

# TRÁCH NHIỆM GIẢI TRÌNH VỀ CÁC QUYẾT ĐỊNH VÀ HỆ QUẢ CỦA HỆ THỐNG TRÍ TUỆ NHÂN TẠO THEO LUẬT TRÍ TUỆ NHÂN TẠO VIỆT NAM NĂM 2025

ThS VĂN THỊ HỒNG NHUNG

NCS Trường Đại học Kinh tế - Luật, ĐHQG-HCM  
PhD Candidate, University of Economics and Law, VNU-HCM  
Email: nhungvth23709@sdh.uel.edu.vn

## Tóm tắt

Luật Trí tuệ nhân tạo năm 2025 của Việt Nam xác lập nguyên tắc “thực hiện trách nhiệm giải trình về các quyết định và hệ quả của hệ thống trí tuệ nhân tạo” tại khoản 3 Điều 4. Bài viết phân tích nội hàm quy định này, làm rõ các vấn đề về nội dung, yêu cầu và giới hạn của trách nhiệm giải trình. Trên cơ sở nghiên cứu các quy định của Việt Nam kết hợp tham khảo các quy định trong pháp luật Liên minh châu Âu và Luật bang Colorado của Hoa Kỳ, bài viết đề xuất các kiến nghị hoàn thiện văn bản hướng dẫn thi hành, nhấn mạnh sự dung hòa giữa bảo vệ quyền con người và thúc đẩy đổi mới sáng tạo.

**Từ khóa:** trách nhiệm giải trình, giải trình quyết định, giải trình hệ quả, Luật Trí tuệ nhân tạo năm 2025  
**Abstract**

Vietnam's 2025 Law on Artificial Intelligence establishes the principle of “implementing accountability for decisions and outcomes of artificial intelligence systems” under clause 3, Article 4. This article analyzes the substantive scope of this provision, elucidating issues regarding the, content, requirements, and limitations of accountability. Based on a study of Vietnamese regulations in conjunction with a comparative analysis of European Union law and the Colorado Privacy Act (USA), the article proposes recommendations for improving the guiding implementation documents, emphasizing the harmonization between the protection of human rights and the promotion of innovation.

**Keywords:** accountability, decision accountability, outcome accountability, 2025 Law on Artificial Intelligence

**DOI:** <https://doi.org/10.70236/khplvn.659>

**Ngày nhận bài:** 05/8/2025

**Ngày duyệt đăng:** 17/03/2026

Tháng 01 năm 2026, tỷ phú Elon Musk đưa ra dự đoán: năm 2026 Trí tuệ nhân tạo tổng quát (*Artificial general intelligence*, AGI) sẽ xuất hiện, robot y tế sẽ giỏi hơn các bác sĩ phẫu thuật giỏi nhất chỉ trong vòng ba năm tới. Thậm chí, vào năm 2030 AGI sẽ vượt trội hơn trí tuệ của toàn bộ nhân loại cộng lại.<sup>1</sup> Nếu dự đoán này trở thành hiện thực thì sẽ có rất nhiều vấn đề pháp lý nghiêm trọng được đặt ra. Trong đó, an toàn và minh bạch của trí tuệ nhân tạo (*Artificial Intelligence*, AI) là một thách thức cốt lõi cần phải giải quyết, đặc biệt khi các hệ thống AI được triển khai trong những lĩnh vực có ảnh hưởng lớn đến quyền con người như y tế, tài chính và tư pháp. Bởi lẽ, các hệ thống AI thường ẩn chứa những tác nhân khó nhận diện trong thuật toán, phần mềm và quá trình ra quyết định khác biệt với con người hoặc máy móc vật lý thông thường. Trong một số trường hợp, thiệt hại xảy ra nhưng khó có thể xác định được nguyên nhân và chứng minh mối quan hệ nhân quả để bảo vệ quyền lợi cho người bị thiệt hại.

Để giải quyết vấn đề này, pháp luật quốc tế và các quốc gia đều tích cực xây dựng và hoàn thiện hệ thống pháp luật về quản trị AI với tính mục tiêu dung hòa đổi mới công nghệ với bảo vệ quyền con người. Trong đó, trách nhiệm giải trình (TNGT) không chỉ là một trong những biện pháp để đảm bảo tính minh bạch của AI mà còn là công cụ giúp cho việc bảo vệ quyền con người hiệu quả hơn trong thời đại công nghệ số.

<sup>1</sup> Công Khải, “Elon Musk và loạt dự đoán tương lai gây sốc: Robot, AI vượt trội nhân loại; tiền không còn ý nghĩa”, *Tuổi Trẻ Online*, <https://tuoitre.vn/elon-musk-va-loat-du-doan-tuong-lai-gay-soc-robot-ai-vuot-troi-nhan-loai-tien-khong-con-y-nghia-20260113095700313.htm>, truy cập ngày 14/01/2026.

Với xu hướng đó, ngày 10 tháng 12 năm 2025,<sup>2</sup> Quốc hội Việt Nam đã thông qua Luật Trí tuệ nhân tạo, số 134/2025/QH15 (sau đây gọi là Luật AI). Đây là đạo luật đầu tiên của Việt Nam điều chỉnh trực tiếp, toàn diện hoạt động AI, trong đó Khoản 3 Điều 4 xác lập nguyên tắc: “Bảo đảm công bằng, minh bạch, không thiên lệch, không phân biệt đối xử và không gây hại cho con người hoặc xã hội; tuân thủ chuẩn mực đạo đức và giá trị văn hóa Việt Nam; thực hiện trách nhiệm giải trình về các quyết định và hệ quả của hệ thống trí tuệ nhân tạo”.

Quy định này đặt ra một số vấn đề cần làm rõ khi xây dựng văn bản hướng dẫn thi hành, đặc biệt là việc giải quyết được hai câu hỏi trọng tâm: (i) Các hệ thống AI hiện đại vẫn chưa đạt được mức độ giải thích được đầy đủ thì TNGT về quyết định và hệ quả của hệ thống AI đặt ra những yêu cầu gì về mặt kỹ thuật? (ii) Ranh giới giữa TNGT và quyền bảo vệ bí mật kinh doanh cần được xác định như thế nào để vừa bảo đảm quyền được thông tin vừa không triệt tiêu động lực đổi mới sáng tạo của doanh nghiệp?

Để trả lời hai câu hỏi trên, bài viết sử dụng phương pháp phân tích luật thực định kết hợp phương pháp so sánh pháp luật theo tiêu chí cụ thể. Việc lựa chọn hai hệ thống pháp luật tham chiếu là Đạo luật AI của Liên minh châu Âu (*Regulation EU 2024/1689* – sau đây gọi là Đạo luật AI EU) và Luật AI của bang Colorado, Hoa Kỳ (*Colo. Rev. Stat.*, sau đây gọi là Luật AI Colorado) xuất phát từ sự khác biệt có tính bổ trợ trong cách tiếp cận. Theo đó, trong khi pháp luật EU đặt TNGT như một nghĩa vụ minh bạch xuyên suốt chuỗi giá trị AI với các yêu cầu kỹ thuật bắt buộc thì Luật AI Colorado lại thiết kế TNGT gắn liền với cơ chế bảo vệ người tiêu dùng và quyền kháng cáo quyết định tự động.

Phạm vi nghiên cứu của bài viết tập trung vào hai vấn đề cốt lõi là (i) mối quan hệ giữa TNGT với khả năng giải thích được của hệ thống AI và (ii) xung đột giữa TNGT với bảo vệ bí mật kinh doanh mà không đi sâu vào toàn bộ các khía cạnh của quản trị AI nói chung.

## 1. Khái niệm trách nhiệm giải trình trong lĩnh vực trí tuệ nhân tạo

Mặc dù Luật AI không quy định thế nào là TNGT nhưng trong lý thuyết về quản trị công, TNGT được khái niệm là trách nhiệm của một chủ thể có nghĩa vụ thông tin, giải thích và biện minh hành vi của mình trước diễn đàn có thẩm quyền giám sát, chất vấn và đưa ra phán quyết.<sup>3</sup> Theo đó, TNGT không đơn thuần là nghĩa vụ giải thích hay báo cáo mang tính hành chính mà là một cơ chế đa tầng điều chỉnh mối quan hệ giữa các chủ thể. Quá trình này thường trải qua ba giai đoạn có sự liên kết chặt chẽ gồm: giai đoạn giải trình và chứng minh (*answerability*); giai đoạn tranh luận và đánh giá (*debate*); giai đoạn xử lý và điều chỉnh (*enforceability*). Ba giai đoạn này không vận hành độc lập mà tạo thành một chu trình phản hồi liên tục, hiệu quả của giai đoạn sau phụ thuộc vào chất lượng của giai đoạn trước: nếu thông tin giải trình không đầy đủ hoặc mơ hồ thì việc tranh luận sẽ thiếu cơ sở và dẫn đến việc thiếu trách nhiệm ở giai đoạn sau.<sup>4</sup>

Với cách hiểu đó, theo tác giả, trong lĩnh vực AI, TNGT nên được xác định là “nghĩa vụ của các chủ thể liên quan trong chuỗi giá trị AI phải cung cấp thông tin rõ ràng, có ý nghĩa về cách thức hệ thống AI hoạt động, đưa ra quyết định và hệ quả của những quyết định đó”.

Trách nhiệm này được đặt ra là do các hệ thống AI thường có tính chất “hộp đen” (*black box*)<sup>5</sup> – tức là tính chất khó giải thích, chứng minh. Thuộc tính này do các hệ thống

<sup>2</sup> Luật Trí tuệ nhân tạo năm 2025, Luật số 134/2025/QH15.

<sup>3</sup> Mark Bovens, “Analysing and assessing accountability: A conceptual framework”, *European Law Journal*, Vol. 13(4), 2007, tr. 447–468.

<sup>4</sup> Yannis Papadopoulos, *Understanding Accountability in Democratic Governance*, Nxb. Cambridge University Press, Cambridge, 2023, tr. 2.

<sup>5</sup> IBM, “What Is Black Box AI and How Does It Work?”, <https://www.ibm.com/think/topics/black-box-ai>, truy cập ngày 29/10/2024.

AI chứa các mô hình mạng neuron sâu (*deep neural networks*) với hàng triệu, thậm chí hàng tỷ tham số. Các mô hình này lại hoạt động theo những cách thức mà ngay cả các nhà phát triển cũng không thể theo dõi và giải thích được một cách rõ ràng từng bước trong quá trình xử lý thông tin. Đặc biệt, khả năng tự học, tự hoàn thiện của thuật toán giúp AI có thể thay đổi hành vi, cách thức ra quyết định dựa trên dữ liệu mới mà không cần can thiệp từ con người càng làm phức tạp thêm vấn đề này.

Ví dụ, mô hình ngôn ngữ nền tảng AnalogSeeker được phát triển đặc biệt cho thiết kế mạch tương tự, được huấn luyện từ mô hình Qwen2.5-32B-Instruct. Mô hình này đạt được độ chính xác 85.04% trên thang đo AMSBench-TQA, vượt trội hơn cả các mô hình thương mại như GPT-4o. Trong nghiên cứu, khi được hỏi về cách giảm nhiễu trong CMOS opamp, AnalogSeeker đã phân tích chính xác các yếu tố và chọn đúng đáp án. Tuy nhiên, các nhà nghiên cứu thấy chỉ thấy được kết quả chứ không thể giải thích được bằng cách nào, với cơ chế tính toán và liên kết khái niệm của mạng neuron như thế nào để AnalogSeeker tạo ra kết quả đó.<sup>6</sup>

Để khắc phục vấn đề hộp đen, lĩnh vực AI giải thích được (*Explainable AI - XAI*) đã phát triển nhiều phương pháp nhằm làm cho quá trình ra quyết định của AI trở nên rõ ràng và dễ hiểu hơn.<sup>7</sup> Tuy nhiên, cần phân biệt rõ giữa XAI với các khái niệm khác về AI thường bị sử dụng lẫn lộn trong thực tiễn lập pháp, bởi sự nhầm lẫn này có thể dẫn đến thiết kế quy phạm không phù hợp với thực tế kỹ thuật.

Theo Khung Quản lý rủi ro AI của Viện Tiêu chuẩn và công nghệ Quốc gia Hoa Kỳ thì (i) minh bạch (*transparency*), (ii) khả năng giải thích (*explainability*) và (iii) khả năng diễn giải (*interpretability*) là những đặc điểm riêng biệt của AI nhưng có mối liên hệ tương hỗ với nhau.<sup>8</sup>

Trong đó, tính minh bạch phản ánh mức độ thông tin về hệ thống AI và đầu ra có sẵn cho những người tương tác, tức là trả lời câu hỏi “điều gì đã xảy ra” trong quá trình vận hành hệ thống. Tính giải thích được đi sâu hơn, mô tả cơ chế vận hành bên trong, trả lời câu hỏi “hệ thống đưa ra quyết định như thế nào”. Còn tính diễn giải được là cấp độ cao nhất, khi ý nghĩa của đầu ra được đặt trong bối cảnh mục đích thiết kế và nhu cầu cụ thể của người dùng, trả lời câu hỏi “tại sao hệ thống đưa ra quyết định đó và điều này có ý nghĩa gì đối với người chịu tác động”.<sup>9</sup>

Việc phân biệt rõ ba thuộc tính này không chỉ là vấn đề kỹ thuật mà có ý nghĩa quan trọng đối với thiết kế khung pháp lý về TNGT, bởi lẽ một hệ thống AI có thể đạt mức minh bạch cao không có nghĩa là người tiếp nhận có thể hiểu được cơ chế đưa ra quyết định cụ thể.

Ví dụ: Ngân hàng triển khai hệ thống AI để xét duyệt hồ sơ vay tín dụng và cung cấp đầy đủ các thông tin minh bạch cho khách hàng theo quy định, tuy nhiên lại không thể trả lời câu hỏi: tại sao hồ sơ X bị từ chối trong khi hồ sơ Y của một người có thu nhập tương tự lại được duyệt? Lý do là nếu chỉ dựa trên thông tin minh bạch nói trên thì chỉ trả lời được rằng hệ thống được xây dựng từ những dữ liệu gì. Còn câu hỏi “hệ thống đã quyết định như thế nào trong trường hợp cụ thể này” sẽ phải dựa trên tính năng giải thích được, đòi hỏi khả năng truy vết quá trình xử lý bên trong mạng neuron để xác định những yếu tố nào đã ảnh hưởng quyết định đến đầu ra cuối cùng.

6 Zihao Chen et al., “AnalogSeeker: An open-source foundation language model for analog circuit design”, *arXiv preprint, arXiv:2508.10409*, 2025, <https://doi.org/10.48550/arXiv.2508.10409>

7 Xiaoming Liu et al., “From black box to glass box: A practical review of explainable artificial intelligence (XAI)”, *AI (MDPI)*, Vol. 6(11), 2025, tr. 2.

8 NIST, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1, 2023, tr. 17.

9 Như trên, tr. 17: “Transparency can answer the question of ‘what happened’ in the system. Explainability can answer the question of ‘how’ a decision was made in the system. Interpretability can answer the question of ‘why’ a decision was made in the system and its meaning or context to the user.”

Do đó, một hệ thống minh bạch không nhất thiết phải là một hệ thống chính xác, được tăng cường quyền riêng tư, an toàn hay công bằng nhưng nếu hệ thống không minh bạch thì gần như không thể đánh giá được liệu nó có sở hữu những đặc điểm như vậy hay không.<sup>10</sup>

Nói cách khác, minh bạch là điều kiện cần nhưng chưa đủ cho TNGT, các thông tin đầu vào chỉ tạo ra nền tảng để các chủ thể có thẩm quyền tiến hành đánh giá, chất vấn và xử lý chứ không tự động cho phép phát hiện được lỗi, thiên kiến hay sự bất hợp lý trong quyết định của hệ thống. Để TNGT vận hành thực chất, hệ thống AI còn phải đạt được mức độ giải thích được nhất định, tức là phải có khả năng chỉ ra mối liên hệ giữa dữ liệu đầu vào, quá trình xử lý và kết quả đầu ra theo cách mà con người có thể kiểm chứng.

Trong ví dụ trên, để có thể kết luận quyết định từ chối của hệ thống đưa ra hợp lý hay có phân biệt đối xử thì cần phải có khả năng chỉ ra trong trường hợp cụ thể của khách hàng đó, những yếu tố nào đã đóng vai trò quyết định và mức độ ảnh hưởng của từng yếu tố là bao nhiêu. Đây mới chính là loại thông tin cho phép khách hàng hiểu được lý do bị từ chối và cho phép cơ quan giám sát đánh giá tính hợp pháp của quyết định.

Trên thực tế, tính “giải thích được đầy đủ” vẫn là thách thức kỹ thuật lớn, đặc biệt đối với các mô hình học sâu và mô hình ngôn ngữ lớn. Các phương pháp XAI hiện nay chủ yếu cung cấp giải thích ở mức xấp xỉ, tức là chỉ tái hiện được một phần logic dẫn đến quyết định chứ không phải toàn bộ quá trình suy luận thực tế bên trong mạng neuron. Do đó, nguyên tắc hệ thống chỉ được vận hành trong các điều kiện mà nó được thiết kế cho và khi đạt được mức độ tin cậy đủ trong đầu ra của mình (*knowledge limits*)<sup>11</sup> có ý nghĩa đặc biệt quan trọng, yêu cầu hệ thống AI cần có cơ chế tự nhận biết khi gặp tình huống nằm ngoài phạm vi thiết kế hoặc khi mức độ tin cậy của đầu ra không đủ cao thì phải từ chối đưa ra kết luận thay vì cố gắng miễn cưỡng đưa ra một quyết định không đáng tin cậy. Đây chính là cơ sở kỹ thuật để pháp luật đặt ra yêu cầu giám sát con người và cơ chế can thiệp khi hệ thống AI tiếp cận giới hạn của mình.<sup>12</sup>

Đáng lưu ý, với trình độ khoa học kỹ thuật hiện tại, trong nhiều trường hợp, việc tăng cường khả năng giải thích có thể làm giảm độ chính xác dự đoán hoặc mâu thuẫn với yêu cầu bảo vệ quyền riêng tư.<sup>13</sup> Sự đánh đổi này không thể được giải quyết bằng một công thức chung mà phải được cân nhắc trong từng bối cảnh cụ thể, tùy thuộc vào mức độ rủi ro, lĩnh vực ứng dụng và đối tượng chịu tác động.

Những giới hạn kỹ thuật trên đặt ra một thách thức pháp lý khó giải quyết khi ban hành các văn bản pháp luật: không thể yêu cầu giải trình ở mức mà công nghệ chưa cho phép, nhưng cũng không thể lấy giới hạn công nghệ làm lý do miễn trừ hoàn toàn TNGT. Để giải quyết thách thức này, pháp luật về TNGT cần được thiết kế theo hai tiêu chí thay vì áp dụng một chuẩn chung cho mọi trường hợp. Cụ thể:

*Thứ nhất:* mức độ nghiêm ngặt của TNGT phải tương xứng với mức độ rủi ro của hệ thống AI, tức là hệ thống có rủi ro càng cao thì yêu cầu giải trình càng chặt chẽ. Ví dụ: một hệ thống AI gợi ý phim đáng xem thì không cần phải đáp ứng tiêu chuẩn giải trình ngang với một hệ thống AI hỗ trợ chẩn đoán y khoa hay tư pháp.

*Thứ hai:* hình thức và mức độ của giải trình phải được thiết kế phù hợp với đối tượng tiếp nhận. Tức là, cùng một quyết định của hệ thống AI thì cách diễn đạt nội dung

10 NIST, *ibid.*, tr. 16.

11 Phillips, P Jonathon, Carina A Hahn, Peter C Fontana, Amy N Yates, Kristen Greene, David A Broniatowski, and Mark A Przybocki, *Four principles of explainable artificial intelligence*, National Institute of Standards and Technology, 2021, tr. ii (Executive Summary), <https://doi.org/10.6028/NIST.IR.8312>

12 Beatriz Botero Arcila, “AI liability in Europe: How does it complement risk regulation and deal with the problem of human oversight?”, *Computer Law & Security Review*, Vol. 54, 2024, 106012, <https://doi.org/10.1016/j.clsr.2024.106012>

13 NIST, *ibid.*, tr. 18.

giải thích và tập trung vào yếu tố cốt lõi cho cơ quan quản lý nhà nước có chuyên môn kỹ thuật (hoặc có thể huy động chuyên gia đánh giá) phải khác với cách giải thích cho một khách hàng phổ thông, không có chuyên môn.

Hai tiêu chí trên xuất phát từ một yêu cầu: TNGT về quyết định của hệ thống AI chỉ có thể được thực hiện một cách thực chất khi hệ thống đạt được một mức độ nhất định về tính giải thích được. Nói cách khác, tính giải thích được là điều kiện kỹ thuật tiên quyết để nghĩa vụ giải trình pháp lý có thể vận hành trên thực tế<sup>14</sup> và giải trình kỹ thuật cần được diễn giải để trở nên dễ tiếp cận, tạo ra sự hiểu biết phù hợp với từng nhóm đối tượng khác nhau.<sup>15</sup>

## 2. Nội dung quy định về trách nhiệm giải trình trong Luật Trí tuệ nhân tạo năm 2025

### 2.1. Nguyên tắc giải trình về quyết định và hệ quả của hệ thống trí tuệ nhân tạo

Trong Luật AI, TNGT được quy định rải rác trong một số điều luật, từ nguyên tắc chung (Điều 4), các hành vi bị cấm (Điều 7) cho đến các quy định về TNGT của các chủ thể là nhà cung cấp (NCC) và bên triển khai (BTK) (Điều 14 và Điều 15). Trong đó, nguyên tắc nền tảng về TNGT trong lĩnh vực AI là cơ sở để áp dụng các quy định cụ thể với hai nội dung là (i) các quyết định của hệ thống AI và (ii) hệ quả của hệ thống AI.

#### 2.1.1. Giải trình về các quyết định của hệ thống trí tuệ nhân tạo

Có thể xác định “quyết định” là kết quả đầu ra của hệ thống AI dưới dạng dự đoán, khuyến nghị, phân loại hoặc hành động cụ thể. Đây là giải trình mang tính phản ứng, được thực hiện sau khi hệ thống AI đã được đưa vào hoạt động và thường tập trung vào việc giải thích quy trình kỹ thuật dẫn đến một kết quả cụ thể. Giải trình về quyết định trả lời cho câu hỏi: “Trí tuệ nhân tạo quyết định như thế nào và vì sao lại quyết định như vậy?”.

Hệ thống AI càng hiện đại, tính tự động càng cao thì các quyết định càng mang tính tự chủ, không phụ thuộc vào sự kiểm soát của con người. Do đó, khi thiệt hại xảy ra, việc làm rõ nguyên nhân gây ra thiệt hại đến từ yếu tố nào và sự hợp lý của quyết định do AI đưa ra trong hoàn cảnh cụ thể đó có vai trò đặc biệt quan trọng.<sup>16</sup>

Có thể hình dung trường hợp một chiếc xe tự lái tham gia giao thông: Dựa vào các dữ liệu đã được huấn luyện, bản thân chiếc xe có thể tự đưa ra quyết định xử lý khi gặp các hoàn cảnh khác nhau mà rất ít hoặc không phụ thuộc vào hành động của người lái. Cụ thể, khi xe tự lái gây tai nạn, giải trình quyết định yêu cầu làm rõ: (i) Dữ liệu nào được hệ thống thu thập và vận dụng tại thời điểm va chạm? (ii) Hệ thống đã xử lý như thế nào, có nhận diện được chướng ngại vật không, nếu không thì tại sao? (iii) Thuật toán đưa ra lệnh gì - phanh, đánh lái, tiếp tục di chuyển? (iv) Có sự can thiệp của người lái vào quyết định đó không?...Việc chứng minh chuỗi nhân quả từ dữ liệu đến xử lý và quyết định giúp xác định lỗi thuộc hệ thống AI, người lái hay yếu tố khách quan. Tuy nhiên, việc này lại không hề đơn giản, thậm chí trong một số trường hợp là bất khả thi.<sup>17</sup>

#### 2.1.2. Về giải trình hệ quả của các hệ thống trí tuệ nhân tạo

Trong khi giải trình về quyết định mang tính phản ứng thì giải trình về hệ quả thường mang tính dự phòng, được thực hiện trước khi hệ thống AI được đưa vào sử dụng và thường mang tính chất hành chính, kỹ thuật để phục vụ công tác quản lý, thanh tra,

14 Matthias Artzt, 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, tr. 44, doi: 10.2478/vjls-2022-0007

15 Jake Van der Laan, “Explainability of artificial intelligence models: Technical foundations and legal principles”, *Vietnamese Journal of Legal Sciences*, Vol. 7(02), 2022, tr. 27-28, doi: 10.2478/vjls-2022-0006

16 Heinrich-Böll-Stiftung, “Artificial Intelligence and Tort Law: Who Should Be Held Liable When AI Causes Damages?”, <https://il.boell.org/en/2021/12/24/artificial-intelligence-ai-tort-law-and-network-theory-who-should-be-held-liable-when-ai>, truy cập ngày 12/01/2026.

17 Madalina Busuioc, “Accountable artificial intelligence: Holding algorithms to account”, *Public Administration Review*, Vol. 81(5), 2021, tr. 825-836.

kiểm tra. Các chủ thể có trách nhiệm làm rõ, đánh giá và công khai các tác động thực tế do quyết định của AI gây ra cho con người, xã hội. Tức là, giải trình về hệ quả sẽ trả lời cho các câu hỏi: “điều gì có thể xảy ra khi sử dụng AI?”

Trở lại với tình huống xe tự lái bên trên, giải trình hệ quả đòi hỏi NCC phải: (i) Đánh giá tác động trước triển khai đối với người người dùng, người bị ảnh hưởng; các rủi ro tiềm ẩn như khả năng nhận diện giảm khi thời tiết xấu, người đi bộ xuất hiện đột ngột hệ thống không kịp phản ứng; (ii) Cơ chế khắc phục các rủi ro tiềm ẩn như quy trình xử lý sự cố, bảo hiểm trách nhiệm, thời hạn cập nhật, thu hồi khi phát hiện lỗi và (iii) Giám sát sau triển khai bằng báo cáo định kỳ, phân tích xu hướng để điều chỉnh trong tương lai.

Những thông tin này vừa là tiêu chí để xác định trách nhiệm của NCC trong việc dự phòng các tình huống rủi ro, vừa giúp bảo vệ quyền lợi của các chủ thể khác trong quá trình ứng dụng AI và là cơ sở để xác định các vấn đề pháp lý khi thiệt hại xảy ra.

Ngoài ra, để bảo đảm cho các quy định TNGT được thực thi trong thực tiễn một cách hiệu quả, khoản 5 Điều 7 Luật AI năm 2025 đã có quy định nghiêm cấm các hành vi: “Che giấu thông tin bắt buộc phải công khai, minh bạch hoặc giải trình; tẩy xóa, làm sai lệch các thông tin, nhãn, cảnh báo bắt buộc trong hoạt động trí tuệ nhân tạo”. Quy định này thể hiện quan điểm của Việt Nam coi tính chính xác của thông tin và nhãn cảnh báo là nền tảng cốt lõi để duy trì trật tự quản lý nhà nước và an toàn xã hội; mọi hành vi cố ý làm sai lệch nền tảng này đều bị coi là vi phạm điều cấm chú không đơn thuần là sự thiếu sót trong tuân thủ quy trình.

Về chế tài đối với hành vi vi phạm, Luật AI năm 2025 quy định cơ chế xử lý mang tính tổng hợp, bao gồm xử lý hành chính, truy cứu trách nhiệm hình sự và trách nhiệm dân sự. Riêng đối với trách nhiệm bồi thường thiệt hại thì khoản 2 và khoản 3 quy định theo hướng nghiêm ngặt: trong trường hợp hệ thống AI rủi ro cao gây thiệt hại về tính mạng, sức khỏe, tài sản, chủ sở hữu hoặc BTK phải bồi thường thiệt hại ngay cả khi họ đã tuân thủ quy định, trừ trường hợp bất khả kháng hoặc lỗi hoàn toàn thuộc về người bị thiệt hại.<sup>18</sup>

## **2.2. Xung đột giữa trách nhiệm giải trình và bảo vệ bí mật kinh doanh**

Như đã phân tích, TNGT chỉ vận hành thực chất và hiệu quả khi chủ thể có nghĩa vụ cung cấp thông tin một cách đầy đủ cho việc giải trình để các bên liên quan có thể đánh giá tính hợp lý của quyết định AI. Tuy nhiên, những thông tin về thuật toán, dữ liệu huấn luyện, bộ tham số và kiến trúc mô hình lại là tài sản trí tuệ có giá trị vô cùng lớn đối với doanh nghiệp. Vì vậy, nếu pháp luật không giải quyết khéo léo bài toán cân bằng lợi ích giữa các chủ thể thì sẽ dẫn đến tình trạng mâu thuẫn: giải trình quá ít thì quyền được thông tin của người sử dụng và người bị ảnh hưởng trở thành hình thức nhưng giải trình quá nhiều thì quyền sở hữu trí tuệ và lợi thế cạnh tranh hợp pháp của doanh nghiệp bị xâm phạm.

Các quy định hiện tại của Luật AI năm 2025 đã thể hiện định hướng giải quyết vấn đề trên với hai nguyên tắc song hành: một bên là vùng cấm tuyệt đối với quy định “việc giải trình và cung cấp thông tin không được yêu cầu tiết lộ mã nguồn, thuật toán chi tiết, bộ tham số hoặc thông tin thuộc bí mật kinh doanh, bí mật công nghệ”<sup>19</sup> và bên kia là nguyên tắc cần thiết, tương xứng với mục đích kiểm tra và không làm lộ bí mật kinh doanh khi cung cấp thông tin cho cơ quan nhà nước có thẩm quyền.<sup>20</sup> Tuy nhiên, những quy định này mới chỉ dừng ở mức nguyên tắc chung, ranh giới giữa “mô tả chức năng” mà NCC phải giải trình và “thuật toán chi tiết” mà NCC được miễn trừ chưa được làm

18 Điều 29 Luật Trí tuệ nhân tạo năm 2025.

19 Điều 14 khoản 1 điểm e Luật Trí tuệ nhân tạo năm 2025.

20 Điều 14 khoản 1 điểm c Luật Trí tuệ nhân tạo năm 2025.

rõ; đồng thời, TNGT đối với ba nhóm đối tượng tiếp nhận khác nhau là cơ quan nhà nước, người sử dụng và người bị ảnh hưởng lại được gộp chung trong Điều 14 với nội dung tương đối giống nhau nên chưa phản ánh sự khác biệt về vai trò, chuyên môn và nhu cầu của từng nhóm.

Đây là vấn đề không phải của riêng Việt Nam hay một quốc gia nào mà là thách thức pháp lý chung trên con đường xây dựng khung pháp lý quản trị AI. Thực tế, các quốc gia đều đã phải cân nhắc các giải pháp để giải quyết bài toán đó và Việt Nam có thể cân nhắc các kinh nghiệm lập pháp của EU và Hoa Kỳ khi thiết kế các quy định về TNGT.

Trước hết, pháp luật EU tiếp cận xung đột giữa TNGT và bảo vệ bí mật kinh doanh bằng cách thiết kế ba cơ chế giải trình riêng biệt tương ứng với ba nhóm chủ thể tiếp nhận, mỗi cơ chế có nội dung, phạm vi và giới hạn rõ ràng. Cụ thể: đối với BTK, Đạo luật AI EU quy định NCC hệ thống AI rủi ro cao phải kèm theo hệ thống bản hướng dẫn sử dụng với nội dung bắt buộc được liệt kê cụ thể, bao gồm mục đích sử dụng, mức độ chính xác và các chỉ số đo lường, các hạn chế đã biết, điều kiện có thể gây rủi ro đến sức khoẻ và quyền cơ bản, biện pháp giám sát con người và các thông tin liên quan đến việc giải thích đầu ra của hệ thống.<sup>21</sup> Các thông tin chi tiết này tạo ra chuẩn mực rõ ràng để xác định nội dung tối thiểu mà NCC phải giải trình thay vì phụ thuộc vào diễn giải chủ quan của từng doanh nghiệp hay cơ quan thực thi.

Đối với người bị ảnh hưởng, EU thiết lập quyền chủ động yêu cầu giải thích về quyết định cá nhân, theo đó bất kỳ cá nhân nào chịu tác động bất lợi có ý nghĩa pháp lý hoặc ảnh hưởng đáng kể tương tự từ quyết định dựa trên đầu ra của hệ thống AI rủi ro cao thuộc các nhóm được Đạo luật AI EU liệt kê đều có quyền yêu cầu BTK cung cấp giải thích rõ ràng và có ý nghĩa về vai trò của hệ thống AI trong quy trình ra quyết định cùng các yếu tố chính của quyết định đó.<sup>22</sup> Như vậy, thay vì chỉ quy định nghĩa vụ của NCC và BTK trong việc cung cấp thông tin công khai ở mức mô tả chức năng, quy định này của EU đã trao cho người bị ảnh hưởng quyền pháp lý chủ động là yêu cầu giải trình đối với chính quyết định đã tác động đến mình.

Đối với cơ quan quản lý, EU yêu cầu NCC phải cung cấp toàn bộ thông tin và tài liệu cần thiết để chứng minh sự tuân thủ, bao gồm cả quyền truy cập nhật ký vận hành tự động,<sup>23</sup> tức là trong trường hợp cần thiết thì cơ quan quản lý có thể yêu cầu quyền được tiếp cận dữ liệu ở mức độ cao nhất. Tuy nhiên, vẫn có cơ chế ràng buộc nghĩa vụ của cơ quan quản lý để đảm bảo sự cân bằng, theo đó, chỉ được yêu cầu dữ liệu thực sự cần thiết cho việc đánh giá rủi ro đi kèm với các biện pháp an ninh mạng phù hợp để bảo vệ thông tin nhận được và phải xóa dữ liệu khi không còn cần thiết cho mục đích thu thập.<sup>24</sup>

Khác với luật EU, Luật AI Colorado tập trung vào quyền của người tiêu dùng hơn là nghĩa vụ kỹ thuật của doanh nghiệp. Khi hệ thống AI rủi ro cao đưa ra quyết định bất lợi, BTK phải cung cấp cho người tiêu dùng các lý do chính của quyết định đó, bao gồm mức độ và cách thức hệ thống AI đóng góp vào quyết định, loại dữ liệu được xử lý và nguồn dữ liệu. Thậm chí, người tiêu dùng có quyền sửa dữ liệu cá nhân không chính xác mà hệ thống đã sử dụng và có quyền kháng cáo quyết định bất lợi, trong đó quy trình kháng cáo phải cho phép xem xét nếu khả thi về mặt kỹ thuật.<sup>25</sup>

Đáng lưu ý, Luật AI Colorado không buộc BTK phải tiết lộ bí mật kinh doanh nhưng cũng không cho phép doanh nghiệp sử dụng lý do giữ bí mật kinh doanh như một lá chắn để né tránh TNGT. Tức là, doanh nghiệp được quyền giữ lại thông tin vì

21 Điều 13 khoản 3 Đạo luật AI EU (Regulation (EU) 2024/1689).

22 Điều 86 Đạo luật AI EU (Regulation (EU) 2024/1689).

23 Điều 21 Đạo luật AI EU (Regulation (EU) 2024/1689).

24 Điều 78 Đạo luật AI EU (Regulation (EU) 2024/1689).

25 Colorado Revised Statutes (Colo. Rev. Stat.), § 6-1-1703(4)(b)(II).

lý do bí mật kinh doanh hoặc theo quy định bảo hộ khác nhưng phải thông báo cho người tiêu dùng biết việc giữ lại đó và nêu rõ căn cứ.<sup>26</sup> Cơ chế này giúp cho người tiêu dùng biết rằng có thông tin bị giữ lại và lý do tại sao, từ đó có thể đánh giá liệu việc giữ lại đó có hợp lý hay không và quyết định có cần khiếu nại hay đưa vấn đề lên cơ quan có thẩm quyền.

Như vậy, so với kinh nghiệm lập pháp quốc tế, hiện tại Luật AI Việt Nam có một số vấn đề cần giải quyết như: (i) gộp chung TNGT cho ba nhóm đối tượng tiếp nhận khác nhau mà chưa phân hóa nội dung và độ sâu thông tin phù hợp với từng đối tượng tiếp nhận; (ii) chưa trao quyền chủ động yêu cầu giải trình cho người bị ảnh hưởng và chưa có cơ chế cụ thể khi NCC hoặc BTK vi phạm dẫn bí mật kinh doanh để từ chối giải trình. Những vấn đề này cần được giải quyết trong văn bản hướng dẫn thi hành Luật AI để đảm bảo TNGT được thực thi một cách hiệu quả trong thực tiễn.

### **3. Kiến nghị hoàn thiện các quy định về trách nhiệm giải trình trong văn bản hướng dẫn thi hành Luật Trí tuệ nhân tạo**

Trên cơ sở phân tích về giới hạn kỹ thuật của XAI cùng với so sánh, đối chiếu các quy định giữa Luật AI Việt Nam, Đạo luật AI EU và Luật AI Colorado, tác giả đề xuất hai kiến nghị trọng tâm cho văn bản hướng dẫn thi hành Luật AI theo hướng mỗi kiến nghị giải quyết một nhóm vấn đề cụ thể.

#### **3.1. Phân hóa nội dung giải trình theo đối tượng tiếp nhận và mức độ rủi ro của hệ thống trí tuệ nhân tạo**

Kiến nghị này xuất phát từ việc Luật AI hiện hành đang gộp chung TNGT cho ba nhóm đối tượng tiếp nhận khác nhau và chưa trao quyền chủ động yêu cầu giải trình cho người bị ảnh hưởng. Do đó, để phù hợp với giới hạn kỹ thuật của AI và đảm bảo quyền lợi của người tiêu dùng, Nghị định hướng dẫn thi hành cần thiết kế TNGT theo hướng phân tầng theo mức độ rủi ro kết hợp với đối tượng tiếp nhận.

*Thứ nhất*, cần xác định ba cấp độ thông tin tương ứng với ba nhóm chủ thể mà Luật AI năm 2025 đã xác định.

Hiện tại, cách quy định của Luật AI năm 2025 khiến cho cùng một nội dung giải trình có thể quá phức tạp đối với người dùng phổ thông nhưng lại quá đơn giản đối với cơ quan quản lý có chuyên môn đánh giá kỹ thuật và chưa đạt được hiệu quả giải trình. Do đó, văn bản hướng dẫn cần phân loại cấp độ thông tin tương ứng với đối tượng tiếp nhận, cách thức diễn giải và mức độ thông tin một cách phù hợp. Cụ thể: Đối với người sử dụng và người bị ảnh hưởng thì thông tin phải được trình bày bằng ngôn ngữ phổ thông, tập trung vào mục đích sử dụng, cách thức vận hành ở mức mô tả chức năng, các hạn chế đã biết và cảnh báo rủi ro. Với cơ quan quản lý nhà nước có thẩm quyền thì thông tin phải bao gồm hồ sơ kỹ thuật chi tiết hơn về dữ liệu huấn luyện, biện pháp kiểm soát rủi ro và kết quả đánh giá hiệu năng. Đặc biệt, ở cấp độ cao nhất, để phục vụ thanh tra, kiểm tra khi có sự cố hoặc dấu hiệu vi phạm thì thông tin có thể được mở rộng đến nhật ký vận hành tự động và quy trình xử lý sự cố.

*Thứ hai*, văn bản hướng dẫn cần làm rõ TNGT chủ động đối với hệ thống AI rủi ro cao.

Với cách phân loại mức độ rủi ro hiện tại, cần xác định rõ với hệ thống AI rủi ro cao thì NCC phải xây dựng và lưu trữ hồ sơ giải trình ngay từ khâu thiết kế trong khi đối với hệ thống rủi ro trung bình và thấp thì TNGT chỉ phát sinh khi có yêu cầu hoặc khi có sự cố. Đặc biệt, đối với các hệ thống AI rủi ro cao hoạt động trong các lĩnh vực nhạy cảm, cần có quy định khuyến khích áp dụng các phương pháp XAI tích hợp ngay

26 Colo. Rev. Stat. § 6-1-1703(8).

từ khâu thiết kế để TNGT không trở thành yêu cầu bổ sung sau mà là thuộc tính vốn có của hệ thống.

Bên cạnh đó, văn bản hướng dẫn cần bổ sung quyền chủ động yêu cầu giải trình cho người bị ảnh hưởng khi họ chịu tác động bất lợi từ quyết định của hệ thống AI rủi ro cao. Cụ thể, NCC và BTK phải thiết lập kênh tiếp nhận yêu cầu giải trình và có nghĩa vụ phản hồi trong một thời hạn xác định, trong đó nội dung giải trình phải nêu rõ vai trò của hệ thống AI trong quy trình ra quyết định và các yếu tố tác động chính đến quyết định đó. Đặc biệt, với các quyết định hoàn toàn tự động có tác động đáng kể đến quyền và lợi ích hợp pháp của người bị ảnh hưởng thì phải có quy định cho phép người bị ảnh hưởng được quyền yêu cầu con người xem xét lại quyết định đó. Cơ chế này vừa khắc phục sự thụ động của người bị ảnh hưởng trong mô hình hiện tại vừa tạo điều kiện cho NCC phát hiện và xử lý sớm các rủi ro trước khi xảy ra các sự cố nghiêm trọng, từ đó giảm tải cho cơ quan nhà nước trong công tác thanh tra, kiểm tra.

Ngoài ra, cần có hướng dẫn cụ thể về định dạng và mức độ chi tiết thông tin ở mỗi cấp độ, có tham chiếu các tiêu chuẩn quốc tế đã được thừa nhận để tạo thuận lợi cho các chủ thể giải trình và thống nhất trong việc áp dụng.

### ***3.2. Xác lập cơ chế giải quyết trong trường hợp nhà cung cấp hoặc bên triển khai viện dẫn bí mật kinh doanh để từ chối giải trình***

Hiện tại, Luật AI năm 2025 chưa làm rõ ranh giới giữa “thông tin thuộc bí mật kinh doanh” và “mô tả chức năng cần công khai” có thể khiến cho việc diễn giải phụ thuộc vào ý chí chủ quan của từng doanh nghiệp và tạo ra nguy cơ ngoại lệ bí mật kinh doanh bị viện dẫn quá rộng. Đồng thời, cũng chưa có cơ chế giải quyết trong tình huống NCC hoặc BTK viện dẫn bí mật kinh doanh để từ chối giải trình. Vì vậy, văn bản hướng dẫn cần bổ sung thêm cơ chế hỗ trợ để giải quyết cả hai nguy cơ này.

Trước hết, nghị định cần liệt kê nội dung tối thiểu của mô tả chức năng mà NCC phải cung cấp trong mọi trường hợp bất kể có viện dẫn bí mật kinh doanh hay không. Nội dung tối thiểu này nên bao gồm mục đích sử dụng của hệ thống, loại dữ liệu đầu vào chủ yếu, phương thức vận hành ở mức mô tả chức năng, các hạn chế đã biết và cảnh báo rủi ro. Đây là những thông tin không đòi hỏi phải tiết lộ bí mật kinh doanh thực sự nhưng người sử dụng và người bị ảnh hưởng lại cần để có cơ sở đưa ra quyết định. Việc liệt kê rõ nội dung tối thiểu sẽ khắc phục khoảng trống hiện tại và ngăn chặn việc viện dẫn bí mật kinh doanh để từ chối cả với những thông tin cơ bản nhất.

Bên cạnh đó, đối với các thông tin nằm ngoài nội dung tối thiểu nói trên, khi NCC hoặc BTK viện dẫn bí mật kinh doanh để từ chối cung cấp, nghị định cần quy định họ phải thông báo cho bên yêu cầu biết về việc có thông tin bị giữ lại và nêu rõ căn cứ pháp lý của việc giữ lại. Đây chính là kinh nghiệm từ mô hình của Luật AI Colorado, vừa tôn trọng quyền bảo vệ bí mật kinh doanh hợp pháp của doanh nghiệp vừa ngăn ngừa việc lạm dụng ngoại lệ bí mật kinh doanh để trốn tránh hoàn toàn TNGT. Với quy định này, dù là cá nhân bị ảnh hưởng hay cơ quan quản lý thì cũng đều biết rõ có thông tin nào đang bị giữ lại và vì sao, từ đó có thể đánh giá tính hợp lý của việc giữ lại đó và quyết định có cần khiếu nại hay yêu cầu can thiệp của cơ quan có thẩm quyền hay không.

Riêng đối với trường hợp cơ quan nhà nước yêu cầu tiếp cận thông tin ở mức cao hơn để phục vụ thanh tra, kiểm tra hoặc xử lý sự cố, văn bản hướng dẫn cần làm rõ nghĩa vụ ràng buộc với cơ quan quản lý để nâng cao hiệu quả áp dụng của Luật AI. Đây là kinh nghiệm mà Việt Nam có thể tham khảo từ pháp luật EU về nghĩa vụ về an ninh mạng, thời hạn xóa dữ liệu hay quy trình xử lý thông tin bí mật để đảm bảo các nghĩa vụ kỹ thuật có thể kiểm chứng được. Cụ thể, cơ quan quản lý chỉ được yêu cầu dữ liệu thực sự

cần thiết cho việc đánh giá rủi ro, phải áp dụng biện pháp an ninh mạng phù hợp để bảo vệ thông tin nhận được và phải xóa dữ liệu khi không còn cần thiết cho mục đích thu thập. Những quy định này không chỉ bảo vệ lợi ích hợp pháp của doanh nghiệp mà còn củng cố niềm tin giữa nhà nước và doanh nghiệp trong quá trình thực thi pháp luật về AI, góp phần khuyến khích doanh nghiệp tuân thủ nghiêm túc thay vì tìm cách né tránh.

### Kết luận

Các quy định TNGT về các quyết định và hệ quả của hệ thống AI trong Luật AI năm 2025 có vai trò quan trọng trong việc xây dựng khung pháp lý AI tại Việt Nam. Quy định này không chỉ góp phần đảm bảo tính minh bạch, công bằng và an toàn trong hoạt động AI mà còn là nền tảng pháp lý để bảo vệ quyền con người trước những thách thức mà công nghệ AI đặt ra.

Tham khảo và đối chiếu với các quy định của pháp luật EU và Hoa Kỳ cho thấy, Luật AI năm 2025 có sự tương đồng, phù hợp với xu hướng tiến bộ trong quản trị AI toàn cầu. Tuy nhiên, để những quy định này phát huy hiệu quả trong thực tiễn, cần có hướng dẫn chi tiết. Trong đó cần tập trung vào việc xác định rõ các tiêu chí đánh giá, hồ sơ kỹ thuật, cấp độ thông tin phù hợp với từng đối tượng tiếp nhận để tạo điều kiện cho các chủ thể thực hiện TNGT. Đồng thời, quy định quy trình, thủ tục giải trình một cách rõ ràng bao gồm cả cơ chế khiếu nại và kháng cáo của người dùng. Ngoài ra, cần hoàn thiện quy định về xử phạt vi phạm hành chính, nâng cao năng lực thực thi và tăng cường hợp tác quốc tế.

Những quy định này sẽ góp phần tạo môi trường pháp lý minh bạch, thu hút đầu tư và thúc đẩy đổi mới sáng tạo, cân bằng giữa bảo vệ quyền lợi của người tiêu dùng với việc tạo điều kiện cho doanh nghiệp phát triển và ứng dụng AI một cách có trách nhiệm. ●

### Tài liệu tham khảo

- [1] Matthias Artzt, 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, doi: 10.2478/vjls-2022-0007
- [2] Beatriz Botero Arcila, “AI liability in Europe: How does it complement risk regulation and deal with the problem of human oversight?”, *Computer Law & Security Review*, Vol. 54, 2024, <https://doi.org/10.1016/j.clsr.2024.106012>
- [3] Mark Bovens, “Analysing and assessing accountability: A conceptual framework”, *European Law Journal*, Vol. 13(4), 2007
- [4] Madalina Busuioc, “Accountable artificial intelligence: Holding algorithms to account”, *Public Administration Review*, Vol. 81(5), 2021
- [5] Zihao Chen et al., “AnalogSeeker: An open-source foundation language model for analog circuit design”, *arXiv preprint, arXiv:2508.10409*, 2025, <https://doi.org/10.48550/arXiv.2508.10409>
- [6] Jake Van der Laan, “Explainability of artificial intelligence models: Technical foundations and legal principles”, *Vietnamese Journal of Legal Sciences*, Vol. 07(02), 2022, doi: 10.2478/vjls-2022-0006
- [7] Xiaoming Liu et al., “From black box to glass box: A practical review of explainable artificial intelligence (XAI)”, *AI (MDPI)*, Vol. 6(11), 2025
- [8] Yannis Papadopoulos, *Understanding accountability in democratic governance*, Nxb. Cambridge University Press, Cambridge, 2023
- [9] Phillips, P Jonathon, Carina A Hahn, Peter C Fontana, Amy N Yates, Kristen Greene, David A Broniatowski, and Mark A Przybocki, *Four principles of explainable artificial intelligence*, National Institute of Standards and Technology, 2021, <https://doi.org/10.6028/NIST.IR.8312>
- [10] Công Khải, “Elon Musk và loạt dự đoán tương lai gây sốc: Robot, AI vượt trội nhân loại; tiền không còn ý nghĩa”, *Tuổi trẻ online* [trans: Công Khải, “Elon Musk and a series of shocking future predictions: Robots and AI will surpass humanity; money will no longer have meaning”, *Tuổi trẻ online*]