

# Xây dựng niềm tin kỹ thuật số trong kỷ nguyên trí tuệ nhân tạo

Nguyễn Văn Hoàng<sup>1</sup>

Ngày nhận bài: 22/02/2026 | Ngày gửi phản biện: 26/01/2026 | Ngày duyệt đăng: 12/02/2026

**Tóm tắt:** Trí tuệ nhân tạo (Artificial Intelligence – AI), đặc biệt là AI tạo sinh (Generative AI), đang trở thành hạ tầng cốt lõi cho đổi mới sáng tạo, tăng trưởng kinh tế và quản trị xã hội. Tuy nhiên, song hành với tiềm năng to lớn là sự suy giảm niềm tin của xã hội vào công nghệ số, xuất phát từ các lo ngại về xâm phạm quyền riêng tư, sai lệch thuật toán, thiếu minh bạch và trách nhiệm giải trình không rõ ràng. Bài viết phân tích khái niệm niềm tin kỹ thuật số trong bối cảnh AI, giải thích tại sao niềm tin kỹ thuật số lại quan trọng, từ đó, đề xuất một số giải pháp để xây dựng niềm tin kỹ thuật số trong kỷ nguyên trí tuệ nhân tạo.

**Từ khóa:** AI tạo sinh, an ninh mạng, niềm tin kỹ thuật số, trí tuệ nhân tạo, quyền riêng tư.

## Building Digital Trust in the Age of Artificial Intelligence

**Abstract:** Artificial intelligence (AI), particularly generative AI, is becoming a core infrastructure for innovation, economic growth, and social governance. However, alongside its immense potential comes a decline in societal trust in digital technology, stemming from concerns about privacy violations, algorithmic biases, lack of transparency, and unclear accountability. This article analyzes the concept of digital trust in the context of AI, explains why digital trust is important, and proposes solutions for building digital trust in the age of artificial intelligence.

**Keywords:** Generative AI, cybersecurity, digital trust, artificial intelligence, privacy.

### 1. Mở đầu

Trong thập kỷ qua, trí tuệ nhân tạo đã chuyển từ vai trò công nghệ hỗ trợ sang vị trí trung tâm của quá trình ra quyết định trong nhiều lĩnh vực then chốt như tài chính - ngân hàng, tuyển dụng, y tế, giáo dục và dịch vụ công. Đặc biệt, sự bùng nổ của AI tạo sinh đã làm gia tăng tốc độ tự động hóa tri thức, cho phép máy móc tạo ra văn bản, hình ảnh, mã nguồn và thậm chí là các khuyến nghị mang tính chiến lược. AI vì vậy không chỉ là công cụ kỹ thuật mà còn là “tác nhân xã hội” có khả năng tác động trực tiếp đến quyền lợi, cơ hội và đời sống của con người.

Tuy nhiên, niềm tin của xã hội vào công nghệ AI không tăng tương xứng với tốc độ ứng dụng. Nhiều khảo sát quốc tế cho thấy người tiêu dùng ngày càng lo ngại về việc dữ liệu cá nhân bị thu thập và sử dụng ngoài tầm kiểm soát; khoảng 75% người được hỏi cho biết họ sẽ không mua sản phẩm hoặc dịch vụ từ những tổ chức mà họ không tin tưởng trong việc bảo vệ dữ liệu cá nhân. Những lo ngại này càng trở nên rõ nét khi AI được triển khai trong các quyết định có ảnh hưởng sâu rộng như đánh giá tín dụng, phân loại hồ sơ xin việc hay chẩn đoán y khoa. Điều này đặc biệt đúng khi các đối tượng xấu lạm dụng trí tuệ nhân tạo tạo sinh - dù là vì mục đích thương mại, tội phạm hay địa chiến lược. Kể từ khi các công cụ deepfake được sử dụng rộng rãi vào năm 2018, và đặc biệt là kể từ khi GPT-3 trở nên phổ biến vào năm 2022, việc tạo ra các nội dung giả mạo hiện nay có thể mở rộng quy mô vượt ngoài sức tưởng tượng. Bằng cách kết hợp các công cụ AI này, các đối tượng xấu có thể tạo ra các “bot” thực sự trên quy mô lớn và triển khai chúng như những người bình luận, người có ảnh hưởng, phóng viên, chuyên gia giả mạo... để đánh lạc hướng dư luận và làm mất niềm tin của người tiêu dùng. Do đó, việc xây dựng lại niềm tin kỹ thuật số trong kỷ nguyên trí tuệ nhân tạo là điều vô cùng cần thiết.

<sup>1</sup> Học viện Ngân hàng, Phân viện Bắc Ninh; Email: hoangnv.bn@hvn.edu.vn

## 2. Khái niệm niềm tin kỹ thuật số

Trong thời đại Cách mạng công nghiệp lần thứ tư, tỷ lệ áp dụng, khả năng truy cập dữ liệu và tính kết nối là những động lực thúc đẩy sự thành công của mọi doanh nghiệp. Đối với mỗi yếu tố đó, sự tin tưởng là yếu tố quyết định thành bại, và điều này đặc biệt đúng đối với những người phát triển và triển khai thể hệ công nghệ kỹ thuật số tiếp theo. Lòng tin không thể tự động hóa. Nó không phải là thứ sẽ xuất hiện từ môi trường được xây dựng, hay tự tìm thấy trong một dòng mã mới, hoặc được đào tạo bởi một mô hình ngôn ngữ lớn mới. Lòng tin là thứ mà những con người đứng sau công nghệ phải lựa chọn để giành được, và sau đó quyết định biến nó thành hiện thực thông qua công việc của họ. Đối với các công ty công nghệ thông tin, lòng tin là một lựa chọn phải được tích hợp vào mọi quyết định khác trong chu kỳ phát triển công nghệ.

Báo cáo “*Xây dựng niềm tin kỹ thuật số*” năm 2022 của Diễn đàn Kinh tế Thế giới định nghĩa niềm tin kỹ thuật số là kỳ vọng của cá nhân rằng các công nghệ và dịch vụ kỹ thuật số và các tổ chức cung cấp chúng sẽ bảo vệ lợi ích của tất cả các bên liên quan và duy trì các kỳ vọng và giá trị xã hội. Các nhà lãnh đạo trong phát triển công nghệ giành được niềm tin khi họ đặt ra các mục tiêu đầy tham vọng về bảo mật và độ tin cậy, để đảm bảo trách nhiệm giải trình và giám sát đối với các sản phẩm của họ, và để thúc đẩy tính toàn diện, đạo đức và trách nhiệm. Họ chứng minh sự tận tâm của mình đối với những mục tiêu đáng tin cậy đó bằng cách thực hiện các hành động rõ ràng và có thể đo lường được trên các khía cạnh của niềm tin kỹ thuật số: an ninh mạng, sự an toàn, khả năng tương tác, sự riêng tư, tính minh bạch, khả năng khắc phục, đạo đức, công bằng.

Điều quan trọng cần nhận ra là lòng tin không phải là một thứ bất biến mà là kỳ vọng của mỗi cá nhân và những yêu cầu hợp lý của họ có thể mở rộng theo thời gian. Đó là lý do tại sao ngày nay chúng ta nhận ra một khía cạnh khác của niềm tin kỹ thuật số là tính bền vững. Khi các công nghệ kỹ thuật số ngày càng tiêu thụ nhiều tài nguyên, chẳng hạn như năng lượng và nước, một công ty công nghệ đáng tin cậy phải xem tính bền vững như một phần nghĩa vụ của mình, được xây dựng thông qua các quyết định thiết kế, quản trị và vận hành công nghệ trong suốt vòng đời của hệ thống AI.

## 3. Vai trò xây dựng niềm tin kỹ thuật số

Trong bối cảnh biến đổi công nghệ nhanh chóng, niềm tin kỹ thuật số nổi lên như một điều kiện nền tảng đối với sự phát triển và chấp nhận xã hội của các công nghệ số thế hệ mới, đặc biệt là các công nghệ thông minh dựa trên thu thập, xử lý và khai thác dữ liệu quy mô lớn. Niềm tin kỹ thuật số là điều kiện tiên quyết của công nghệ thông minh dựa trên dữ liệu: Các công nghệ số thế hệ mới, bao gồm trí tuệ nhân tạo, không chỉ phụ thuộc vào năng lực tính toán mà còn dựa chủ yếu vào khả năng tiếp cận, liên thông và phân tích khối lượng dữ liệu lớn. Trong bối cảnh này, niềm tin của cá nhân và tổ chức vào cách thức dữ liệu được thu thập, sử dụng và bảo vệ trở thành yếu tố quyết định đối với mức độ chấp nhận và hiệu quả triển khai công nghệ.

Sự suy giảm niềm tin xã hội phản ánh những thất bại có hệ thống trong quản trị công nghệ. Thực tiễn những năm gần đây cho thấy, các sự cố an ninh mạng, vi phạm quyền riêng tư và thiếu trách nhiệm giải trình trong quá trình phát triển và triển khai công nghệ đã làm xuất hiện các “vết nứt” trong lòng tin xã hội. Những thất bại này không chỉ gây tổn hại trực tiếp đến người dùng mà còn làm xói mòn niềm tin vào năng lực tự điều chỉnh của ngành công nghệ nói chung. Sự xói mòn niềm tin này cũng đã làm suy yếu quan hệ giữa công nghệ và các chủ thể xã hội. Hệ quả của khủng hoảng niềm tin không chỉ giới hạn trong mối quan hệ giữa người dùng và sản phẩm công nghệ, mà còn lan rộng sang quan hệ giữa doanh nghiệp công nghệ và chính phủ, giữa các nhà đổi mới và tổ chức triển khai, cũng như giữa công nghệ với công chúng. Khi niềm tin bị suy giảm, công nghệ không còn được nhìn nhận như một công cụ trung lập phục vụ lợi ích xã hội, mà trở thành nguồn rủi ro cần được kiểm soát. Các dữ liệu khảo sát cho thấy sự suy giảm đáng kể mức độ tin cậy đối với ngành công nghệ. Trong thập kỷ qua, tại Mỹ, lĩnh vực công nghệ từng được xem là ngành đáng tin cậy nhất đã tụt hạng nhanh chóng về mức độ tin nhiệm xã hội. Giai đoạn 2022 - 2023, từ một phần ba đến một phần tư người dân Mỹ đánh giá ngành công nghệ kém đáng tin cậy hơn các lĩnh vực doanh nghiệp khác. Báo cáo *Edelman Trust Barometer Global Report*

2024 cho thấy 39% người được khảo sát trên toàn cầu cho rằng, việc quản lý đổi mới còn yếu kém, đồng thời nhiều cá nhân bày tỏ sự hoài nghi sâu sắc đối với sự phát triển của trí tuệ nhân tạo.

Bên cạnh đó, khủng hoảng niềm tin cũng đang thúc đẩy xã hội tái định hình quan hệ với công nghệ. Trước sự suy giảm niềm tin, cả cá nhân lẫn chính phủ ngày càng tìm cách thiết lập lại các nguyên tắc chi phối mối quan hệ giữa xã hội và công nghệ, thông qua các yêu cầu nghiêm ngặt hơn về minh bạch, trách nhiệm giải trình và bảo vệ quyền lợi người dùng. Việc ngành công nghệ đánh mất “sự chấp thuận xã hội” không chỉ ảnh hưởng đến uy tín mà còn trực tiếp đe dọa khả năng duy trì đổi mới và tăng trưởng dài hạn. Khi các mô hình và ứng dụng AI ngày càng đòi hỏi quyền truy cập sâu rộng vào các kho dữ liệu, sự hoài nghi của cá nhân và khách hàng doanh nghiệp lại thúc đẩy xu hướng hạn chế chia sẻ thông tin. Trong bối cảnh niềm tin suy giảm, việc giữ kín dữ liệu được xem là lựa chọn an toàn hơn, dẫn đến tình trạng “thiếu dữ liệu” – một yếu tố có thể kìm hãm chính sự phát triển của các hệ thống AI tiên tiến.

Niềm tin kỹ thuật số không phải là điều tất yếu mất đi, mà là kết quả của lựa chọn quản trị. Mặc dù xu hướng ngờ vực gia tăng là rõ rệt, ngành công nghệ không buộc phải chấp nhận sự suy giảm niềm tin như một kết cục không thể tránh khỏi. Thông qua các quyết định có trách nhiệm trong thiết kế, triển khai và quản trị công nghệ, đặc biệt là việc đặt niềm tin kỹ thuật số làm mục tiêu cốt lõi, các tổ chức công nghệ hoàn toàn có khả năng khôi phục và củng cố niềm tin của xã hội.

#### **4. Giải pháp xây dựng niềm tin kỹ thuật số trong kỷ nguyên trí tuệ nhân tạo**

Trong kỷ nguyên trí tuệ nhân tạo, niềm tin kỹ thuật số không còn là yếu tố mang tính hỗ trợ, mà đã trở thành điều kiện nền tảng cho sự chấp nhận xã hội, tính chính danh và tính bền vững của đổi mới công nghệ. Khi AI ngày càng tham gia sâu vào các quyết định ảnh hưởng trực tiếp đến quyền, cơ hội và đời sống của con người, việc xây dựng niềm tin kỹ thuật số đòi hỏi một hệ thống giải pháp toàn diện, kết hợp giữa thiết kế công nghệ, quản trị tổ chức và khung chính sách công.

##### *- Tích hợp niềm tin ngay từ giai đoạn thiết kế công nghệ*

Một trong những nguyên nhân chính dẫn đến khủng hoảng niềm tin đối với công nghệ AI là việc các hệ thống được thiết kế ưu tiên hiệu suất và tốc độ triển khai, trong khi các yếu tố như an toàn, quyền riêng tư và tác động xã hội chỉ được xem xét ở giai đoạn muộn. Cách tiếp cận này làm gia tăng rủi ro hệ thống và khiến niềm tin trở nên mong manh. Việc tích hợp các nguyên tắc như an ninh, quyền riêng tư, kiểm soát truy cập, và đánh giá tác động AI ngay từ giai đoạn thiết kế giúp hạn chế các rủi ro khó khắc phục về sau. Điều này tạo ra nền tảng kỹ thuật vững chắc cho niềm tin, bởi niềm tin chỉ có thể tồn tại khi người dùng tin rằng hệ thống được xây dựng với mục tiêu bảo vệ họ, không chỉ tối ưu hóa lợi ích tổ chức.

##### *- Thúc đẩy minh bạch và khả năng giải thích của hệ thống AI*

Trong nhiều trường hợp, sự hoài nghi đối với AI không xuất phát từ kết quả sai, mà từ việc người dùng không hiểu hệ thống hoạt động như thế nào và vì sao một quyết định được đưa ra. Các hệ thống AI “hộp đen” làm suy yếu khả năng đánh giá, phản biện và kiểm soát của con người. Việc phát triển và triển khai các công cụ AI có khả năng giải thích giúp người dùng, nhà quản lý và cơ quan giám sát hiểu được các yếu tố ảnh hưởng đến quyết định của AI. Điều này đặc biệt quan trọng trong các lĩnh vực có tác động lớn đến quyền con người như tài chính, y tế, tuyển dụng và tư pháp, nơi niềm tin không thể dựa trên “niềm tin mù quáng” mà phải dựa trên khả năng kiểm chứng.

##### *- Bảo vệ quyền riêng tư và thiết lập quản trị dữ liệu có trách nhiệm*

Trong kỷ nguyên AI dựa trên dữ liệu lớn, niềm tin của cá nhân gắn chặt với cách dữ liệu cá nhân được thu thập, xử lý và sử dụng. Các hành vi thu thập dữ liệu vượt quá mục đích, thiếu minh bạch hoặc không có sự đồng thuận đã làm suy giảm nghiêm trọng niềm tin xã hội đối với công nghệ. Các cơ chế cho phép cá nhân biết dữ liệu của mình được sử dụng vào mục đích gì, có thể kiểm soát phạm vi chia sẻ và thực hiện quyền rút lại sự đồng thuận không chỉ đáp ứng yêu cầu pháp lý, mà còn có ý nghĩa xã hội sâu sắc. Khi con người giữ vai trò chủ thể thay vì đối tượng bị khai thác, niềm tin kỹ thuật số được củng cố một cách bền vững.

### - Thiết lập cơ chế trách nhiệm giải trình rõ ràng và hiệu quả

Một trong những thách thức lớn nhất của AI là sự phân tán trách nhiệm giữa nhiều chủ thể: nhà phát triển mô hình, đơn vị tích hợp, tổ chức vận hành và người sử dụng. Khi xảy ra sai sót hoặc thiệt hại, việc không xác định rõ ai chịu trách nhiệm làm suy yếu niềm tin của xã hội. Việc thiết lập cơ chế phân định trách nhiệm rõ ràng, kết hợp với quy trình giám sát và báo cáo minh bạch, giúp chuyển niềm tin từ kỳ vọng mơ hồ sang sự đảm bảo có thể kiểm chứng. Điều này cũng tạo động lực để các tổ chức đầu tư nghiêm túc vào quản trị AI có trách nhiệm.

### - Tăng cường giám sát liên tục và kiểm toán độc lập

Các hệ thống AI có thể thay đổi hành vi theo thời gian do học từ dữ liệu mới hoặc thay đổi bối cảnh sử dụng. Vì vậy, việc đánh giá niềm tin không thể dừng lại ở thời điểm triển khai ban đầu. Các cơ chế giám sát liên tục, kiểm toán định kỳ và đánh giá độc lập giúp phát hiện sớm sai lệch, thiên kiến hoặc suy giảm hiệu năng. Sự hiện diện của các cơ chế này gửi tín hiệu mạnh mẽ tới xã hội rằng tổ chức sẵn sàng chịu trách nhiệm và điều chỉnh khi cần thiết.

### - Xây dựng cơ chế khắc phục và bảo vệ người bị ảnh hưởng

Không có hệ thống AI nào hoàn hảo. Do đó, điều quan trọng không phải là tránh mọi sai sót, mà là có cơ chế hiệu quả để khắc phục khi sai sót xảy ra. Việc đảm bảo rằng các quyết định AI có thể bị thách thức, xem xét lại và điều chỉnh bởi con người giúp duy trì vai trò trung tâm của con người trong hệ thống xã hội – yếu tố then chốt để duy trì niềm tin kỹ thuật số.

### - Hoàn thiện khung pháp lý và nâng cao năng lực xã hội

Khung pháp lý rõ ràng về AI, dữ liệu cá nhân và an ninh mạng tạo ra ranh giới tối thiểu cho hành vi công nghệ, giúp giảm bất định và tăng niềm tin xã hội. Nâng cao nhận thức của công chúng về AI, dữ liệu và quyền kỹ thuật số giúp giảm sự sợ hãi không cần thiết, đồng thời tăng khả năng tham gia có ý nghĩa của xã hội vào quá trình giám sát và định hướng công nghệ.

## Kết luận

Xây dựng niềm tin kỹ thuật số trong kỷ nguyên trí tuệ nhân tạo đòi hỏi một chiến lược tổng thể, trong đó công nghệ, quản trị và chính sách không thể tách rời. Niềm tin không đối lập với đổi mới, mà là điều kiện để đổi mới được xã hội chấp nhận và duy trì lâu dài. Chỉ khi các hệ thống AI được thiết kế và vận hành một cách có trách nhiệm, minh bạch và lấy con người làm trung tâm, niềm tin kỹ thuật số mới có thể trở thành nền tảng cho một xã hội số an toàn, công bằng và bền vững.

## Tài liệu tham khảo

1. Cisco (2024). *Cisco consumer privacy survey 2024*. Cisco Systems. <https://www.cisco.com/c/en/us/about/trust-center/consumer-privacy-survey.html>
2. Edelman (2024). *Edelman trust barometer 2024: Innovation in peril*. <https://www.edelman.com/trust/2024/trust-barometer>
3. European Commission (2024). *The EU Artificial Intelligence Act*. European Union. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
4. National Institute of Standards and Technology (NIST) (2023). *AI risk management framework (AI RMF 1.0)*. U.S. Department of Commerce. <https://www.nist.gov/itl/ai-risk-management-framework>
5. OECD. (2019). *OECD principles on artificial intelligence*. OECD Publishing.
6. UNESCO (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>
7. World Economic Forum (2022). *Earning digital trust: Decision-making for trustworthy technologies*. World Economic Forum. <https://www.weforum.org/reports/earning-digital-trust-decision-making-for-trustworthy-technologies>