

Yêu cầu sửa đổi Bộ luật Hình sự Việt Nam và tham khảo tái cấu trúc trách nhiệm hình sự một số tội danh trong kỷ nguyên trí tuệ nhân tạo

18/06/2026 15:46

Từ khi có các hệ thống trí tuệ nhân tạo (AI) ra đời đã tác động thay đổi to lớn đến mọi quan hệ xã hội mà luật hình sự bảo vệ, phát sinh yêu cầu cấp thiết phải thay đổi để thích ứng phòng ngừa và giải quyết tốt các hành vi phạm tội mới trong kỷ nguyên số. Bài viết phân tích bất cập hiện hành về cấu thành tội phạm; tham khảo kinh nghiệm một số trường hợp điển hình có hành vi phạm tội liên quan AI ở Liên minh Châu Âu (EU), Hoa Kỳ và Trung Quốc. Qua đó, đề xuất xây dựng khung lý thuyết trách nhiệm hình sự trong môi trường số, làm cơ sở sửa đổi, bổ sung cấu trúc trách nhiệm pháp lý ở một số tội danh của Bộ luật Hình sự (BLHS).

Đặt vấn đề

Khung lý thuyết cơ bản về các yếu tố cấu thành tội phạm (khoản 1 Điều 8 BLHS), mà pháp luật hình sự Việt Nam xây dựng là dựa trên nền tảng hành vi của con người, bao gồm: yếu tố khách quan (Hành vi nguy hiểm cho xã hội); yếu tố chủ thể (Cá nhân/pháp nhân thương mại); yếu tố chủ quan (Lỗi cố ý/vô ý) và yếu tố khách thể (Quan hệ xã hội được luật hình sự bảo vệ). Nói cách khác, cơ sở xác định một hành vi phạm tội theo lý luận truyền thống phải là hành vi do con người thực hiện; mang tính vật lý; có thể xác định rõ về thời gian, không gian và mối quan hệ nhân quả. Tuy nhiên, sự phát triển và ứng dụng rộng rãi các hệ thống AI đã làm lung lay nền tảng của khung lý thuyết hiện thời - hoạt động với những đặc điểm vượt ra ngoài khuôn khổ lập luận truyền thống:

Thứ nhất, AI vận hành với mức độ tự chủ cao, thực hiện hành vi mà không cần sự kiểm soát trực tiếp, liên tục của con người, qua đó loại trừ cơ sở xác định “lỗi” - yếu tố “phi chủ quan”;

Thứ hai, quá trình ra quyết định của hầu hết AI mang tính “hộp đen”, nên tính minh bạch sẽ khó giải thích, khó xác định hành vi khách quan và mối quan hệ nhân quả với hậu quả - yếu tố “phi khách quan”;

Thứ ba, AI hoạt động với tốc độ cao, phạm vi rất lớn và tạo ra hệ quả thiệt hại lan rộng ngay, phi vật lý hoặc khó định lượng. Rất trở ngại cho việc xác định quan hệ xã hội nào bị xâm hại - yếu tố “phi khách thể”;

Thứ tư, AI không phải là chủ thể pháp luật theo nghĩa truyền thống, không có ý

chỉ pháp lý và không thể chịu trách nhiệm hình sự độc lập, đưa đến thách thức trong việc quy buộc trách nhiệm cụ thể ở các trường hợp có sự tham gia của nhiều chủ thể, như: nhà phát triển AI, vận hành AI, người sử dụng AI - thể hiện yếu tố “phi chủ thể”¹.

Như vậy, các hành vi tố tụng liên quan AI sẽ gây lúng túng trong việc vận dụng các yếu tố cấu thành tội phạm cụ thể của BLHS, như: thiệt hại trực tiếp là do hệ thống tự động gây ra; hành vi thao túng hoặc lợi dụng thuật toán; hành vi tội phạm mới về giả mạo bằng công nghệ deepfake; cũng như vấn đề quy buộc trách nhiệm pháp lý các chủ thể trong chuỗi ứng dụng phát triển và vận hành AI. Sự bất cập không chỉ đặt ra thách thức đối với hiệu quả thực thi pháp luật mà còn ảnh hưởng đến yêu cầu bảo đảm công bằng, minh bạch và tính vững chắc pháp lý trong xử lý tội phạm². Do đó, trọng tâm bài viết đặt ra là việc định hình lại cách thức tiếp cận tội danh trong môi trường số và tái cấu trúc trách nhiệm hình sự đối với những hành vi phạm tội đang “núp bóng” dưới các giao thức hoạt động công nghệ số.

1. So sánh pháp luật và kinh nghiệm pháp lý ở các quốc gia phát triển và hiện trạng pháp luật hình sự ở Việt Nam

1.1. Liên minh Châu Âu (EU)

“AI Act”³ (Luật Trí tuệ Nhân tạo) của EU được xem là khuôn khổ pháp lý toàn diện điều chỉnh AI, đánh dấu bước ngoặt quan trọng trong quản trị công nghệ số. Đây là đạo luật có ý nghĩa toàn cầu sâu rộng tương tự như quy định “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016” - GDPR⁴ - nhằm thúc đẩy các quốc gia EU xây dựng khung pháp lý AI⁵ theo hướng minh bạch, có trách nhiệm và lấy con người làm trung tâm; chú trọng cơ chế phòng ngừa, yêu cầu đánh giá tác động, minh bạch thuật toán và giám sát liên tục đối với AI rủi ro cao. Điểm đặc biệt của mô hình EU là chú trọng quản trị khả năng gây hại tiềm tàng; ví dụ: AI nhận diện khuôn mặt, hệ thống chấm điểm tín dụng, AI trong y tế... phải tuân thủ nghiêm ngặt quy tắc trong đánh giá tác động, sự minh bạch và cơ chế giám sát.

Trường hợp điển hình, như vụ việc “Clearview AI”⁶ liên quan đến các hệ thống AI sinh trắc học, nổi lên vào năm 2020 - đã dẫn đến việc thiết lập nhiều chương trình bảo mật thông tin cá nhân của các cơ quan bảo vệ dữ liệu EU (từ năm 2021 - 2024). Cụ thể, Clearview AI (Hoa Kỳ) là công ty chuyên phát triển hệ thống nhận diện khuôn mặt sử dụng AI, đã thu thập hàng tỷ hình ảnh khuôn mặt từ nguồn công khai trên Internet mà không có sự đồng ý của cá nhân liên quan, sau đó xây dựng một cơ sở dữ liệu phục vụ cho việc nhận diện và truy vết danh tính. Tuy nhiên, hoạt động này gây ra tranh chấp pháp lý nghiêm trọng, đặc biệt liên quan đến quyền riêng tư và bảo vệ dữ liệu cá nhân. Nhiều cơ quan quản lý tại EU (Pháp, Ý, Anh)⁷ đã kết luận rằng Clearview AI vi phạm các quy định của GDPR, do xử lý dữ liệu sinh trắc học mà không có căn cứ pháp lý. Công ty đã bị xử phạt hành chính nhiều lần với mức phạt hàng chục triệu euro/lần và bị yêu cầu xóa dữ

liệu của công dân EU.

Vụ việc cho thấy, cách EU áp dụng nguyên tắc bảo vệ dữ liệu và quyền riêng tư đối với AI: không cần chứng minh thiệt hại cụ thể, chỉ cần hệ thống có nguy cơ xâm phạm quyền cơ bản là đã đủ để áp dụng chế tài - minh họa rõ hơn phương pháp tiếp cận của mô hình EU có ưu điểm ở tính dự phòng và khả năng bảo vệ quyền con người. Tuy nhiên, như nêu trên, nó cũng bị phê phán là có thể làm chậm đổi mới công nghệ và tạo gánh nặng tuân thủ cho doanh nghiệp.

1.2. Hoa Kỳ

Khác với cách tiếp cận thống nhất và mang tính quy phạm cao của EU, Hoa Kỳ thực thi cơ chế điều chỉnh AI mang tính phân tán, dựa trên hệ thống án lệ và các quy định theo ngành. Theo *National Institute of Standards and Technology* (NIST, 2023), Hoa Kỳ ưu tiên xây dựng các khung quản trị mềm nhằm định hướng doanh nghiệp thay vì áp đặt các nghĩa vụ pháp lý cứng nhắc⁸; thể hiện cơ chế trách nhiệm pháp lý mang tính phân tán, có thể gồm nhiều chủ thể cùng chịu trách nhiệm. Pháp luật Hoa Kỳ không quy định AI là chủ thể trách nhiệm hình sự, việc truy tố cá nhân/pháp nhân do Bộ Tư pháp Hoa Kỳ (DOJ) quyết định dựa trên sự cáo buộc trách nhiệm là người tạo ra, huấn luyện hoặc sử dụng AI để phạm tội; Hoặc, coi việc sử dụng AI là yếu tố tăng nặng dẫn đến mức án tù dài hơn⁹.

Năm 2018, tại Hoa Kỳ đã xảy ra vụ tai nạn xe tự lái của Uber tại Bang Arizona: Một phương tiện tự lái đang trong quá trình thử nghiệm đã va chạm và gây tử vong cho một người đi bộ. Cuộc điều tra cho thấy hệ thống AI của xe đã không nhận diện kịp thời nạn nhân, đồng thời người giám sát an toàn trong xe không can thiệp đúng lúc (National Transportation Safety Board - NTSB, 2019)¹⁰. Việc xử lý không quy trách nhiệm hình sự cho hệ thống AI, mà cáo buộc đối với cá nhân người vận hành - giám sát an toàn, bị truy tố về tội sơ suất gây chết người. Từ góc độ pháp lý, vụ việc phản ánh rõ cách tiếp cận: AI không được công nhận là chủ thể chịu trách nhiệm hình sự, chỉ con người/pháp nhân mới có thể mang lỗi và do đó mới có thể bị quy trách nhiệm hình sự - AI dù có vai trò trực tiếp gây hậu quả, trách nhiệm pháp lý vẫn phải được truy xét về các chủ thể con người.

1.3. Trung Quốc

Cuối năm 2022, sự bùng nổ ChatGPT với khả năng tạo ra và phổ biến thông tin của công nghệ tiềm ẩn gây bất ổn cho quản lý nhà nước về bảo vệ quyền riêng tư, an toàn và công bằng¹¹. Vì lý do đó, Trung Quốc (TQ) có thể được xem là một trong những quốc gia tiên phong trong việc xây dựng khung pháp lý chuyên biệt đối với công nghệ Deepfake, với “Các quy định về quản lý dịch vụ thông tin Internet tổng hợp sâu” năm 2022¹² - đặt ra nghĩa vụ toàn diện đối với nhà cung cấp dịch vụ AI, bao gồm quản lý nội dung, bảo vệ dữ liệu và đặc biệt là yêu cầu gắn nhãn mác rõ ràng đối với nội dung được tạo ra từ AI. Cụ thể, mọi nội dung như hình ảnh, video, âm thanh hoặc văn bản do AI “sáng tác” đều phải đánh dấu,

giúp người dùng nhận diện và tránh bị nhầm lẫn với thông tin thật. Các doanh nghiệp kinh doanh AI tại TQ như công ty “Douyin”¹³ buộc phải triển khai cơ chế nhận diện, hiển thị nhãn mác về nội dung AI, đồng thời lưu trữ metadata để đảm bảo khả năng truy vết nguồn gốc thông tin. Cho thấy, TQ không giới hạn vào xử lý hậu quả, mà kiểm soát từ khi tạo lập nội dung, thông qua các nghĩa vụ kỹ thuật và quản trị chặt chẽ đối với nhà cung cấp dịch vụ.

Tháng 5/2023, TQ tiếp tục ban hành “Các biện pháp tạm thời để quản lý dịch vụ trí tuệ nhân tạo tạo sinh” nhằm tác động sự phát triển có trách nhiệm và điều chỉnh việc sử dụng công nghệ; đồng thời, bảo vệ lợi ích quốc gia và quyền của công dân TQ¹⁴. Đây là mô hình mà nhà nước đóng vai trò trung tâm trong việc định hình tiêu chuẩn công nghệ và hành vi thị trường. Tóm lại, TQ điều chỉnh doanh nghiệp công nghệ theo khuôn khổ của ba trụ cột: an ninh, quyền riêng tư và cạnh tranh. Mỗi trụ cột đều có luật cơ bản và bộ quy định riêng, phản ánh rõ cách điều tiết cân bằng giữa đổi mới và kiểm soát đối với AI¹⁵.

Khảo sát ba mô hình trên, cho thấy không chỉ khác biệt về kỹ thuật pháp lý, mà còn phản ánh ba hướng tiếp cận nền tảng - EU ưu tiên quyền con người và quản trị rủi ro; Hoa Kỳ đề cao trách nhiệm cá nhân và tính linh hoạt; TQ nhấn mạnh kiểm soát nhà nước và ổn định xã hội. Tuy nhiên, thực tế hiện nay sẽ khó có thể có mô hình quản trị nào hoàn toàn tối ưu. EU có thể phải đối mặt với nguy cơ làm chậm đổi mới; Hoa Kỳ thiếu tính thống nhất; TQ tuy có hiệu quả cao trong kiểm soát rủi ro nhanh chóng, nhưng có thể va chạm mạnh quyền tự do cá nhân, làm tăng chi phí xã hội do sự can thiệp sâu của nhà nước. Điểm chung quan trọng là cả ba hệ thống đều là đang thay đổi mô hình trách nhiệm cá nhân thuần túy, hướng tới hình thức trách nhiệm đa dạng hơn - xác định rõ chủ thể chịu trách nhiệm trong bối cảnh AI ngày càng tự động và phức tạp.

1.4. Hiện trạng pháp luật hình sự ở Việt Nam

Ở Việt Nam, các nghiên cứu gần đây đã bước đầu nêu lên những bất cập pháp lý mới. Theo một số tác giả nghiên cứu¹⁶, pháp luật hiện chưa có quy định cụ thể về trách nhiệm pháp lý đối với hành vi gây thiệt hại do hệ thống AI “tạo sinh”, dẫn đến khó khăn trong việc áp dụng các chế định hiện hành như trách nhiệm dân sự hoặc hình sự. Tương tự, nhiều ý kiến cũng cho rằng việc thiếu các tiêu chuẩn về minh bạch thuật toán và kiểm soát rủi ro, có thể làm phát sinh tăng hơn việc xâm hại quyền con người trong môi trường số¹⁷.

Mặc dù, gần đây Việt Nam đã có những nỗ lực đáng kể sửa đổi, bổ sung một số luật nhằm thích ứng với yêu cầu phát triển trí tuệ nhân tạo, như: Luật An ninh mạng 2025; Luật Trí tuệ nhân tạo 2025 và Luật Bảo vệ cơ sở dữ liệu cá nhân 2025; nhưng, đều là các luật chuyên biệt điều chỉnh một nhóm quan xã hội nhất định thuộc lĩnh vực quản lý nhà nước, ít có tính đấu tranh trực tiếp về xử lý và phòng ngừa vi phạm, tội phạm trong lĩnh vực hình sự liên quan AI. BLHS sửa đổi, bổ sung năm 2025 cũng chỉ tập trung vào những nội dung: Bộ án tử hình ở 8

tội danh, xem xét chính sách đối với nhóm tội tham nhũng đã khắc phục hậu quả, bổ sung tội phạm về ma túy.v.v. Đáng lưu ý, các tội danh liên quan đến môi trường số tại BLHS hiện hành không hề thay đổi so với lần sửa đổi của 10 năm trước (2017). Trong 424 tội danh của BLHS, hiện chỉ ghi nhận 13 tội danh liên quan Internet, gồm: 9 tội danh trong lĩnh vực công nghệ thông tin, mạng viễn thông tại Mục 2, Chương XXI (Điều 285 đến Điều 294) và 4 tội danh rải rác ở các chương khác là các Điều 117, 155, 156 và 326. Đặc biệt, BLHS với 13 tội danh trên hoàn toàn không có yếu tố tương tác và đồng bộ với Luật An ninh mạng - Luật này ghi nhận khoảng 70 hành vi tội phạm trên môi trường số¹⁸ - cho thấy, BLHS hiện nay khó có thể đấu tranh ngang tầm những tội phạm tinh vi về AI.

2. Đề xuất định hình lại khung lý thuyết về trách nhiệm hình sự và tái cấu trúc trách nhiệm pháp lý ở một số tội danh

2.1. Về khung lý thuyết

Công nghệ AI và tội phạm sử dụng AI đã làm lung lay những tiền đề lý luận cốt lõi của luật hình sự truyền thống. *Trước hết*, ở phương diện *hành vi khách quan*, nhận thức truyền thống đối với hành vi phạm tội là do con người thực hiện, có thể xác định không gian, thời gian, quan hệ nhân quả. Tuy nhiên, trong môi trường AI, hành vi gây thiệt hại thường mang tính gián tiếp, trung gian và phân tán. Một quyết định gây hậu quả pháp lý có thể là kết quả của chuỗi xử lý dữ liệu phức tạp, kết hợp giữa thuật toán học, dữ liệu đầu vào và điều kiện vận hành. Do đó, hành vi phạm tội không phải đơn lẻ mà là kết quả của “chuỗi” thao tác hệ thống.

Thứ hai, ở phương diện *lỗi*, AI không có ý chí, nhận thức hay động cơ theo nghĩa pháp lý truyền thống, mà ý chí con người lại bị phân mảnh trong vòng đời thiết kế thuật toán; khác với cấu trúc “một chủ thể - một hành vi - một hậu quả” như trước đây, hệ thống AI liên quan nhiều chủ thể, và, mỗi chủ thể đóng một vai trò khác nhau trong việc tạo lập rủi ro/hậu quả thiệt hại. Vì vậy, nếu tiếp tục mô hình trách nhiệm đơn tuyến, luật hình sự sẽ khó khăn khi xác định chủ thể trách nhiệm và mức độ gánh chịu hậu quả pháp lý tương ứng.

Như vậy, yêu cầu định hình lại về khung lý thuyết trách nhiệm hình sự trong môi trường số cần nhận thức nguyên tắc nền tảng, đó là: (i) Nguyên tắc trách nhiệm pháp lý là thuộc con người - AI không thể là chủ thể của quan hệ pháp luật hình sự. (ii) Nguyên tắc đánh giá trách nhiệm theo chức năng, tránh lối xem xét trách nhiệm chung chung, mà cần căn cứ vào vai trò từng chủ thể trong vòng đời AI để cá thể hóa trách nhiệm pháp lý hình sự, như: tội phạm thuộc nhóm thiết kế, phát triển; tội phạm nhóm vận hành, sử dụng; nhóm thu thập dữ liệu và nhóm cài đặt hệ thống tự động.

2.2. Về cấu trúc trách nhiệm hình sự ở một số tội danh

Từ những phân tích trên đây, bài viết đề xuất tái cấu trúc trách nhiệm hình sự một số tội danh của BLHS hiện hành, như sau:

2.2.1. Nhóm tội xâm phạm an ninh quốc gia (Chương XIII)

- Tội phản bội tổ quốc (Điều 108):

Bổ sung hành vi sử dụng AI giả mạo thông tin tình báo, tài liệu ngoại giao; dùng deepfake gây nhiễu loạn chiến lược về nhận thức xã hội làm sai lệch nhận thức. Mở rộng khái niệm “Cấu kết” bao gồm cả việc hợp tác thông qua nền tảng công nghệ AI xuyên biên giới.

- Tội hoạt động nhằm lật đổ chính quyền (Điều 109):

Bổ sung hành vi huấn luyện, sử dụng AI để tổ chức, điều phối hoạt động chống đối mà không cần thành lập tổ chức truyền thống; Tạo bot AI tự động hóa việc lan truyền thông tin bịa đặt, kích động; đồng thời, kích hoạt hệ thống AI cá nhân hóa nội dung tuyên truyền nhằm gia tăng hiệu quả lật đổ chính quyền. Bổ sung loại hình mới “Tổ chức ảo” như một hình thức cấu thành tội phạm.

- Tội làm, tàng trữ, phát tán hoặc tuyên truyền thông tin, tài liệu, vật phẩm nhằm chống Nhà nước CHXHCNVN (Điều 117):

Bổ sung hành vi dùng Deepfake, nội dung giả mạo do AI tạo dựng, ứng dụng thuật toán tự động phát tán nội dung chống nhà nước. Xác định khái niệm “phát tán” là bao gồm cả việc thiết lập, vận hành hệ thống có khả năng tự động lan truyền nội dung nhằm thao túng dư luận xã hội.

Việc bổ sung các hành vi khách quan liên quan đến AI ở nhóm tội xâm phạm an ninh quốc gia là cần thiết, nhằm chuyển góc độ tiếp cận chỉ dựa trên hành vi cơ học sang hành vi số hóa/tự động hóa và mang tính hệ thống.

2.2.2. Nhóm tội xâm phạm sở hữu (Chương XVI)

- Tội trộm cắp tài sản (Điều 173):

Nếu đặc trưng của hành vi trộm cắp truyền thống là “lén lút” và nhằm chiếm đoạt tài sản vật lý (tiền, vật có giá trị), thì trong thời đại số, tài sản không thuần túy “vật” mà chuyển sang dạng “số”, như tài khoản mạng, ví điện tử, tài sản số. Bổ sung hành vi phạm tội sử dụng AI vượt lỗ hổng bảo mật, chiếm quyền truy cập tài khoản; chiếm đoạt dữ liệu cá nhân có giá trị thương mại. Mở rộng khái niệm tài sản bao gồm: tài sản số, dữ liệu có giá trị vật chất; đồng thời, cần thừa nhận quyền sở hữu đối với dữ liệu là một quyền tài sản được bảo vệ. Bổ sung hành vi “Chiếm đoạt quyền truy cập tài khoản số, ví điện tử”. Cấu trúc lại hành vi “lén lút” là bao gồm cả hành vi truy cập trái phép thông qua hệ thống AI.

- Tội lừa đảo chiếm đoạt tài sản (Điều 174):

Nếu hành vi truyền thống là dùng thủ đoạn gian dối chiếm đoạt tài sản thì hành vi

mới của tội danh này, bao gồm: Deepfake giọng nói, video giả danh người thân, lãnh đạo doanh nghiệp để yêu cầu chuyển tiền vào tài khoản trung gian giả danh; Hoặc, AI tạo kịch bản mạo nhận cơ quan nhà nước/ngân hàng thực hiện lừa đảo.

Đề xuất bổ sung tình tiết “Sử dụng công nghệ trí tuệ nhân tạo để thực hiện hành vi gian dối” - bao gồm việc tạo lập danh tính số giả mạo, sử dụng AI tự động tương tác.

- Tội lạm dụng tín nhiệm chiếm đoạt tài sản (Điều 175):

Đặc trưng của dấu hiệu “lạm dụng” theo khái niệm truyền thống là dựa vào sự tín nhiệm đã có để vay, mượn, thuê hoặc nhận được tài sản bằng hình thức hợp pháp và chiếm đoạt. Hành vi mới là lạm dụng quyền truy cập vào hệ thống số để chiếm đoạt tài sản hoặc dữ liệu có giá trị. Hoặc, người sử dụng AI để sao chép, khai thác dữ liệu nội bộ nhằm trục lợi, chiếm đoạt tài sản số sau khi được ủy quyền quản lý. Yếu tố khó khăn trong xác định hành vi này là phân định rõ ranh giới giữa “sử dụng hợp pháp” và “chiếm đoạt” trong môi trường số.

Cần bổ sung hành vi mới “Chiếm đoạt dữ liệu, tài sản số được giao quản lý hợp pháp”; và, làm rõ khái niệm “tín nhiệm” trong môi trường số, như: quyền truy cập, quyền quản trị dữ liệu. Thêm tình tiết tăng nặng về sử dụng AI để khai thác, sao chép mã nguồn hoặc chuyển dịch tài sản số trái phép.

Có thể thấy, trong bối cảnh chuyển đổi số, khái niệm “tài sản” và “quyền tài sản” đã thay đổi cơ bản từ hình thái hữu hình sang dạng phi vật chất (dữ liệu, tài khoản số). Trong khi đó, BLHS hiện hành được thiết kế chủ yếu dựa vào mô hình tài sản vật lý; Do vậy, đã phát sinh kẽ hở cho các hành vi phạm tội mới về chiếm đoạt quyền tài sản. Thừa nhận quyền “sở hữu dữ liệu” và chiếm đoạt “dữ liệu giá trị” là bước phát triển cần thiết về định hình lại nhận thức tội phạm học và luật thực định, góp phần hiện đại hóa pháp luật hình sự, tăng khả năng tương thích và bảo vệ hiệu quả quyền sở hữu trong nền kinh tế số.

2.2.3. Nhóm tội xâm phạm trật tự quản lý kinh tế (Chương XVIII)

- Tội buôn lậu (Điều 188):

Là hành vi buôn bán trái phép qua biên giới. Trong môi trường số, phát sinh hành vi mới là sử dụng AI để tối ưu hóa đường dây buôn lậu. Hoặc, giao dịch hàng hóa bất hợp pháp qua nền tảng số, blockchain và AI hỗ trợ dự đoán, né tránh kiểm tra Hải quan. Do vậy, cần bổ sung tình tiết “Sử dụng công nghệ AI tổ chức, điều phối hoạt động buôn lậu”; đồng thời, thêm vào khái niệm “vận chuyển” là bao gồm cả giao dịch “phi vật lý” - tối ưu hóa đường dây buôn lậu bằng thuật toán.

- Tội sản xuất, buôn bán hàng giả (Điều 192):

Quy định hiện hành là hành vi sản xuất, kinh doanh hàng giả. Hành vi mới là AI

tạo hàng giả số, sản phẩm số, nội dung số giả mạo. Hoặc, dùng AI sao chép thiết kế, nhãn hiệu với độ chính xác cao; Deepfake thương hiệu, sản phẩm ảo và phân phối hàng giả qua nền tảng số tự động. Vì vậy, bổ sung khái niệm hàng giả gồm hàng hóa số giả, tài sản số giả. Thêm hành vi phạm tội mới “Tạo lập, phân phối sản phẩm giả mạo bằng AI”; tăng nặng định khung đối với hành vi sử dụng AI sản xuất hàng giả quy mô lớn. Bên cạnh đó, cần quy định trách nhiệm quản trị nền tảng số trong việc kiểm soát hàng giả do AI tạo dựng.

Về phương thức phòng ngừa, cần chuyển từ phương thức “hậu kiểm” sang phương thức chủ động “tiền kiểm” trong quản lý công nghệ. Kết hợp trách nhiệm hình sự và nghĩa vụ tuân thủ kỹ thuật trong không gian số.

Do khuôn khổ có giới hạn, bài viết khái quát chung hướng sửa đổi đối với các nhóm tội danh khác, như sau:

- Đối với nhóm tội xâm phạm nhân thân (Chương XIV): Bổ sung tội danh mới “Tội tạo dựng nội dung giả bằng công nghệ”; Hoặc, là tình tiết tăng nặng ở các tội danh “Tội làm nhục người khác” (Điều 155); “Tội vu khống” (Điều 156),

- Đối với nhóm tội xâm phạm an toàn, trật tự công cộng (Chương XIX): Bổ sung khái niệm “Hệ thống AI tự động gây hậu quả thiệt hại” nhằm quy định đối với trách nhiệm của người thiết kế và người sử dụng AI.

- Đối với nhóm tội về công nghệ thông tin, mạng viễn thông (Chương XXI):

Nhóm tội hiện hành gồm 9 tội danh (Điều 285 đến Điều 294) là nhóm cần sửa đổi nhiều nhất. Cụ thể, có những khái niệm định tội đã không phù hợp nữa, như “Phần mềm có tính năng tấn công mạng máy tính”; “Chương trình tin học”, nên cần thay thế thuật ngữ mới: “AI”, “Thuật toán”. Đồng thời, xác định hành vi có yếu tố công nghệ: AI tự động hóa tấn công lừa đảo quy mô lớn; hoặc, sử dụng AI xâm nhập hệ thống xác thực, tấn công ví điện tử/tài sản số. Ghi nhận thêm tình tiết tăng nặng về sử dụng AI để che giấu, xóa vết điện tử hay xâm nhập hệ thống bảo mật.

Bổ sung Điều 290 hiện hành về “Tội sử dụng mạng máy tính, mạng viễn thông, phương tiện điện tử thực hiện hành vi chiếm đoạt tài sản” - sửa theo tội danh mới “Tội sử dụng AI tự động hóa tấn công; xâm nhập hệ thống xác thực chiếm đoạt tài sản ví điện tử, tài sản số”.

Bổ sung nhóm tội danh mới: “Tội huấn luyện AI để phạm tội”; “Tội sử dụng AI tạo nội dung bất hợp pháp”; “Tội mua bán mô hình AI nguy hiểm” và phân loại khung xử phạt theo chức năng AI cấp độ thấp hoặc AI tự chủ.

- Đối với nhóm tội xâm phạm hoạt động tư pháp (Chương XXIV):

Ở hai hành vi phạm tội tại Điều 375 hiện hành là “Tội làm sai lệch hồ sơ vụ án, vụ việc”, gồm phạm tội đối với hồ sơ vụ án và phạm tội đối với hồ sơ vụ việc. Bổ

sung hành vi phạm tội mới là sử dụng AI tạo chứng cứ giả; Deepfake trong tổ tụng. Hoặc, dùng “Chứng cứ điện tử giả tạo lập bằng công nghệ”; đồng thời, quy định rõ thuật ngữ “Chứng cứ số” và “Bằng chứng do AI tạo lập”.

- Bổ sung khái niệm pháp lý mới:

Thêm vào phần chung BLHS một số thuật ngữ pháp lý của môi trường số: “Trí tuệ nhân tạo”; “Dữ liệu”; “Tài sản số”; “Hệ thống tự động”. Đồng thời, bổ sung những yếu tố mới thuộc mặt chủ quan phạm tội (lỗi, động cơ, mục đích) phái sinh từ hành vi liên quan AI tự hành của người lập trình, người sử dụng, người tổ chức đối với chủ thể doanh nghiệp công nghệ, chủ sở hữu nền tảng số.

- Xác lập quan hệ pháp luật mới về thẩm quyền tài phán đối với tội phạm xuyên biên giới là máy chủ (server) ở nước ngoài thực hiện phạm tội ở Việt Nam hoặc ngược lại.

Kết luận

Yêu cầu sửa đổi BLHS và tái cấu trúc trách nhiệm hình sự một số tội danh phát sinh mới liên quan AI, đặt ra yêu cầu cấp thiết chuyển đổi nhận thức hình thái tội phạm mang tính vật lý/cơ học sang hình thái tội phạm hành vi số/thuật toán; mở rộng cách tiếp cận tội phạm cá nhân sang tội phạm hệ thống. Đáp ứng yêu cầu chuyển đổi đó, không đơn thuần là kỹ thuật lập pháp mà là sự định hình lại bản chất của hành vi nguy hiểm trong kỷ nguyên số; nếu không kịp thời điều chỉnh, pháp luật hình sự sẽ có những kẽ hở cho đa phần hành vi gây hại thấu lọt mà không bị phát hiện, đưa đến một tình trạng tội phạm trở thành hợp pháp trên danh nghĩa luôn tồn tại trong xã hội pháp quyền.

TS. HOÀNG MINH KHÔI (GV Khoa Luật - Đại học Hoa Sen)

Tài liệu tham khảo

1. Anderson, J. Rainie, L. and Luchsinger, A. (2018) “Artificial Intelligence and the Future of Humans”. <http://www.pewinternet.org/2018/12/10/artificial-intelligence-and-the-future-of-humans/> 10 December 2018
2. European Commission. “Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence”. Publication 21 April 2021. <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>.
3. Emmanuel Barthe (2017). “Intelligence artificielle en droit: derrière la “hype”, la réalité”. <https://www.precisement.org/blog/Intelligence-artificielle-en-droit-derriere-la-hype-la-realite.html>.
4. Đào Bá Minh, “Thực trạng áp dụng

phương thức giải quyết tranh chấp bằng trọng tài và hòa giải trực tuyến ở EU”, *Nghề luật*, số 7, 2021. <https://sti.vista.gov.vn/publication/view/c2e3f5dfaa7479c2731d656cffeab433-5f4253b204f0c7cd4fe28c9b148fea24-346032.html>.

5. Nguyễn Tấn Hoàng Hải, Trần Bảo Khanh, Trần thị Hạnh Nhân. “Công nghệ và pháp lý của việc ứng dụng trí tuệ nhân tạo và giải quyết tranh chấp trực tuyến trong giải quyết tranh chấp bằng trọng tài thương mại”. *Khoa học pháp lý Việt Nam* 9(81) (2024). <https://doi.org/10.70236/tckhplvn.116> Truy cập 15/3/2026.

6. Gabriel Hallevy (2016), “The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control”, *The University of Akron IdeaExchange@UAkron*. <https://ideaexchange.uakron.edu/cgi/viewcontent.cgi?article=1037&context=akronintellectualproperty>.

1 Hoàng Minh Khôi, *Giới hạn quyền tự do thông tin theo quy định của Luật An ninh mạng năm 2018*, <https://www.quanlynhanuoc.vn/2023/05/05/gioi-han-quyen-tu-do-thong-tin-theo-quy-dinh-cua-luat-an-ninh-mang-nam-2018/>, truy cập ngày 15/3/2026.

2 Nora Osmani, “*The Complexity of Criminal Liability of AI Systems*”, *Masaryk University Journal of Law and Technology*, Vol.14,No.1(2020), <https://journals.muni.cz/mujlt/article/view/12440>, truy cập ngày 15/3/2026.

3 EU Artificial Intelligence Act (2024), “*The EU Artificial Intelligence Act*”, <https://artificialintelligenceact.eu/>, truy cập ngày 15/3/2026.

4 “GDPR”: General Data Protection Regulation - Quy định (EU) 2016/679 của Nghị viện Châu Âu và Hội đồng 27/4/2016. <http://data.europa.eu/eli/reg/2016/679/2016-05-04>, truy cập ngày 15/3/2026.

5 Cộng hoà Pháp: không quy định AI là chủ thể trách nhiệm hình sự, nhưng sử dụng AI phạm tội sẽ bị xử lý theo các tội danh của “Đạo luật không gian kỹ thuật số” (SREN) - từ Điều 211 đến Điều 228, https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000049571542, truy cập 16/6/2026.

6 VTV.VN (2024), “Clearview AI bị phạt nặng tại Châu Âu”, <https://vtv.vn/cong-nghe/clearview-ai-bi-phat-nang-tai-chau-au-20240904064938418.htm>, truy cập ngày 15/3/2026.

7 European Data Protection Board, “The French SA Fines Clearview AI EURO

20 milion”, https://www.edpb.europa.eu/news/national-news/2022/french-sa-fines-clearview-ai-eur-20-million_en 20 October 2022, truy cập ngày 15/3/2026.

8 National Institute of Standards and Technology (2023), Artificial Intelligence Risk Management Framework (AI RMF 1.0), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>, truy cập ngày 20/12/2025.

9 Dynamis (2026), “AI-powered Fraud and DOJ’s Expanding Enforcement Toolkit: What Businesses and Executives Need to Know Now”. <https://www.dynamisllp.com/knowledge/ai-fraud-enforcement-doj-defense-strategies-2025>, truy cập ngày 16/6/2026.

10 National Transportation Safety Board, “Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian”, Tempe, Arizona March 18, 2018,

<https://www.nts.gov/investigations/accidentreports/reports/har1903.pdf>, truy cập ngày 20/12/2025.

11 Yirong Sun và Jingxian Zeng (2024), “China’s interim measure for the management of generative ai service: a comparison between the final and draft versions of the text”, <https://fpf.org/blog/chinas-interim-measures-for-the-management-of-generative-ai-services-a-comparison-between-the-final-and-draft-versions-of-the-text/>, truy cập ngày 20/12/2025.

12 China Law Translate, “Provisions on the Administration of Deep Synthesis Internet Information Services” on 2022/12/11, <https://www.chinalawtranslate.com/en/deep-synthesis/>, truy cập ngày 15/3/2026.

13 Douyin: ứng dụng mạng xã hội video ngắn tại Trung Quốc, được phát triển bởi ByteDance, còn gọi là “TikTok Trung Quốc”, độc lập với TikTok quốc tế. Tập trung vào thị trường nội địa với các tính năng vượt trội như mua sắm trực tiếp, nhận diện khuôn mặt và kho hiệu ứng, âm nhạc đa dạng, <https://cellphones.com.vn/sforum/douyin-la-gi>, truy cập ngày 20/12/2025.

- Xem thêm: Sai, S., & Wang, Z. (2026). “Các phương pháp điều chỉnh hình sự đối với các tội phạm liên quan đến Deepfake: Tập trung vào tội gian lận”. *Tạp chí Nghiên cứu Khoa học Xã hội Châu Á Quốc tế*, 3 (1), 20-31. <https://doi.org/10.70267/ijassr.260301.2031>, truy cập 16/6/2026.

14 https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm, truy cập ngày 20/12/2025.

15 Qiheng Chen (2024), “China’s Emerging Approach to Regulating General-Purpose Artificial Intelligence: Balancing Innovation and Control”,

<https://asiasociety.org/policy-institute/chinas-emerging-approach-regulating-general-purpose-artificial-intelligence-balancing-innovation-and>, truy cập ngày 20/12/2025.

16 Nguyễn Thị Quế Anh, Vũ Công Giao, Mai Văn Thắng, Trí tuệ nhân tạo với pháp luật và quyền con người, Nxb. Tư pháp, Hà Nội, 2019.

17 Matthias Artzt and Tran Viet Dung, “Artificial Intelligence and Data Protection: How to Reconcile Both Areas from the European Law Perspective”, Vietnamese Journal of Legal Sciences, Vol. 7(2), 2022, <https://reference-global.com/article/10.2478/vjls-2022-0007>, truy cập ngày 20/12/2025.

18 Hoàng Minh Khôi, Tlđd (1).