

XỬ LÝ HÀNH VI GIẢ MẠO GIỌNG NÓI, HÌNH ẢNH BẰNG TRÍ TUỆ NHÂN TẠO - KHOẢNG TRỐNG PHÁP LUẬT VÀ KIẾN NGHỊ HOÀN THIỆN

■ **NGUYỄN THÀNH PHƯƠNG, ĐINH TRẦN NGỌC HUYỀN* & NGUYỄN TRƯỜNG DUY****

Tóm tắt: Quá trình chuyển đổi số và phát triển mạnh mẽ của các giao dịch tài chính số tại Việt Nam đang diễn ra nhanh chóng, đồng thời, cũng làm gia tăng các hành vi lừa đảo công nghệ cao, đặc biệt, các thủ đoạn lợi dụng trí tuệ nhân tạo (AI), trong đó, nổi bật là công nghệ Deepfake. Công nghệ này cho phép giả mạo giọng nói, hình ảnh với độ chân thực cao, đe dọa trực tiếp cơ chế xác thực sinh trắc học và an toàn ngân hàng số. Nghiên cứu tập trung phân tích: (i) sự thiếu thống nhất trong định tội danh giữa Điều 174 và Điều 290 Bộ luật Hình sự năm 2015 sửa đổi, bổ sung các năm 2017, 2025; (ii) thách thức đối với giá trị chứng minh của chứng cứ điện tử và sự thiếu quy chuẩn giám định AI chuyên biệt; (iii) khoảng trống trong cơ chế bảo vệ dữ liệu sinh trắc học, chế tài đối với hành vi thương mại hóa dữ liệu và cơ chế phong tỏa tài khoản trong giao dịch nghi ngờ gian lận. Trên cơ sở đó, nghiên cứu đề xuất hoàn thiện pháp luật theo hướng: làm rõ ranh giới định tội danh; xây dựng quy chuẩn giám định AI; hình sự hóa hành vi chiếm đoạt dữ liệu sinh trắc học; tăng cường cơ chế phong tỏa tài khoản có điều kiện... Các giải pháp này nhằm nâng cao năng lực ứng phó của hệ thống pháp lý trước tội phạm Deepfake trong kỷ nguyên số.

Từ khóa: Deepfake; giả mạo giọng nói; dữ liệu điện tử; xác thực sinh trắc học; phong tỏa tài khoản.

Nhận bài: 19/01/2026

Hoàn thành biên tập: 09/4/2026

Duyệt đăng: 15/4/2026

HANDLING OF VOICE AND IMAGE FORGERY USING ARTIFICIAL INTELLIGENCE - LEGAL GAPS AND RECOMMENDATIONS FOR IMPROVEMENT

Abstract: The rapid digital transformation and development of digital financial transactions in Vietnam are also increasing high-tech fraud, particularly schemes exploiting artificial intelligence (AI), most notably Deepfake technology. This technology allows for highly realistic voice and image forgery, directly threatening biometric authentication mechanisms and digital banking security. This research focuses on analyzing: (i) the inconsistency in defining offenses between Article 174 and Article 290 of the 2015 Penal Code, as amended in 2017 and 2025; (ii) challenges to the probative value of electronic evidence and the lack of specialized AI-based forensic standards; and (iii) gaps in the mechanism for protecting biometric data, sanctions for the commercialization of data, and mechanisms for freezing accounts in suspected fraudulent transactions. Based on this, the study proposes improving the law in the following directions: clarifying the boundaries of criminal offenses; developing standards for AI authentication; criminalizing the act of stealing biometric data; strengthening the mechanism for conditional account freezing... These solutions aim to enhance the legal system's ability to respond to Deepfake crimes in the digital age.

Keywords: Deepfake; voice spoofing; electronic data; biometric authentication; account freezing.

Article received: 19/01/2026

Editing completed: 09/4/2026

Approved for publication: 15/4/2026

* Khoa Luật, Trường Đại học Nam Cần Thơ

** Trường Đại học An Giang

Đặt vấn đề

Trong tiến trình chuyển đổi số, các dịch vụ tài chính, ngân hàng tại Việt Nam đang phụ thuộc vào các phương thức định danh điện tử (eKYC) và xác thực sinh trắc học như nhận diện khuôn mặt, giọng nói hoặc video trực tuyến. Những công nghệ này giúp đơn giản hóa quy trình giao dịch, mở rộng khả năng tiếp cận dịch vụ tài chính và thúc đẩy phát triển của hệ sinh thái ngân hàng số. Tuy nhiên, cùng với sự phát triển của trí tuệ nhân tạo (AI), đặc biệt là công nghệ giả mạo khuôn mặt, giọng nói hoặc video trực tuyến (Deepfake), các đặc điểm sinh trắc học được xem là công cụ xác thực đáng tin cậy có thể bị tái tạo với độ chân thực rất cao. Điều này tạo ra nguy cơ mất an toàn đối với giao dịch và làm thay đổi bản chất của các hành vi gian lận trong môi trường số. Về bản chất, Deepfake là công nghệ sử dụng các mô hình học sâu để tổng hợp hoặc tái tạo hình ảnh, âm thanh và video của cá nhân dựa trên dữ liệu sinh trắc học sẵn có. Khi sử dụng cho mục đích xấu, công nghệ này có thể tạo ra nội dung giả mạo giọng nói hoặc hình ảnh với mức độ tương đồng cao so với người thật. Trong lĩnh vực tài chính số, Deepfake thường được sử dụng để giả mạo danh tính lãnh đạo doanh nghiệp, nhân viên ngân hàng hoặc người thân của nạn nhân nhằm yêu cầu chuyển tiền hoặc để vượt qua các bước xác thực sinh trắc học trong quy trình eKYC và giao dịch trực tuyến. Khác với các phương thức lừa đảo truyền thống, Deepfake không chỉ tác động đến nhận thức của nạn nhân mà còn có khả năng làm suy yếu chính các cơ chế bảo mật của hệ thống tài chính số.

Pháp luật Việt Nam bước đầu điều chỉnh các hành vi này thông qua các quy định về an ninh mạng, bảo vệ dữ liệu cá nhân và các tội danh chiếm đoạt tài sản trong Bộ luật Hình sự năm 2015, sửa đổi, bổ sung các năm 2017, 2025 (Bộ luật Hình sự năm 2015). Tuy nhiên, thực tiễn áp dụng bộc lộ nhiều khoảng trống và sự giao thoa phức tạp. Trước hết, việc định tội danh giữa Điều 174 và Điều 290 Bộ luật Hình sự năm 2015 thiếu rõ ràng; trong khi chưa có quy chuẩn kỹ thuật chuyên biệt cho hoạt động giám định AI. Đồng thời, cơ chế bảo vệ dữ liệu sinh trắc học là nguồn dữ liệu đầu vào quan trọng để tạo ra Deepfake chưa được thiết kế tương xứng với mức độ rủi ro đối với an toàn giao dịch tài chính số. Vì vậy, việc nghiên cứu, đánh giá thực trạng pháp luật về xử lý hành vi giả mạo giọng nói, hình ảnh bằng AI và đề xuất giải pháp hoàn thiện pháp luật là cần thiết. Trên cơ sở đó, nghiên cứu phân tích hai nhóm vấn đề chính: (i) thực trạng các quy định pháp luật liên quan đến xử lý hành vi giả mạo giọng nói, hình ảnh bằng AI trong giao dịch tài chính số; (ii) các kiến nghị hoàn thiện pháp luật để nâng cao hiệu quả phòng ngừa và xử lý loại tội phạm Deepfake trong bối cảnh chuyển đổi số.

1. Một số vấn đề về công nghệ Deepfake

Về mặt thuật ngữ, Deepfake là sự kết hợp giữa Deep learning (học sâu) và Fake (giả mạo), xuất hiện lần đầu vào năm 2017 trên các diễn đàn công nghệ quốc tế để chỉ kỹ thuật tổng hợp hình ảnh, âm thanh và video dựa trên trí tuệ nhân tạo¹. Tại Việt Nam, thuật ngữ Deepfake chưa được quy định trong Bộ luật Hình sự năm 2015 và

1. Chesney, B., & Citron, D. (2019), *Deep fakes: A looming challenge for privacy, democracy, and*

Luật An ninh mạng năm 2025, nhưng bản chất pháp lý của hành vi này được bao hàm trong các quy định về sử dụng công nghệ thông tin và giả mạo thông tin.

Dưới góc độ pháp lý, Deepfake là hành vi giả mạo thông tin được quy định tại Điều 13 Luật An ninh mạng năm 2025. Theo khoản 1 Điều 2 Luật Bảo vệ dữ liệu cá nhân năm 2025, dữ liệu cá nhân là thông tin dưới dạng số hoặc dạng khác có khả năng xác định hoặc giúp xác định một con người cụ thể, qua đó cho thấy, khuôn mặt và giọng nói là những yếu tố nhận diện trực tiếp một cá nhân trong môi trường số, do đó thuộc phạm vi dữ liệu cá nhân theo định nghĩa này.

Đồng thời, khoản 3 Điều 2 Luật Bảo vệ dữ liệu cá nhân năm 2025 xác định, dữ liệu cá nhân nhạy cảm là loại dữ liệu gắn liền với quyền riêng tư, khi bị xâm phạm có thể ảnh hưởng trực tiếp đến quyền và lợi ích hợp pháp của chủ thể dữ liệu. Cụ thể hóa quy định này, điểm d khoản 1 Điều 4 Nghị định số 356/2025/NĐ-CP ngày 31/12/2025 của Chính phủ quy định chi tiết một số điều và biện pháp thi hành Luật Bảo vệ dữ liệu cá nhân (Nghị định số 356/2025/NĐ-CP) liệt kê dữ liệu sinh trắc học, đặc điểm di truyền là dữ liệu cá nhân nhạy cảm. Với tính chất là dữ liệu sinh trắc học phục vụ nhận diện cá nhân, khuôn mặt và giọng nói thuộc nhóm dữ liệu cá nhân nhạy cảm và phải được áp dụng cơ chế bảo vệ nghiêm ngặt.

Về điều kiện xử lý, khoản 1 Điều 11 Luật

Bảo vệ dữ liệu cá nhân năm 2025 quy định dữ liệu cá nhân chỉ được thu thập khi có sự đồng ý của chủ thể dữ liệu, trừ trường hợp pháp luật có quy định khác; đồng thời, khoản 3 Điều 9 Luật Bảo vệ dữ liệu cá nhân năm 2025 yêu cầu sự đồng ý phải được thể hiện rõ ràng, cụ thể và có thể kiểm chứng. Do đó, hành vi thu thập, phân tích và sử dụng khuôn mặt, giọng nói của cá nhân để tạo Deepfake mà không có sự đồng ý hợp lệ theo các quy định trên là hành vi xử lý dữ liệu cá nhân nhạy cảm trái pháp luật, vi phạm trực tiếp nghĩa vụ bảo vệ dữ liệu cá nhân và xâm phạm quyền kiểm soát dữ liệu của chủ thể dữ liệu².

Trên cơ sở các phân tích trên, có thể hiểu: Deepfake là việc sử dụng công nghệ thuộc lĩnh vực trí tuệ nhân tạo để tạo lập, chỉnh sửa hoặc tái tạo hình ảnh, giọng nói hoặc video của một cá nhân theo hướng giả mạo, làm sai lệch nhận thức của chủ thể tiếp nhận, đồng thời, xâm phạm đến dữ liệu cá nhân, đặc biệt là dữ liệu sinh trắc học của cá nhân đó.

2. Thực trạng pháp luật điều chỉnh về hành vi giả mạo giọng nói, hình ảnh bằng trí tuệ nhân tạo trong giao dịch tài chính số

Trong bối cảnh giao dịch tài chính số phát triển mạnh mẽ, sự xuất hiện của các công nghệ thuộc lĩnh vực trí tuệ nhân tạo, đặc biệt là Deepfake làm gia tăng nguy cơ giả mạo giọng nói, hình ảnh với mức độ tinh vi cao, đe dọa trực tiếp đến an toàn và tính xác thực của giao dịch.

national security, Calif. L. Rev., 107, 1753.

2. Tambiama Madiega (2023), *Generative AI and watermarking*, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI\(2023\)757583_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf), truy cập ngày 26/02/2026.

Tuy nhiên, pháp luật Việt Nam hiện đang thiếu quy định cụ thể để điều chỉnh trực tiếp hành vi này trong môi trường tài chính số. Qua phân tích các vấn đề có liên quan đến hành vi giả mạo giọng nói, hình ảnh bằng AI trong giao dịch tài chính số, cho thấy tồn tại một số vấn đề sau:

Thứ nhất, việc định tội danh đối với hành vi sử dụng công nghệ Deepfake để chiếm đoạt tài sản hiện còn thiếu nhất quán khi so sánh quy định giữa Điều 174 và Điều 290 Bộ luật Hình sự năm 2015. Xét về cấu thành tội phạm, tội lừa đảo chiếm đoạt tài sản theo Điều 174 Bộ luật Hình sự năm 2015 được xây dựng đặc trưng bởi hành vi dùng thủ đoạn gian dối làm cho người bị hại tin tưởng và tự nguyện chuyển giao tài sản cho người phạm tội, qua đó, xâm phạm trực tiếp đến quyền sở hữu. Trong bối cảnh công nghệ số, thủ đoạn gian dối này có thể được thực hiện thông qua nhiều phương thức khác nhau. Ví dụ, người phạm tội có thể giả mạo giọng nói của lãnh đạo doanh nghiệp, người thân hoặc nhân viên ngân hàng nhằm tạo lòng tin và yêu cầu nạn nhân chuyển tiền. Trong trường hợp này, Deepfake đóng vai trò là công cụ, phương tiện thực hiện hành vi gian dối, còn bản chất pháp lý của hành vi vẫn phù hợp với cấu thành tội lừa đảo chiếm đoạt tài sản.

Trong khi đó, Điều 290 Bộ luật Hình sự năm 2015 quy định tội sử dụng mạng máy tính, mạng viễn thông, phương tiện điện tử thực hiện hành vi chiếm đoạt tài sản, nhấn mạnh yếu tố sử dụng hạ tầng công nghệ thông

tin như phương thức phạm tội đặc thù. Theo đó, hành vi chiếm đoạt tài sản được thực hiện thông qua lợi dụng hệ thống mạng, thiết bị điện tử hoặc môi trường giao dịch trực tuyến. Trong bối cảnh này, Deepfake là phương tiện kỹ thuật nhằm thao túng quy trình xác thực điện tử hoặc đánh lừa người dùng trong giao dịch trực tuyến. Sự khác biệt trong cấu trúc cấu thành của hai điều luật cho thấy, cùng một hành vi sử dụng Deepfake có thể thỏa mãn dấu hiệu của Điều 174 và Điều 290, tùy thuộc vào cách xác định vai trò của công nghệ trong thực hiện hành vi phạm tội. Nếu Deepfake được sử dụng như thủ đoạn gian dối tác động trực tiếp đến nhận thức của nạn nhân để họ tự nguyện chuyển giao tài sản, hành vi có xu hướng được định tội theo Điều 174. Ngược lại, nếu nhấn mạnh việc sử dụng môi trường mạng và phương tiện điện tử như phương thức chiếm đoạt tài sản, thì Điều 290 có thể được áp dụng. Tuy nhiên, cấu trúc pháp lý của Điều 290 bộc lộ một số hạn chế làm cho ranh giới áp dụng Điều 174 trở nên phức tạp. Cụ thể, điều luật được thiết kế theo hướng loại trừ “nếu không thuộc quy định tại Điều 173 và Điều 174”, trong khi bản chất của Điều 290 là hành vi chiếm đoạt tài sản. Việc đặt dấu hiệu loại trừ giữa các tội danh có khách thể khác nhau, khi một bên là quyền sở hữu thuần túy, một bên là khách thể kép gồm an toàn hoạt động của mạng máy tính và quyền sở hữu tài sản có thể gây lúng túng trong thực tiễn định tội danh³.

3. Lê Tuấn Tú, Nguyễn Thanh Thảo Nhi (2024), *Tội “Sử dụng mạng viễn thông, mạng máy tính, phương tiện điện tử để chiếm đoạt tài sản” theo quy định của Bộ luật Hình sự 2015*, <https://lsvn.vn/>

Bên cạnh đó, Điều 290 Bộ luật Hình sự năm 2015 xây dựng theo phương pháp liệt kê các hành vi cụ thể, chủ yếu gắn với lĩnh vực tài chính - ngân hàng, thương mại điện tử hoặc thanh toán điện tử. Cách tiếp cận này làm thu hẹp phạm vi áp dụng của điều luật. Ngoài ra, một số dấu hiệu như hành vi “truy cập trái phép vào tài khoản” chưa được giải thích thống nhất về phạm vi khái niệm “tài khoản”, chỉ gồm tài khoản ngân hàng, hay có thể mở rộng sang tài khoản mạng xã hội, ví điện tử hoặc các nền tảng số khác. Điều này đặt ra yêu cầu cần có hướng dẫn giải thích thống nhất hoặc hoàn thiện nhằm bảo đảm tính rõ ràng, khả năng dự báo và nhất quán trong định tội danh đối với các hành vi sử dụng công nghệ mới như Deepfake trong môi trường số.

Thứ hai, trong các vụ án liên quan đến hành vi giả mạo giọng nói, hình ảnh bằng AI nhằm chiếm đoạt tài sản trong giao dịch tài chính số, đang xuất hiện nguy cơ bị cáo phủ nhận tính xác thực của chứng cứ điện tử bằng cách cho rằng đó là sản phẩm giả tạo của trí tuệ nhân tạo⁴. Trong bối cảnh tội phạm Deepfake có thể tái tạo giọng nói chủ tài khoản để yêu cầu

chuyển tiền hoặc tạo video xác thực giả nhằm vượt qua cơ chế eKYC, các dữ liệu như bản ghi âm cuộc gọi, video xác thực giao dịch hoặc file ghi nhận thao tác trên ứng dụng ngân hàng trở thành chứng cứ trọng yếu trong tố tụng hình sự. Tuy nhiên, khi bị can viện dẫn khả năng “giả mạo bằng AI”, giá trị chứng minh của các dữ liệu lập tức bị đặt vào trạng thái tranh luận⁵. Theo đó, pháp luật Việt Nam hiện chưa có quy trình chuyên biệt cho giám định AI, tạo ra khoảng trống pháp lý, để các chủ thể có thể khai thác nhằm loại trừ giá trị chứng minh của chứng cứ điện tử⁶.

Thứ ba, về mặt kỹ thuật, các mô hình Deepfake chỉ có thể đạt độ chân thực cao khi được huấn luyện trên tập dữ liệu sinh trắc học quy mô lớn, gồm hình ảnh khuôn mặt, mẫu giọng nói và các đặc trưng nhận dạng cá nhân khác. Do đó, dữ liệu sinh trắc học giữ vai trò là điều kiện đầu vào bắt buộc trong quá trình tạo sinh nội dung giả mạo⁷. Tuy nhiên, cơ chế bảo vệ loại tài sản số này chưa hoàn thiện. Cụ thể, dữ liệu sinh trắc học đang bị thương mại hóa và thu thập thông qua các hoạt động như xác thực eKYC, chấm công, các ứng dụng chỉnh

toi-su-dung-mang-vien-thong-mang-may-tinh-phuong-tien-dien-tu-de-chiem-doat-tai-san-theo-quy-dinh-cua-bo-luat-hinh-su-2015-a152321.html, truy cập ngày 20/02/2026.

4. Schiff, K. J., Schiff, D. S., & Bueno, N. S. (2025), *The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?* American Political Science Review, 119(1), 71-90. doi:10.1017/S0003055423001454, truy cập ngày 26/02/2026.

5. Westerlund, M. 2019, *The Emergence of Deepfake Technology: A Review*, Technology Innovation Management Review, 9(11): 40-53, <http://doi.org/10.22215/timreview/1282>, truy cập ngày 26/02/2026.

6. Lưu Quang Anh, Bùi Lê Hiếu (2024), *Tội sử dụng mạng máy tính, mạng viễn thông, phương tiện điện tử thực hiện hành vi chiếm đoạt tài sản - Vướng mắc và một số kiến nghị*, <https://tapchitoaan.vn/toi-su-dung-mang-may-tinh-mang-vien-thong-phuong-tien-dien-tu-thuc-hien-hanh-vi-chiem-doat-tai-san-vuong-mac-va-mot-so-kien-nghi12972.html>, truy cập ngày 26/02/2026.

7. Westerlund, M., 2019, *The Emergence of Deepfake Technology: A Review*, Technology Innovation Management Review, Vol. 9, No. 11, pp.40-53.

sửa ảnh trên mạng xã hội, sau đó được tổng hợp và rao bán⁸. Theo Hiệp hội An ninh quốc gia, trong năm 2024 có hơn 66% người dùng xác nhận thông tin của họ từng được sử dụng trái phép⁹. Mặc dù Điều 7 Luật Bảo vệ dữ liệu cá nhân năm 2025 chính thức ghi nhận hành vi “mua, bán dữ liệu cá nhân” là hành vi bị nghiêm cấm và Điều 8 thiết lập cơ chế xử phạt hành chính nghiêm khắc. Tuy nhiên, trong môi trường ngân hàng số, dữ liệu sinh trắc học như khuôn mặt, giọng nói, đặc điểm nhận dạng... không chỉ là thông tin định danh mà còn là yếu tố xác thực giao dịch trong hệ thống eKYC, xác thực đa lớp và thanh toán trực tuyến. Do đó, việc mua bán hoặc chiếm đoạt dữ liệu không chỉ xâm phạm quyền riêng tư, mà còn trực tiếp đe dọa an toàn hệ thống tài chính và tính toàn vẹn của giao dịch. Qua đó, Điều 8 Luật Bảo vệ dữ liệu cá nhân năm 2025 đề ra mức phạt hành chính tối đa có thể lên đến 10 lần khoản thu từ hành vi vi phạm hoặc 03 tỷ đồng. Tuy nhiên, Luật chưa thiết kế cơ chế tăng nặng riêng đối với hành vi thương mại hóa dữ liệu có khả năng trực tiếp làm suy yếu cơ chế bảo mật giao dịch tài chính. Việc đặt hành vi mua bán dữ liệu sinh trắc học ngang hàng với các hành vi vi phạm khác mà không phân tầng rủi ro có thể dẫn đến sự thiếu tương xứng trong chính sách hình sự và quản lý rủi ro tài chính. Bộ luật Hình sự năm 2015 chưa có tội danh riêng cho hành vi “Trộm cắp danh tính sinh trắc học”. Các

hành vi này thường được xử lý theo Điều 288 về Tội đưa hoặc sử dụng trái phép thông tin mạng máy tính, mạng viễn thông, nhưng cấu thành tội tập trung vào hậu quả gây thiệt hại vật chất, chưa bao quát được những rủi ro lâu dài của việc mất dữ liệu sinh trắc học.

Thứ tư, mặc dù, pháp luật đặt ra yêu cầu nghiêm ngặt đối với hoạt động bảo vệ dữ liệu cá nhân trong lĩnh vực tài chính, ngân hàng và thông tin tín dụng, tuy nhiên, vẫn tồn tại một số bất cập. Cụ thể: (i) khoản 1 Điều 8 Nghị định số 356/2025/NĐ-CP yêu cầu các tổ chức phải áp dụng “tiêu chuẩn, quy chuẩn kỹ thuật bảo vệ dữ liệu cá nhân” và quy chuẩn về khứ nhận dạng, ẩn danh dữ liệu, nhưng không xác định cụ thể hệ thống tiêu chuẩn nào là bắt buộc, mức độ tuân thủ và cơ chế kiểm tra việc thực hiện. Bên cạnh đó, nghĩa vụ “ghi lại nhật ký toàn bộ hoạt động xử lý dữ liệu cá nhân” được quy định rộng nhưng chưa làm rõ thời gian lưu trữ, phạm vi nội dung và cơ chế bảo mật đối với chính dữ liệu nhật ký, trong khi đây là nguồn dữ liệu nhạy cảm; (ii) khoản 2 Điều 8 Nghị định số 356/2025/NĐ-CP đặt trọng tâm vào cơ chế “sự đồng ý” của chủ thể dữ liệu, yêu cầu phải thông tin rõ mục đích xử lý, nguồn thu thập, thời gian lưu trữ và cơ chế rút lại sự đồng ý. Tuy nhiên, trong lĩnh vực tài chính, tín dụng, nhiều hoạt động xử lý dữ liệu được thực hiện dựa trên nghĩa vụ pháp lý bắt buộc theo luật chuyên ngành như pháp luật về các tổ chức tín dụng hoặc phòng, chống rửa

8. Thành Luân, Anh Quân (2021), *Vụ rò rỉ 17 GB dữ liệu người dùng Việt: Cách bảo vệ thông tin cá nhân*, <https://thanhluen.vn/vu-ro-ri-17-gb-du-lieu-nguoi-dung-viet-cach-bao-ve-thong-tin-ca-nhan-1851069078.htm>, truy cập ngày 21/02/2026.

9. Bộ Tư pháp (2025), *Mua cá nhân dữ liệu vẫn tràn lan trên mạng*, <https://dx.moj.gov.vn/mua-ban-du-lieu-ca-nhan-van-tran-lan-tren-mang-903.htm>, truy cập ngày 20/02/2026.

tiền, mà không hoàn toàn dựa trên sự đồng ý. Việc cho phép rút lại sự đồng ý, không làm rõ giới hạn áp dụng có thể tạo ra xung đột giữa pháp luật bảo vệ dữ liệu cá nhân và pháp luật chuyên ngành. Đồng thời, điều luật chưa đề cập đến các vấn đề quan trọng như minh bạch thuật toán chấm điểm tín dụng hay cơ chế phản đối quyết định tự động; (iii) khoản 3 Điều 8 Nghị định số 356/2025/NĐ-CP quy định nghĩa vụ thông báo trong vòng 72 giờ khi phát hiện lộ, mất dữ liệu nhạy cảm, song, chưa làm rõ thời điểm bắt đầu tính thời hạn này, tiêu chí xác định mức độ sự cố phải thông báo, cũng như trách nhiệm phối hợp giữa các chủ thể trong chuỗi xử lý dữ liệu (ngân hàng, công ty fintech, trung gian thanh toán). Trong thực tiễn, dữ liệu tài chính thường được xử lý qua nhiều bên liên quan, nên việc chỉ quy định trách nhiệm cho tổ chức “trực tiếp thu thập” dữ liệu có thể dẫn đến khoảng trống trách nhiệm khi xảy ra sự cố.

Thứ năm, trong bối cảnh giả mạo giọng nói và hình ảnh bằng AI ngày càng được sử dụng như công cụ vượt qua cơ chế xác thực sinh trắc học trong giao dịch tài chính số, phong tỏa tài khoản trở thành biện pháp quan trọng nhằm phòng ngừa và hạn chế thiệt hại do tội phạm công nghệ cao gây ra. Mặc dù, Điều 11 Nghị định số 52/2024/NĐ-CP ngày 15/5/2024 của Chính phủ quy định về thanh toán không dùng tiền mặt cho phép ngân hàng phong tỏa tài khoản khi có nghi ngờ gian lận, nhưng cơ chế này trên thực tế chưa phát huy hiệu quả trong các

tình huống liên quan đến Deepfake. Nguyên nhân không chỉ xuất phát từ sự thận trọng của tổ chức tín dụng trước nguy cơ khiếu kiện dân sự nếu phong tỏa nhầm, mà còn từ việc pháp luật chưa thiết kế cơ chế bảo vệ pháp lý rõ ràng cho ngân hàng khi hành động trên cơ sở đánh giá rủi ro công nghệ cao. Thực tiễn cho thấy, tốc độ thực hiện hành vi gian lận bằng Deepfake vượt xa khả năng phản ứng của các cơ chế kiểm soát truyền thống trong hệ thống ngân hàng. Ví dụ, năm 2024 tại Hong Kong, nhân viên tài chính của tập đoàn đa quốc gia bị lừa chuyển hơn 25 triệu USD sau khi tham gia cuộc họp trực tuyến với những người mà anh tin là giám đốc tài chính và các lãnh đạo cấp cao của công ty. Theo kết quả điều tra của cảnh sát Hong Kong, toàn bộ hình ảnh và giọng nói của những người tham gia cuộc họp đều được tạo ra bằng công nghệ Deepfake dựa trên dữ liệu video công khai trước đó. Tin rằng đây là cuộc họp hợp pháp, nhân viên này đã thực hiện nhiều lệnh chuyển tiền theo yêu cầu của các “lãnh đạo” và số tiền nhanh chóng được chuyển qua nhiều tài khoản trung gian trước khi cơ quan chức năng có thể can thiệp¹⁰. Vụ việc này cho thấy, dòng tiền có thể bị tẩu tán chỉ trong thời gian rất ngắn, làm cho các cơ chế kiểm soát hành chính truyền thống khó có thể phản ứng kịp thời.

Thứ sáu, về công tác hợp tác quốc tế và tương trợ tư pháp điện tử. Bản chất của tội phạm Deepfake trong giao dịch tài chính là xuyên biên giới, thể hiện ở việc nguồn nguyên liệu là dữ liệu sinh trắc học được thu thập từ các nền tảng

10. Shannon Murphy (2024), *A Deepfake Scammed a Bank out of \$25M - Now What?*, https://www.trendmicro.com/en_us/research/24/b/deepfake-video-calls.html, truy cập ngày 20/02/2026.

mạng xã hội toàn cầu, mô hình AI được huấn luyện ở nước ngoài và đặc biệt là dòng tiền chiếm đoạt thường được tẩu tán nhanh thông qua các sàn giao dịch tiền mã hóa phi tập trung. Tuy nhiên, cơ chế phòng, chống hiện tại của Việt Nam đang gặp thách thức lớn về mặt thể chế quốc tế như: (i) tốc độ tương trợ tư pháp còn chậm so với tốc độ vận động nhanh và phi biên giới của loại tội phạm này. Trong khi các kênh tương trợ tư pháp hình sự truyền thống cần ít nhất một tháng hoặc nhiều thời gian hơn để xử lý yêu cầu thu thập chứng cứ điện tử, truy vết dòng tiền hoặc phong tỏa tài sản thì đối với tội phạm Deepfake, tài sản bị chiếm đoạt có thể được tẩu tán trong thời gian rất ngắn, thậm chí chỉ vài phút; (ii) khó khăn trong truy vết tiền mã hóa, bởi việc truy vết và thu hồi tài sản khi đã được quy đổi thành tài sản ảo trên các sàn giao dịch phi tập trung đòi hỏi cơ chế hợp tác tức thời với các cơ quan thực thi pháp luật quốc tế chuyên trách về tội phạm mạng như Europol, Interpol hoặc các đơn vị của Hoa Kỳ để nhanh chóng xác định và làm việc với các sàn giao dịch đó. Ngoài ra, do Việt Nam chưa có các thỏa thuận song phương hoặc đa phương chuyên biệt, cho phép chia sẻ dữ liệu sinh trắc học và chứng cứ điện tử xuyên biên giới nhanh chóng, tạo ra rào cản pháp lý lớn trong việc chứng minh chuỗi hành vi phạm tội¹¹.

3. Giải pháp hoàn thiện pháp luật điều chỉnh hành vi giả mạo giọng nói, hình ảnh bằng AI trong giao dịch tài chính số

Trong bối cảnh chuyển đổi số trong lĩnh vực

tài chính, công nghệ AI, đặc biệt là Deepfake, đang bị lợi dụng để giả mạo giọng nói, hình ảnh nhằm thực hiện các giao dịch gian lận ngày càng tinh vi. Tuy nhiên, pháp luật hiện hành chưa theo kịp sự phát triển này và còn thiếu quy định cụ thể cũng như cơ chế kiểm soát hiệu quả. Vì vậy, việc hoàn thiện khung pháp lý điều chỉnh hành vi giả mạo bằng AI là yêu cầu cấp thiết nhằm bảo đảm an toàn giao dịch và bảo vệ các chủ thể trong môi trường tài chính số, thông qua các giải pháp cụ thể sau:

Thứ nhất, làm rõ ranh giới áp dụng giữa Điều 174 và Điều 290 Bộ luật Hình sự năm 2015 trong các trường hợp sử dụng công nghệ Deepfake để chiếm đoạt tài sản. Sự giao thoa giữa hai tội danh này có thể dẫn đến cách hiểu và áp dụng pháp luật không thống nhất trong thực tiễn. Vì vậy, Hội đồng Thẩm phán Tòa án nhân dân tối cao cần nghiên cứu ban hành nghị quyết hướng dẫn áp dụng pháp luật theo hướng: nếu Deepfake chỉ đóng vai trò là thủ đoạn gian dối làm cho nạn nhân tự nguyện chuyển giao tài sản thì áp dụng Điều 174 Bộ luật Hình sự năm 2015; ngược lại, nếu hành vi chiếm đoạt gắn với việc lợi dụng hệ thống mạng, tài khoản hoặc hạ tầng số thì áp dụng Điều 290 Bộ luật Hình sự năm 2015. Đồng thời, khái niệm “tài khoản” trong Điều 290 cần được giải thích theo nghĩa rộng, gồm tài khoản ngân hàng, ví điện tử và các nền tảng số có giá trị tài sản.

Thứ hai, xây dựng khung phân tích đánh giá tính dữ liệu điện tử trên cơ sở kết hợp Điều 99

11. Congressional Record (2010), *Proceedings and debates of the 111th*, <https://www.congress.gov/111/crec/2010/11/15/CREC-2010-11-15.pdf>, truy cập ngày 26/02/2026.

và Điều 107 Bộ luật Tố tụng hình sự năm 2015, sửa đổi, bổ sung các năm 2021, 2025 (Bộ luật Tố tụng hình sự năm 2015). Cụ thể, có thể triển khai xây dựng quy trình theo 03 bước: (i) giai đoạn thu thập và bảo toàn nguồn dữ liệu sẽ căn cứ dựa trên khoản 1 và khoản 2 Điều 107 quy định phương tiện điện tử phải được thu giữ kịp thời, mô tả đúng thực trạng, niêm phong và lập biên bản; nếu không thể thu giữ thì phải sao lưu và bảo quản như vật chứng. Đây là nền tảng pháp lý để bảo đảm tính nguyên vẹn ban đầu của dữ liệu; (ii) giai đoạn sao lưu và giám định trên bản sao, theo đó tại khoản 4 Điều 107 yêu cầu việc phục hồi, tìm kiếm, giám định dữ liệu điện tử chỉ được thực hiện trên bản sao. Điều này cho phép thiết lập cơ chế đối chiếu giữa dữ liệu gốc đã niêm phong và bản sao phục vụ giám định. Trong các vụ việc có liên quan đến Deepfake, giám định cần tập trung vào: phân tích siêu dữ liệu; kiểm tra tính liên tục của chuỗi dữ liệu; so sánh đặc điểm kỹ thuật số; đối chiếu với nguồn độc lập (camera khác, nhật ký hệ thống, dữ liệu máy chủ); (iii) giai đoạn đánh giá giá trị chứng minh, căn cứ khoản 3 Điều 107, việc giám định được thực hiện theo quyết định trưng cầu của cơ quan tiến hành tố tụng. Kết quả giám định phải chuyển sang dạng có thể đọc, nghe, nhìn được và được đánh giá cùng các chứng cứ khác. Như vậy, việc bị cáo viện dẫn khả năng Deepfake không mặc nhiên làm mất giá trị chứng cứ; gánh nặng chứng minh thuộc về cơ quan tố tụng, nhưng được thực hiện thông qua cơ chế giám định tư

pháp đã được pháp luật quy định. Hiện, pháp luật Việt Nam chưa có quy chuẩn kỹ thuật chuyên biệt cho giám định Deepfake. Tuy nhiên, Điều 107 Bộ luật Tố tụng hình sự năm 2015 thiết lập quy trình bảo toàn - sao lưu - giám định tương đối đầy đủ để kiểm soát rủi ro giả mạo dữ liệu. Theo đó, nhà làm luật cần bổ sung hướng dẫn kỹ thuật chuyên sâu, ví dụ: quy chuẩn phân tích AI, chuẩn xác thực dữ liệu số, tiêu chuẩn mã băm (hàm băm), chuẩn lưu vết điện tử... nhằm nâng cao năng lực phát hiện Deepfake.

Thứ ba, xem xét hình sự hóa hành vi chiếm đoạt hoặc mua bán trái phép dữ liệu sinh trắc học. Trong môi trường tài chính số, dữ liệu sinh trắc học như khuôn mặt hoặc giọng nói không chỉ là thông tin định danh, mà còn là yếu tố xác thực giao dịch. Tuy nhiên, pháp luật hình sự hiện hành chưa quy định tội danh độc lập điều chỉnh hành vi chiếm đoạt hoặc mua bán trái phép loại dữ liệu này. Vì vậy, cần nghiên cứu bổ sung quy định hình sự đối với hành vi khai thác trái phép dữ liệu sinh trắc học, đặc biệt, trong trường hợp hành vi này phục vụ gian lận tài chính hoặc xâm nhập hệ thống thanh toán¹².

Thứ tư, hoàn thiện quy định về bảo vệ dữ liệu cá nhân trong lĩnh vực tài chính - ngân hàng. Các quy định tại Điều 8 Nghị định số 356/2025/NĐ-CP cần được cụ thể hóa hơn nhằm bảo đảm tính minh bạch, cụ thể, rõ ràng và khả thi trong thực tiễn áp dụng. Theo đó, pháp luật nên xác định rõ các tiêu chuẩn bảo mật bắt buộc đối với dữ liệu sinh trắc học, phạm vi và thời hạn

12. Trần Nam Anh, Thái Anh Tấn, Phạm Hồng Thắng, *Phòng ngừa tội phạm sử dụng trí tuệ nhân tạo (AI) ở một số quốc gia trên thế giới và kinh nghiệm đối với Việt Nam*, Kỷ yếu diễn đàn khoa học trẻ “Pháp luật Việt Nam trong thời đại pháp quyền và kỷ nguyên công nghệ số”, Trường Đại học Luật - Đại học Quốc gia Hà Nội, tr.129-141.

lưu trữ dữ liệu, cũng như giới hạn quyền rút lại sự đồng ý của chủ thể dữ liệu trong trường hợp việc xử lý dữ liệu được thực hiện trên cơ sở nghĩa vụ pháp lý theo luật chuyên ngành.

Thứ năm, thiết lập cơ chế pháp lý cho việc phong tỏa tài khoản trong trường hợp nghi ngờ gian lận bằng công nghệ AI. Trong đó, lập bộ tiêu chí rủi ro chuyên biệt để nhận diện giao dịch nghi ngờ deepfake như: sai lệch giữa mẫu sinh trắc học gốc và dữ liệu xác thực thời gian thực; thay đổi bất thường về thiết bị hoặc vị trí truy cập; giao dịch giá trị lớn phát sinh ngay sau khi cập nhật thông tin nhận dạng hoặc giao dịch liên quan đến tài khoản đã có cảnh báo gian lận. Khi các tiêu chí này được đáp ứng, ngân hàng có thể kích hoạt cơ chế phong tỏa tạm thời trong thời hạn ngắn, đồng thời, tiến hành xác minh đa kênh và báo cáo cơ quan quản lý.

Ngoài ra, nên quy định cơ chế miễn trừ trách nhiệm có điều kiện đối với tổ chức tín dụng nếu việc phong tỏa được thực hiện trên cơ sở tuân thủ quy trình đánh giá rủi ro theo quy định pháp luật. Hiện, pháp luật chưa ghi nhận cơ chế “miễn trừ có điều kiện” cho ngân hàng khi phong tỏa tài khoản trên cơ sở đánh giá rủi ro hợp lý trong trường hợp nghi ngờ giả mạo bằng AI. Vì vậy, việc phong tỏa có thể phát sinh trách nhiệm bồi thường nếu đánh giá sai, trong khi không phong tỏa sẽ tạo điều kiện cho dòng tiền bị tẩu tán, gia tăng thiệt hại cho nạn nhân và hệ thống tài chính¹³.

Thứ sáu, hoàn thiện khuôn khổ pháp lý và tăng cường hợp tác quốc tế trong xử lý tội phạm công nghệ cao. Do tội phạm sử dụng Deepfake trong giao dịch tài chính thường mang tính

xuyên biên giới, để việc xử lý hiệu quả các vụ việc này đòi hỏi cơ chế tương trợ tư pháp điện tử linh hoạt và kịp thời. Vì vậy, cần tiếp tục hoàn thiện các quy định pháp luật về thu thập chứng cứ điện tử, truy vết dòng tiền và phong tỏa tài sản trong các vụ án có yếu tố nước ngoài, đồng thời, mở rộng các cơ chế hợp tác quốc tế trong phòng, chống tội phạm công nghệ cao.

Kết luận

Sự phát triển của giao dịch tài chính số trong tiến trình chuyển đổi số quốc gia tạo ra không gian mới cho tội phạm sử dụng AI, đặc biệt là Deepfake. Khi giọng nói và hình ảnh là nền tảng của xác thực sinh trắc học có thể bị tái tạo với độ chân thực cao, ranh giới giữa giả mạo thông tin và xâm phạm trực tiếp tài sản ngày càng thu hẹp, đặt ra thách thức cho pháp luật hình sự trong định tội danh, chứng minh và phòng ngừa rủi ro. Nghiên cứu cho thấy một số khoảng trống như sự thiếu thống nhất giữa Điều 174 và Điều 290 Bộ luật Hình sự năm 2015, cơ chế bảo vệ dữ liệu sinh trắc học và chế tài chưa tương xứng với rủi ro trong môi trường tài chính số, cùng với hạn chế trong phong tỏa tài khoản và hợp tác quốc tế. Những vấn đề này phản ánh sự chênh lệch giữa tốc độ phát triển công nghệ và khả năng phản ứng của pháp luật. Vì vậy, cần tiếp cận theo hướng hệ thống: làm rõ cấu thành tội phạm trong không gian mạng, xây dựng quy chuẩn giám định AI, tăng cường bảo vệ dữ liệu sinh trắc học, thiết lập cơ chế phong tỏa tài khoản có điều kiện và thúc đẩy tương trợ tư pháp điện tử xuyên biên giới. Chỉ khi bảo đảm tính minh định, khả năng dự báo và phản ứng

13. Trần Nam Anh, Thái Anh Tấn, Phạm Hồng Thắng, tldđ, tr.129-141.

kip thời của pháp luật, hệ thống tài chính số mới có thể phát triển bền vững trong kỷ nguyên trí tuệ nhân tạo □

TÀI LIỆU THAM KHẢO

1. Abdulazeez Alali & George Theodorakopoulos, *Partial Fake Speech Attacks in the Real World Using Deepfake Audio*, Journal of Cybersecurity and Privacy, Vol. 5, Article 6, 2025.

2. Bộ Tư pháp (2025), *Mua cá nhân dữ liệu vẫn tràn lan trên mạng*, <https://dx.moj.gov.vn/mua-ban-du-lieu-ca-nhan-van-tran-lan-tren-mang-903.htm>, truy cập ngày 20/02/2026.

3. Chesney, B., & Citron, D. (2019), *Deepfakes: A looming challenge for privacy, democracy, and national security*, Calif. L. Rev., 107, 1753.

4. Congressional Record (2010), *Proceedings and debates of the 111th Congress*, <https://www.congress.gov/111/crec/2010/11/15/CREC-2010-11-15.pdf>, truy cập ngày 26/02/2026.

5. Lê Tuấn Tú, Nguyễn Thanh Thảo Nhi (2024), *Tội “Sử dụng mạng viễn thông, mạng máy tính, phương tiện điện tử để chiếm đoạt tài sản” theo quy định của Bộ luật Hình sự năm 2015*, <https://lsvn.vn/toi-su-dung-mang-vien-thong-mang-may-tinh-phuong-tien-dien-tu-de-chiem-doat-tai-san-theo-quy-dinh-cua-bo-luat-hinh-su-2015-a152321.html>, truy cập ngày 20/02/2026.

6. Lưu Quang Anh, Bùi Lê Hiếu (2024), *Tội sử dụng mạng máy tính, mạng viễn thông, phương tiện điện tử thực hiện hành vi chiếm đoạt tài sản - Vương mắc và một số kiến nghị*, [https://tapchitoaan.vn/toi-su-dung-mang-may-tinh-mang-vien-thong-phuong-tien-dien-tu-thuc-hien-hanh-vi-chiem-doat-tai-san-vuong-](https://tapchitoaan.vn/toi-su-dung-mang-may-tinh-mang-vien-thong-phuong-tien-dien-tu-thuc-hien-hanh-vi-chiem-doat-tai-san-vuong-mac-va-mot-so-kiem-nghi-12972.html)

[mac-va-mot-so-kiem-nghi12972.html](https://tapchitoaan.vn/toi-su-dung-mang-may-tinh-mang-vien-thong-phuong-tien-dien-tu-thuc-hien-hanh-vi-chiem-doat-tai-san-vuong-mac-va-mot-so-kiem-nghi-12972.html), truy cập ngày 26/02/2026.

7. Schiff, K. J., Schiff, D. S., & Bueno, N. S. (2025), *The Liar’s Dividend: Can Politicians Claim Misinformation to Evade Accountability?*, American Political Science Review, 119(1), 71 - 90, doi:10.1017/S0003055423001454, truy cập ngày 26/02/2026.

8. Tambiama Madiaga (2023), *Generative AI and watermarking*, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI\(2023\)757583_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf), truy cập ngày 26/02/2026.

9. Thành Luân, Anh Quân (2021), *Vụ rò rỉ 17 GB dữ liệu người dùng Việt: Cách bảo vệ thông tin cá nhân*, <https://thanhnien.vn/vu-ro-ri-17-gb-du-lieu-nguoi-dung-viet-cach-bao-ve-thong-tin-ca-nhan-1851069078.htm>, truy cập ngày 21/02/2026.

10. Shannon Murphy (2024), *A Deepfake Scammed a Bank out of \$25M - Now What?*, https://www.trendmicro.com/en_us/research/24/b/deepfake-video-calls.html, truy cập ngày 20/02/2026.

11. Trần Nam Anh, Thái Anh Tấn, Phạm Hồng Thắng, *Phòng ngừa tội phạm sử dụng trí tuệ nhân tạo (AI) ở một số quốc gia trên thế giới và kinh nghiệm đối với Việt Nam*, Kỷ yếu diễn đàn khoa học trẻ “Pháp luật Việt Nam trong thời đại pháp quyền và kỷ nguyên công nghệ số”, Trường Đại học Luật - Đại học Quốc gia Hà Nội.

12. Westerlund, M. (2019), *The Emergence of Deepfake Technology: A Review*, Technology Innovation Management Review, 9(11), 40 - 53, <http://doi.org/10.22215/timreview/1282>, truy cập ngày 26/02/2026.