Khung quản lý rủi ro trí tuệ nhân tạo trong các tổ chức tài chính- ngân hàng: Đề xuất mô hình và hướng tiếp cận

Triệu Thu Hương
Học viện Ngân hàng, Việt Nam

Ngày nhận: 24/07/2025 Ngày nhận bản sửa: 31/10/2025 Ngày duyệt đăng: 19/11/2025

Tóm tắt: *Trí tuệ nhân tạo (AI) đang đóng vai trò ngày càng quan trọng trong lĩnh vực tài chính- ngân hàng, nhưng cũng đặt ra nhiều thách thức liên quan đến minh bạch, công bằng, khả năng giải thích, và tuân thủ pháp lý. Bài báo này đề xuất một khung quản lý rủi ro AI dành riêng cho các tổ chức tài chính-ngân hàng tại Việt Nam, được xây dựng trên cơ sở kết hợp bộ chỉ số KAIRI (Knowledge, Accuracy, Interpretability, Robustness, Impact) và các khung quản lý rủi ro AI đã được công bố, nhằm nhận diện, đo lường và kiểm soát rủi ro trong toàn bộ vòng đời vận hành của hệ thống AI. Khung này được hình thành từ việc tổng hợp các nghiên cứu trong và ngoài nước, đồng thời tuân theo nguyên tắc 5 trụ cột bao gồm: thiết kế minh bạch, đo lường định lượng rủi*

Artificial intelligence risk management framework in Financial and Banking Organizations: Proposed model and approach

Abstract: Artificial Intelligence (AI) is playing an increasingly important role in the financial and banking sector, but also raises multiple challenges related to transparency, fairness, explainability, and regulatory compliance. This study proposes an AI risk management framework specifically designed for financial institutions, grounded in the integration of the KAIRI index (Knowledge, Accuracy, Interpretability, Robustness, Impact) with established AI risk management frameworks. The framework is intended to facilitate the identification, quantification, and mitigation of risks across the entire operational lifecycle of AI systems. Its development is informed by a synthesis of domestic and international research, and it is structured around five foundational pillars: transparent design, quantitative risk assessment, lifecycle monitoring, independent auditing, and compliance–ethics. By leveraging KAIRI as a core foundation in conjunction with existing frameworks, this approach provides a comprehensive and practicable pathway for the safe, sustainable, and governance-aligned adoption of AI within financial institutions.

Keywords: Artificial Intelligence (AI), AI Risk Management, Finance and Banking, Risk Governance Framework, KAIRI Index

Link Doi: https://doi.org/10.59276/JELB.2026.03.3037

Trieu Thu Huong
Email: huongtrieu@hvnh.edu.vn
Organization: Banking Academy of Viet Nam

© Học viện Ngân hàng
ISSN 3030 - 4199

ro, giám sát vòng đời, kiểm toán độc lập và tuân thủ- đạo đức. Cách tiếp cận này giúp cung cấp một giải pháp toàn diện và khả thi, hỗ trợ các tổ chức tài chính- ngân hàng triển khai AI an toàn, bền vững và phù hợp với các chuẩn mực quản trị hiện đại.

Từ khóa: *Trí tuệ nhân tạo (AI), Quản lý rủi ro AI, Tài chính-ngân hàng, Khung quản trị rủi ro, Chỉ số KAIRI*

Trích dẫn: Triệu Thu Hương. (2026). Khung quản lý rủi ro trí tuệ nhân tạo trong các tổ chức tài chính-ngân hàng: Đề xuất mô hình và hướng tiếp cận. *Tạp chí Kinh tế - Luật và Ngân hàng, 28(3)*, 1-14. https://doi.org/10.59276/JELB.2026.03.3037

1. Đặt vấn đề

AI đang trở thành công nghệ chiến lược thúc đẩy đổi mới và nâng cao năng lực cạnh tranh trong ngành tài chính- ngân hàng. AI hỗ trợ tự động hóa, phân tích dữ liệu lớn, tối ưu hoá ra quyết định tín dụng, quản trị rủi ro và cá nhân hóa dịch vụ (Bahoo và cộng sự, 2024; Giudici và cộng sự, 2024). Tại Việt Nam, nhiều ngân hàng thương mại và các công ty tài chính đã ứng dụng AI trong chấm điểm tín dụng, quản lý đầu tư, thẩm định hồ sơ, chăm sóc khách hàng… (Đào Lê Kiều Oanh & Huỳnh Lê Xuân Uyên, 2023).

Tuy nhiên, AI cũng tiềm ẩn rủi ro đáng kể như tính khó giải thích của mô hình, sai lệch dữ liệu đầu vào, tấn công giả mạo và vi phạm quyền riêng tư (NIST, 2023; Giudici và cộng sự, 2024). Đặc biệt trong lĩnh vực tài chính, nơi yêu cầu cao về minh bạch và tuân thủ- các rủi ro này có thể dẫn đến sai lệch quyết định và mất lòng tin khách hàng. Tại Việt Nam, chưa có khung pháp lý chuyên biệt cho AI nhưng đã bước đầu thiết lập nền tảng kiểm soát rủi ro thông qua Chiến lược quốc gia về AI (Chính phủ, 2021), Luật An ninh mạng (2018), Nghị định 13/2023/NĐ-CP và các chính sách của Ngân hàng Nhà nước. Trong khi đó, ở nhiều nước, các khung quản lý rủi ro AI chuyên biệt như FREE-AI của Ấn Độ (Manikanda, 2025) hay EDGE của Canada (Government of Canada, 2024) đã và đang dần trở nên phổ biến. Các tổ chức tài chính-ngân hàng chủ yếu dựa trên chuẩn mực quốc tế (Basel II/III về quản trị rủi ro tín dụng và vốn), các quy định pháp luật liên quan như Luật An toàn thông tin mạng, Luật An ninh mạng, cùng với một số sáng kiến nội bộ. Những công cụ này chỉ gián tiếp điều chỉnh AI, chưa đưa ra các nguyên tắc cụ thể về minh bạch, khả năng giải thích (explainability), kiểm thử mô hình, giảm thiểu thiên kiến (bias mitigation) hay trách nhiệm pháp lý khi AI gây sai sót.

Các quy định quản trị rủi ro hiện nay vẫn còn lúng túng với hệ thống AI, và dù nhiều tổ chức đã tích hợp ESG trong báo cáo phát triển bền vững, quá trình này còn sơ khai, gặp thách thức về pháp lý, nguồn lực, dữ liệu và thiếu khung quản lý rủi ro AI thống nhất. Do đó, xây dựng và áp dụng khung quản trị rủi ro AI cho ngành tài chính-ngân hàng tại Việt Nam là hết sức cần thiết nhằm đảm bảo hiệu quả triển khai, tuân thủ pháp lý và bảo vệ quyền lợi người dùng.

Mục tiêu của nghiên cứu là đề xuất một khung quản trị rủi ro AI phù hợp với bối cảnh vận hành, yêu cầu pháp lý và đặc thù ngành tài chính tại các nước mới nổi, đặc biệt là Việt Nam. Khung được xây dựng trên cơ sở kết hợp kinh nghiệm quốc tế (NIST- National Institute of

Standards and Technology, EU AI Act, Basel, ISO- International Organization for Standardization) với đặc thù ngành tài chính- ngân hàng và điều kiện thực tiễn trong nước, qua đó tạo nên đóng góp chính của bài báo. Khung đề xuất bao gồm 5 trụ cột chính, tích hợp bộ chỉ số KAIRI nhằm lượng hóa và cảnh báo rủi ro, hướng đến việc hỗ trợ tổ chức tài chính triển khai AI một cách an toàn, minh bạch, công bằng và có khả năng giải thích, đồng thời đáp ứng tiêu chuẩn quốc tế và định hướng phát triển bền vững. Nghiên cứu được thực hiện bằng phương pháp hỗn hợp, kết hợp giữa phân tích tài liệu và khảo sát- phỏng vấn chuyên gia trong lĩnh vực ngân hàng, fintech. Về cấu trúc, bài báo được cấu trúc như sau: (i) đặt vấn đề, (ii) cơ sở lý thuyết và khoảng trống nghiên cứu; (iii) phương pháp nghiên cứu; (iv) kết quả nghiên cứu và thảo luận; và (v) kết luận hạn chế và hướng nghiên cứu.

2. Cơ sở lý thuyết và khoảng trống nghiên cứu

2.1. Các loại rủi ro của AI trong lĩnh vực tài chính- ngân hàng

Rủi ro mô hình (Model risk)

Rủi ro mô hình là khả năng mô hình toán học hoặc AI tạo ra kết quả sai lệch, không ổn định hay khó lý giải, dẫn đến quyết định tài chính sai (Anke, 2016). Trong ngân hàng, AI được ứng dụng rộng rãi trong chấm điểm tín dụng, phát hiện gian lận, định giá tài sản, nhưng cũng làm tăng rủi ro do tính "hộp đen" của học sâu, vốn khó giải thích và kiểm chứng hơn so với các phương pháp thống kê truyền thống (Bhattacharyya và cộng sự, 2025).

Rủi ro mô hình bắt nguồn từ dữ liệu đầu vào sai lệch, lựa chọn mô hình không phù hợp hoặc quy trình phát triển thiếu kiểm soát (Giudici và cộng sự, 2024; NIST, 2023). Mô hình còn dễ bị drift khi thị trường thay đổi nếu không được cập nhật thường xuyên. Trong tài chính, các sai sót này có thể dẫn đến thiệt hại lớn, từ quyết định tín dụng sai đến quản trị đầu tư kém hiệu quả. Hơn nữa, tính khó giải thích làm tăng rào cản tuân thủ, nhất là trước các quy định mới như EU AI Act (2021) yêu cầu minh bạch và trách nhiệm giải trình (Finocchiaro, 2024).

Rủi ro đạo đức và thiên vị

Rủi ro đạo đức trong AI xảy ra khi hệ thống gây thiên vị hoặc vi phạm quyền con người, dù hoạt động kỹ thuật chính xác (Raji và cộng sự, 2022). Trong tài chính, điều này thể hiện qua các quyết định tín dụng hay đầu tư không công bằng, thường do dữ liệu thiếu đại diện hoặc thiết kế mô hình bỏ qua yếu tố công bằng, ví dụ loại trừ phụ nữ hoặc cư dân nông thôn (Pessach & Shmueli, 2022).

Hiệu suất cao là chưa đủ nếu mô hình tạo ra chênh lệch giữa các nhóm người dùng (Giudici và cộng sự, 2024). Do đó, các tổ chức tài chính cần áp dụng cách tiếp cận công bằng theo thiết kế (fairness by design), tích hợp chỉ số công bằng như Demographic Parity, Equal Opportunity, đồng thời tuân thủ các quy định quốc tế như OECD AI Principles (2019) và EU AI Act yêu cầu giám sát, đánh giá định kỳ với hệ thống AI rủi ro cao.

Rủi ro bảo mật (security)

Rủi ro bảo mật trong AI là nguy cơ mô hình bị tấn công hoặc khai thác, gây rò rỉ dữ liệu, suy giảm hiệu suất hay quyết định sai lệch, dẫn đến thiệt hại nghiêm trọng (Nguyễn Minh Hải, 2025). Trong các tổ chức tài chính, ngân hàng, nơi xử lý dữ liệu tín dụng và tài chính nhạy cảm, các lỗ hổng có thể ảnh hưởng lớn đến tài chính, pháp lý và uy tín. Các mối đe dọa phổ biến như: tấn công dữ liệu đầu vào (adversarial attacks),

rò rỉ thông tin huấn luyện, lỗ hổng từ API/ bên thứ ba, và thiếu giám sát, cập nhật mô hình (Goodfellow và cộng sự, 2014; Shokri và cộng sự, 2017; NIST, 2023).

Các cuộc tấn công AI tài chính có thể gây hậu quả nghiêm trọng: thao túng giao dịch, gián đoạn hoạt động, đánh cắp dữ liệu dẫn đến gian lận và tổn thất cho khách hàng. Điều này không chỉ làm giảm hiệu quả vận hành mà còn khiến ngân hàng mất uy tín, vi phạm quy định như GDPR (General Data Protection Regulation) hay Basel III. Theo Global Risk (2025), hệ thống AI tài chính là mục tiêu tấn công ưu tiên do giá trị kinh tế và dữ liệu mà chúng nắm giữ.

Rủi ro về hoạt động

Rủi ro hoạt động trong AI là nguy cơ tổn thất do lỗi quy trình, hệ thống, con người hoặc yếu tố bên ngoài khi triển khai và vận hành mô hình. Trong ngân hàng, sai sót nhỏ trong các nghiệp vụ như thẩm định tín dụng, phát hiện gian lận hay giao dịch tự động có thể gây gián đoạn dịch vụ và mất lòng tin khách hàng (Basel Committee, 2011). Nguyên nhân chính là: mô hình khó giải thích (đặc biệt với deep learning), tự động hóa thiếu giám sát, cập nhật không kiểm soát và phụ thuộc bên thứ ba (Rudin, 2019; Giudici và cộng sự, 2024; NIST, 2023). Hậu quả có thể gây ra gián đoạn tín dụng, lỗi giao dịch, cảnh báo gian lận sai, thiệt hại tài sản, chi phí khắc phục và mất uy tín. Đặc biệt, tính khó giải thích của AI còn cản trở tuân thủ pháp lý.

Rủi ro tuân thủ

Rủi ro tuân thủ trong AI xuất hiện khi tổ chức không đáp ứng yêu cầu pháp lý, đạo đức hay giám sát, đặc biệt trong tài chính-ngân hàng với các ứng dụng nhạy cảm như chấm điểm tín dụng và phát hiện gian lận. Nếu AI thiếu minh bạch, bảo mật hoặc công bằng, tổ chức có thể bị phạt, mất giấy phép hoặc kiện tụng; nguyên nhân thường do không cập nhật quy định mới, thiếu đánh giá rủi ro pháp lý, hoặc lạm dụng tự động hóa mà thiếu giám sát con người (European Commission, 2021).

Hậu quả không tuân thủ có thể rất nghiêm trọng: phạt đến 30 triệu Euro hoặc 6% doanh thu toàn cầu, thậm chí cấm dùng AI cho ứng dụng "rủi ro cao" nếu thiếu giải thích, giám sát con người hay đánh giá rủi ro (European Commission, 2021). Ngoài ra, tổ chức còn mất niềm tin từ khách hàng, đối tác; vì vậy Basel Committee (2023) khuyến nghị tích hợp rủi ro AI vào quản lý tuân thủ và rủi ro hoạt động.

2.2. Các khung quản lý rủi ro AI hiện hành

Để có cái nhìn toàn diện, dưới đây trình bày một số khung quản lý rủi ro AI tiêu biểu đã và đang được các tổ chức quốc tế và khu vực phát triển.

OECD AI Principles (OECD, 2019)

Năm 2019, OECD (Organisation for Economic Co-operation and Development) là tổ chức quốc tế đầu tiên ban hành bộ nguyên tắc toàn diện về AI mang tên OECD Principles on Artificial Intelligence. Dù không mang tính ràng buộc pháp lý, bộ nguyên tắc này đã được hơn 40 quốc gia, kể cả các nước G20, công nhận và áp dụng, trở thành nền tảng tham chiếu cho nhiều khung pháp lý và tiêu chuẩn kỹ thuật như EU AI Act, ISO 42001. OECD đề xuất 5 nguyên tắc chính để hướng tới phát triển AI có trách nhiệm: (1) Lấy con người làm trung tâm; (2) Đảm bảo công bằng, không phân biệt đối xử; (3) Minh bạch và có thể giải thích; (4) Vận hành an toàn, tin cậy; (5) Rõ ràng về trách nhiệm và quản trị. Tuy nhiên, do chỉ mang tính khuyến nghị, thiếu hướng dẫn kỹ thuật và chưa bao quát thách thức mới như AI tạo sinh hay LLMs, nên

cần được bổ sung bằng các quy định chi tiết và ràng buộc hơn.

NIST AI RMF 1.0 (NIST, 2023)

Khung quản trị rủi ro AI (AI RMF) do Viện Tiêu chuẩn và Công nghệ Quốc gia Hoa Kỳ (NIST) phát triển nhằm hỗ trợ các tổ chức nhận diện, đánh giá, quản lý và giám sát rủi ro AI xuyên suốt vòng đời hệ thống. Đây là khung tự nguyện, phi điều tiết và linh hoạt, phù hợp với mọi loại hình tổ chức- từ khu vực công đến tư, từ doanh nghiệp nhỏ đến tập đoàn lớn. Cấu trúc lõi của khung được cấu thành bởi 4 chức năng chính: (1) Map (Lập bản đồ)- Xác định bối cảnh triển khai, các bên liên quan và rủi ro tiềm ẩn; (2) Measure (Đo lường)- Đề cập đến đánh giá rủi ro AI thông qua các chỉ số định tính và định lượng như độ tin cậy, khả năng giải thích, công bằng, quyền riêng tư ở mức độ chưa chi tiết; (3) Manage (Quản lý)- Thực hiện các biện pháp kiểm soát, kiểm thử và phản ứng để giảm thiểu rủi ro; (4) Govern (Quản trị)- Thiết lập chính sách nội bộ, trách nhiệm giải trình và tiêu chuẩn đạo đức để đảm bảo AI vận hành minh bạch và tuân thủ pháp luật.

Khung AI RMF của NIST được thiết kế linh hoạt, thực tiễn, tùy chỉnh theo quy mô và áp dụng cho cả khu vực công lẫn tư, đồng thời tương thích với các chuẩn quốc tế như ISO 23894, EU AI Act và OECD AI Principles. NIST còn cung cấp Playbook và Roadmap hỗ trợ triển khai, nhấn mạnh các giá trị như minh bạch, công bằng và quyền con người. Tuy nhiên, khung còn hạn chế về tính đặc thù ngành (đặc biệt tài chính), thiếu công cụ định lượng và đòi hỏi năng lực nội bộ cao, gây khó khăn cho tổ chức vừa và nhỏ.

EU AI Act (European Commission, 2021)

EU AI Act là khung pháp lý toàn diện đầu tiên trên thế giới do Ủy ban châu Âu đề xuất nhằm điều chỉnh việc phát triển và sử dụng hệ thống AI. Mục tiêu chính là đảm bảo AI hoạt động an toàn, minh bạch, có thể kiểm tra và tôn trọng quyền con người. Đạo luật phân loại hệ thống AI theo bốn mức rủi ro, từ thấp đến cao (Bảng 1), và đưa ra các yêu cầu pháp lý tương ứng để kiểm soát tác động tiềm ẩn đến xã hội (European Commission, 2021).

Theo Điều 6 và Phụ lục III, EU AI Act xếp các hệ thống AI trong tài chính (chấm điểm tín dụng, đánh giá rủi ro, quyết định tài chính cá nhân) vào nhóm "rủi ro cao", với 5 yêu cầu: minh bạch và giải thích được, có giám sát con người, quản lý dữ liệu chất lượng, đánh giá tuân thủ trước triển khai, và đăng ký trong cơ sở dữ liệu AI kèm dấu CE. Đây là khung pháp lý ràng buộc đầu tiên nhằm bảo đảm minh bạch, an toàn và quyền con người trong toàn bộ vòng đời AI. Tuy nhiên, EU AI Act còn hạn chế: phân loại rủi ro cứng nhắc, gánh nặng tuân thủ cho SME (International Organization for Standardization), thiếu hướng dẫn kỹ

Bảng 1. Các mức độ rủi ro của hệ thống AI

Mức rủi ro	Mô tả
Không thể chấp nhận	AI vi phạm quyền cơ bản → bị cấm hoàn toàn (ví dụ: chấm điểm công dân, thao túng hành vi trẻ em).
Rủi ro cao	AI dùng trong các lĩnh vực nhạy cảm như tài chính, y tế, giáo dục, tư pháp → chịu giám sát nghiêm ngặt.
Rủi ro giới hạn	Yêu cầu minh bạch, ví dụ như AI tạo sinh phải thông báo là "do máy tạo ra".
Rủi ro tối thiểu	Miễn trừ kiểm soát, ví dụ như AI lọc thư rác hoặc chơi game.

Nguồn: European Commission, 2021

thuật đồng bộ; do đó, cần các khung pháp lý linh hoạt hơn, ngữ cảnh hóa rủi ro và giám sát cả nhà phát triển mô hình nền.

ISO/IEC 42001:2023 (ISO/IEC, 2023)
ISO/IEC 42001:2023 là tiêu chuẩn quốc tế đầu tiên về hệ thống quản lý AI (AIMS), do ISO và IEC ban hành, nhằm hỗ trợ tổ chức xác định, kiểm soát và giảm thiểu rủi ro trong suốt vòng đời AI. Tiêu chuẩn này bao hàm yêu cầu về phạm vi áp dụng, cam kết lãnh đạo, quản lý rủi ro, năng lực đội ngũ, kiểm soát vận hành, đánh giá hiệu quả và cải tiến liên tục, đồng thời có thể tích hợp với ISO 27001 hoặc 9001 để đồng bộ quy trình và tiết kiệm chi phí. Tuy nhiên, ISO/IEC 42001 mang tính tự nguyện, thiếu hướng dẫn kỹ thuật cụ thể và đòi hỏi chuyên môn cao, gây khó khăn cho tổ chức nhỏ, và cần điều chỉnh để phù hợp với các quy định ràng buộc như EU AI Act.

Basel Committee AI Guidelines (BIS, 2023)
Khung quản trị rủi ro AI của Basel hỗ trợ ngân hàng kiểm soát rủi ro khi ứng dụng AI, dựa trên nguyên tắc thay vì mô hình cụ thể, tích hợp với chuẩn mực quản lý rủi ro truyền thống như Basel II/III. Khung nhận diện các rủi ro như thiên lệch dữ liệu, thiếu minh bạch, hạn chế giải thích và rủi ro đạo đức, nhấn mạnh giám sát lãnh đạo, đánh giá mô hình độc lập và truy vết vòng đời AI. Ưu điểm là tính toàn diện, linh hoạt và dễ tích hợp với hệ thống quản lý hiện có, góp phần bảo vệ người tiêu dùng và tăng kiểm soát. Hạn chế là chỉ mang tính khuyến nghị, không ràng buộc pháp lý, thiếu hướng dẫn kỹ thuật, đòi hỏi nguồn lực lớn và mức độ áp dụng phụ thuộc chính sách từng quốc gia.

WEF AI Governance Framework (WEF, 2019)
WEF đề cập khung quản trị rủi ro AI toàn diện trong báo cáo "AI Governance: A Holistic Approach to Implementing Ethical and Responsible AI" (2019), do nhóm chuyên gia đa ngành xây dựng, được cấu thành bởi 4 trụ cột: chiến lược và vận hành, kiến trúc công nghệ và dữ liệu, con người và văn hóa, đánh giá và giám sát. Khung hướng dẫn thực tiễn, tích hợp toàn diện giữa công nghệ, văn hóa và đạo đức, nhấn mạnh đào tạo và năng lực nội bộ, phù hợp với cả tổ chức lớn và nhỏ. Tuy nhiên, do không ràng buộc pháp lý và thiếu chỉ số định lượng, việc triển khai trong các lĩnh vực yêu cầu giám sát nghiêm ngặt như tài chính- ngân hàng có thể gặp khó khăn.

2.3. Bộ chỉ số KAIRI

KAIRI do Giudici và cộng sự (2024) đề xuất là bộ chỉ số định lượng tổng hợp giúp đánh giá mức độ rủi ro của mô hình AI. KAIRI cụ thể là: Knowledge (tri thức mô hình- K), Accuracy (độ chính xác- A), Interpretability (khả năng giải thích- I1), Robustness (tính bền vững- R) và Impact (tác động xã hội- I_2), nhằm cân bằng hiệu suất kỹ thuật với minh bạch, khả năng chịu lỗi và rủi ro đạo đức/pháp lý.
Công thức tính toán tổng hợp của KAIRI:
$KairiScore = w_1.K + w_2.A + w_3.I_1 + w_4.R + w_5.I_2$
Trong đó:

○ $w_1 + w_2 + w_3 + w_4 + w_5 = 1$ (trọng số tùy chỉnh theo loại mô hình)

○ Mỗi thành phần được chuẩn hóa về thang 0-1

○ KAIRI Score càng cao thì rủi ro càng thấp
Chỉ số KAIRI là bước tiến quan trọng để lượng hóa rủi ro AI toàn diện, vượt xa các chỉ số kỹ thuật thuần túy như độ chính xác. Bằng cách kết hợp định lượng và định tính, KAIRI đánh giá hiệu suất, minh bạch, khả năng giải thích và tác động xã hội, giúp

Bảng 2. Khung phân loại rủi ro dựa trên KAIRI

KAIRI Score	Mức độ rủi ro	Hành động yêu cầu
0,8 - 1,0	Thấp (Green)	Giám sát thường xuyên
0,6 - 0,8	Trung bình (Yellow)	Tăng cường giám sát, xem xét cải tiến
0,4 - 0,6	Cao (Orange)	Yêu cầu hành động khắc phục ngay
< 0,4	Rất cao (Red)	Tạm dừng sử dụng, đánh giá lại toàn bộ

Nguồn: Giudici và cộng sự, 2024

kiểm toán, so sánh và lựa chọn mô hình phù hợp, đặc biệt trong các lĩnh vực nhạy cảm như tài chính và tuân thủ các quy định như EU AI Act.

2.4. Đánh giá so sánh và khoảng trống nghiên cứu

Căn cứ theo các đặc điểm của từng khung, tiêu chuẩn quản lý rủi ro, nhóm nghiên cứu đã phân tích, đưa bảng đánh giá tổng hợp (Bảng 3).

Các khung và tiêu chuẩn hiện hành như NIST AI RMF, EU AI Act, ISO/IEC 42001 hay Basel Committee tập trung vào nguyên

Bảng 3. Đánh giá Khung/tiêu chuẩn quản lý rủi ro AI phổ biến hiện hành

Tên khung/ tiêu chuẩn	Tổ chức/ quốc gia	Lĩnh vực áp dụng	Ưu điểm	Hạn chế
OECD AI Principles	OECD (Quốc tế)	Toàn ngành, bao gồm tài chính	Bộ nguyên tắc phổ quát, được nhiều nước áp dụng làm cơ sở luật pháp và chính sách AI; nhấn mạnh các giá trị nhân văn	Mang tính định hướng chung; không cung cấp hướng dẫn chi tiết hay tiêu chuẩn đo lường định lượng rủi ro
NIST AI Risk Management Framework (AI RMF 1.0-2023)	NIST- Hoa Kỳ	Đa ngành, trong đó có tài chính	Minh bạch, dễ tiếp cận, giúp đánh giá và phân loại rủi ro AI; hỗ trợ xây dựng chính sách và quy trình quản lý rủi ro	Chưa cung cấp bộ tiêu chuẩn đo lường hay các chỉ số định lượng cụ thể để giám sát, đo đạc rủi ro AI chi tiết
EU AI Act	Liên minh châu Âu (EU)	Nhạy cảm, bao gồm tài chính	Cơ sở pháp lý bắt buộc, bảo vệ người dùng, khung chuẩn hóa cho toàn EU, tăng niềm tin và sự tuân thủ của hệ thống AI	Tập trung vào tuân thủ và giám sát pháp lý, thiếu bộ công cụ đo lường định lượng rủi ro AI cụ thể
ISO/IEC 42001:2023	ISO/IEC (Quốc tế)	Toàn ngành, đặc biệt tài chính, y tế	Tích hợp với ISO 27001, 9001; thúc đẩy minh bạch, công bằng, trách nhiệm giải trình theo chuẩn quốc tế	Tập trung vào hệ thống quản lý; chưa phát triển bộ chỉ số đo lường định lượng rủi ro AI chi tiết
Khung Basel Committee for AI Risk Management	Ủy ban Basel (toàn cầu)	Tài chính-ngân hàng	Linh hoạt, dễ tích hợp với hệ thống rủi ro hiện hữu; tập trung bảo vệ người tiêu dùng, tăng cường giải thích được AI	Khuyến nghị mang tính nguyên tắc, thiếu mô hình định lượng và bộ chỉ số đo lường cụ thể
Khung quản lý AI của World Economic Forum (WEF)	WEF (Quốc tế)	Đa ngành, có thể áp dụng tài chính	Toàn diện, dễ tiếp cận, nhấn mạnh con người và văn hóa tổ chức, phù hợp với các tổ chức nhỏ và lớn	Hướng dẫn không mang tính pháp lý; thiếu công cụ đo lường định lượng rủi ro AI chuẩn hóa
Bộ chỉ số định lượng KAIRI	Đề xuất từ các tổ chức nghiên cứu/ quản lý AI	Có thể áp dụng trong tài chính	Cung cấp đo lường định lượng cụ thể, hỗ trợ cảnh báo sớm, giám sát rủi ro AI hiệu quả	Phụ thuộc việc xây dựng và hiệu chỉnh liên tục; chưa chuẩn hóa rộng rãi

Nguồn: Tác giả tổng hợp

tắc và hướng dẫn quản trị, nhưng còn thiếu công cụ đo lường định lượng rủi ro, minh bạch dữ liệu- mô hình- kết quả và cơ chế kiểm toán độc lập, những yếu tố quan trọng với AI trong tài chính-ngân hàng. Khoảng trống này tạo nhu cầu về khung quản lý đặc thù, kết hợp chỉ số định lượng như KAIRI, minh bạch hóa chuỗi dữ liệu và mô hình, đồng thời thiết lập cảnh báo động và kiểm toán độc lập, giúp tổ chức kiểm soát, giám sát và chứng minh năng lực quản lý rủi ro AI hiệu quả, đồng thời tuân thủ quy định pháp lý và đặc thù ngành.

3. Phương pháp nghiên cứu

Để triển khai nghiên cứu, nhóm nghiên cứu áp dụng phương pháp nghiên cứu hỗn hợp (mixed methods), kết hợp giữa nghiên cứu định tính và định lượng. Quy trình nghiên cứu được chia thành ba giai đoạn chính:

Giai đoạn 1- Tổng hợp và phân tích tài liệu

Thực hiện thu thập và phân tích tài liệu giai đoạn 2014- 2025 từ các nguồn học thuật và chính thức như Scopus, Google Scholar, NIST, EU, OECD và Basel Committee. Quá trình phân tích bao gồm phân loại rủi ro AI theo lĩnh vực, so sánh các khung quản lý hiện hành và xác định khoảng trống nghiên cứu. Trên cơ sở này, nhóm đề xuất khung đánh giá rủi ro sơ lược với 5 trụ cột.

Giai đoạn 2- Thu thập ý kiến chuyên gia

Thực hiện khảo sát 12 chuyên gia giàu kinh nghiệm trong AI/fintech và quản lý rủi ro, thông qua phỏng vấn bán cấu trúc và khảo sát định lượng sử dụng thang đo Likert. Các chuyên gia làm việc tại ngân hàng, công ty fintech, cơ quan quản lý và học thuật, cụ thể là: 3 chuyên gia ngân hàng thương mại, 3 chuyên gia fintech, 2 chuyên gia thuộc cơ quan quản lý, 4 chuyên học thuật.

Kết quả cho thấy hơn 12/12 chuyên gia xác nhận AI đã và đang được ứng dụng rộng rãi trong tài chính- ngân hàng, đặc biệt trong quản trị rủi ro và phát hiện gian lận và đánh giá rủi ro AI là vấn đề trọng yếu, trong khi 11/12 cho rằng khung hiện tại chưa hiệu quả, thiếu cơ chế định lượng và giám sát vòng đời mô hình. Các khung quốc tế như NIST AI RMF, EU AI Act và Basel Guidelines được 12/12 chuyên gia đánh giá là phù hợp định hướng toàn cầu nhưng cũng nhận thấy khó áp dụng trọn vẹn tại Việt Nam do khác biệt về hạ tầng và quy định. Với khung 5 trụ cột đề xuất (Thiết kế minh bạch, Đánh giá định lượng rủi ro, Giám sát vòng đời, Kiểm định độc lập, Tuân thủ & đạo đức), 12/12 chuyên gia nhận định phù hợp và toàn diện và đánh giá khả thi khi áp dụng thực tế, nhất là khi có hướng dẫn rõ ràng về chỉ số KAIRI. Một số khuyến nghị chính gồm: chuẩn hóa tài liệu mô hình, hoàn thiện chỉ số KAIRI, tăng cường kiểm toán độc lập và đào tạo đạo đức AI. Nhìn chung, khảo sát cho thấy sự đồng thuận cao về nhu cầu xây dựng khung quản lý rủi ro AI đặc thù, và khung đề xuất là khả thi và phù hợp bối cảnh Việt Nam.

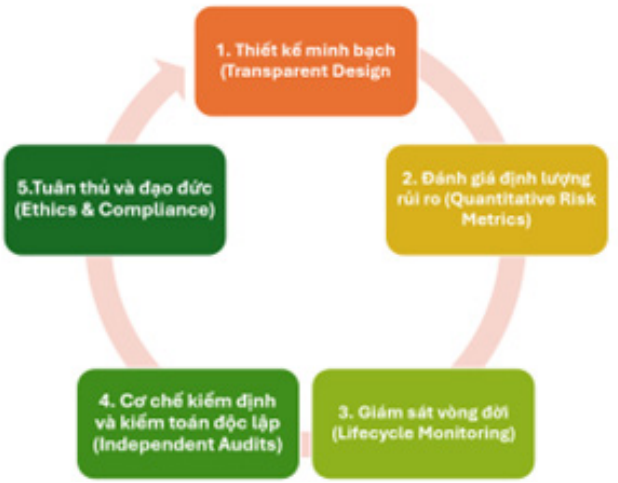
Giai đoạn 3- Quy trình đánh giá và hiệu chỉnh khung

Kết quả từ Giai đoạn 2 được sử dụng để thiết kế chi tiết khung quản lý rủi ro AI với định lượng rõ ràng, sau đó tiếp tục tiến hành đánh giá và hiệu chỉnh thông qua phản hồi từ các chuyên gia, đảm bảo tính phù hợp với bối cảnh thực tiễn và chuẩn mực quốc tế.

4. Kết quả nghiên cứu và thảo luận

4.1. Đề xuất khung quản lý rủi ro AI cho lĩnh vực tài chính- ngân hàng

Khung quản lý rủi ro AI cho lĩnh vực tài chính- ngân hàng được đề xuất trên cơ sở: (1) lấp đầy khoảng trống các khung truyền thống (NIST, EU AI Act, ISO/IEC 42001, Basel) bằng công cụ đo lường định lượng



Hình 1. Khung quản lý rủi ro AI cho lĩnh vực tài chính-ngân hàng

rủi ro AI thực tiễn cho tổ chức tài chính-ngân hàng; (2) đáp ứng yêu cầu minh bạch và giải trình ngày càng cao, nâng cao niềm tin và bảo vệ quyền lợi của các đối tượng chịu tác động; (3) thúc đẩy quản trị AI hiện đại, chủ động thông qua giám sát, cảnh báo và cải tiến liên tục mà các mô hình quản lý truyền thống khó đạt được.

Nhóm nghiên cứu đề xuất khung quản lý rủi ro AI dựa trên 5 trụ cột như Hình 1.

Trụ cột 1. Thiết kế minh bạch (Transparent Design)

Dựa trên yêu cầu minh bạch và khả năng giải thích trong EU AI Act, đồng thời làm rõ các nguyên tắc của ISO/IEC 42001, khung đề xuất tập trung vào nguyên tắc Thiết kế minh bạch (Transparency by Design). Nguyên tắc này là nền tảng xây dựng lòng tin và tăng cường giải trình cho hệ thống AI, bao gồm khả năng truy xuất nguồn gốc dữ liệu huấn luyện, làm rõ giả định mô hình và cung cấp đầy đủ tài liệu về quá trình xây dựng, điều chỉnh, kiểm thử mô hình. Việc ghi nhận chi tiết các bước xử lý dữ liệu- từ thu thập, làm sạch, chọn mẫu đến chuẩn hóa- giúp giảm nguy cơ thiên lệch và sai lệch đầu ra. Trong lĩnh vực tài chính- ngân hàng, nơi dữ liệu nhạy cảm và quyết định mô hình ảnh hưởng trực tiếp đến quyền lợi người dùng, truy vết toàn bộ chuỗi dữ liệu- mô hình- kết quả không chỉ là yêu cầu đạo đức mà còn bắt buộc để tuân thủ các quy định pháp lý như GDPR và Basel

Trụ cột 2. Đánh giá định lượng rủi ro (Quantitative Risk Metrics)

Dựa trên yêu cầu phân loại rủi ro của NIST AI RMF và nhu cầu minh bạch, giám sát liên tục trong tài chính, việc bổ sung phương pháp đánh giá định lượng rủi ro AI là cấp thiết. Đánh giá không thể chỉ dựa vào cảm quan mà cần bộ chỉ số đo lường rõ ràng để giám sát, so sánh và kiểm soát theo thời gian. Các khung như KAIRI cho phép lượng hóa thiên lệch, rò rỉ thông tin, khả năng giải thích và độ tin cậy thống kê, giúp xác định mức rủi ro chấp nhận được, thiết lập ngưỡng kiểm soát và tự động kích hoạt biện pháp ứng phó. Đây là bước trung gian quan trọng giữa đánh giá đạo đức định tính và kiểm toán kỹ thuật, đưa AI vào tầm quản trị tương đương các rủi ro tài chính truyền thống.

Trụ cột 3. Giám sát vòng đời (Lifecycle

Monitoring)

Dựa trên yêu cầu quản lý vòng đời AI trong WEF AI Governance, trụ cột 3 đề xuất thiết lập hệ thống giám sát liên tục với ngưỡng cảnh báo linh hoạt để phát hiện sớm rủi ro mới. AI là hệ thống học động, thay đổi theo dữ liệu và môi trường, nên quản trị rủi ro hiệu quả đòi hỏi theo dõi toàn bộ vòng đời từ thiết kế, huấn luyện, kiểm thử, triển khai đến ngừng sử dụng. Việc đặt các điểm kiểm soát (control points) giúp phát hiện trôi dữ liệu (data drift), suy giảm hiệu suất mô hình hoặc rò rỉ thông tin cá nhân, đồng thời đáp ứng yêu cầu pháp lý và chuẩn mực như Basel, ISO/IEC 38507 hay NIST AI RM.

Trụ cột 4. Cơ chế kiểm định và kiểm toán độc lập (Independent Audits)

Dựa trên yêu cầu đánh giá tuân thủ của EU AI Act, trụ cột 4 đề xuất cơ chế kiểm toán độc lập đa ngành (công nghệ, pháp lý, tài chính) để đảm bảo khách quan và minh bạch trong quản trị AI. Với hệ thống AI rủi ro cao như tài chính, y tế hay tư pháp, kiểm định định kỳ bởi bên thứ ba giúp đánh giá toàn diện dữ liệu, mô hình, logic ra quyết định, khả năng kháng tấn công và tác động xã hội. Cơ chế này tương tự kiểm toán tài chính, phát hiện và ngăn chặn sai lệch, thiên lệch hoặc sử dụng sai mục đích, đồng thời đáp ứng yêu cầu quy trình đánh giá sự phù hợp (conformity assessment) trong EU AI Act và các tiêu chuẩn quốc tế như ISO/IEC 42001 hay AI Assurance Framework.

Trụ cột 5. Tuân thủ và đạo đức (Compliance and Ethics)

Căn cứ EU AI Act, ISO/IEC 42001 và WEF Framework, trụ cột 5 đề xuất tích hợp kiểm soát tuân thủ với xây dựng văn hóa và đào tạo nhận thức, nhằm triển khai AI minh bạch, có trách nhiệm và phù hợp pháp luật cũng như chuẩn mực xã hội, đặc biệt trong tài chính- ngân hàng. Hệ thống AI bền vững không chỉ tuân thủ GDPR, EU AI Act hay ISO 23894 mà còn thể hiện các giá trị cốt lõi như công bằng, minh bạch và tôn trọng con người. Tuân thủ vừa là yêu cầu pháp lý, vừa là lớp bảo vệ trước rủi ro pháp lý, khủng hoảng niềm tin và thiệt hại uy tín, được thực hiện thông qua hội đồng đạo đức, đánh giá tác động (AIA) và đào tạo liên tục cho đội ngũ phát triển.

Khung quản lý rủi ro AI được đề xuất gồm 5 trụ cột: Thiết kế minh bạch, Đánh giá định lượng rủi ro (với chỉ số KAIRI), Giám sát vòng đời, Kiểm định- kiểm toán độc lập và Tuân thủ- đạo đức. Khung này đảm bảo sự hài hòa giữa kỹ thuật, pháp lý và đạo đức, hướng tới quản trị AI toàn diện, minh bạch và có trách nhiệm, phù hợp với xu hướng toàn cầu và bối cảnh tài chính- ngân hàng tại Việt Nam.

Khung quản lý rủi ro AI đề xuất nổi bật ở khả năng tổng hợp hệ thống các phương pháp tốt nhất từ NIST, EU AI Act và Basel, đồng thời bổ sung những yếu tố còn thiếu trong bối cảnh tài chính- ngân hàng Việt Nam, nơi các ứng dụng AI đang ngày càng phổ biến nhưng thiếu nền tảng quản trị rủi ro bài bản. Điểm khác biệt chính là chuyển từ đánh giá định tính sang định lượng, sử dụng chỉ số KAIRI để đo lường rủi ro mô hình một cách toàn diện và khách quan, biến quản trị rủi ro AI từ "nghệ thuật" thành "khoa học" có thể kiểm chứng bằng dữ liệu. Khung này đáp ứng nhu cầu cấp thiết của ngành, giúp các tổ chức quản lý rủi ro đạo đức, pháp lý và vận hành, đồng thời thích ứng với các quy định quốc tế khắt khe như EU AI Act. Trong lĩnh vực nhạy cảm như tài chính, nơi quyết định AI ảnh hưởng trực tiếp đến tài sản và cơ hội của khách hàng, khung minh bạch và có thể kiểm toán là nền tảng để xây dựng niềm tin với khách hàng, nhà đầu tư và cơ quan quản lý.

4.2. Thảo luận mức độ phù hợp của

khung đề xuất với các loại rủi ro đặc thù của AI tài chính- ngân hàng

Để làm rõ cơ sở lý thuyết của khung đề xuất, nhóm nghiên cứu đánh giá mức độ bao phủ các loại rủi ro đặc thù của AI trong tài chính- ngân hàng (Bảng 4).

4.3. Đối sánh khung đề xuất với các khung/tiêu chuẩn quốc tề và ràng buộc pháp lý tại Việt Nam

Nhằm xác định khả năng bổ sung của khung đề xuất đối với các khung hiện hành, đồng thời đánh giá mức độ phù hợp của khung này với các ràng buộc pháp lý tại Việt Nam, tác giả đã xây dựng Bảng đối sánh số 5 và Bảng đối sánh số 6.

Khung đề xuất 5 trụ cột là một mô hình phù hợp, thực tiễn và toàn diện, được xây dựng có cơ sở từ các khung quốc tế uy tín và bám sát đặc thù của lĩnh vực tài chính. Việc đưa chỉ số định lượng (KAIRI) vào đánh giá là một bước tiến mới giúp nâng cao tính khách quan và khả năng kiểm toán, điều mà nhiều khung hiện hành chưa thực sự làm rõ. Mô hình này có tiềm năng trở thành nền tảng cho việc xây dựng chính sách nội bộ, công cụ đánh giá hoặc khuyến nghị triển khai AI tại các tổ chức tài chính toàn cầu, đặc biệt các nước mới nổi như Việt Nam.

5. Kết luận, hạn chế và hướng nghiên cứu

Bảng 4. Mức độ phù hợp của khung đề xuất với các loại rủi ro của hệ thống AI trong lĩnh vực tài chính- ngân hàng

Loại rủi ro	Trụ cột trong khung đề xuất phản ánh tốt nhất	Đánh giá mức độ bao phủ	Luận giải
Rủi ro mô hình	Trụ cột 2- Định lượng rủi ro, Trụ cột 3	Cao- Đã bao quát cả độ chính xác, khả năng giải thích, độ bền mô hình.	Mức độ cao vì khung không chỉ định hướng mà còn cung cấp công cụ cụ thể (chỉ số định lượng và điểm kiểm soát), giúp tổ chức tài chính kiểm soát rủi ro mô hình tương tự như rủi ro tín dụng truyền thống. Không có khoảng trống lớn, nên đánh giá "Cao" là hợp lý.
Rủi ro đạo đức và thiên vị	Trụ cột 5- Tuân thủ và đạo đức, Trụ cột 1	Cao- Nhấn mạnh yêu cầu minh bạch và không vi phạm quy tắc công bằng.	Mức độ cao vì khung không chỉ dừng ở định lượng mà còn liên kết với minh bạch (Trụ cột 1- yêu cầu theo dõi nguồn gốc và quá trình xử lý dữ liệu để giảm thiên lệch, tuân thủ EU AI Act (Điều 13) và ISO/IEC 42001 (mục 5), giúp ngăn thiên vị trong quyết định tài chính và bảo vệ người dùng) và đào tạo liên tục (Trụ cột 5- mở rộng bằng văn hóa đạo đức, hội đồng đạo đức và đánh giá tác động AI (AIA), nhấn mạnh công bằng và tôn trọng con người, phù hợp EU AI Act (Điều 9-10) và WEF Framework, bảo vệ niềm tin và chống rủi ro) , giúp phát hiện và khắc phục thiên vị sớm. So với NIST AI RMF (mục 1.2 về đạo đức).
Rủi ro bảo mật	Trụ cột 3- Giám sát vòng đời	Trung bình- phù hợp nhưng cần bổ sung rõ yêu cầu bảo mật kỹ thuật.	Đánh giá "Trung bình" vì đề xuất giám sát vòng đời để phát hiện "rò rỉ thông tin cá nhân" và thay đổi môi trường, phù hợp với Basel (yêu cầu giám sát rủi ro hoạt động) và NIST AI RMF (mục 4 về bảo mật). Điều này giúp phát hiện sớm các vấn đề như data leakage hoặc model vulnerabilities trong tài chính (ví dụ: bảo vệ dữ liệu khách hàng theo GDPR). Khung vẫn cần công cụ định lượng cụ thể cho vấn đề bảo mật.

Rủi ro hoạt động (gián đoạn, khó giải thích)	Trụ cột 3- Giám sát vòng đời, Trụ cột 1	Cao- Có bao gồm theo dõi vận hành và thiết kế có thể giải thích.	Mức độ cao vì khung không chỉ theo dõi vận hành mà còn tích hợp thiết kế từ đầu, giúp tổ chức tài chính duy trì hoạt động liên tục và giảm thiểu gián đoạn. Cụ thể: Trụ cột 3 nhấn mạnh giám sát liên tục, bao gồm drift dữ liệu và model decay, với ngưỡng cảnh báo để tránh gián đoạn, phù hợp ISO/IEC 38507 và Basel về rủi ro hoạt động, đặc biệt với AI dự báo thị trường theo dữ liệu thời gian thực. Trụ cột 1 bổ sung "Thiết kế minh bạch" để đảm bảo khả năng giải thích mô hình, thông qua tài liệu quá trình xây dựng, giải quyết rủi ro "khó giải thích" theo EU AI Act (Điều 13) và NIST AI RMF (mục 3). Điều này vượt trội hơn các khung truyền thống thiếu cơ chế chủ động.
Rủi ro tuân thủ	Trụ cột 4- Kiểm toán độc lập, Trụ cột 5	Cao- Liên quan trực tiếp đến yêu cầu pháp lý và quy định kiểm tra định kỳ.	Mức độ cao vì khung biến tuân thủ thành quy trình liên tục, qua kiểm toán định kỳ, giúp tránh rủi ro pháp lý và thiệt hại uy tín- một vấn đề lớn trong ngành ngân hàng. Cụ thể: Trụ cột 4 khuyến nghị kiểm toán độc lập đa ngành (công nghệ, pháp lý, tài chính), tương tự conformity assessment theo EU AI Act (Điều 43) và ISO/IEC 42001 (mục 8), giúp phát hiện vi phạm sớm, đặc biệt với Basel và GDPR. Trụ cột 5 kết hợp tuân thủ với văn hóa tổ chức (hội đồng đạo đức, đào tạo), nhấn mạnh minh bạch và công bằng, phù hợp WEF Framework và ISO 23894.

Nguồn: Tác giả tổng hợp

5.1. Kết luận

Sự phát triển nhanh chóng của AI mang lại cơ hội đổi mới cho tài chính- ngân hàng, nhưng đồng thời tạo ra rủi ro kỹ thuật, pháp lý và đạo đức. Bài báo tổng hợp các rủi ro đặc thù như rủi ro mô hình, thiên vị, bảo mật, vận hành và tuân thủ, chỉ ra nhu cầu cấp thiết về một khung quản trị rủi ro toàn diện, thích ứng cao. Các khung hiện hành như NIST AI RMF, EU AI Act, OECD AI Principles hay ISO/IEC 42001 cung cấp nền tảng quan trọng nhưng còn thiếu công cụ định lượng, kiểm soát vòng đời và bối cảnh hóa cho tài chính. Trên cơ sở đó, bài báo đề xuất khung 5 trụ cột: minh bạch thiết kế, đánh giá định lượng (KAIRI), giám sát vòng đời, kiểm định độc lập và tuân thủ đạo đức- pháp lý, kế thừa tinh thần quốc tế đồng thời khả thi với thực tiễn Việt Nam. Trong tương lai, các tổ chức cần nội địa hóa các nguyên tắc này và xây dựng

Bảng 5. Đối sánh khung đề xuất với các khung quốc tế

Trụ cột đề xuất	Đối sánh với khung quốc tế	Đánh giá
1. Thiết kế minh bạch (Transparent Design	Có trong NIST, EU AI Act, OECD	Không mới nhưng cần thiết, tạo nền tảng.
2. Đánh giá định lượng rủi ro (Quantitative Risk Metrics)	Xuất hiện trong nghiên cứu Giudici và cộng sự	Mới- Gợi mở hướng đo lường cụ thể.
3. Giám sát vòng đời (Lifecycle Monitoring)	NIST RMF (Manage), EU AI Act	Chuẩn quốc tế, vận dụng đúng cách.
4. Cơ chế kiểm định và kiểm toán độc lập (Independent Audits)	EU AI Act, ISO 42001, Basel	Cần thiết trong tài chính-ngân hàng.
5.Tuân thủ và đạo đức (Ethics & Compliance)	OECD, EU AI Act, WEF	Bao phủ các nguyên tắc quan trọng.

Nguồn: Tác giả tổng hợp

Bảng 6. Đối sánh khung đề xuất với một số các quy định pháp lý tại Việt Nam về AI

Quy định pháp luật	Trụ cột phản ánh tốt nhất	Luận giải
1. Luật An ninh mạng (2018) và Luật An toàn thông tin (2015)	Trụ cột 3 (Giám sát vòng đời), Trụ cột 1 (Thiết kế minh bạch), Trụ cột 4 (Kiểm toán độc lập)	Phù hợp với giám sát và minh bạch để phát hiện rò rỉ dữ liệu, nhưng thiếu chi tiết kỹ thuật chống tấn công mạng (cần bổ sung để tuân thủ yêu cầu báo cáo sự cố). Căn cứ: Tập trung bảo vệ hạ tầng, khung AI cần cụ thể hơn về bảo mật.
2. Chiến lược Quốc gia về nghiên cứu, phát triển và ứng dụng AI đến 2030 (Quyết định 127/QĐ-TTg, 2021)	Trụ cột 5 (Tuân thủ và đạo đức), Trụ cột 3 (Giám sát vòng đời), Trụ cột 1 (Thiết kế minh bạch), Trụ cột 2 (Đánh giá định lượng rủi ro)	Bao quát toàn diện AI có trách nhiệm và hạn chế rủi ro qua giám sát chủ động và minh bạch. Phù hợp mục tiêu quốc gia về an toàn dữ liệu và xây dựng khung pháp luật.
3. Nghị định 13/2023/ND-CP (Bảo vệ dữ liệu cá nhân, hiệu lực 1/7/2023)	Trụ cột 1 (Thiết kế minh bạch), Trụ cột 3 (Giám sát vòng đời), Trụ cột 5 (Tuân thủ và đạo đức)	Trực tiếp hỗ trợ bảo vệ quyền riêng tư qua truy xuất dữ liệu và giám sát rò rỉ. Khung giúp giảm thiểu rủi ro dữ liệu cá nhân trong AI tài chính.
4. Quyết định số 1290/QĐ-BKHCN (2024) và Ủy ban đạo đức AI (VINASA, 2024)	Trụ cột 5 (Tuân thủ và đạo đức), Trụ cột 1 (Thiết kế minh bạch), Trụ cột 2 (Đánh giá định lượng rủi ro)	Tích hợp đạo đức cốt lõi với 9 nguyên tắc an toàn, minh bạch và bảo vệ quyền con người, lượng hóa được các tiêu chí để đo lường tuân thủ.

Nguồn: Tác giả tổng hợp

quy trình kiểm soát rủi ro phù hợp với pháp luật, đạo đức và điều kiện kỹ thuật; khung đề xuất có thể là nền tảng tham chiếu hữu ích cho mục tiêu đó.

5.2. Hạn chế và hướng nghiên cứu tiếp theo

Mặc dù nghiên cứu này đã đề xuất một khung quản lý rủi ro AI toàn diện về mặt lý thuyết, với cấu trúc rõ ràng và khả năng định lượng cho các tổ chức tài chính- ngân hàng, nơi mà yêu cầu cao về sự rõ ràng, minh bạch, khả năng giải thích. Tuy nhiên, hạn chế chính của đề xuất là chưa được kiểm chứng trong thực tiễn triển khai. Trong thời gian tới, các hướng nghiên cứu tiếp theo cần tập trung vào ba phương diện chính. Thứ nhất, tiến hành nghiên cứu tình huống thực tế thông qua hợp tác với một hoặc nhiều tổ chức tài chính tại Việt Nam nhằm áp dụng thử nghiệm khung vào hệ thống AI cụ thể, từ đó điều chỉnh các thành phần chưa phù hợp. Thứ hai, cần tiếp tục phát triển và nội địa hóa bộ chỉ số KAIRI, điều chỉnh các thành phần đo lường để phù hợp với đặc thù dữ liệu, môi trường pháp lý và mức độ trưởng thành công nghệ tại Việt Nam. Những nghiên cứu này sẽ là bước đi cần thiết để hoàn thiện khung và tăng cường tính khả thi, góp phần đưa các nguyên tắc quản trị rủi ro AI vào thực tiễn một cách hiệu quả. ■

Tài liệu tham khảo

Anke, N. (2016). Model risk management. KPMG. https://www.scribd.com/document/464628109/KPMG-Whitepaper-Model-Risk-Management-2016

Bahoo, S., Cucculelli, M., Goga, X., & Mondolo, J. (2024). Artificial intelligence in finance: A comprehensive review through bibliometric and content analysis. *SN Business & Economics*. https://doi.org/10.1007/s43546-023-00618-x

Basel Committee on Banking Supervision. (2023). *Disclosure of climate-related financial risks* (BCBS Consultation Document No. d560). Bank for International Settlements. https://www.bis.org/bcbs/publ/d560.pdf

Bhattacharyya, A., Yu, Y., Yang, H., Singh, R., Joshi, T., Chen, J., & Yalavarthy, K. (2025). Model risk management for generative ai in financial institutions. *arXiv preprint arXiv:2503.15668*. https://arxiv.org/pdf/2503.15668?

Bộ Khoa học và Công nghệ. (2024). Quyết định số 1290/QĐ-BKHCN về việc hướng dẫn một số nguyên tắc về nghiên cứu, phát triển các hệ thống trí tuệ nhân tạo có trách nhiệm. Truy cập từ https://sokhcn.haiphong.gov.vn/chien-luoc-quy-hoach-ke-hoach/quyet-dinh-so-1290-qd-bkhcn-ve-viec-huong-dan-mot-so-nguyen-tac-ve-nghien-cuu-phat-trien-cac-he--692957

Chính phủ. (2021). *Chiến lược quốc gia về trí tuệ nhân tạo đến năm 2030*. Hà Nội. https://chinhphu.vn/?pageid=27160&docid=202565&tagid=6&type=1

Chính phủ. (2023). *Nghị định số 13/2023/NĐ-CP về bảo vệ dữ liệu cá nhân*. https://vanban.chinhphu.vn/?pageid=27160&docid=207759

Đào Lê Kiều Oanh, & Huỳnh Lê Xuân Uyên (2023). Xu hướng ứng dụng trí tuệ nhân tạo trong sự phát triển của ngành Ngân hàng. *Tạp chí Ngân hàng*. https://tapchinganhang.gov.vn/xu-huong-ung-dung-tri-tue-nhan-tao-trong-su-phat-trien-cua-nganh-ngan-hang-10697.html

European Commission. (2021). *Proposal for a regulation on artificial intelligence (AI Act)*. Retrieved September 19, 2025, from https://artificialintelligenceact.eu

Finocchiaro, G. (2024). The regulation of artificial intelligence. *AI & Society, 39*, 1961- 1968. https://doi.org/10.1007/s00146-023-01650-z

Giudici, P., Centurelli, M., & Turchetta, S. (2024). Artificial intelligence risk measurement. *Expert Systems with Applications, 235*, 121220. https://doi.org/10.1016/j.eswa.2023.121220

Global Risk Institute. (2025). *Responsible use of AI in financial services: Balancing risks and opportunities*. Retrieved from https://globalriskinstitute.org/publication/responsible-use-of-ai-in-financial-services/

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*. https://arxiv.org/pdf/1412.6572

Government of Canada. (2024). *OSFI–FCAC risk report: AI uses and risks at federally regulated financial institutions*. Retrieved September 19, 2025, from https://www.osfi-bsif.gc.ca/en/about-osfi/reports-publications/osfi-fcac-risk-report-ai-uses-risks-federally-regulated-financial-institutions

LNT & Partners. (2024). *Twin transition in Vietnam's financial sector: Green and digital transitions*. LNT & Partners. Retrieved September 19, 2025, from https://www.lntpartners.com/vi/legal-briefing/en-twin-transition-in-vietnams-financial-sector---green-and-digital-transitions-vi-chuyen-doi-kep-trong-linh-vuc-tai-chinh-viet-nam---chuyen-doi-xanh-va-chuyen-

Manikandan, A. (2025, August 13). I*ndia central bank committee recommends AI framework for finance sector.* Reuters. Retrieved September 19, 2025, from https://www.reuters.com/sustainability/boards-policy-regulation/india-cenbank-committee-recommends-ai-framework-finance-sector-2025-08-13/

NIST. (2023). *AI Risk Management Framework (AI RMF 1.0).* National Institute of Standards and Technology. Retrieved September 19, 2025 from https://www.nist.gov/itl/ai-risk-management-framework

Nguyễn Minh Hải. (2025). Ứng dụng AI trong ngân hàng: Đổi mới dịch vụ và trải nghiệm khách hàng. *VNPT A*I. Truy cập 19/09/2025 từ https://vnptai.io/vi/blog/detail/ung-dung-ai-trong-ngan-hang

OECD. (2019). OECD Principles on Artificial Intelligence. Retrieved from https://www.oecd.org/en/topics/artificial-intelligence.html

Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. *ACM Computing Surveys, 55*(3), 1-44. https://doi.org/10.1145/3494672

Quốc hội. (2018). *Luật An ninh mạng số 24/2018/QH14*. Hà Nội. https://vanban.chinhphu.vn/?pageid=27160&docid=206114

Raji, I. D., Kumar, I. E., Horowitz, A., & Selbst, A. (2022). The fallacy of AI functionality. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 959-972. https://doi.org/10.1145/3531146.3533158

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence, 1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy* (SP) (pp. 3-18). IEEE. https://doi.org/10.1109/SP.2017.41

Vietnam Briefing. (2025). Vietnam's AI Sector in 2025: Regulatory Frameworks and Opportunities for Investors. Truy cập từ https://www.vietnam-briefing.com/news/vietnams-ai-sector-in-2025-regulatory-frameworks-and-opportunities-for-investors.html

World Economic Forum. (2019). *AI Governance: A Holistic Approach to Implement Ethics into A*I [White paper]. Truy cập từ https://www.weforum.org/publications/ai-governance-a-holistic-approach-to-implement-ethics-into-ai/