# AN NINH MẠNG TRONG KỶ NGUYÊN AI: RỦI RO, NGUYÊN NHÂN VÀ GIẢI PHÁP ĐỐI VỚI NGƯỜI DÙNG TRONG DOANH NGHIỆP

● **PHẠM TUẤN TRINH**
Khoa Công nghệ - Kỹ thuật, Trường Đại học Bà Rịa - Vũng Tàu

**TÓM TẮT:**

Trong những năm gần đây, trí tuệ nhân tạo (tiếng Anh là artificial intelligence, viết tắt là AI) đã trở thành động lực quan trọng thúc đẩy chuyển đổi số trong doanh nghiệp, giúp nâng cao năng suất, tối ưu hóa chi phí và cải thiện trải nghiệm khách hàng. Bài viết phân tích các nguy cơ an ninh mạng liên quan AI, đồng thời đề xuất các giải pháp phòng ngừa. Bao gồm nâng cao nhận thức và đào tạo định kỳ cho người dùng, xây dựng chính sách sử dụng AI rõ ràng, ứng dụng công nghệ phòng thủ thông minh và kiểm tra bảo mật định kỳ. Sự kết hợp giữa con người, quy trình và công nghệ được xác định là chiến lược hiệu quả giúp doanh nghiệp chủ động bảo vệ người dùng, giảm thiểu thiệt hại và duy trì hoạt động ổn định trong môi trường số hóa, đáp ứng thách thức và cơ hội của kỷ nguyên AI.

**Từ khóa:** an ninh mạng doanh nghiệp, trí tuệ nhân tạo, rủi ro mạng AI, quản trị rủi ro công nghệ, đào tạo nhận thức an ninh.

## 1. Đặt vấn đề

Trong những năm gần đây, trí tuệ nhân tạo (AI) nổi lên như một động lực quan trọng thúc đẩy chuyển đổi số trong doanh nghiệp. Các khảo sát quốc tế cho thấy, hơn 70% doanh nghiệp đã ứng dụng AI vào nhiều quy trình vận hành nhằm nâng cao năng suất, giảm chi phí và cải thiện trải nghiệm khách hàng (McKinsey, 2023). Tại Việt Nam, Bộ Khoa học và Công nghệ xác định AI là công nghệ trọng tâm trong chiến lược chuyển đổi số quốc gia, góp phần đổi mới mô hình kinh doanh và tối ưu hóa hoạt động doanh nghiệp (Bộ Khoa học và Công nghệ, 2023). Tuy nhiên, bên cạnh lợi ích, AI cũng mở ra nhiều rủi ro mới cho an ninh mạng. Khả năng tạo sinh nội dung, mô phỏng hành vi và tự động phân tích của AI có thể bị tội phạm mạng lợi dụng để thực hiện các cuộc tấn công tinh vi hơn. Báo cáo của Cục An toàn thông tin (NCSC) ghi nhận sự gia tăng rõ rệt của các hình thức tấn công sử dụng AI như deep fake, email lừa đảo cá nhân hóa hay mã độc tự biến đổi (NCSC, 2024). Trong bối cảnh đó, người dùng trong doanh

nghiệp, từ nhân viên văn phòng đến cán bộ quản lý trở thành mục tiêu tấn công hàng đầu. Sự tinh vi của các chiêu thức lợi dụng AI khiến người dùng dễ bị đánh lừa, dẫn đến rò rỉ dữ liệu, gián đoạn hoạt động và thiệt hại tài chính.

**2. Các rủi ro an ninh mạng mới đối với người dùng doanh nghiệp trong kỷ nguyên AI**

***2.1. Deep fake - giả mạo giọng nói, hình ảnh***

Sự phát triển mạnh mẽ của công nghệ AI tạo sinh trong những năm gần đây đã làm gia tăng đáng kể các cuộc tấn công deepfake nhắm vào doanh nghiệp. Các mô hình AI hiện có thể tạo ra giọng nói, khuôn mặt hoặc toàn bộ đoạn video giả mạo với độ chân thực rất cao, khiến người dùng khó phân biệt thật - giả (Chesney & Citron, 2019). Dựa trên công nghệ này, tội phạm mạng triển khai hình thức tấn công mạo danh lãnh đạo (CEO Fraud), trong đó chúng sử dụng video hoặc cuộc gọi deepfake để yêu cầu nhân viên chuyển tiền, cung cấp thông tin đăng nhập hoặc chia sẻ tài liệu mật. Ngoài ra, deepfake còn bị lợi dụng để tạo ra các video hoặc nội dung truyền thông giả mạo, trong đó lãnh đạo hoặc doanh nghiệp bị gán ghép phát ngôn, hành vi sai lệch, từ đó gây tổn hại nghiêm trọng tới uy tín và hình ảnh thương hiệu trên thị trường. Một số báo cáo an ninh toàn cầu ghi nhận nhiều doanh nghiệp đã thiệt hại từ vài trăm nghìn đến hàng triệu USD bởi các vụ lừa đảo sử dụng deepfake nhằm thao túng nhân viên tài chính hoặc nhân viên quản trị hệ thống (Symantec, 2023).

Tại Việt Nam, Cục An toàn thông tin cũng cảnh báo mức độ gia tăng của tấn công deepfake, đặc biệt trong bối cảnh các nền tảng mạng xã hội và ứng dụng AI tạo sinh được sử dụng rộng rãi. Theo các chuyên gia, rủi ro này sẽ tiếp tục tăng khi tội phạm mạng ngày càng dễ dàng tiếp cận công cụ tạo giọng nói và video giả mạo (NCSC, 2024). Vì vậy, người dùng trong doanh nghiệp trở thành mục tiêu đặc biệt dễ bị khai thác do thường xuyên phải xử lý các yêu cầu từ cấp quản lý và những thông tin nhạy cảm.

***2.2. Spear-phishing AI***

Sự phát triển của các mô hình ngôn ngữ lớn (LLMs) đã làm thay đổi đáng kể mức độ tinh vi của các cuộc tấn công spear phishing. Thay vì các email lừa đảo có nội dung đơn giản, sai chính tả và dễ nhận diện như trước đây, tội phạm mạng ngày nay có thể sử dụng AI để tạo ra email với văn phong tự nhiên, ngữ nghĩa logic và bối cảnh phù hợp với vị trí công việc của từng người nhận (Europol, 2023). Một số mô hình AI thậm chí có khả năng phân tích hành vi người dùng, lịch sử tương tác hoặc dữ liệu công khai trên mạng xã hội để cá nhân hóa nội dung email, từ đó tăng đáng kể xác suất thành công của cuộc tấn công (IBM Security, 2024). Bên cạnh đó, AI cũng giúp tin tặc tự động hóa quy trình gửi hàng loạt email nhắm tới từng nhóm người dùng cụ thể trong doanh nghiệp mà chi phí gần như bằng không. Điều này khiến số lượng và mức độ tinh vi của các cuộc tấn công spear phishing gia tăng mạnh trên toàn cầu. Tại Việt Nam, Cục An toàn thông tin (NCSC, 2024) cũng cảnh báo xu hướng lạm dụng AI tạo sinh để sản xuất email mạo danh bộ phận tài chính, nhân sự hoặc đối tác nhằm đánh lừa nhân viên doanh nghiệp.

***2.3. Tấn công tự động quy mô lớn***

Sự phát triển của trí tuệ nhân tạo (AI) đã giúp tội phạm mạng nâng cấp khả năng tấn công tự động hóa trên quy mô lớn. Nhờ AI, hacker có thể tự động dò quét các điểm yếu trong hệ thống, dự đoán mật khẩu dựa trên hành vi và thói quen của người dùng, cũng như phát hiện sai sót trong cấu hình hệ thống nhanh hơn nhiều lần so với con người (Symantec, 2023; Gartner, 2023). Các công cụ AI này cho phép thực hiện các cuộc tấn công liên tục, có mục tiêu cụ thể và đa dạng hóa phương thức xâm nhập mà không cần sự can thiệp trực tiếp của hacker, dẫn đến nguy cơ thông tin nhạy cảm của người dùng bị đánh cắp hoặc bị khai thác, ngay cả khi họ tuân thủ đầy đủ các quy tắc bảo mật cơ bản (IBM Security, 2024).

Tại Việt Nam, các báo cáo của Cục An toàn thông tin (NCSC, 2024) cũng ghi nhận xu hướng tăng các cuộc tấn công tự động hóa, đặc biệt nhắm vào hệ thống doanh nghiệp vừa và nhỏ, nơi các biện pháp kiểm soát an ninh còn hạn chế. Điều này

nhấn mạnh nhu cầu áp dụng các giải pháp phòng thủ thông minh, như giám sát hành vi người dùng và kiểm tra lỗ hổng định kỳ, nhằm giảm thiểu rủi ro từ các cuộc tấn công AI quy mô lớn.

***2.4. Rò rỉ dữ liệu từ việc sử dụng AI thiếu kiểm soát***

Việc áp dụng các công cụ AI, đặc biệt là chatbot và nền tảng tạo sinh nội dung, đang trở nên phổ biến trong doanh nghiệp, hỗ trợ soạn thảo email, phân tích tài liệu và tóm tắt dữ liệu nội bộ. Tuy nhiên, nếu không có cơ chế kiểm soát chặt chẽ, thông tin nhạy cảm của doanh nghiệp có thể bị chia sẻ vô tình cho các nền tảng AI bên ngoài (Microsoft Security, 2024; Gartner, 2023). Nhiều nghiên cứu chỉ ra, việc nhập dữ liệu nhạy cảm vào các công cụ AI công cộng có thể tạo ra rủi ro rò rỉ dữ liệu, từ thông tin khách hàng đến kế hoạch kinh doanh, đặc biệt khi nền tảng lưu trữ và xử lý dữ liệu trên điện toán đám mây hoặc bên thứ ba (IBM Security, 2024).

Tại Việt Nam, Cục An toàn thông tin (NCSC, 2024) cũng cảnh báo nhân viên doanh nghiệp thường chưa được đào tạo đầy đủ về quy tắc sử dụng AI an toàn, dẫn đến nguy cơ vô tình tiết lộ dữ liệu nội bộ. Do đó, doanh nghiệp cần thiết lập các chính sách và hướng dẫn rõ ràng về việc sử dụng AI, đồng thời áp dụng các giải pháp kiểm soát dữ liệu, giám sát truy cập và mã hóa để giảm thiểu rủi ro rò rỉ thông tin nhạy cảm.

***2.5. Tấn công bằng mã độc được tạo bởi AI***

Sự phát triển của AI đã làm thay đổi đáng kể cách thức tạo ra và phân phối mã độc. Các công cụ AI cho phép tin tặc tạo mã độc nhanh hơn, đa dạng hơn và khó bị phát hiện bởi các hệ thống phòng vệ truyền thống (Symantec, 2023). Nhờ khả năng phân tích mẫu và tự động học từ dữ liệu, AI có thể tạo ra những biến thể mã độc liên tục thay đổi chữ ký, từ đó vượt qua các cơ chế phát hiện dựa trên mẫu (signature based detection) đang được nhiều doanh nghiệp sử dụng (IBM Security, 2024).

Đặc biệt, một số mô hình AI độc hại còn có khả năng tự học từ các bản vá bảo mật và kỹ thuật phát hiện mới, từ đó tạo ra biến thể mã độc tối ưu để né tránh cơ chế phòng thủ của doanh nghiệp (Gartner, 2023). Điều này khiến tấn công bằng mã độc ngày càng tinh vi, có khả năng tự động mở rộng quy mô và tấn công đồng thời nhiều hệ thống.

Tại Việt Nam, Cục An toàn thông tin (NCSC, 2024) ghi nhận xu hướng gia tăng các chiến dịch phát tán mã độc có ứng dụng AI, đặc biệt trong lĩnh vực tài chính, thương mại điện tử và doanh nghiệp vừa và nhỏ, những nơi thường có hệ thống phòng vệ chưa đủ mạnh. Các chuyên gia cảnh báo mã độc do AI tạo ra sẽ tiếp tục trở thành mối đe dọa hàng đầu nếu doanh nghiệp không cập nhật kịp thời các giải pháp an ninh mạng hiện đại.

**3. Nguyên nhân khiến người dùng doanh nghiệp dễ bị tổn thương**

***3.1. Thiếu nhận thức về rủi ro từ AI***

Một trong những thách thức lớn nhất đối với doanh nghiệp trong kỷ nguyên AI là nhận thức hạn chế của nhân viên về các rủi ro an ninh mạng liên quan đến AI. Nhiều nghiên cứu chỉ ra phần lớn nhân viên chưa được đào tạo bài bản để nhận biết và xử lý các mối đe dọa từ AI, bao gồm deepfake, lừa đảo spear-phishing do AI tạo ra, hoặc các phương thức khai thác dữ liệu nhạy cảm qua công cụ AI tạo sinh (Europol, 2023; IBM Security, 2024).

Sự thiếu hiểu biết này dẫn đến nguy cơ nhân viên vô tình trở thành mắt xích yếu trong chuỗi an ninh mạng, dễ bị lừa đảo, chia sẻ thông tin nhạy cảm hoặc thực hiện các hành động làm tăng rủi ro bảo mật. Tại Việt Nam, Cục An toàn thông tin (NCSC, 2024) cũng cảnh báo nhiều nhân viên doanh nghiệp chưa nắm được cách sử dụng AI một cách an toàn, từ việc kiểm soát dữ liệu nhập vào các nền tảng AI đến nhận diện email hoặc cuộc gọi giả mạo. Việc nâng cao nhận thức và đào tạo bài bản về rủi ro AI trở thành yêu cầu cấp thiết để giảm thiểu các sự cố an ninh mạng liên quan AI trong doanh nghiệp.

***3.2. Áp lực công việc và sự phụ thuộc vào công nghệ***

Trong môi trường doanh nghiệp hiện đại, áp lực công việc cao khiến người dùng thường phải xử lý thông tin nhanh chóng, dẫn đến việc không

đủ thời gian để xác thực nội dung và dễ tin vào các email, công văn hoặc thông báo "trông có vẻ chính xác" (Europol, 2023). Sự phụ thuộc ngày càng tăng vào công cụ số và AI tạo sinh còn làm gia tăng tính thuyết phục của các thông tin giả mạo, từ đó giảm khả năng nhận diện rủi ro của nhân viên (IBM Security, 2024; Microsoft Security, 2024).

Các nghiên cứu chỉ ra rằng khi người dùng phải xử lý khối lượng lớn thông tin trong thời gian ngắn, khả năng nhận biết các dấu hiệu cảnh báo như email giả mạo, nội dung deep fake hoặc các yêu cầu truy cập trái phép sẽ giảm đáng kể (Symantec, 2023). Tại Việt Nam, Cục An toàn thông tin (NCSC, 2024) cũng nhấn mạnh áp lực công việc kết hợp với thiếu đào tạo về AI và bảo mật là nguyên nhân khiến người dùng dễ trở thành mục tiêu tấn công mạng, từ đó làm tăng nguy cơ rò rỉ dữ liệu và tổn thất tài chính trong doanh nghiệp.

***3.3. Thiếu quy trình quản trị rủi ro liên quan AI***

Một thách thức quan trọng khác trong doanh nghiệp là thiếu các quy trình quản trị rủi ro cụ thể liên quan đến AI. Nhiều doanh nghiệp hiện nay chỉ áp dụng chính sách bảo mật truyền thống, tập trung vào kiểm soát truy cập, bảo vệ hệ thống và dữ liệu theo các tiêu chuẩn cũ, mà chưa xây dựng hướng dẫn chi tiết về việc sử dụng AI trong công việc, kiểm soát dữ liệu đưa vào các công cụ tạo sinh, hay quản lý rủi ro từ tự động hóa (Gartner, 2023; IBM Security, 2024).

Theo báo cáo của Cục An toàn thông tin (NCSC, 2024), việc thiếu quy trình quản trị rủi ro AI dẫn đến nhiều hệ quả như rò rỉ dữ liệu nhạy cảm, tấn công spear-phishing hiệu quả hơn, và mã độc AI khó phát hiện. Đồng thời, Europol (2023) cũng chỉ ra rằng các doanh nghiệp không có chính sách AI rõ ràng thường trở thành mục tiêu của các chiến dịch tấn công tự động hóa và mạo danh. Do đó, việc xây dựng hướng dẫn, chính sách và quy trình quản lý rủi ro AI là bước cần thiết để bảo vệ người dùng và giảm thiểu thiệt hại tiềm ẩn từ các cuộc tấn công mạng hiện đại.

**4. Giải pháp nâng cao an ninh mạng cho người dùng doanh nghiệp trong kỷ nguyên AI**

***4.1. Đào tạo nhận thức an ninh mạng gắn với AI***

Để giảm thiểu rủi ro từ các mối đe dọa AI, doanh nghiệp cần triển khai chương trình đào tạo nhận thức an ninh mạng chuyên biệt, tích hợp các nội dung liên quan đến AI. Các nội dung cần bao gồm:

❖ Nhận diện deepfake: hướng dẫn nhân viên phân biệt video, hình ảnh và giọng nói giả mạo do AI tạo ra.

❖ Xử lý email hoặc tin nhắn nghi ngờ sử dụng AI: cung cấp kỹ năng phân tích và kiểm chứng thông tin trước khi thực hiện hành động.

❖ Nguyên tắc sử dụng AI an toàn: xác định phạm vi và cách thức sử dụng các công cụ AI trong công việc, tránh lạm dụng hoặc nhập dữ liệu nhạy cảm.

❖ Quy tắc bảo mật khi nhập dữ liệu vào chatbot/AI: hướng dẫn cách bảo vệ thông tin nhạy cảm, mã hóa dữ liệu và kiểm soát truy cập.

❖ Đào tạo định kỳ và mô phỏng tấn công: tổ chức các bài tập thực hành, giả lập tình huống tấn công AI để nâng cao khả năng phản ứng và nhận diện mối đe dọa của nhân viên (IBM Security, 2024; NCSC, 2024; Europol, 2023).

Việc đào tạo liên tục và thực hành định kỳ không chỉ nâng cao nhận thức, mà còn giúp xây dựng văn hóa an ninh mạng trong doanh nghiệp, giảm nguy cơ rò rỉ dữ liệu, tấn công spear#phishing và các hình thức tấn công dựa trên AI.

***4.2. Xây dựng chính sách sử dụng AI trong doanh nghiệp***

Để quản trị rủi ro liên quan đến AI, doanh nghiệp cần thiết lập chính sách sử dụng AI rõ ràng và chi tiết, bao gồm các nội dung chính sau:

❖ Danh mục công cụ AI được phép sử dụng: xác định các nền tảng, phần mềm hoặc dịch vụ AI được phê duyệt để đảm bảo an toàn và tuân thủ pháp luật.

❖ Quy định về dữ liệu nhập vào AI: nêu rõ loại dữ liệu nhạy cảm hoặc nội dung bí mật không được phép đưa vào các công cụ AI, nhằm tránh rò rỉ thông tin doanh nghiệp.

❖ Kiểm soát, lưu trữ và theo dõi truy cập: áp dụng các biện pháp giám sát, phân quyền và mã hóa dữ liệu khi sử dụng AI, đảm bảo chỉ người dùng được phép mới có thể truy cập và chỉnh sửa dữ liệu.

❖ Quy trình xử lý sự cố: xây dựng hướng dẫn chi tiết về cách phản ứng khi xảy ra rò rỉ dữ liệu hoặc các sự cố an ninh liên quan đến AI, bao gồm thông báo, đánh giá rủi ro và khắc phục (Gartner, 2023; IBM Security, 2024; NCSC, 2024).

Việc áp dụng chính sách này giúp doanh nghiệp giảm thiểu rủi ro bảo mật, kiểm soát việc sử dụng AI trong môi trường làm việc, đồng thời nâng cao hiệu quả vận hành mà vẫn đảm bảo an toàn thông tin.

***4.3. Ứng dụng AI để phòng thủ***

Trong bối cảnh các mối đe dọa mạng sử dụng AI ngày càng tinh vi, doanh nghiệp có thể tận dụng AI như một công cụ phòng thủ chủ động. Các giải pháp phòng thủ dựa trên AI bao gồm:

❖ Hệ thống phát hiện bất thường dựa trên hành vi (Behavioral Analytics): sử dụng các mô hình AI/ML để phân tích hành vi truy cập, nhận diện các hoạt động bất thường hoặc đáng ngờ trong hệ thống (IBM Security, 2024).

❖ Công cụ kiểm soát truy cập thông minh: áp dụng AI để phân quyền tự động, theo dõi truy cập và phát hiện truy cập trái phép nhằm ngăn chặn rủi ro từ bên trong (Gartner, 2023).

❖ Hệ thống cảnh báo sớm tấn công (Early Threat Detection): triển khai AI để nhận diện các dấu hiệu tấn công mạng, spear-phishing hoặc mã độc trước khi chúng gây ra thiệt hại thực tế (Symantec, 2023; Microsoft Security, 2024).

Sử dụng AI cho phòng thủ không chỉ là giải pháp kỹ thuật cần thiết, mà còn là xu hướng tất yếu khi các cuộc tấn công dựa trên AI ngày càng mạnh mẽ và đa dạng. Việc tích hợp AI vào hệ thống phòng vệ giúp doanh nghiệp tăng khả năng phản ứng, giảm thiểu thiệt hại và nâng cao mức độ bảo mật tổng thể.

***4.4. Tăng cường xác thực đa lớp (MFA)***

Việc triển khai xác thực đa lớp (Multi-Factor Authentication - MFA) là một trong những biện pháp bảo mật cơ bản nhưng cực kỳ hiệu quả. MFA giúp giảm đáng kể rủi ro từ các cuộc tấn công đoán mật khẩu, đánh cắp danh tính hoặc truy cập trái phép, ngay cả trong bối cảnh các mối đe dọa dựa trên AI ngày càng tinh vi (IBM Security, 2024; Gartner, 2023). Bằng cách yêu cầu nhiều yếu tố xác thực - chẳng hạn mật khẩu, mã OTP, sinh trắc học, MFA đảm bảo rằng chỉ người dùng hợp pháp mới có thể truy cập hệ thống, nâng cao mức độ bảo vệ dữ liệu và thông tin nhạy cảm của doanh nghiệp.

***4.5. Kiểm tra, đánh giá bảo mật định kỳ***

Để duy trì an ninh hệ thống, doanh nghiệp cần thực hiện kiểm tra và đánh giá bảo mật định kỳ, bao gồm:

❖ Rà soát lỗ hổng hệ thống để phát hiện các điểm yếu mới.

❖ Kiểm tra quyền truy cập và xác nhận người dùng chỉ có quyền cần thiết.

❖ Đánh giá nhật ký hoạt động nhằm phát hiện các hành vi bất thường.

❖ Thử nghiệm tấn công giả lập (penetration testing) để mô phỏng các kịch bản xâm nhập thực tế và kiểm tra khả năng phòng thủ của hệ thống (Symantec, 2023; Microsoft Security, 2024).

Hoạt động kiểm tra định kỳ không chỉ giúp phát hiện và khắc phục lỗ hổng kịp thời, mà còn tăng khả năng ứng phó trước các cuộc tấn công dựa trên AI, bảo vệ người dùng và dữ liệu nhạy cảm trong doanh nghiệp.

**5. Kết luận**

Trong kỷ nguyên trí tuệ nhân tạo (AI), doanh nghiệp đang đứng trước cơ hội và thách thức song song. AI mang lại nhiều lợi ích quan trọng, giúp tối ưu hóa quy trình vận hành, nâng cao năng suất lao động, giảm chi phí và cải thiện trải nghiệm khách hàng. Các công cụ AI tạo sinh, phân tích dữ liệu lớn và tự động hóa cho phép doanh nghiệp ra quyết định nhanh chóng và chính xác hơn, từ đó thúc đẩy quá trình chuyển đổi số và nâng cao năng lực cạnh tranh. Tuy nhiên, bên cạnh những lợi ích này, AI cũng tạo ra các rủi ro an ninh mạng với mức độ tinh vi chưa từng có. Những cuộc tấn công dựa trên AI, bao gồm deepfake,

spear-phishing được cá nhân hóa, mã độc tự học và tấn công tự động hóa quy mô lớn, đang gia tăng mạnh mẽ, tác động trực tiếp đến người dùng trong doanh nghiệp - những nhân viên trực tiếp thao tác dữ liệu và vận hành hệ thống. Nhân viên trở thành "mắt xích yếu" nếu thiếu nhận thức về các rủi ro này, dẫn đến khả năng bị lừa đảo, rò rỉ dữ liệu nhạy cảm hoặc gây gián đoạn hoạt động doanh nghiệp.

Để chủ động bảo vệ người dùng và giảm thiểu thiệt hại, doanh nghiệp cần triển khai chiến lược toàn diện, kết hợp giữa con người, quy trình và công nghệ.

***Trước hết,*** nâng cao nhận thức của nhân viên thông qua đào tạo định kỳ về AI và an ninh mạng, hướng dẫn nhận diện deepfake, xử lý email nghi ngờ sử dụng AI và áp dụng nguyên tắc bảo mật khi tương tác với công cụ AI, là bước đi quan trọng để giảm thiểu rủi ro.

***Thứ hai,*** xây dựng chính sách sử dụng AI rõ ràng giúp kiểm soát công cụ, dữ liệu và quyền truy cập, đồng thời thiết lập quy trình xử lý sự cố khi xảy ra rò rỉ hoặc tấn công.

***Thứ ba,*** ứng dụng công nghệ phòng thủ dựa trên AI như hệ thống phát hiện bất thường, cảnh báo sớm và kiểm soát truy cập thông minh giúp doanh nghiệp phát hiện và ngăn chặn các mối đe dọa tự động.

***Cuối cùng,*** việc kiểm tra, đánh giá bảo mật định kỳ, bao gồm rà soát lỗ hổng, thử nghiệm tấn công giả lập và đánh giá nhật ký hoạt động, đảm bảo các điểm yếu mới được phát hiện và khắc phục kịp thời. Thông qua việc kết hợp đồng bộ các biện pháp trên, doanh nghiệp không chỉ giảm thiểu rủi ro mà còn xây dựng văn hóa an ninh mạng bền vững, duy trì hoạt động ổn định và nâng cao khả năng ứng phó trước các mối đe dọa AI ngày càng tinh vi, từ đó đảm bảo an toàn dữ liệu và lợi ích lâu dài trong môi trường số hóa hiện đại (IBM Security, 2024; NCSC, 2024; Gartner, 2023) ■

**TÀI LIỆU THAM KHẢO:**

Bộ Khoa học và Công nghệ (2023). Báo cáo chuyển đổi số quốc gia năm 2023. Nhà xuất bản Thông tin và Truyền thông. Truy cập tại https://dx.gov.vn.

Cục An toàn thông tin (NCSC) (2024). Báo cáo hiện trạng an toàn thông tin mạng Việt Nam năm 2024. Bộ Khoa học và Công nghệ. Truy cập tại https://ncsc.gov.vn.

Chesney R., & Citron D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753-1820. Available at https://doi.org/10.2139/ssrn.3213954.

Europol (2023). Internet organised crime threat assessment (IOCTA) 2023: AI and cybercrime. European Union Agency for Law Enforcement Cooperation. Available at https://www.europol.europa.eu/iocta-report.

Gartner (2023). Top security and risk management trends: Generative AI challenges for enterprise security. Gartner Research. Available at https://www.gartner.com/en/documents/.

Google Cloud (2024). Threat horizons report: AI and emerging cyber risks. Google Cybersecurity Action Team. Available at https://cloud.google.com/security/threat-horizons/ai-emerging-cyber-risks.

IBM Security. (2024). X-Force threat intelligence index 2024: AI-powered phishing, malware, and impersonation attacks. IBM Corporation. Available at https://www.ibm.com/think/x-force/2024-x-force-threat-intelligence-index.

McKinsey & Company (2023). The state of AI in 2023: Generative AI's breakout year. McKinsey Global Institute. Available at https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year.

Microsoft Security (2024). Microsoft digital defense report 2024: AI-powered cyber threats. Microsoft Security Division. Available at https://www.microsoft.com/en-us/security/business-trends/2024-digital-defense-report-ai-powered-cyber-threats.

National Institute of Standards and Technology (NIST) (2023). AI risk management framework (AI RMF 1.0). U.S. Department of Commerce. Available at https://www.nist.gov/itl/ai-risk-management-framework.

Symantec (2023). Internet security threat report 2023: AI-driven fraud, malware, and spear-phishing attacks. Available at https://www.symantec.com.

World Economic Forum (2024). Global cybersecurity outlook 2024. World Economic Forum. Available at https://www.weforum.org/reports/global-cybersecurity-outlook-2024.

**Ngày nhận bài: 9/9/2025**
**Ngày phản biện đánh giá và sửa chữa: 20/9/2025**
**Ngày chấp nhận đăng bài: 14/10/2025**

# CYBERSECURITY IN THE AGE OF ARTIFICIAL INTELLIGENCE: RISKS, DRIVERS, AND SOLUTIONS FOR BUSINESSES

● **PHAM TUAN TRINH**
Faculty of Engineering Technology,
Ba Ria - Vung Tau University

**ABSTRACT:**

In recent years, artificial intelligence (AI) has emerged as a central driver of enterprise digital transformation, enhancing productivity, optimizing costs, and improving customer experiences. At the same time, the growing adoption of AI has introduced increasingly complex cybersecurity risks related to data privacy, system integrity, and operational continuity. This study examines the major cybersecurity risks associated with AI deployment in enterprises and analyzes corresponding preventive approaches. The analysis emphasizes the importance of integrating human awareness, organizational processes, and technological solutions, including user education and training, clearly defined AI governance policies, the application of intelligent security technologies, and regular security assessments. Such an integrated approach enables organizations to proactively protect digital assets, reduce potential damage from cyber threats, and maintain stable and resilient operations in an increasingly digitalized business environment. By addressing cybersecurity challenges alongside technological innovation, enterprises can better leverage AI while ensuring data protection and long-term business sustainability.

**Keywords:** enterprise cybersecurity, artificial intelligence, AI-related cyber risks, technology risk management, security awareness training.