tiếp hiệu quả hơn - một ngôn

ngữ mới hình thành. Những điều này khiến những câu chuyện viễn tưởng về người máy (rô-bốt) thay thế con người và ngày càng khẳng định và sự lo lắng là điều có thể hiểu được. Ngay từ thừa AI còn sơ khai, con người đã được bao quanh những câu chuyện viễn tưởng mà nơi đó rô-bốt thống trị và loại người diệt vong. Những câu chuyện đó dù không có thực nhưng đã khiến chúng ta mặc định sẵn những thành kiến về AI [17; tr.1231]³⁰. Hay công ty bảo hiểm từ chối thanh toán các khoản bảo hiểm sức khỏe dựa trên thuật toán đánh giá rủi ro, hay cảnh sát quyết định bắt giữ cá nhân nào đó do nghi ngờ lên kế hoạch phạm tội dựa trên một mô hình dự báo từ siêu máy tính xử lý dữ liệu... Ngoài ra, một xu hướng mới trong ngành công nghiệp rô-bốt là học tăng cường sâu (*deep reinforcement learning*) khi mà nhà thiết kế chỉ đơn giản vạch ra các mục tiêu hành vi của hệ thống và nó tự động học hỏi bằng các tương tác trực tiếp với môi trường. Điều này dẫn tới hệ quả một hệ thống còn khó hiểu hơn. Mặt khác, khi AI được sử dụng để hỗ trợ đưa ra án phạt tù giam, hỗ trợ thẩm phán hay để thực hiện hỗ trợ thi hành án trong trại giam... rõ ràng cách thức nhận định và đưa ra quyết định phải rõ ràng và nhiều tiến bộ, phát triển vượt bậc, siêu việt khác nữa... Gần đây nhất, theo tác giả Toby Walsh, với sự phát triển của khoa học và công nghệ, ông dự báo đến năm 2062 là thời đại của *trí thông minh nhân tạo toàn năng* [39; tr.35], đồng thời đặt ra hàng loạt vấn đề mà thế giới phải đối mặt như:

(1) Người số³¹;

- (2) Không còn “*chúng ta*”;
- (3) Không còn ý thức;
- (4) Không còn công việc;
- (5) Không còn chiến tranh;
- (6) Không còn giá trị nhân văn;
- (7) Không còn bình đẳng;
- (8) Không còn riêng tư;
- (9) Không còn chính trị;
- (10) Phương Tây sụp đổ [39; tr.29-216].

Đáng chú ý, chúng ta cũng cần dự báo khả năng loài người bị máy móc tàn sát một cách vô tình hơn là cố ý,... [39; tr.57]. Do đó, cùng với thay đổi của thế giới, những chính sách, pháp luật liên quan đến lĩnh vực AI cần được đặc biệt đầu tư, quan tâm và nghiên cứu, nhất là trong lĩnh vực luật hình sự - việc sử dụng AI hoặc bản thân AI gắn trong thực thể thực hiện hành vi nguy hiểm cho xã hội, gây ra thiệt hại (hậu quả) nguy hiểm cho xã hội, xâm phạm đến quyền con người và trật tự, an toàn xã hội³².

3.3. Thay đổi lớn tư duy trong chính sách, pháp luật hình sự Việt Nam liên quan đến tội phạm, các dấu hiệu trong từng yếu tố cấu thành tội phạm mà do hành vi của thực thể AI thực hiện

Mặc dù trong thời gian qua, xét riêng lĩnh vực chính sách, PLHS nước ta còn phản ứng còn chậm đối với sự phát triển của AI trong nghiên cứu, đào tạo, giảng dạy,... Do đó, từ những biến đổi liên tục trên, với “*giả định*” thực thể AI là chủ thể của tội phạm và phải chịu TNHS, thì xét riêng trong chính sách hình sự và PLHS Việt Nam đòi hỏi phải có sự thay đổi tư duy trong tương lai như sau³³:

³⁰ Ví dụ: Trong một lần giao tiếp, Bob và Alice (hai AI của Facebook) đã từ bỏ sử dụng các cấu trúc ngữ pháp tiếng Anh và bắt đầu giao tiếp bằng cách sử dụng thứ ngôn ngữ mà chúng tự tạo ra. Bob đã mở đầu bằng câu “*I can i i everything else*”, Alice đáp lại “*balls have zero to me to me to me...*”. Và cuộc trò chuyện cứ thế tiếp tục theo cách đó. Cuộc hội thoại này bao hàm nhiều thứ có nghĩa chứ không đơn thuần là những gì chúng ta thấy.

³¹ *Người số* (Homodigitalis - sự tiến hóa của loài người Homo dưới dạng kỹ thuật số). Các hoạt động và không gian sinh hoạt của chúng ta sẽ dần được số hóa, thậm chí có khi chỉ có thể là trải nghiệm được dưới dạng số hóa mà thôi. Suy nghĩ của con người sẽ được thay thế bằng tư duy của máy móc. Cuộc sống của con người trong đời sống thực sẽ được mã hóa thành thế giới nhân tạo và thực thể ảo... Xem: Toby

Walsh, *Năm 2062 - Thời đại của trí thông minh nhân tạo*, Đỗ Tôn Minh Khoa dịch, Nxb. Tổng hợp thành phố Hồ Chí Minh, 2019, tr.10.

³² Chẳng hạn, hiện nay, trên thế giới ngày càng xuất hiện tội phạm sử dụng giọng nói AI để đánh cắp tiền. Cụ thể, theo Nhật báo The Wall Street Journal đưa tin, tội phạm mạng đã giả mạo thành công giọng nói của giám đốc điều hành của một công ty năng lượng để yêu cầu chuyển khoản khẩn cấp 243.000 USD trong một cuộc tấn công lừa đảo. Các chuyên gia về tội phạm mạng gọi đây là một trong những thủ đoạn bất thường nhất từng được sử dụng liên quan tới trí tuệ nhân tạo (AI)...

³³ Về các nội dung này, chúng tôi sẽ phân tích cụ thể trong một số nghiên cứu khác sắp tới (TG).

(1) Về chính sách hình sự, đòi hỏi phải xây dựng chủ thuyết, hệ quan điểm để lý giải, giải thích hàng loạt vấn đề định hướng về TNHS của thực thể AI, thực thể AI có gì khác biệt so với pháp nhân trong việc truy cứu TNHS. Liệu có cần quy định chủ thể của tội phạm mới trong PLHS không, nếu có quy định như thế nào. Hiện nay, các thực thể AI dưới dạng rô-bốt tham gia ngày càng nhiều vào hoạt động của con người tương tự như pháp nhân [4; tr.267]. Các hành vi phạm tội được thực hiện bởi thực thể AI hoặc thông qua thực thể AI [4; tr.273]. Các thực thể AI không có linh hồn và một số thực thể AI không có cơ thể hay không có linh hồn hoặc nếu AI phát triển thì cơ thể và linh hồn của AI sẽ như thế nào... Điều kiện chủ thể của thực thể AI bao gồm các điều kiện nào trên cơ sở khái niệm mới về tội phạm do thực thể AI thực hiện,...

Cùng với đó, cơ sở lý luận, thực tiễn của những loại tội phạm nào thực thể AI sẽ phải chịu TNHS, vấn đề đồng phạm giữa thực thể AI và các chủ thể khác (pháp nhân, cá nhân)... từ đó đưa ra *chính sách xử lý* và *hệ quan điểm* khi giải quyết vấn đề này và hệ thống chính sách, pháp luật chung cùng điều chỉnh,...

(2) Về PLHS, trên cơ sở chính sách hình sự trên, trong PLHS cần điều chỉnh hàng loạt vấn đề liên quan đến tội phạm, cấu thành tội phạm như: khái niệm tội phạm, các dấu hiệu đặc trưng trong các yếu tố cấu thành tội phạm do thực thể AI thực hiện được cụ thể hóa trong PLHS, vấn đề chủ thể của tội phạm và điều kiện chịu TNHS, đồng phạm và TNHS trong đồng phạm, hệ thống các biện pháp cưỡng chế về hình sự áp dụng đối với thực thể AI bên cạnh áp dụng với pháp nhân, cá nhân phạm tội,...³⁴.

3.4. Thay đổi lớn tư duy trong việc quy định chế tài hình sự đối với thực thể AI và việc quy định thủ tục, trình tự bắt giữ, điều tra, truy tố, xét xử, thi hành...

³⁴ Từ đây, dẫn đến các thay đổi lớn trong pháp luật tố tụng hình sự, pháp luật thi hành án hình sự và nhiều đạo luật khác có liên quan (TG).

Tiếp theo “*giả định*” trên, giả định tiếp rằng thực thể AI phải chịu TNHS, đồng thời bị truy tố, xét xử và buộc tội thì việc bắt giữ, điều tra, truy tố và xét xử thế nào³⁵. Sau khi buộc tội, Tòa án phải đưa ra bản án đối với thực thể AI thì hệ thống biện pháp cưỡng chế về hình sự sẽ áp dụng là gì (có nhiều loại biện pháp cưỡng chế về hình sự khác nhau, trong đó quan trọng nhất là hình phạt gì, thi hành thế nào). Tất cả những vấn đề này cần được dự liệu trong chính sách hình sự và trong xây dựng, áp dụng PLHS.

Ở đây, chúng tôi ví dụ một loại hình phạt - biện pháp cưỡng chế về hình sự nghiêm khắc nhất hiện đang áp dụng đối với *người phạm tội* là *tử hình* (còn đối với pháp nhân là *đình chỉ hoạt động vĩnh viễn*). Theo đó, tử hình là hình phạt đặc biệt chỉ áp dụng đối với người phạm tội đặc biệt nghiêm trọng thuộc một trong các nhóm tội xâm phạm an ninh quốc gia, xâm phạm tính mạng con người, các tội phạm về ma túy, tham nhũng và một số tội phạm đặc biệt nghiêm trọng khác do BLHS quy định. Vì vậy, câu hỏi đặt ra ở đây, trường hợp thực thể AI phạm tội cũng tương tự, thì hình phạt áp dụng đối với thực thể AI này tương tự như hình phạt tử hình là gì. Rõ ràng, “*mạng sống*” của một thực thể AI chính là *sự tồn tại độc lập* như là một thực thể. Đôi khi thực thể AI có bề ngoài thực thể (ví dụ, như là một con rô-bốt), đôi khi nó chỉ là sự tồn tại một cách trừu tượng (ví dụ, như là một phần mềm được cài đặt trên một hệ thống máy tính hay trên một máy chủ mạng lưới)... [1; tr.200]. Do đó, hình phạt tử hình đối với thực thể AI có thể ví dụ là: *xóa bỏ (hủy bỏ) phần mềm AI kiểm soát thực thể AI* (các nhà công nghệ học cần kết hợp với các nhà luật học). Tuy nhiên, việc quy định trình tự, thủ tục thi hành hình phạt này sẽ như thế nào (chưa nói đến các hình phạt khác và việc quy định thủ tục, trình tự bắt giữ, điều tra, truy tố, xét xử, thi hành ra sao,...). Những vấn đề này cũng rất quan trọng khi quy định TNHS đối với thực thể AI thì việc đặt ra hệ thống chế tài hình phạt dự

³⁵ Về nội dung này, chúng tôi sẽ tiếp tục đề cập trong một số nghiên cứu khác sắp tới (TG).

kiến áp dụng cũng là nhiệm vụ không kém phần quan trọng.

3.5. Phát sinh vấn đề đổi mới tư duy về chính sách, pháp luật tố tụng hình sự, pháp luật thi hành án hình sự, hoàn thiện hệ thống pháp luật có liên quan

Đây là vấn đề phát sinh từ các vấn đề trên. Đổi mới chính sách, PLHS cũng có nghĩa đồng thời phải đổi mới tư duy về xây dựng chính sách, pháp luật của các luật hình thức,... (luật tố tụng hình sự và luật thi hành án hình sự) trên cơ sở những định hướng cơ bản trong chính sách, PLHS, qua đó, tạo hệ thống tổng thể đáp ứng yêu cầu mới của đất nước và dự báo tương lai.

Ngoài ra, chính sách hình sự đặt ra yêu cầu hoàn thiện hệ thống pháp luật Việt Nam trong lĩnh vực công nghệ thông tin và có liên quan đến AI như:

(1) Luật Công nghệ cao năm 2008, sửa đổi năm 2013, 2014;

(2) Luật Công nghệ thông tin năm 2006;

(3) Luật An ninh thông tin mạng năm 2015;

(4) Luật An ninh mạng năm 2018;

(5) Và các văn bản tạo hàng lang pháp lý điều chỉnh các lĩnh vực cụ thể như: AI, tiền ảo, công nghệ in 3D,... tạo hệ thống toàn diện, tổng thể trong công tác đấu tranh phòng, chống tội phạm do thực thể AI thực hiện trước thách thức CMCN 4.0.

6. Nhu cầu lớn trong việc tăng cường hợp tác quốc tế, cũng như công tác đào tạo, bồi dưỡng cán bộ, đội ngũ chuyên môn tạo tổng thể hệ thống đáp ứng yêu cầu trước thách thức của thực thể AI có hoạt động gây ra các hậu quả nguy hiểm cho xã hội

Như vậy, cùng với việc đổi mới tư duy, hoàn thiện chính sách, PLHS thì cơ chế bảo đảm đòi hỏi thường xuyên tăng cường hợp tác quốc tế trong phòng, chống tội phạm với khu vực và toàn cầu như: hợp tác song phương, đa phương, tham gia các tổ chức phòng, chống tội phạm quốc tế nhằm từng bước tiếp cận, chuyên giao các công nghệ, kỹ năng phòng, chống tội phạm, nắm bắt kiến thức, công nghệ về sử dụng

các phương tiện công nghệ, kinh nghiệm phòng ngừa, phát hiện và đấu tranh chống tội phạm do thực thể AI thực hiện... Đặc biệt, việc có chính sách liên quan đến xây dựng đội ngũ cán bộ làm công tác lập pháp, hành pháp và tư pháp tinh thông về AI trong giải quyết, ứng phó cũng là xu hướng cần triển khai sớm.

Kết hợp nghiên cứu giữa các nhà luật học với các nhà tâm lý học, khoa học thần kinh, công nghệ học, ngôn ngữ học, giáo dục học và kinh tế học trong nghiên cứu về AI, chính sách để xử lý, kiểm soát thực thể AI và phòng ngừa thực thể AI vi phạm.

Cùng với việc đẩy mạnh nghiên cứu khoa học, phát triển và sử dụng thành thạo các ứng dụng khoa học và công nghệ trong công tác phòng, chống tội phạm, kết hợp với đầu tư kinh phí mua sắm, trang bị các phương tiện hiện đại để phát hiện, thu thập các loại tài liệu, chứng cứ điện tử... phục vụ công tác phát hiện, xử lý thì yêu cầu nâng cao trình độ, chuyên môn cho cán bộ là nhiệm vụ quan trọng để chủ động ứng phó, xử lý...

Tóm lại, nghiên cứu bước đầu về các mô hình TNHS đặt ra dự kiến có thể áp dụng đối với thực thể AI là một vấn đề vô cùng khó khăn và mới tại Việt Nam và cũng trở lên khoa học viễn tưởng khi “*thực thể AI*” vẫn do con người tạo ra và điều kiện chịu TNHS đối với một chủ thể vẫn đòi hỏi các yếu tố khách quan và chủ quan. Xã hội luôn biến đổi không ngừng, thay đổi trong lĩnh vực khoa học và công nghệ thì ngày càng trở lên như vũ bão và cũng khó dự đoán khó lường, không thể thấy trước được. Tuy nhiên, nghiên cứu này bước đầu dự liệu tương lai viễn cảnh ở nước ta, cũng như có thêm kiến nghị giúp cho việc nghiên cứu, hoạch định chính sách hình sự và xây dựng, hoàn thiện hệ thống PLHS trong bối cảnh CMCN 4.0³⁶ [40]; [41; tr.17-18]; [42; tr.204-216]; [43; tr.9-21], qua đó, không chỉ bảo vệ con người trước những hành vi nguy hiểm cho thực thể AI

³⁶ Ngày 27/9/2019, Bộ Chính trị đã ban hành Nghị quyết số 52-NQ/TW về một số chủ trương, chính sách chủ động tham gia cuộc CMCN lần thứ tư, trong đó tại mục III có đề ra một số chủ trương, chính sách chủ động tham gia cuộc CMCN lần thứ tư (TG).

thực hiện, mà còn có cơ chế kiểm soát đối với chủ thể này, nếu tương lai nó là chủ thể của tội phạm và phải chịu TNHS [39; tr.3-7]³⁷. Mặc dù vậy, là nghiên cứu đầu tiên ở Việt Nam nên chúng tôi không tránh khỏi có nhận xét, đánh giá mang tính phỏng đoán, còn mơ hồ và chưa toàn diện về cơ sở lý luận và thực tiễn vững chắc, nhưng chúng tôi tin rằng, sẽ có nhiều nghiên cứu sâu, rộng và liên ngành [2]; [44, 45; tr.1200-1236] về chủ đề này kết hợp trong tương lai bởi các nhà luật học, dự báo học, khoa học thần kinh, ngôn ngữ học, công nghệ học và tâm lý học.

Tài liệu tham khảo

- [1] Gabriel Hallevy, The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control, Akron Intellectual Property Journal, 4 (2), (2010).
<https://ideaexchange.uakron.edu/akronintellectualproperty/vol4/iss2/1>.
- [2] I.N. Mosechkin, Artificial Intelligence and Criminal Responsibility: Problems of the Formation of a New Type of Crime Subject, Vestnik SPbGU. Right, 3, 10 (2019).
- [3] Klaus Schwab, Fourth Industrial Revolution, National Political Publishing House - Truth, Hanoi, 2018 (in Vietnamese).
- [4] Yueh-Hsuan Weng, Chien-Hsun Chen, Chuen-Tsai Sun, Toward the Human-Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots, INT.J.SOC, ROBOT, 2009.
- [5] Klaus Schwab, The fourth industrial revolution: What it means, how to respond, World Economic Forum, Geneva, 2016.
- [6] Kevin Kelly, 12 technology trends in the 4.0 era, National Economics University Press, Hanoi, 2018 (in Vietnamese).
- [7] Matilda Claussén-Karlsson, Artificial Intelligence and the External Element of the Crime: An Analysis of the Liability Problem, JU101A, Final Thesis for the Law Program, Second Cycle, 30 Credits, Spring 2017.
- [8] Tien Phuc, Five frightening warnings about future AI disasters, <http://kenh14.vn/>, accessed May 1, 2019 (in Vietnamese).
- [9] Amir Husain, The Sentient Machine: The Coming Age of Artificial Intelligence, translated by Hoang Lan, Industrial and Commercial Publishing, Hanoi, 2019.
- [10] Trinh Tien Viet, Policy and Criminal Law of Vietnam against the challenge of Industrial Revolution 4.0, Proceedings of the National Scientific Conference: "Fourth Industrial Revolution and legal issues posed for the construction and completion of Vietnam legal system", Ministry of Justice, Hanoi, June 24, 2019 (in Vietnamese).
- [11] Trinh Tien Viet, From the Beginning to the End of the Criminal Law, Journal of Law, VNU, Episode 1 (34) (2018) (in Vietnamese).
- [12] Trinh Tien Viet, Crime and criminal responsibility, National Political Publishing House, Hanoi, 2013 (in Vietnamese).
- [13] Lisa M Storm, Criminal Law, Chapter 4 - "Elements of a crime", Lulu Publishing Services, United States, 2015.
- [14] Hoang Phe (chief editor), Vietnamese dictionary, Institute of Linguistics, Danang Publishing House and Dictionary Center, 2006 (in Vietnamese).
- [15] Nguyen Thi Que Anh, Ngo Huy Cuong (co-editor), The Fourth Industrial Revolution and issues on Vietnam's legal reform, National Political Publishing House - Truth, Hanoi, 2018 (in Vietnamese).
- [16] Klaus Schwab, Shaping the Fourth Industrial Revolution, translated by Nguyen Van and ThanhThep, World Publishing House, Hanoi, 2019 (in Vietnamese).
- [17] Lawrence B. Solum, Legal Personhood for Artificial Intelligences, 70 N.C.L.REV, 1992.
- [18] Roger C. Schank, What Is AI Anyway?, AI MAG., Winter 1987.
- [19] George R. Cross, Cary G. Debessonnet, An Artificial Intelligence Application in the Law: CCLIPS, A Computer Program that Processes Legal Information, 1 HIGH TECH. L.J., 1986.

³⁷ Tác giả Toby Walsh không chỉ trình bày các thành tựu của công nghệ máy tính bằng văn phong và ngôn ngữ truyền cảm hứng, tập sách có cách dẫn dắt tuyệt vời để người đọc - dù không chuyên về lĩnh vực máy tính - cảm nhận được rằng cái ngày trí thông minh nhân tạo hiện diện trong nhiều cảnh huống của con người truyền thống, là đang gần đến như đưa tay sờ thấy được. Tác giả đặt một hình dung về nhân loại từ Người tinh khôn đến *Người số* - một giống loài được tạo ra bằng kỹ thuật số và sống trong môi trường kỹ thuật số. Trí thông minh nhân tạo sẽ phá hủy hay đóng góp cho sự phát triển của xã hội loài người... Xem: Toby Walsh, *Năm 2062 - Thời đại của trí thông minh nhân tạo*, Đỗ Tôn Minh Khoa dịch, Nxb. tổng hợp thành phố Hồ Chí Minh, 2019.

- [20] Ajay Agrawal, Joshua Gans, Avi Goldfarb, *Artificial Intelligence in the 4.0 technology revolution*, translated by Dao Loi, Labor Publishing House, Hanoi, 2018 (in Vietnamese).
- [21] Robert P. Fine, Gary M. Cohen, *Is Criminal Negligence a Defensible Basic for Criminal Liability?* 16 BUFF. L. REV., 1966.
- [22] Donald Stuart, *Mens Rea, Negligence and Attempts*, CRIM. L. REV., 1968.
- [23] Steven J. Frank, *Tort Adjudication and the Emergence of Artificial Intelligence Software*, 21 SUFFOLK U. L. REV., 1987.
- [24] Maruerite E. Gerstner, *Liability Issues with Artificial Intelligence Software*, 33 SANTA CLARA L. REV., 1993.
- [25] Margaret A. Boden, *Has AI Helped Psychology?, The Foundations of Artificial Intelligence*, Derek Partridge & Yorick Wilks eds., 2006.
- [26] Daniel C. Dennett, *Evolution, Error, and Intentionality, The Foundations of Artificial Intelligence* Derek Partridge & Yorick Wilks eds., 2006.
- [27] Robert Batey, *Judicial Exploration of Mens Rea Confusion at Common Law and Under the Model Penal Code*, 18 GA. ST. U. L. REV. 341, 2001.
- [28] David Levy, *The Ethical Treatment of Artificially Conscious Robots*, I INT'L J. SOC. ROBOTICS, 2009.
- [29] Dylan Evans, *Can robots have emotions?*, Oxford University Press, 2001, http://www.inf.ed.ac.uk/events/hotseat/dylan_position.pdf, truy cập ngày 19/9/2019.
- [30] Tang Howe Hing, Musa Mailah, *Simulator for Control of Control of Autonomous Nonholomich Wheeled Robot*, 8 J. OF APPLIED SCI. 2534, 2536, 2008.
- [31] Le Cam, *Postgraduate Monograph: Basic issues in Criminal Law science (General Part)*, Hanoi National University Press, 2005 (in Vietnamese).
- [32] Dinh Van Que, *Commenting on Criminal Code 2015 - Part 1: General provisions*, Information and Communications Publishing House, Hanoi, 2017 (in Vietnamese).
- [33] *AI accompanies the threat*, <http://www.pcworld.com.vn/>, accessed September 15, 2019 (in Vietnamese).
- [34] Andrew Weissmann, David Newman, *Rethinking Criminal Corporate Liability*, 82 IND. L.J., 2007.
- [35] Trinh Quoc Toan, *Criminal Responsibility of Legal Entities in Criminal Law*, National Political Publishing House, Hanoi, 2011 (in Vietnamese).
- [36] Nguyen Ngoc Hoa, *The concept of crime and the determination of criminal liability of commercial legal entities in Vietnam's Criminal Code 2015*, Juricprudence Journal, 2, 2016 (in Vietnamese).
- [37] Nguyen Ngoc Hoa, *Scientific commentary Criminal Code 2015, amended and supplemented in 2017 (General Part)*, Justice Publishing House, Hanoi, 2017 (in Vietnamese).
- [38] Le Cam, *Vietnam's criminal law from the tenth century to the present: History and reality*, Hanoi National University House, 2018 (in Vietnamese).
- [39] Toby Walsh, 2062 - *The era of artificial intelligence*, translated by Do Ton Minh Khoa, Ho Chi Minh City General Publishing House, 2019 (in Vietnamese).
- [40] Politburo, *Resolution No.52-NQ/TW on a number of guidelines and policies to actively participate in the Fourth Industrial Revolution*, Hanoi, September 27 (2019) (in Vietnamese).
- [41] Trinh Tien Viet, *Continuing to innovate legal thinking in Vietnamese Criminal Law before the challenge of Industrial Revolution 4.0*, Journal of Procuratorate Studies, 4 (31) (2019) (in Vietnamese).
- [42] Trinh Tien Viet, *Vietnam's criminal law before to non-traditional security challenges*, National Politice Publishing House, Hanoi, 2019 (in Vietnamese).
- [43] Trinh Tien Viet, *Vietnam's Criminal Policy in the Context of the Fourth Industrial Revolution*, Journal of the People's Court, 7 (4) (2019) (in Vietnamese).
- [44] <http://www-formal.stanford.edu/jmc/whatisai/node1.html>, truy cập ngày 10/10/2019.
- [45] Houbing Song, Ravi Srinivasan, Tamim Sookor, Sabina Jeschke (ed.), *Smart Cities: Foundations, principles and applications*, translated by Ngo Thanh Nam, National Political Publishing House - Truth, Hanoi, 2019 (in Vietnamese).