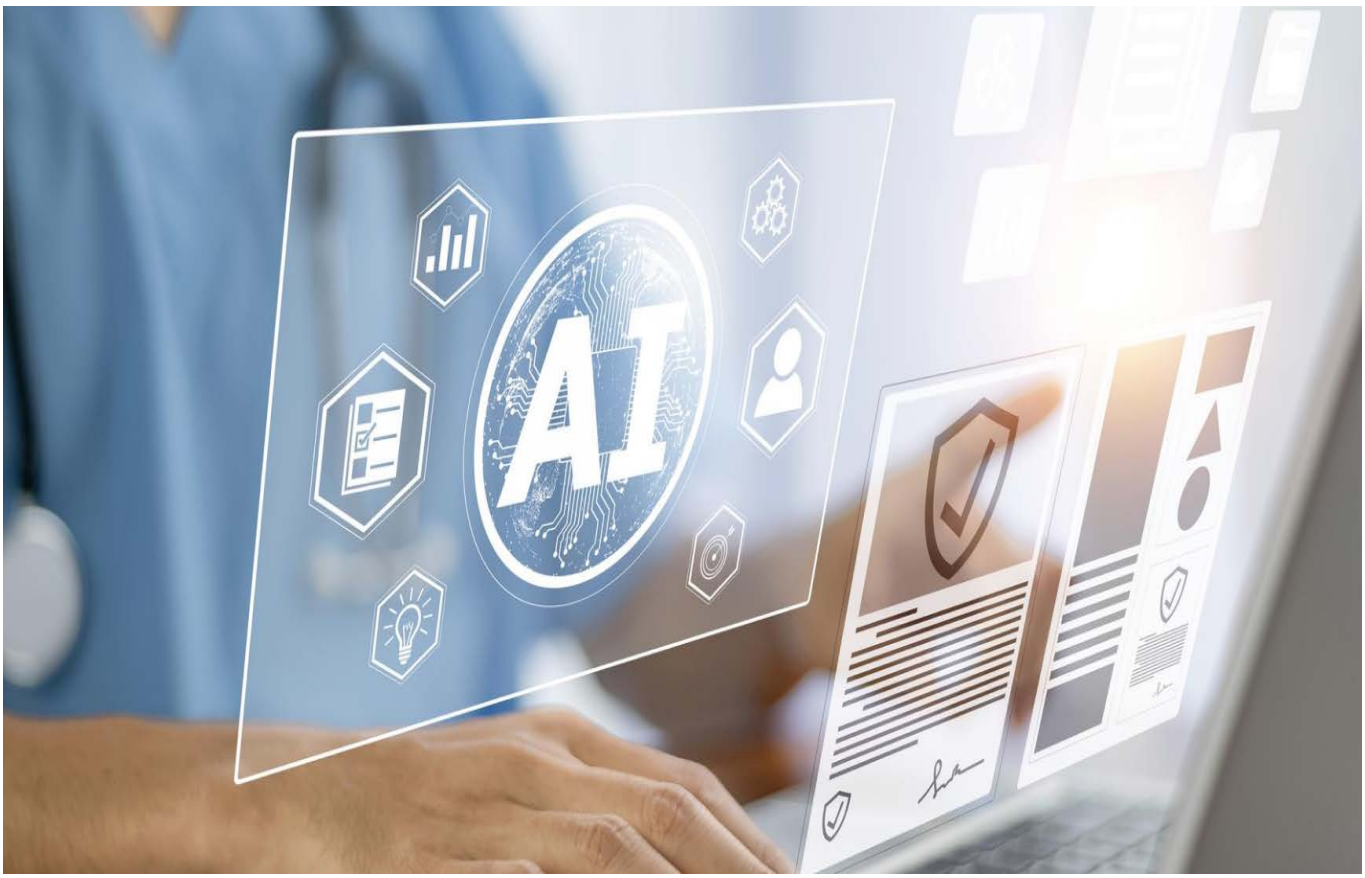


QUẢN TRỊ TRÍ TUỆ NHÂN TẠO VÀ CÁC THÁCH THỨC ĐẠO ĐỨC TRONG NGHỀ KIỂM TOÁN: KHUNG KHỔ LÝ THUYẾT VÀ GIẢI PHÁP THỰC TIỄN TRONG BỐI CẢNH CHUYỂN ĐỔI SỐ

● HÀ MINH TUẤN*

Trong kỷ nguyên Cách mạng Công nghiệp 4.0, trí tuệ nhân tạo (AI) đang tái cấu trúc sâu sắc quy trình kiểm toán, chuyển dịch từ phương pháp chọn mẫu truyền thống sang giám sát liên tục dựa trên dữ liệu lớn. Tuy nhiên, sự phụ thuộc vào thuật toán làm nảy sinh các thách thức đạo đức nghiêm trọng như thiên kiến tự động hóa, tính minh bạch của "hộp đen" và rủi ro bảo mật dữ liệu. Bài báo này nghiên cứu mối quan hệ giữa Quản trị AI (AI Governance) và đạo đức nghề nghiệp, phân tích các khung pháp lý quốc tế (EU AI Act, NIST, ISO 42001) và Luật Trí tuệ nhân tạo số 134/2025/QH15 tại Việt Nam. Kết quả nghiên cứu đề xuất một khung quản trị đa tầng nhằm bảo vệ tính chính trực và sự tin cậy của công chúng đối với nghề kiểm toán trong môi trường số.

Từ khóa: Quản trị AI, đạo đức kiểm toán, kiểm toán 4.0, ISO/IEC 42001, Luật AI 2025.

*Trường Đào tạo và Bồi dưỡng nghiệp vụ kiểm toán; Email: tuanhm@sav.gov.vn

AI governance and ethical challenges in the auditing profession: A theoretical framework and practical solutions in the context of digital transformation

In the Industry 4.0 era, Artificial Intelligence (AI) is profoundly restructuring auditing processes, shifting from traditional sampling methods to continuous monitoring based on Big Data. However, reliance on algorithms gives rise to serious ethical challenges, such as automation bias, "black box" transparency issues and data security risks. This paper examines the relationship between AI Governance and professional ethics, analyzing international legal frameworks (EU AI Act, NIST, ISO 42001) and Vietnam's Law on Artificial Intelligence No. 134/2025/QH15. The research findings propose a multi-layered governance framework to safeguard integrity and public trust in the auditing profession in a digital environment.

Keywords: AI Governance, audit ethics, Audit 4.0, ISO/IEC 42001, Law on AI 2025.

JEL classification: M42, O33, D83

https://doi.org/10.65771/ati-jas.03202603

1. Đặt vấn đề

Sự trỗi dậy của AI đã trở thành động lực then chốt vận hành hệ thống tài chính toàn cầu. Trong lĩnh vực kiểm toán, công nghệ này cho phép phân tích 100% dữ liệu giao dịch trong thời gian thực, đánh dấu bước ngoặt từ "kiểm toán sau" sang "giám sát liên tục".

Tuy nhiên, đi cùng với làn sóng đổi mới đó là không ít mối lo ngại. AI có thể vô tình tái tạo định kiến xã hội thông qua dữ liệu thiên lệch, can thiệp quá sâu vào đời tư cá nhân, hoặc bị lợi dụng để thao túng thông tin, tạo ra tin giả. Một số hệ thống AI phức tạp thậm chí vận hành theo cách mà ngay cả nhà phát triển cũng khó lý giải hoặc kiểm soát hoàn toàn. Khi thuật toán thay thế phán quyết của con người, các câu hỏi về tính độc lập, khách quan và trách nhiệm giải trình trở nên hóc búa hơn bao giờ hết. Do đó, nghiên cứu về quản trị AI là yêu cầu cấp thiết để đảm bảo công nghệ phục vụ lợi ích con người một cách an toàn và minh bạch.

2. Quản trị trí tuệ nhân tạo

2.1. Khái niệm

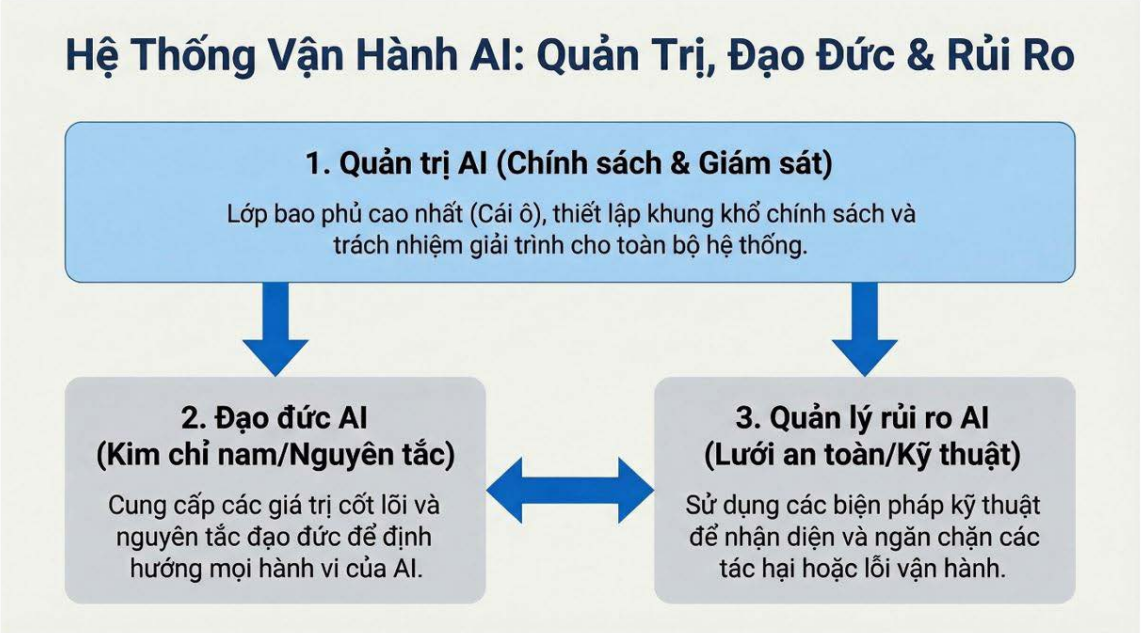
Quản trị AI không dựa trên một khái niệm đơn nhất mà là một hệ sinh thái đa tầng bao gồm chính sách, quy trình và các cân nhắc đạo đức cần thiết để giám sát việc phát triển, triển khai và duy trì các hệ thống AI. Nó đóng vai trò như một "hệ điều hành của niềm tin", thiết lập các rào chắn (guardrails) để đảm bảo AI vận hành trong giới hạn pháp lý và giá trị tổ chức.

Một định nghĩa mang tính thực thi tổ chức thường xem quản trị AI là hệ thống các quy tắc, thực hành và quy trình được sử dụng để đảm bảo việc sử dụng công nghệ AI của một tổ chức phù hợp với chiến lược, mục tiêu và giá trị của tổ chức đó, đồng thời đáp ứng các yêu cầu pháp lý, nguyên tắc đạo đức và kỳ vọng của các bên liên quan.

Khung quản trị AI cung cấp một cách tiếp cận có cấu trúc để giải quyết vấn đề minh bạch, trách nhiệm giải trình và công bằng, cũng như thiết lập các tiêu chuẩn về xử lý dữ liệu, khả năng giải thích mô hình và quy trình ra quyết định. Thông qua quản trị AI, các tổ chức thúc đẩy đổi mới AI có trách nhiệm đồng thời giảm thiểu rủi ro liên quan đến thiên kiến, vi phạm quyền riêng tư và các mối đe dọa an ninh.

Ngoài ra, cũng cần phân biệt giữa quản trị AI với đạo đức AI hoặc quản lý rủi ro AI. Trên thực tế, ba khái niệm này tồn tại trong một mối quan hệ thứ bậc và tương hỗ chặt chẽ. Đạo đức AI cung cấp nền tảng giá trị; quản trị AI thiết lập khung chiến lược và thực thi và quản lý rủi ro AI cung cấp các công cụ kỹ thuật để kiểm soát các tác hại cụ thể.

Hình 1: So sánh các khái niệm



2.2. Các trụ cột cốt lõi

Một khung quản trị AI vững chắc dựa trên 5 trụ cột then chốt: (1) Tính minh bạch, (2) Công bằng, (3) Trách nhiệm giải trình, (4) Bảo mật & Quyền riêng tư, và (5) An toàn & Bền vững. Đặc biệt, trong kiểm toán, "khả năng giải thích" (Explainability) là yêu cầu tiên quyết để có thể thẩm định logic đằng sau các quyết định của máy móc nhằm đảm bảo trách nhiệm giải trình của kiểm toán viên khi sử dụng các kết quả do hệ thống AI phân tích.

Hình 2: Các trụ cột của khung quản trị AI

Nguồn: Tổng hợp

2.3. Các khung pháp lý và tiêu chuẩn quốc tế

Hiện nay, có nhiều mô hình quản trị khác nhau, mỗi mô hình đều có những ưu điểm và hạn chế riêng tuỳ thuộc vào mục tiêu chiến lược của mỗi quốc gia hoặc khu vực. Theo Diễn đàn kinh tế thế giới, có 4 mô hình chủ yếu sau:

Bảng 1: Ma trận so sánh các cách tiếp cận quản trị AI toàn cầu

Cách tiếp cận	Định nghĩa	Ưu điểm	Hạn chế
Dựa trên rủi ro (Risk-based)	Phân loại và ưu tiên các rủi ro liên quan đến tác hại tiềm ẩn của hệ thống AI.	Tối ưu hóa nguồn lực giám sát vào các lĩnh vực nguy hiểm nhất.	Khó khăn trong việc định lượng rủi ro mới nổi.
Dựa trên quy tắc (Rules-based)	Đưa ra các quy tắc, tiêu chuẩn và yêu cầu chi tiết, cụ thể cho các hệ thống AI.	Độ tin cậy pháp lý cao, dễ dàng kiểm tra tuân thủ.	Có thể kìm hãm đổi mới, nhanh chóng trở nên lỗi thời.
Dựa trên nguyên tắc (Principles-based)	Thiết lập các hướng dẫn cơ bản, để các tổ chức tự diễn giải và triển khai chi tiết.	Linh hoạt, thích ứng tốt với sự thay đổi công nghệ.	Thiếu tính cưỡng chế, dễ dẫn đến "rửa đạo đức" (ethics washing).
Dựa trên kết quả (Outcomes-based)	Tập trung vào việc đạt được các kết quả mong muốn mà không quy định quy trình cụ thể.	Khuyến khích sự sáng tạo trong cách giải quyết vấn đề.	Khó khăn trong việc đo lường và quy trách nhiệm giải trình.

Nguồn: Tổng hợp

Việc lựa chọn mô hình quản trị thường phản ánh triết lý chính trị và kinh tế của quốc gia đó. Liên minh châu Âu (EU) đã tiên phong với Đạo luật AI dựa trên rủi ro, trong khi Vương quốc Anh ưu tiên mô hình dựa trên nguyên tắc để duy trì tính cạnh tranh về công nghệ và Hoa Kỳ kết hợp các tiêu chuẩn tự nguyện với các sắc lệnh hành pháp tập trung vào an ninh quốc gia.

2.3.1. Đạo luật AI của Liên minh châu Âu (EU AI Act)

Đạo luật AI của Liên minh châu Âu, được đưa ra vào năm 2024, đã trở thành văn bản pháp lý toàn diện đầu tiên trên thế giới điều chỉnh AI. Đạo luật này không chỉ giới hạn trong phạm vi EU mà còn áp dụng cho các nhà cung cấp và triển khai bên ngoài EU nếu đầu ra của hệ thống AI của họ được sử dụng trong khối. Điều này tạo ra một tiêu chuẩn toàn cầu mới, buộc các công ty công nghệ lớn phải tuân thủ nếu muốn tiếp cận thị trường châu Âu.

Hệ thống phân tầng rủi ro của EU AI Act là tâm điểm của kiến trúc quản trị này, nhằm cân bằng giữa việc bảo vệ quyền cơ bản và thúc đẩy đổi mới sáng tạo, phân loại rủi ro AI thành 4 cấp độ:

- Rủi ro không thể chấp nhận được: Bị cấm hoàn toàn (ví dụ: chấm điểm xã hội).
- Rủi ro cao: Cần kiểm soát nghiêm ngặt (ví dụ: AI trong hạ tầng trọng yếu, tài chính, kiểm toán).
- Rủi ro hạn chế: Yêu cầu tính minh bạch (ví dụ: Chatbots).
- Rủi ro tối thiểu: Không bị điều chỉnh đáng kể.

2.3.2. Khung quản trị rủi ro AI của NIST (Hoa Kỳ)

Khung này tập trung vào kết quả đầu ra thay vì đưa ra các yêu cầu kỹ thuật cứng nhắc, giúp nó không bị lỗi thời trước tốc độ thay đổi nhanh chóng của công nghệ AI. Khung cung cấp một phương pháp tiếp cận linh hoạt cho các tổ chức để nhận diện, đo lường và quản lý rủi ro AI dựa trên bốn chức năng chính giúp các tổ chức giải quyết các rủi ro hệ thống trong suốt vòng đời của AI:

• Quản trị (Govern): Chức năng này tập trung vào việc thiết lập văn hóa quản lý rủi ro trong tổ chức. Nó bao gồm việc xác định vai trò, trách nhiệm, xây dựng các chính sách, quy trình và phân bổ nguồn lực cần thiết để đảm bảo tính trách nhiệm giải trình.

• Ánh xạ (Map): Giúp tổ chức hiểu rõ bối cảnh sử dụng AI. Bước này yêu cầu xác định các bên liên quan, mục đích dự kiến của hệ thống và các rủi ro tiềm ẩn cụ thể cho từng trường hợp sử dụng.

• Đo lường (Measure): Sử dụng các phương pháp định lượng và định tính để đánh giá hiệu suất, độ tin cậy, thiên kiến và an ninh của hệ thống AI. Chức năng này giúp kiểm chứng các đặc tính đáng tin cậy của mô hình thông qua kiểm thử thường xuyên.

• Quản lý (Manage): Tập trung vào việc triển khai các biện pháp kiểm soát để ưu tiên phản ứng với các rủi ro đã được ánh xạ và đo lường. Tổ chức cần phân bổ đủ nguồn lực để xử lý các mối nguy hiểm và thực hiện giám sát liên tục hệ thống đang vận hành.

Điểm mạnh của NIST là sự tập trung vào "độ tin cậy" (trustworthiness), bao gồm 7 đặc tính: tính hợp lệ và độ tin cậy, an toàn, an ninh và khả năng phục hồi, tính trách nhiệm và minh bạch, khả năng giải thích, quyền riêng tư và tính công bằng.

2.3.3. Các nguyên tắc AI của OECD

Các nguyên tắc này tập trung vào cách thiết kế và triển khai AI để tối ưu hóa lợi ích và giảm thiểu rủi ro cho xã hội:

• Tăng trưởng bao trùm, phát triển bền vững và hạnh phúc: AI cần được phát triển để mang lại lợi ích cho con người và hành tinh bằng cách thúc đẩy sự thịnh vượng, giảm bất bình đẳng và bảo vệ môi trường.

• Các giá trị nhân văn và sự công bằng: Các hệ thống AI phải tuân thủ thượng tôn pháp luật, quyền con người và các giá trị dân chủ, bao gồm tính tự chủ, phẩm giá, quyền riêng tư và không phân biệt đối xử.

• Minh bạch và khả năng giải thích: Cần đảm bảo các bên liên quan hiểu được khi nào họ đang tương tác với AI và có thể hiểu được cách thức hệ thống đưa ra kết quả để có thể thách thức các đầu ra của AI.

• Độ tin cậy, an ninh và an toàn: Các hệ thống AI phải hoạt động một cách mạnh mẽ, an toàn và bảo mật trong suốt vòng đời của chúng, đảm bảo các rủi ro tiềm ẩn được đánh giá và quản lý liên tục.

• Trách nhiệm giải trình: Những người tham gia vào chu kỳ sống của AI phải chịu trách nhiệm về việc vận hành đúng đắn các hệ thống này dựa trên các nguyên tắc nêu trên.

2.3.4. Tiêu chuẩn ISO/IEC 42001 (AIMS)

Đây là tiêu chuẩn quốc tế đầu tiên về Hệ thống Quản lý AI (AI Management System - AIMS) có thể được chứng nhận. Khác với NIST mang tính hướng dẫn, ISO/IEC 42001 cung cấp một "hệ điều hành" cho quản trị, cho phép các tổ chức xây dựng các quy trình có thể kiểm toán được. Việc kết hợp EU AI Act (như một bộ quy tắc) và ISO/IEC 42001 (như một hệ điều hành) đang trở thành chiến lược tối ưu cho các doanh nghiệp toàn cầu để chuyển từ trạng thái "loay hoay tuân thủ" sang "vận hành ổn định".

2.3.5. Luật AI tại Việt Nam

Luật Trí tuệ nhân tạo 2025, được thiết kế dựa trên 3 nguyên tắc: Quản lý dựa trên rủi ro, trung lập về công nghệ và luật khung linh hoạt.

Khía cạnh	Chi tiết Luật AI 2025	Ý nghĩa
Phân loại rủi ro	3 mức: Rủi ro cao, Rủi ro trung bình, Rủi ro thấp.	Tương thích với tiêu chuẩn EU và OECD.
Minh bạch nội dung	Bắt buộc gắn nhãn và watermark cho âm thanh, hình ảnh do AI tạo ra.	Chống tin giả và bảo vệ quyền tác giả.
Cơ quan điều phối	Thành lập Cục Trí tuệ nhân tạo quốc gia	Thống nhất quản lý nhà nước về AI.
Cơ chế Sandbox	Thử nghiệm có kiểm soát trong y tế, giáo dục, tài chính.	Thúc đẩy đổi mới trong môi trường an toàn.
Hệ thống cũ	Phải hoàn thành nghĩa vụ tuân thủ trước 01/03/2027.	Đảm bảo tính kế thừa và chuyển tiếp pháp lý.

3. Sự chuyển dịch mô hình kiểm toán và thách thức đạo đức

3.1. Từ kiểm toán truyền thống đến kiểm toán 4.0

AI đã thay đổi gốc rễ cách tiếp cận kiểm toán, từ dựa trên chứng từ giấy và kiểm tra thủ công trên cơ sở chọn mẫu để phát hiện các gian lận và sai sót và sử dụng máy tính và các phần mềm hỗ trợ (CAATs) để xử lý dữ liệu số hóa sang sử dụng thuật toán để phân tích 100% dữ liệu giao dịch trong thời gian thực, cho phép chuyển từ "kiểm toán sau" sang "giám sát liên tục".

Sự giao thoa giữa AI và kiểm toán không chỉ là việc thay đổi công cụ, mà là một sự tái định nghĩa về quy trình tạo ra sự tin cậy tài chính.

Bảng 2: Các ứng dụng chủ đạo của AI trong quy trình kiểm toán

Công nghệ	Ứng dụng cụ thể trong kiểm toán	Lợi ích mang lại
Xử lý ngôn ngữ tự nhiên (NLP)	Đọc và tóm tắt hàng ngàn hợp đồng kinh tế, điều khoản, bảng kê.	Phát hiện nhanh các điều khoản bất lợi hoặc sai lệch chuẩn mực kế toán.
Học máy (Machine Learning)	Phân định các giao dịch bất thường dựa trên quy luật lịch sử.	Nhận diện gian lận tinh vi mà mắt thường hoặc các hàm Excel không thấy được.
Thị giác máy tính (Computer Vision)	Kiểm kê hàng tồn kho qua drone hoặc hình ảnh vệ tinh.	Giảm thiểu sai sót con người và tiết kiệm thời gian kiểm kê thực địa.
Phân tích dự báo (Predictive Analytics)	Đánh giá khả năng hoạt động liên tục (going concern) của doanh nghiệp.	Đưa ra cảnh báo sớm về rủi ro phá sản dựa trên biến động thị trường.

3.2. Các thách thức đạo đức trọng yếu

Việc áp dụng Trí tuệ nhân tạo (AI) vào kiểm toán không đơn thuần là áp dụng một công cụ làm việc mới, mà là một sự thay đổi về tư duy nghiệp vụ. Theo Liên đoàn Kế toán Quốc tế (IFAC), các nguyên tắc đạo đức cốt lõi - Chính trực, Khách quan, Năng lực chuyên môn và Thận trọng, Tính bảo mật và Tư cách nghề nghiệp - đang đứng trước những tác động đến từ những thay đổi mạnh mẽ trong mô hình rủi ro kiểm toán, từ việc thay đổi cách nhận diện rủi ro truyền thống đến việc phát sinh các loại rủi ro mới đặc thù của công nghệ.

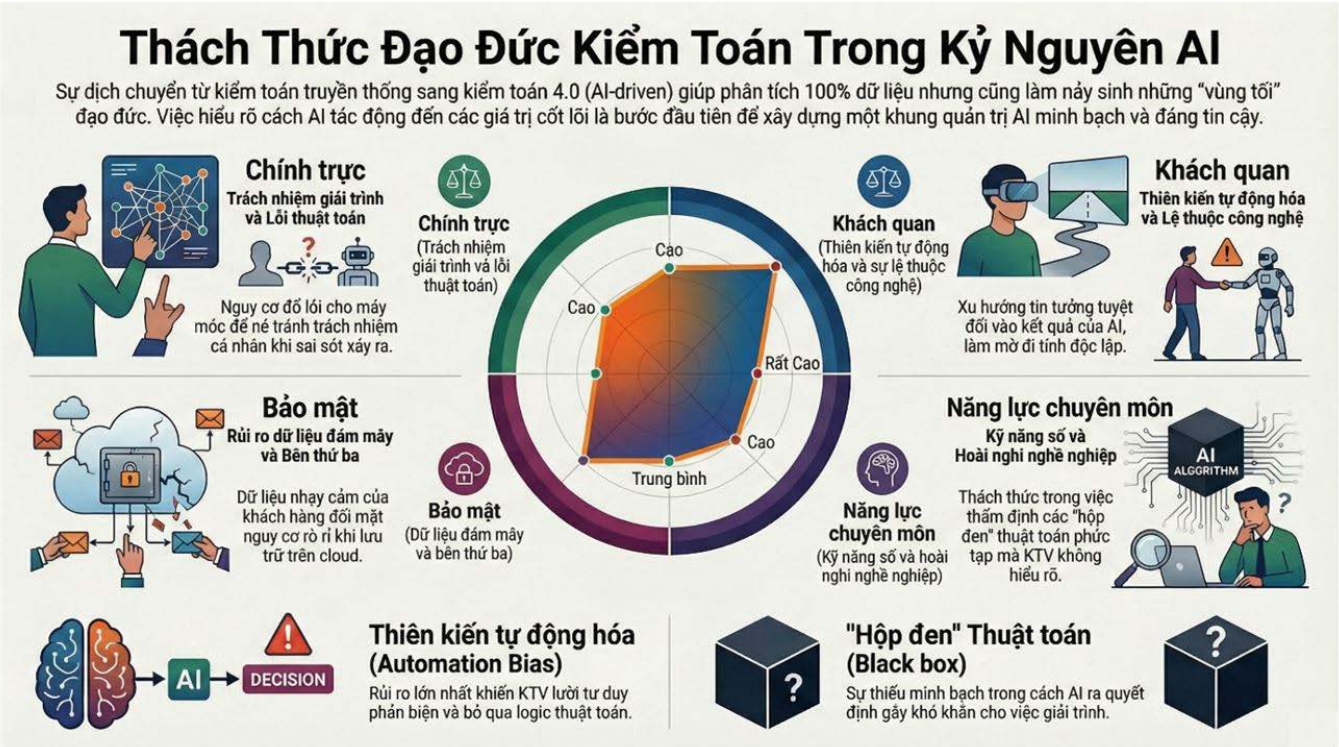
● Xói mòn tính hoài nghi nghề nghiệp: Thiên kiến tự động hóa (Automation Bias) khiến kiểm toán viên tin tưởng tuyệt đối vào kết quả của thuật toán mà không kiểm tra lại logic.

● Hộp đen và trách nhiệm giải trình: Khi AI không thể giải thích tại sao một giao dịch bị gắn cờ, kiểm toán viên rơi vào thế tiến thoái lưỡng nan về bằng chứng.

● Xung đột lợi ích: Việc các công ty kiểm toán vừa cung cấp dịch vụ tư vấn AI vừa thực hiện kiểm toán đơn vị đó tạo ra nguy cơ "vừa đá bóng vừa thổi còi".

● Bảo mật dữ liệu: Rủi ro rò rỉ thông tin nhạy cảm của đơn vị được kiểm toán khi đưa dữ liệu lên nền tảng đám mây của bên thứ ba.

Hình 3: Thách thức đạo đức kiểm toán trong kỷ nguyên AI



4. Khung quản trị AI trong nghề kiểm toán tại Việt Nam

4.1. Định hướng quản trị AI trong lĩnh vực kiểm toán

Với những rủi ro đã nêu, quản trị AI trong nghề kiểm toán sẽ đóng vai trò là một "hệ điều hành của niềm tin", cung cấp khung khổ để chuyển hóa các nguyên tắc đạo đức trừu tượng thành các cơ chế có thể kiểm tra, xác thực và cưỡng chế thực thi và do đó đảm bảo tính có thể kiểm tra, kiểm toán được. Điều này cần được thể hiện qua các nguyên tắc sau:

(i) Tạo lập nền tảng cho sự minh bạch và khả năng truy xuất: Các hệ thống AI thường vận hành như những "hộp đen" (black boxes) khó giải thích, gây thách thức lớn cho việc kiểm tra truyền thống. Quản trị AI cần thiết lập các dấu vết kiểm tra không thể thay đổi (immutable audit trails), ghi lại mọi hành động, quyết định và sự chuyển đổi của mô hình. Điều này đảm bảo tính truy xuất nguồn gốc (traceability) không thể chối cãi, một yêu cầu then chốt để thực hiện chức năng giám sát.

(ii) Đáp ứng các nghĩa vụ pháp lý và tuân thủ bắt buộc: Với sự ra đời của các khung pháp lý như Đạo luật AI của EU (EU AI Act) hay Luật AI 2025 của Việt Nam, việc kiểm toán không còn là hoạt động tự nguyện mà là nghĩa vụ bắt buộc đối với các hệ thống AI rủi ro cao. Luật AI 2025 yêu cầu các hệ thống rủi ro cao phải được các tổ chức chứng nhận đánh giá sự phù hợp trước khi vận hành.

(iii) Cung cấp bộ công cụ quản lý rủi ro kỹ thuật: Khung khổ quản trị cần quy định các công cụ để đánh giá các rủi ro đặc thù mà công nghệ thông tin truyền thống không có như: Phát hiện thiên kiến (Bias detection), thiết lập hạ tầng kiểm thử để đo lường tính công bằng và ngăn chặn các đầu ra phân biệt đối xử; Giám sát sự sai lệch (Model drift) đảm bảo mô hình được theo dõi liên tục để phát hiện sự sụt giảm hiệu suất theo thời gian; Đảm bảo an ninh mô hình: đánh giá khả năng chống lại các cuộc tấn công đối nghịch (adversarial attacks) và bảo vệ dữ liệu nhạy cảm.

(iv) Tiêu chuẩn hóa quy trình theo các khuôn khổ được chấp thuận như ISO/IEC 42001. Sự ra đời của tiêu chuẩn ISO/IEC 42001 là một bước ngoặt lớn cho nghề kiểm toán. Đây là tiêu chuẩn quốc tế đầu tiên về "Hệ thống quản lý AI" (AIMS) có thể được chứng nhận, cho phép các kiểm toán viên xây dựng các quy trình kiểm tra có cấu trúc, lặp lại được và dựa trên bằng chứng cụ thể thay vì chỉ kiểm tra tài liệu trên giấy tờ. Việc kết hợp tiêu chuẩn này với các khung pháp lý giúp chuyển đổi trạng thái từ "loay hoay tuân thủ" sang "vận hành ổn định".

Như vậy, quản trị AI là nền tảng không thể thiếu để nghề kiểm toán thích nghi với sự phát triển của công nghệ. Trong nghề kiểm toán, quản trị AI không chỉ là vấn đề tuân thủ công nghệ mà là bảo vệ giá trị giá trị cốt lõi của nghề kiểm toán, đó là niềm tin của công chúng dựa trên bảo vệ uy tín nghề nghiệp và phù hợp với xu hướng chuyển đổi số.

4.2. Tổ chức thực hiện

Để AI thực sự trở thành công cụ hỗ trợ đắc lực hoạt động kiểm toán thay vì tạo ra các rủi ro đạo đức, cần có một cách tiếp cận đa tầng từ cấp độ vĩ mô, tổ chức cho đến từng cá nhân kiểm toán viên.

Một là, cập nhật chuẩn mực đạo đức nghề nghiệp. Các tổ chức nghề nghiệp cần tham gia thiết lập "hành lang pháp lý đạo đức".

- Bổ sung hướng dẫn về "Đạo đức số" (Digital Ethics): Cần ban hành các văn bản hướng dẫn chi tiết cách áp dụng 5 nguyên tắc đạo đức cơ bản trong bối cảnh sử dụng AI. Ví dụ: Định nghĩa lại thế nào là "Năng lực chuyên môn" khi kiểm toán viên sử dụng thuật toán phức tạp.

- Thiết lập tiêu chuẩn kiểm chứng AI: Xây dựng bộ tiêu chuẩn đánh giá tính khách quan và tin cậy của các phần mềm kiểm toán dựa trên AI. Một công cụ AI chỉ nên được phép sử dụng trong kiểm toán nếu nó đạt được các chứng chỉ về tính minh bạch và bảo mật dữ liệu.

- Hợp tác xuyên ngành: Các tổ chức kiểm toán cần làm việc chặt chẽ với các chuyên gia công nghệ và luật sư để dự báo các rủi ro pháp lý phát sinh từ AI, từ đó cập nhật kịp thời các quy định về trách nhiệm giải trình.

Hai là, xây dựng khung quản trị AI tại cấp độ tổ chức. Cần thiết lập một hệ thống kiểm soát nội bộ nghiêm ngặt đối với các công cụ AI. Trong đó:

● Áp dụng mô hình "Con người kiểm soát" (Human-in-the-loop): Quy trình kiểm toán phải được thiết kế sao cho mọi kết luận quan trọng do AI đưa ra đều phải qua sự phê duyệt cuối cùng của một kiểm toán viên có kinh nghiệm. AI cung cấp bằng chứng, nhưng con người đưa ra phán quyết.

● Thẩm định mô hình định kỳ (Model Validation): Các công cụ AI không được coi là "bất biến". Công ty cần có bộ phận độc lập thực hiện kiểm định lại thuật toán định kỳ để phát hiện hiện tượng "lệch mô hình" (model drift) hoặc sự xuất hiện của các định kiến mới trong dữ liệu.

● Thiết lập hồ sơ theo dõi (audit trail) để ghi lại các quyết định của AI, đảm bảo có thể truy xuất nguồn gốc và xác định trách nhiệm khi có vấn đề phát sinh đảm bảo việc kiểm chứng tính giải thích được của kết quả.

● Thiết lập quy định về công bố mức độ phụ thuộc vào AI của ý kiến kiểm toán tại báo cáo kiểm toán để đảm bảo minh bạch, xây dựng niềm tin bền vững, thực hiện trách nhiệm giải trình và đảm bảo tuân thủ các khuôn khổ pháp lý.

Ba là, nâng cao năng lực và thay đổi tư duy của ở cấp độ cá nhân kiểm toán viên. Kiểm toán viên không cần trở thành lập trình viên, nhưng họ cần có "năng lực số" để không bị lệ thuộc vào công nghệ.

● Kỹ năng "Hoài nghi thuật toán": Đào tạo kiểm toán viên cách đặt câu hỏi ngược lại với kết quả của AI. Thay vì chấp nhận kết quả, kiểm toán viên phải biết đặt câu hỏi: "Tại sao AI lại chọn giao dịch này?" hoặc "Dữ liệu nào có thể đã bị AI bỏ sót?".

● Đào tạo về AI giải thích được (Explainable AI - XAI): kiểm toán viên cần được trang bị kiến thức để sử dụng các công cụ giải thích thuật toán, giúp biến các "hộp đen" thành các quy trình có thể hiểu và diễn giải được cho khách hàng hoặc các bên điều tra.

● Giáo dục đạo đức trong kỷ nguyên số: Các chương trình đào tạo không chỉ tập trung vào kỹ năng sử dụng phần mềm mà phải nhấn mạnh vào các tình huống đạo đức khi máy móc và con người có ý kiến trái chiều.

5. Kết luận

AI là "cánh cửa" mở ra hiệu suất nhưng cũng là "thách thức" đối với giá trị truyền thống. Các nghiên cứu chỉ ra rằng thách thức đạo đức lớn nhất không đến từ thuật toán mà đến từ sự lơ là của con người. Mô hình "kiểm toán viên lai" (Hybrid Auditor) sẽ là chuẩn mực mới, nơi AI là bộ não nhưng con người là yếu tố quyết định. Nghề kiểm toán sẽ không biến mất, nhưng nó sẽ được tái định nghĩa. Để duy trì vai trò là "người gác cổng" cho niềm tin của thị trường tài chính, các kiểm toán viên và các tổ chức nghề nghiệp phải chủ động xây dựng một hành lang đạo đức số vững chắc. Quản trị AI không nhằm hạn chế công nghệ, mà để đảm bảo đạo đức của nghề nghiệp không bị bỏ lại phía sau trong tương lai kỹ thuật số.❑

TÀI LIỆU THAM KHẢO

1. Palo Alto Networks, *What Is AI Governance?,* 2026;
2. Lumenova AI, *AI Governance Platform vs AI Risk Management Tool,* 2026;
3. World Economic Forum, *Generative AI Governance: Shaping a Collective Global Future,* 2024;
4. ModelOp, *EU AI Act: Summary & Compliance Requirements,* 2026;
5. NIST, *AI Risk Management Framework,* 2021;
6. Binhdanhocvuso.lamdong.gov.vn, *Luật Trí tuệ nhân tạo trên thế giới và Việt Nam,* 2026;
7. Thư viện Pháp luật, *Luật Trí tuệ nhân tạo 2025 số 134/2025/QH15;*
8. A. Taeihagh, *Governance of Generative AI,* Policy Soc., 2025;
9. LuatVietnam, *Luật Trí tuệ nhân tạo 2025 và 10+ điểm đáng chú ý;*
10. ISACA, *2025 ISOIEC 42001 and EU AI Act: A Practical Pairing.*

Ngày nhận bài: 11/02/2026
Ngày chỉnh sửa: 12/02/2026
Ngày duyệt đăng: 13/03/2026