

THUẬT TOÁN AI DƯỚI GÓC ĐỘ BÍ MẬT KINH DOANH VÀ MINH BẠCH XỬ LÝ DỮ LIỆU CÁ NHÂN: SỰ ĐỐI LẬP CẦN ĐƯỢC DUNG HÒA

THS LÊ TRẦN QUỐC CÔNG

Khoa Luật Quốc tế, Trường Đại học Luật TP. HCM
Faculty of International Law, Ho Chi Minh City University of Law

Email: ltqcong@hcmulaw.edu.vn

VÕ NGUYỄN QUỲNH NHƯ;⁽¹⁾ NGUYỄN THỊ THUY VÂN⁽²⁾

Trường Đại học Luật TP. HCM

Ho Chi Minh City University of Law

Email: vnnqnhu0401@gmail.com; nguyenthithuyvan5704@gmail.com

Tóm tắt

Trong bối cảnh các hệ thống trí tuệ nhân tạo (AI), quyền được giải thích về cách thức xử lý dữ liệu cá nhân của chủ thể dữ liệu có thể xung đột với nhu cầu bảo vệ bí mật kinh doanh khi thuật toán AI được xem là tài sản của doanh nghiệp. Bài viết gợi mở 03 phương hướng để hài hòa nghịch lý trên: (i) tích hợp tính minh bạch từ giai đoạn thiết kế, (ii) mô hình minh bạch có mục tiêu, và (iii) cơ chế trung gian kiểm soát. Từ kinh nghiệm pháp lý châu Âu, bài viết gợi mở những cách tiếp cận mới cho Việt Nam.

Từ khóa: bí mật kinh doanh, quyền của chủ thể dữ liệu, trí tuệ nhân tạo, thuật toán AI

Abstract

In the context of artificial intelligence (AI) systems, the data subject's right to an explanation of personal data processing may conflict with the need to protect trade secrets when AI algorithms are treated as corporate assets. This article explores this tension and proposes three approaches to reconcile it: (i) integrating transparency from the design stage, (ii) adopting targeted transparency models, and (iii) establishing intermediary control mechanisms. Drawing on European legal experience, the article offers new perspectives and potential regulatory approaches for Vietnam.

Keywords: trade secrets, data subject rights, artificial intelligence, AI algorithms

DOI: <https://doi.org/10.70236/khplvn.742>

Ngày nhận bài: 13/10/2025

Ngày duyệt đăng: 02/4/2026

Trong bối cảnh cạnh tranh kinh doanh khốc liệt, thuật toán AI có giá trị thương mại và được coi là tài sản vô giá của doanh nghiệp,¹ từ đó dẫn đến nhu cầu bảo hộ thuật toán AI như một bí mật kinh doanh nhằm bảo vệ lợi ích thương mại hợp pháp và khuyến khích đổi mới sáng tạo. Đối lập với nhu cầu bảo hộ bí mật kinh doanh là quyền được giải thích về cách thức xử lý dữ liệu cá nhân của chủ thể dữ liệu. Quyền này xuất phát từ nguyên tắc tự quyết thông tin, vốn là một quyền cơ bản được công nhận trong nhiều hệ thống pháp luật trên thế giới.² Đặc biệt, Liên minh châu Âu

1 Tanya Aplin, Alfred Radauer, Martin A. Bader, "The role of EU trade secrets law in the data economy: An empirical analysis", *IIC - International Review of Intellectual Property and Competition Law*, Vol. 54(6), 2023, tr. 826–858. DOI: <https://doi.org/10.1007/s40319-023-01325-8>; Lê Trần Quốc Công, "Quy định về tự do dữ liệu trong các hiệp định thương mại tự do thế hệ mới – tác động đối với pháp luật Việt Nam", *Tạp chí Khoa học Pháp lý Việt Nam*, số 07(146), 2021, tr. 94 – 95.

2 Sandra Wachter, Brent Mittelstadt, "A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI", *Columbia Business Law Review*, Vol. 2, 2019, tr. 5. DOI: <https://doi.org/10.7916/cblr.v2019i2.3424>

(European Union, EU) đã cụ thể hóa quyền này thông qua quyền không phải chịu quyết định tự động (Điều 22) và quyền nhận “thông tin có ý nghĩa về logic liên quan” (Điều 15(1)(h)). Có thể thấy, từ đây đặt ra nhu cầu hài hòa giữa một bên là bảo hộ bí mật kinh doanh của thuật toán AI (đại diện cho lợi ích thương mại và động lực đổi mới sáng tạo) và bên kia là quyền minh bạch thông tin về cách thức xử lý dữ liệu cá nhân và quyền tự quyết dữ liệu (đại diện cho quyền cơ bản của cá nhân). Nghịch lý này ngày càng phức tạp hơn khi AI tiếp tục phát triển và các thuật toán đã vượt xa khả năng giải thích thông thường.³

1. Bảo hộ bí mật kinh doanh và quyền minh bạch cách thức xử lý dữ liệu cá nhân

1.2. Thuật toán AI là bí mật kinh doanh

Theo Hiệp định về các khía cạnh liên quan đến thương mại của quyền sở hữu trí tuệ (*Trade-Related Aspects of Intellectual Property Rights*, TRIPS), bí mật kinh doanh là thông tin không được biết đến một cách phổ biến, có giá trị thương mại và đã được áp dụng các biện pháp hợp lý để giữ bí mật.⁴ Tại EU, Chỉ thị về Bảo vệ bí quyết kinh doanh 2016/943/EU đã tạo ra khuôn khổ pháp lý thống nhất cho việc bảo hộ đối tượng này. Theo đó, thuật toán AI được bảo hộ như bí mật kinh doanh khi thỏa mãn ba điều kiện: (i) bí mật, (ii) có giá trị thương mại do tính bí mật, và (iii) chủ sở hữu đã áp dụng các biện pháp hợp lý để giữ bí mật.⁵ Với đặc điểm của các mô hình học sâu là tính “hộp đen” (*black box*), chính người tạo ra thuật toán cũng khó có thể hiểu đầy đủ cách thức hệ thống ra quyết định cụ thể khi xử lý dữ liệu cá nhân⁶ và việc truy vết quá trình ra quyết định gần như bất khả thi. “Hộp đen” tạo ra thách thức kép: một mặt làm tăng giá trị của thuật toán như một bí mật kinh doanh; mặt khác, làm tăng lo ngại về tính minh bạch của cách thức xử lý dữ liệu cá nhân, đặc biệt khi thuật toán đưa ra những quyết định ảnh hưởng trực tiếp đến quyền và lợi ích của chủ thể dữ liệu.⁷

1.2. Quyền minh bạch cách thức xử lý dữ liệu cá nhân trong môi trường AI

Khi xử lý dữ liệu cá nhân, tính minh bạch và trách nhiệm giải trình có mối liên hệ chặt chẽ nhưng không đồng nhất vì ta không thể truy cứu trách nhiệm pháp lý đối với một quyết định xử lý dữ liệu cá nhân khi không hiểu

3 Agathe Balayn, Lorenzo Corti, Fanny Rancourt, Fabio Casati, Ujwal Gadiraju, “Understanding stakeholders’ perceptions and needs across the LLM supply chain”, *Cornell University Repository*, 2024; Matthias Artzt, Trần Việt Dũng, “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. 40.

4 Điều 39 Hiệp định về các khía cạnh liên quan đến thương mại của quyền sở hữu trí tuệ (TRIPS Agreement).

5 Điều 2 Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, [2016] OJ L157/1.

6 Jenna Burrell, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms”, *Big Data & Society*, Vol. 3(1), 2016, tr. 1-12. DOI: <https://doi.org/10.1177/2053951715622512>

7 Dennis Crouch, “Using intellectual property to regulate artificial intelligence”, *Missouri Law Review*, Vol. 89(3), 2024, tr. 836.

cách thức đưa ra quyết định đó.⁸ Đối với chủ thể dữ liệu, việc không có thông tin đầy đủ về cách thức dữ liệu của họ được xử lý cũng đồng nghĩa với việc cá nhân không thể thực hiện các quyền như quyền truy cập, quyền sửa đổi, quyền xóa dữ liệu và quyền phản đối việc xử lý hiệu quả.⁹

Sự minh bạch của cách thức xử lý dữ liệu cá nhân đòi hỏi sự cởi mở, công khai, trong khi đó, tính bảo mật của thuật toán AI lại là sự hạn chế, kiểm soát và bảo vệ thông tin. Hay nói một cách khác, trong khi tính minh bạch đảm bảo lợi thế của chủ thể dữ liệu dưới góc độ quyền riêng tư, thì bảo hộ bí mật kinh doanh duy trì lợi thế cho chủ sở hữu bí mật ở góc độ lợi ích vật chất. Điều này đặt ra thách thức trong việc thiết kế khung pháp lý về chỉnh AI.¹⁰ Khung pháp lý của EU cũng phản ánh xung đột này. Quy định chung về bảo vệ dữ liệu (*General Data Protection Regulation*, GDPR) đòi hỏi tính minh bạch trong xử lý dữ liệu cá nhân, trong khi chỉ thị về bảo vệ bí quyết kinh doanh 2016/943/EU lại bảo vệ quyền giữ bí mật thông tin mang giá trị kinh tế. Sự tồn tại song song của hai khung pháp lý này đặt ra thách thức lớn khi xác định ranh giới giữa công khai và bảo mật thông tin về cách thức xử lý dữ liệu.¹¹

2. Bảo hộ bí mật kinh doanh là thuật toán AI và minh bạch xử lý dữ liệu cá nhân theo pháp luật EU

2.1. Bảo hộ bí mật kinh doanh tại EU

Chỉ thị về Bảo vệ bí quyết kinh doanh 2016/943/EU của EU là văn bản quan trọng cung cấp sự bảo hộ pháp lý chống lại việc thu thập, sử dụng hoặc tiết lộ trái phép bí mật kinh doanh. Theo Điều 4, thu thập bí mật kinh doanh mà không có sự đồng ý của chủ sở hữu bí mật được coi là bất hợp pháp. Điều này xảy ra khi tiếp cận, chiếm đoạt hoặc sao chép trái phép tài liệu, đồ vật, vật liệu, chất liệu hoặc tập tin điện tử có chứa bí mật kinh doanh hoặc từ đó có thể suy ra bí mật kinh doanh. Chỉ thị cũng thừa nhận một số ngoại lệ nhằm đảm bảo lợi ích công có thể đòi hỏi việc tiết lộ thông tin bí mật. Như vậy, dựa trên tinh thần của Điều 5(d) Chỉ thị này, trong trường hợp việc công bố cách thức xử lý dữ liệu cá nhân là cần thiết để bảo vệ lợi ích công, như phòng ngừa phân biệt đối xử, đảm bảo công bằng khi ứng dụng AI để xử lý dữ liệu cá nhân thì lúc này thuật toán AI không còn có thể nằm trong vỏ bọc của “bí mật kinh doanh”.

Bên cạnh đó, Đạo luật AI của EU cũng thừa nhận việc bảo hộ bí mật kinh doanh liên quan đến AI. Điều 53(7) của Đạo luật quy định rằng khi thực hiện quyền hạn của mình, các cơ quan quản lý AI phải xem xét đầy đủ

8 IEEE, “Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (version 2)”, 2019, tr. 29.

9 Sandra Wachter, Brent Mittelstadt, *tlđđ*, 2019, tr. 38.

10 Kristin Foss-Solbrekk, “Searchlights across the black box: Trade secrecy versus access to information”, *Computer Law & Security Review*, Vol. 50, 2023. DOI: <https://doi.org/10.1016/j.clsr.2023.105811>

11 Perry Keller, Tanya F. Aplin, “Reconciling trade secrets and AI public transparency”, *Cambridge Law Journal*, forthcoming, 2026, tr. 2 – 3. DOI: <http://dx.doi.org/10.2139/ssrn.5044278>

quyền và lợi ích của nhà cung cấp thông tin, bao gồm cả việc bảo vệ thông tin bí mật như bí mật kinh doanh.¹² Theo cách tiếp cận của pháp luật EU, kể cả việc minh bạch cũng không được gây phương hại một cách không cân xứng đến quyền sở hữu trí tuệ và bí mật kinh doanh.

2.2. Quy định về minh bạch xử lý dữ liệu cá nhân

Điều 22 của GDPR trao cho chủ thể dữ liệu quyền không bị ràng buộc bởi quyết định hoàn toàn dựa trên xử lý tự động. Quy định này áp dụng trực tiếp đối với các hệ thống AI sử dụng phương thức ra quyết định tự động dựa trên dữ liệu cá nhân. Tuy nhiên, ba ngoại lệ mà quyết định tự động vẫn được cho phép là: (a) khi cần thiết để thực hiện hoặc ký kết hợp đồng, (b) khi được luật pháp Liên minh hoặc quốc gia cho phép, hoặc (c) khi dựa trên sự đồng ý rõ ràng của chủ thể dữ liệu. Trong những trường hợp này, người kiểm soát dữ liệu phải thực hiện “các biện pháp thích hợp để bảo vệ quyền, tự do và lợi ích hợp pháp của chủ thể dữ liệu”, bao gồm “ít nhất quyền có sự can thiệp của con người từ phía người kiểm soát” (Điều 22(3)). Theo đó, GDPR yêu cầu người kiểm soát dữ liệu cung cấp cho chủ thể dữ liệu “thông tin có ý nghĩa về logic liên quan, cũng như ý nghĩa và hậu quả dự kiến của việc xử lý” trong trường hợp ra quyết định tự động.¹³

Để làm rõ vấn đề này, Hội đồng Bảo vệ dữ liệu châu Âu (*European Data Protection Board*, EDPB) đã ban hành Hướng dẫn diễn giải Điều 22 và Điều 15(1)(h). Cụ thể, “thông tin có ý nghĩa về logic liên quan” nên bao gồm giải thích về lý do đưa ra quyết định xử lý dữ liệu và các tiêu chí được sử dụng, nhưng không cần cung cấp giải thích phức tạp về thuật toán hoặc tiết lộ toàn bộ thuật toán.¹⁴ Tương tự, ý kiến tư vấn của Tổng biện lý (Advocate General) Richard de la Tour về vụ *CK v Dun & Bradstreet Austria* (Case C-203/22) đã chỉ ra rằng giới hạn của quyền được giải thích là không nên mở rộng đến thông tin kỹ thuật, như thuật toán mà chủ thể dữ liệu không thể hiểu được nếu không có chuyên môn “cụ thể” và “thông thường” không nên buộc phải tiết lộ thông tin được bảo vệ như bí mật kinh doanh.¹⁵

Như vậy, tính minh bạch bao gồm mọi hình thức tiết lộ thông tin về cách thức hệ thống hoạt động. Điều này lệ thuộc vào việc liệu rằng có thể

12 Điều 53(7) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (EU AI Act), [2024] OJ L1689/1.

13 Điều 15(1)(h) GDPR.

14 EDPB, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*, European Data Protection Board, 2017.

15 Case C-203/22 *CK v Dun & Bradstreet Austria GmbH*, Opinion of AG Richard De La Tour, EU:C:2024:738, tr. 80. Cụ thể là một cá nhân (được gọi là CK) đã bị một nhà mạng viễn thông tại Áo từ chối ký hợp đồng cung cấp dịch vụ điện thoại di động với lý do không đủ khả năng tín dụng. Quyết định này hoàn toàn dựa trên hệ thống chấm điểm tín dụng tự động (*automated credit scoring*) do công ty Dun & Bradstreet Austria (D&B) cung cấp. Dựa vào Điều 15(1)(h) của GDPR (Quyền truy cập của chủ thể dữ liệu đối với quy trình ra quyết định tự động), CK đã yêu cầu D&B cung cấp “thông tin có ý nghĩa về logic liên quan” (*meaningful information about the logic involved*) để hiểu tại sao mình bị đánh giá thấp. Tuy nhiên, D&B từ chối cung cấp thông tin chi tiết về các thông số và thuật toán chấm điểm, lập luận rằng những thông tin này là bí mật kinh doanh (*trade secrets*) được pháp luật bảo vệ. Sự việc được đưa ra tòa án Áo và sau đó được đệ trình lên CJEU để xin giải thích sơ bộ.

cung cấp lời giải thích có thể hiểu được cho quyết định xử lý dữ liệu của hệ thống AI hay không.¹⁶ Chính vì thế, chủ thể dữ liệu cũng sẽ được quyền nhận được lời giải thích về các quyết định tự động khi dữ liệu của họ được xử lý.¹⁷ Một phương án khả thi được ủng hộ là tập trung vào “khả năng biện minh” thay vì mô tả chính xác cơ chế hoạt động của thuật toán, cụ thể hơn là giải thích tại sao các tiêu chí ra quyết định xử lý dữ liệu được coi là hợp lý và phù hợp cho mục đích xử lý.¹⁸

2.3. Các xu hướng nhằm hài hòa giữa bảo hộ bí mật kinh doanh là thuật toán AI và tính minh bạch cách thức xử lý dữ liệu cá nhân

2.3.1. Xu hướng thứ nhất: Tích hợp tính minh bạch từ giai đoạn thiết kế (transparency by design)

Transparency by design (TbD) là phương pháp tiếp cận tích hợp tính minh bạch vào quá trình thiết kế hệ thống AI ngay từ đầu. Mô hình TbD được đề xuất bởi Felzmann và đồng nghiệp, bao gồm ba giai đoạn chính:¹⁹

(1) Giai đoạn thiết kế AI: Đặt ra các yêu cầu thiết kế chung để tăng cường tính minh bạch khi phát triển hệ thống mới.

(2) Giai đoạn thông tin về xử lý và phân tích dữ liệu: Cung cấp thông tin giúp quá trình xử lý dữ liệu, các thủ tục ra quyết định và rủi ro trở nên minh bạch hơn.

(3) Giai đoạn trách nhiệm giải trình: Các khía cạnh minh bạch theo định hướng tổ chức và liên quan đến các bên liên quan về khả năng kiểm tra, phản hồi và báo cáo thường xuyên.

Trong khi quyền riêng tư từ giai đoạn thiết kế (*Privacy by Design*, PbD) đã được luật hóa trong GDPR như một nghĩa vụ pháp lý tại Điều 25, yêu cầu người kiểm soát dữ liệu thực hiện “các biện pháp kỹ thuật và tổ chức thích hợp” để đảm bảo bảo vệ dữ liệu “theo thiết kế và theo mặc định”, TbD vẫn chưa được công nhận chính thức trong luật pháp của EU, mặc dù khái niệm này đang dần được ghi nhận trong các cuộc thảo luận chính sách.²⁰ Cả PbD

16 Sandra Wachter, Brent Mittelstadt, Chris Russell, “Counterfactual explanations without opening the black box: Automated decisions and the GDPR”, *Harvard Journal of Law & Technology*, Vol. 31(2), 2018, tr. 841-887. DOI: <http://dx.doi.org/10.2139/ssrn.3063289>

17 Lilian Edwards, Michael Veale, “Enslaving the algorithm: From a ‘right to an explanation’ to a ‘right to better decisions?’”, *IEEE Security & Privacy*, Vol. 16(3), 2018, tr. 46-54. DOI: <https://doi.org/10.1109/MSP.2018.2701152>

18 Quyền này đòi hỏi bên kiểm soát dữ liệu phải đưa ra biện minh (*justification*) từ trước (*ex-ante*) để thiết lập xem liệu một suy luận có hợp lý (*reasonable*) hay không: (i) Tại sao dữ liệu đầu vào là một cơ sở được chấp nhận về mặt chuẩn mực để đưa ra suy luận, (ii) Tại sao các suy luận này liên quan với nhau và được chấp nhận về mặt chuẩn mực cho mục đích xử lý đã chọn hoặc loại quyết định tự động, (iii) Liệu dữ liệu và các phương pháp được sử dụng để đưa ra suy luận có chính xác và đáng tin cậy về mặt thống kê hay không. Xem Nicolas P. Suzor, Sarah M. West, Andrew Quodling, Jillian York, “What do we mean by ‘accessibility’? A framework for understanding technology and rights”, *Journal of Information Policy*, Vol. 9, 2019.

19 Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz & Aurelia Tamò-Larrieux, “Towards transparency by design for artificial intelligence”, *Science and Engineering Ethics*, Vol. 26(6), 2020, tr. 3333-3361. DOI: <https://doi.org/10.1007/s11948-020-00276-4>

20 Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz & Aurelia Tamò-Larrieux, *tldd*, 2020, tr. 3343.

và TbD đều nhấn mạnh tầm quan trọng của việc tích hợp các giá trị từ giai đoạn thiết kế, thay vì coi chúng là các tính năng được thêm vào sau.²¹ Tuy nhiên, PbD tập trung vào việc hạn chế các nguy cơ liên quan đến việc xử lý dữ liệu cá nhân, trong khi TbD lại nhằm cung cấp thông tin nhiều hơn về cách dữ liệu được xử lý.

2.3.2. Xu hướng thứ hai: Mô hình minh bạch có mục tiêu (*targeted transparency*)

Minh bạch có mục tiêu là xu hướng phát triển từ mô hình minh bạch tổng quát truyền thống, như quyền tự do thông tin, sang các quyền minh bạch được thiết kế riêng cho các mục đích, quy trình và các bên liên quan cụ thể. Thay vì áp dụng một giải pháp cho tất cả các vấn đề, minh bạch có mục tiêu điều chỉnh yêu cầu minh bạch để đáp ứng nhu cầu thông tin về cách thức xử lý dữ liệu của các đối tượng cụ thể.²² Mô hình có thể cung cấp các loại giải thích khác nhau về cách thức xử lý dữ liệu cho các bên liên quan khác nhau: giải thích kỹ thuật chi tiết cho cơ quan quản lý, giải thích về tác động cá nhân cho chủ thể dữ liệu.²³

Đạo luật Dữ liệu của EU (*Data Act*) cung cấp một ví dụ rõ ràng về minh bạch có mục tiêu. Theo Điều 4(6) và 5(9), bất kỳ dữ liệu sử dụng nào được bảo vệ bởi yêu cầu bí mật kinh doanh không cần phải tiết lộ, trừ khi việc đó “hoàn toàn cần thiết” cho mục đích của Đạo luật Dữ liệu và bên nhận đồng ý với bất kỳ biện pháp nào cần thiết để bảo vệ tính bảo mật. Cách tiếp cận này cho người dùng đủ điều kiện vào trong rào cản bảo mật bí mật kinh doanh, áp dụng nghĩa vụ không tiết lộ để ngăn chặn việc chia sẻ trái phép.²⁴ Ngoài ra, Điều 40 của Đạo luật Dịch vụ số của EU tạo ra giải pháp khác là cung cấp quyền truy cập vào dữ liệu do các nền tảng trực tuyến rất lớn hoặc công cụ tìm kiếm trực tuyến rất lớn nắm giữ cho các nhà nghiên cứu được xác minh.²⁵

2.3.3. Xu hướng thứ ba: Cơ chế trung gian kiểm soát

Thay vì yêu cầu tiết lộ trực tiếp bí mật kinh doanh cho công chúng, cách tiếp cận này dựa vào các trung gian đáng tin cậy để đánh giá thuật toán và cung cấp một mức độ giám sát, trách nhiệm giải trình.²⁶ Cơ quan quản lý như cơ quan bảo vệ dữ liệu, cơ quan cạnh tranh hoặc cơ quan giám sát AI chuyên biệt đóng vai trò quan trọng trong việc kiểm tra và giám sát các hệ thống AI mà không công khai thuật toán. Đạo luật AI của EU cũng giới thiệu “cơ quan đánh giá sự phù hợp” để kiểm tra sự tuân thủ các yêu cầu của Đạo luật (Điều 33) của các hệ thống AI rủi ro cao.

21 Olya Kudina, Peter-Paul Verbeek, “Ethics from within: Google glass, the collingridge dilemma, and the mediated value of privacy”, *Science, Technology, & Human Values*, Vol. 44(2), 2019, tr. 291-314. DOI: <https://doi.org/10.1177/0162243918793711>

22 Perry Keller, Tanya F. Aplin, *tldd*, 2024, tr. 23-25.

23 Tal Z. Zarsky, “Transparent predictions”, *University of Illinois Law Review*, Vol. 2013(4), 2013, tr. 1503-1570.

24 Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz & Aurelia Tamò-Larrieux, *tldd*, 2020, tr. 3347.

25 Điều 40 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act), [2022] OJ L277/1.

26 Perry Keller, Tanya F. Aplin, *tldd*, 2024, tr. 32-35.

Tuy nhiên, cần lưu ý là không có giải pháp nào là trọn vẹn, mỗi hướng tiếp cận sẽ có những điểm mạnh và điểm yếu riêng. Những vấn đề này sẽ được thể hiện thông qua bảng so sánh sau:

Tiêu chí	Tính minh bạch từ giai đoạn thiết kế	Minh bạch có mục tiêu	Cơ chế trung gian kiểm soát
Điểm mạnh	- Tích hợp tính minh bạch ngay từ đầu - Giảm chi phí và khó khăn kỹ thuật về sau - Thừa nhận tính đa dạng của minh bạch	- Tùy chỉnh theo nhu cầu các bên liên quan - Giảm tiết lộ thông tin không cần thiết - Trình bày thông tin phù hợp từng nhóm	- Giải quyết xung đột giữa minh bạch và bảo mật - Đảm bảo độc lập thông qua bên thứ ba - Tận dụng chuyên môn kỹ thuật
Điểm yếu	- Thiếu hướng dẫn và tiêu chuẩn cụ thể - Yêu cầu thay đổi lớn trong thực tiễn hiện tại - Chưa khắc phục vấn đề “hộp đen”	- Gây bất bình đẳng trong quyền tiếp cận thông tin - Phân loại/ phân cấp thông tin phức tạp - Không đáp ứng minh bạch công cộng	- Phụ thuộc vào sự tin tưởng trung gian - Tạo điểm tập trung dễ bị lạm dụng - Có thể chưa đáp ứng yêu cầu minh bạch toàn diện
Điều kiện áp dụng	Cần cải cách quy trình từ sớm, ngay khi thiết kế thuật toán AI.	Dễ áp dụng theo từng tình huống cụ thể	Cần xây dựng lòng tin của công chúng và khung pháp lý rõ ràng
Mức độ đáp ứng kỳ vọng công chúng	Có tiềm năng cao nhưng còn hạn chế với mô hình phức tạp	Tùy vào đối tượng được cung cấp thông tin mà tính minh bạch có độ rộng hẹp khác nhau	Minh bạch mang tính gián tiếp, phụ thuộc bên thứ ba

Từ những kinh nghiệm ở EU, tác giả cho rằng không có một giải pháp duy nhất có thể giải quyết hoàn toàn nghịch lý giữa bảo hộ bí mật kinh doanh và quyền minh bạch về cách thức xử lý dữ liệu cá nhân mà thay vào đó gợi ý cách tiếp cận đa chiều và kết hợp.

3. Gợi mở cho Việt Nam

3.1. Hiện trạng pháp luật Việt Nam

Khung pháp lý Việt Nam về bảo hộ bí mật kinh doanh nhìn chung phù hợp với các tiêu chuẩn quốc tế theo Hiệp định TRIPS và thông lệ quốc tế,²⁷ trong đó Luật Sở hữu trí tuệ năm 2005 (sửa đổi, bổ sung năm 2009, 2019 và 2022) quy định bí mật kinh doanh được bảo hộ khi đáp ứng ba điều kiện tại Điều 85. Tuy nhiên, vấn đề đặt ra không phải thuật toán AI có được bảo hộ như bí mật kinh doanh không²⁸ mà ở chỗ các ngoại lệ cho phép sử dụng hoặc tiết lộ bí mật kinh doanh có dung hòa được lợi ích giữa chủ sở hữu và quyền của chủ thể dữ liệu hay không?

27 Trương Thị Thanh Tuyết, *Bảo hộ quyền sở hữu công nghiệp đối với bí mật kinh doanh theo pháp luật Việt Nam*, Đại học Quốc gia Hà Nội, Hà Nội, 2011, tr. 105.

28 Thuật toán AI hoàn toàn có thể được bảo hộ như là bí mật kinh doanh đã được phân tích trong các phần trước.

Ở khía cạnh bảo vệ dữ liệu cá nhân, khung pháp lý hiện hành được quy định chủ yếu Luật Bảo vệ Dữ liệu cá nhân năm 2025, bắt đầu có hiệu lực từ ngày 01/01/2026.²⁹ Quyền được biết về hoạt động xử lý dữ liệu cá nhân của chủ thể dữ liệu đã được ghi nhận, đồng thời tại Điều 30 cũng đã chỉ ra những nguyên tắc xử lý dữ liệu cá nhân bằng AI.³⁰ Tuy nhiên, Luật này cũng chưa quy định một cách trực tiếp về tính minh bạch của cách thức xử lý dữ liệu của thuật toán và về khả năng tiếp cận thông tin, giải trình liên quan đến cách thức xử lý thông tin của AI. Đáng chú ý, sự ra đời của Luật Trí tuệ nhân tạo năm 2025 đã ghi nhận việc thực hiện trách nhiệm giải trình về các quyết định và hệ quả của hệ thống AI như là một nguyên tắc cơ bản trong hoạt động AI. Vấn đề là nguyên tắc này sẽ được thực thi như thế nào cũng cần được làm rõ và hướng dẫn chi tiết.

Nhằm hướng dẫn cụ thể hơn quy trình bảo đảm quyền được biết của chủ thể dữ liệu, Điều 5 Nghị định số 356/2025/NĐ-CP đã đặt ra một số yêu cầu đối với bên kiểm soát, xử lý dữ liệu. Theo đó, ngay từ khâu thiết kế, bên kiểm soát, xử lý dữ liệu cá nhân phải tích hợp tính minh bạch, rõ ràng vào quy trình, thủ tục, nhằm bảo đảm quyền được biết của chủ thể dữ liệu cá nhân. Trong các trường hợp chủ thể dữ liệu yêu cầu rút lại sự đồng ý cho phép xử lý dữ liệu cá nhân, được xem, sửa hoặc xóa dữ liệu cá nhân thì bên kiểm soát và quản lý dữ liệu cá nhân phải công khai thông tin và xử lý theo quy định của pháp luật. Bên cạnh đó, vấn đề bảo vệ dữ liệu cá nhân trong hệ thống AI cũng được đề cập. Cụ thể, Điều 10 đặt ra việc thiết lập một cơ chế giám sát hoạt động của hệ thống AI dựa trên hai phương diện: (i) giám sát của cơ quan nhà nước có thẩm quyền và (ii) trách nhiệm giải trình của bên kiểm soát dữ liệu cá nhân, bên kiểm soát và xử lý dữ liệu cá nhân. Chính vì thế, các tổ chức, cá nhân áp dụng các biện pháp bảo vệ dữ liệu cá nhân trong hệ thống AI phải thực hiện đánh giá tuân thủ các quy định về bảo vệ dữ liệu cá nhân định kỳ 01 năm/lần.

Với Luật Trí tuệ nhân tạo năm 2025, để đáp ứng các yêu cầu về tính minh bạch của hệ thống AI nhằm đảm bảo khả năng giải thích thuật toán, nhà cung cấp phải bảo đảm AI tương tác trực tiếp với con người được thiết kế và vận hành để người sử dụng nhận biết khi đang tương tác với hệ thống.³¹ Đặc biệt, Nhà nước khuyến khích các tổ chức, cá nhân áp dụng Khung đạo đức trí tuệ nhân tạo quốc gia, trong đó nhấn mạnh yêu cầu về khả năng giải

29 Trước đây là Nghị định số 13/2023, có thể xem thêm Trần Việt Dũng, Lê Trần Quốc Công, “Trách nhiệm của chủ thể quản lý nền tảng kỹ thuật số (platforms) – Kinh nghiệm của EU”, trong *Kỷ yếu Hội thảo quốc tế Bảo vệ dữ liệu cá nhân trong môi trường số*, Nxb. Đại học Quốc gia TP. Hồ Chí Minh, 2023; Lê Trần Quốc Công, Nguyễn Đào Phương Thúy, “The impact of the European Union’s personal data protection ‘standards’ on the evolution of laws in ASEAN countries”, *Proceedings of Chiang Mai International Conference 2024*, Chiang Mai University, 2024.

30 Lê Công Trần Quốc, Trần Việt Dũng và Nguyễn Đào Phương Thúy, “Global norms and regional innovation: GDPR, evolution of data protection in ASEAN and the legal trajectory of AI in Vietnam”, *TalTech Journal of European Studies*, Vol. 15(3), 2025, tr. 273-278. DOI: <https://doi.org/10.2478/bjes-2025-0039>

31 Khoản 1 Điều 11 Luật Trí tuệ nhân tạo năm 2025.

thích và trách nhiệm giải trình.³² Để hiện thực hóa quy định này, các chủ thể phát triển và cung cấp hệ thống AI cần xác định rõ các tác động mà hệ thống có thể gây ra, đồng thời chuẩn bị tài liệu giải thích và các bằng chứng liên quan đến quá trình thiết kế, huấn luyện và kiểm thử của hệ thống.³³ Tuy nhiên, cần lưu ý là Khung đạo đức trí tuệ nhân tạo quốc gia hiện chỉ mang tính chất lựa chọn mà không phải là quy định bắt buộc phải tuân thủ.

Có thể thấy, kể cả pháp luật về bảo vệ dữ liệu cá nhân và Luật Trí tuệ nhân tạo 2025 đã bước đầu đặt ra một số quy định nền tảng về tính minh bạch trong xử lý dữ liệu cá nhân bởi AI, nhưng hoàn toàn bối rối trước việc bảo hộ thuật toán AI như một bí mật kinh doanh dưới góc độ pháp luật sở hữu trí tuệ. Thực vậy, trong số năm ngoại lệ mà bí mật kinh doanh được sử dụng, tiết lộ không bị xem là hành vi bị cấm tại khoản 3 Điều 125, các trường hợp tại điểm a, d và đ hầu như không gây xung đột khi áp dụng với thuật toán AI do người dùng thường không đủ khả năng để khai thác, phân tích bí mật hoặc tái tạo thuật toán, và nếu có thì chính họ đã hiểu cách thức xử lý dữ liệu nên không còn nhu cầu tiếp cận. Tuy vậy, ngoại lệ tại điểm b và c liên quan đến bảo vệ công chúng và không nhằm mục đích thương mại đặt ra vấn đề là liệu việc chủ thể dữ liệu yêu cầu được biết về thuật toán vì cho rằng thuật toán xâm phạm quyền riêng tư hoặc gây phân biệt đối xử có được coi là vì lợi ích công chúng hay không? Đồng thời, dù việc giải trình thuật toán không mang tính thương mại, nhưng việc cho phép mọi chủ thể dữ liệu tiếp cận các thông tin này cũng có thể làm mất đi lợi thế kinh tế của chủ sở hữu bí mật kinh doanh. Chưa kể, việc bảo hộ thuật toán AI dưới dạng một sáng chế³⁴ tại Việt Nam có thể khiến các doanh nghiệp ngần ngại do lo ngại chưa đáp ứng điều kiện đăng ký bảo hộ, thay vào đó là nắm giữ những tài sản trí tuệ này trong vòng bảo mật dưới dạng bí mật kinh doanh.³⁵ Do đó, cần cách tiếp cận linh hoạt hơn để dung hòa hai lợi ích đối lập này.

3.2. Gợi mở hướng tiếp cận mới cho Việt Nam

Để giải quyết nghịch lý giữa bảo hộ bí mật kinh doanh và quyền minh bạch về cách thức xử lý dữ liệu cá nhân, Việt Nam có thể xem xét áp dụng mô hình tích hợp cả ba xu hướng tại EU đã thảo luận trước đó với điều kiện cụ thể của Việt Nam.

Giai đoạn 1: Xây dựng khung pháp lý cho tích hợp tính minh bạch từ giai đoạn thiết kế - Xây dựng các nguyên tắc pháp lý về tích hợp tính minh bạch từ giai đoạn thiết kế đối với các hệ thống AI, tương tự như cách GDPR đã luật hóa quyền riêng tư từ giai đoạn thiết kế, và có thể được đưa vào các văn bản hướng dẫn thực thi Luật Bảo vệ dữ liệu cá nhân nhằm quy định chuyên biệt về AI.

³² Khoản 4 Điều 26 Luật Trí tuệ nhân tạo năm 2025.

³³ Điểm d khoản 2 Điều 3 Thông tư số 05/2026/TT-BKHCN Ban hành Khung đạo đức trí tuệ nhân tạo quốc gia.

³⁴ Xem thêm khả năng bảo hộ AI như một sáng chế trong Lưu Minh Sang, Trần Đức Thành, “Trí tuệ nhân tạo và những thách thức pháp lý”, *Tạp chí Khoa học và Công nghệ Việt Nam*, số 08, 2020, tr. 21-22..

³⁵ Nguyễn Hữu Cẩn, “Thực trạng phát triển tài sản trí tuệ của doanh nghiệp Việt Nam và một số gợi ý chiến lược”, *Tạp chí Khoa học và Công nghệ Việt Nam*, số 12, 2017, tr. 65.

Giai đoạn 2: Áp dụng tiếp cận minh bạch có mục tiêu – Thay vì áp dụng các yêu cầu minh bạch đồng nhất cho tất cả các hệ thống AI, có thể phân loại hệ thống AI dựa trên mức độ rủi ro và áp dụng các yêu cầu minh bạch khác nhau tùy theo từng cấp độ rủi ro. Dưới góc độ Luật Bảo vệ dữ liệu cá nhân, cần quy định về quyền của chủ thể dữ liệu không bị ràng buộc bởi quyết định của hệ thống tự động,³⁶ gồm các ngoại lệ hạn chế và yêu cầu biện pháp bảo vệ phù hợp, như quyền xử lý dữ liệu phải có can thiệp của con người và quyền phản đối quyết định.

Giai đoạn 3: Phát triển cơ chế trung gian kiểm soát – Khác với cách tiếp cận của châu Âu, đề xuất của nhóm tác giả về cơ chế trung gian kiểm soát riêng về bí mật kinh doanh liên quan đến thuật toán AI không phải là hoạt động của cơ quan có thẩm quyền về bảo vệ dữ liệu mà là nhiệm vụ của các cơ quan đang thực thi các nhiệm vụ chuyên trách nhằm bảo hộ các đối tượng của quyền sở hữu trí tuệ vì chỉ có các tổ chức này mới đủ điều kiện để thực hiện kiểm toán thuật toán và đánh giá tính tuân thủ của các hệ thống AI trong việc xử lý dữ liệu nói chung và về bản chất, thuật toán AI nguồn gốc vẫn là sản phẩm được tạo ra bởi trí tuệ con người. Để làm rõ hơn về ý tưởng này tại Việt Nam, nhóm tác giả đề xuất quy định nhiệm vụ liên quan đến việc kiểm soát, giải trình hay giải thích cách vận hành thuật toán AI trong quá trình xử lý dữ liệu (bất kể dữ liệu cá nhân hay dữ liệu phi cá nhân) mà doanh nghiệp xem là bí mật kinh doanh là một hoạt động thuộc thẩm quyền của Cục Sở hữu trí tuệ. ●

Tài liệu tham khảo

- [1] Tanya Aplin, Alfred Radauer, Martin A. Bader, “The role of EU trade secrets law in the data economy: An empirical analysis”, *IIC - International Review of Intellectual Property and Competition Law*, Vol. 54(6), 2023. DOI: <https://doi.org/10.1007/s40319-023-01325-8>
- [2] Matthias Artzt, Trần Việt Dũng, “Artificial intelligence and data protection: How to reconcile both areas from the European Law Perspective”, *Vietnamese Journal of Legal Sciences*, Vol. 7(2), 2022
- [3] Agathe Balayn, Lorenzo Corti, Fanny Rancourt, Fabio Casati, Ujwal Gadiraju, “Understanding stakeholders’ perceptions and needs across the LLM supply chain”, *Cornell University Repository*, 2024. DOI: <https://doi.org/10.48550/arXiv.2403.11718>
- [4] Jenna Burrell, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms”, *Big Data & Society*, Vol. 3(1), 2016. DOI: <https://doi.org/10.1177/2053951715622512>
- [5] Nguyễn Hữu Cẩn, “Thực trạng phát triển tài sản trí tuệ của doanh nghiệp Việt Nam và một số gợi ý chiến lược”, *Tạp chí Khoa học và Công nghệ Việt Nam*, số 12, 2017 [trans: Nguyen Huu Can, “Current state of intellectual property asset development in Vietnamese enterprises and strategic recommendations”, *Vietnam Journal of Science and Technology*, No. 12, 2017]
- [6] Lê Trần Quốc Công, “Quy định về tự do dữ liệu trong các hiệp định thương mại tự do thế hệ mới – tác động đối với pháp luật Việt Nam”, *Tạp chí Khoa học Pháp lý Việt Nam*, số 07(146), 2021 [trans: Le Tran Quoc Cong, “Provisions on data freedom in new-generation free trade agreements – impacts on Vietnamese law”, *Vietnamese Journal of Legal Sciences*, Issue 07(146), 2021]

36 Tương tự Điều 22 GDPR.

- [7] Lê Trần Quốc Công, Nguyễn Đào Phương Thúy, “The impact of the European Union’s personal data protection ‘standards’ on the evolution of laws in ASEAN countries”, *Proceedings of Chiang Mai International Conference 2024*, Chiang Mai University, Chiang Mai, 2024
- [8] Lê Công Trần Quốc, Trần Việt Dũng và Nguyễn Đào Phương Thúy., “Global norms and regional innovation: GDPR, evolution of data protection in ASEAN and the legal trajectory of AI in Vietnam”, *TalTech Journal of European Studies*, Vol. 15(3), 2025. DOI: <https://doi.org/10.2478/bjes-2025-0039>
- [9] Dennis Crouch, “Using intellectual property to regulate artificial intelligence”, *Missouri Law Review*, Vol. 89(3), 2024
- [10] Trần Việt Dũng, Lê Trần Quốc Công, “Trách nhiệm của chủ thể quản lý nền tảng kỹ thuật số (Platforms) – Kinh nghiệm của EU”, trong *Kỷ yếu Hội thảo quốc tế Bảo vệ dữ liệu cá nhân trong môi trường số*, Nxb. Đại học Quốc gia TP. Hồ Chí Minh, TP. Hồ Chí Minh, 2023 [trans: Tran Viet Dung, Le Tran Quoc Cong, “Liability of digital platform managers – EU experience”, in *International Conference Proceedings on Personal Data Protection in the Digital Environment*, Vietnam National University Ho Chi Minh City Press, Ho Chi Minh City, 2023]
- [11] Lilian Edwards, Michael Veale, “Enslaving the algorithm: From a ‘right to an explanation’ to a ‘right to better decisions?’”, *IEEE Security & Privacy*, Vol. 16(3), 2018. DOI: <https://doi.org/10.1109/MSP.2018.2701152>
- [12] Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz, Aurelia Tamò-Larrieux, “Towards transparency by design for artificial intelligence”, *Science and Engineering Ethics*, Vol. 26(6), 2020. DOI: <https://doi.org/10.1007/s11948-020-00276-4>
- [13] Kristin Foss-Solbrekk, “Searchlights across the black box: Trade secrecy versus access to information”, *Computer Law & Security Review*, Vol. 50, 2023. DOI: <https://doi.org/10.1016/j.clsr.2023.105811>
- [14] IEEE, *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*, 1st edn, IEEE, 2019
- [15] Perry Keller, Tanya F. Aplin, “Reconciling trade secrets and AI public transparency”, *Cambridge Law Journal*, forthcoming, 2026. DOI: <http://dx.doi.org/10.2139/ssrn.5044278>
- [16] Olya Kudina, Peter-Paul Verbeek, “Ethics from within: Google glass, the collingridge dilemma, and the mediated value of privacy”, *Science, Technology, & Human Values*, Vol. 44(2), 2019. DOI: <https://doi.org/10.1177/0162243918793711>
- [17] Lưu Minh Sang, Trần Đức Thành, “Trí tuệ nhân tạo và những thách thức pháp lý”, *Tạp chí Khoa học và Công nghệ Việt Nam*, số 08, 2020 [trans: Luu Minh Sang, Tran Duc Thanh, “Artificial intelligence and legal challenges”, *Vietnam Journal of Science and Technology*, No. 08, 2020]
- [18] Nicolas P. Suzor, Sarah M. West, Andrew Quodling, Jillian York, “What do we mean by ‘accessibility’? A framework for understanding technology and rights”, *Journal of Information Policy*, Vol. 9, 2019
- [19] Trương Thị Thanh Tuyết, *Bảo hộ quyền sở hữu công nghiệp đối với bí mật kinh doanh theo pháp luật Việt Nam*, Luận văn thạc sĩ luật học, Đại học Quốc gia Hà Nội, Hà Nội, 2011 [trans: Trương Thị Thanh Tuyet, *Protection of industrial property rights for trade secrets under Vietnamese law*, Master’s thesis in law, Vietnam National University, Hanoi, 2011]
- [20] Sandra Wachter, Brent Mittelstadt, “A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI”, *Columbia Business Law Review*, Vol. 2019(2), 2019. DOI: <https://doi.org/10.7916/cblr.v2019i2.3424>
- [21] Sandra Wachter, Brent Mittelstadt, Chris Russell, “Counterfactual explanations without opening the black box: Automated decisions and the GDPR”, *Harvard Journal of Law & Technology*, Vol. 31(2), 2018. DOI: <https://doi.org/10.2139/ssrn.3063289>
- [22] Tal Z. Zarsky, “Transparent predictions”, *University of Illinois Law Review*, Vol. 2013(4), 2013