Ký bởi: CÔNG BÁO NƯỚC CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Cơ quan: VĂN PHÒNG CHÍNH PHỦ
Ngày ký: 2026-10-16 17:04:33 +07:00

NGÂN HÀNG NHÀ NƯỚC VIỆT NAM	CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM Độc lập - Tự do - Hạnh phúc
Số: 53/2026/TT-NHNN	*Hà Nội, ngày 30 tháng 9 năm 2026*

THÔNG TƯ
Quy định về an toàn, quản lý rủi ro và điều kiện triển khai đối với việc ứng dụng hệ thống trí tuệ nhân tạo trong ngành Ngân hàng

Căn cứ Luật Ngân hàng Nhà nước Việt Nam số 46/2010/QH12;

Căn cứ Luật Các tổ chức tín dụng số 32/2024/QH15 được sửa đổi, bổ sung bởi Luật số 96/2025/QH15;

Căn cứ Luật Trí tuệ nhân tạo số 134/2025/QH15;

Căn cứ Nghị định số 142/2026/NĐ-CP của Chính phủ quy định chi tiết một số điều và biện pháp để tổ chức, hướng dẫn thi hành Luật Trí tuệ nhân tạo;

Căn cứ Nghị định số 26/2025/NĐ-CP của Chính phủ quy định chức năng, nhiệm vụ, quyền hạn và cơ cấu tổ chức của Ngân hàng Nhà nước Việt Nam được sửa đổi, bổ sung bởi Nghị định số 198/2026/NĐ-CP;

Căn cứ Quyết định số 33/2026/QĐ-TTg của Thủ tướng Chính phủ ban hành Danh mục hệ thống trí tuệ nhân tạo có rủi ro cao;

Theo đề nghị của Cục trưởng Cục Công nghệ thông tin;

Thống đốc Ngân hàng Nhà nước Việt Nam ban hành Thông tư quy định về an toàn, quản lý rủi ro và điều kiện triển khai đối với việc ứng dụng hệ thống trí tuệ nhân tạo trong ngành Ngân hàng.

Chương I
QUY ĐỊNH CHUNG

Điều 1. Phạm vi điều chỉnh và đối tượng áp dụng

1. Phạm vi điều chỉnh

a) Thông tư này quy định về an toàn, quản lý rủi ro và điều kiện triển khai đối với việc ứng dụng hệ thống trí tuệ nhân tạo để ra quyết định hoặc hỗ trợ ra quyết định trong hoạt động nghiệp vụ ngành Ngân hàng cung cấp dịch vụ cho

khách hàng (sau đây gọi tắt là hoạt động nghiệp vụ), bao gồm:

(i) Hoạt động ngân hàng và các hoạt động kinh doanh khác của tổ chức tín dụng, chi nhánh ngân hàng nước ngoài;

(ii) Hoạt động cung ứng dịch vụ trung gian thanh toán của tổ chức cung ứng dịch vụ trung gian thanh toán;

(iii) Hoạt động thông tin tín dụng của công ty thông tin tín dụng;

(iv) Hoạt động cung ứng dịch vụ Tiền di động của tổ chức cung ứng dịch vụ Tiền di động.

b) Khuyến khích áp dụng các quy định tại Thông tư này đối với các hệ thống trí tuệ nhân tạo phục vụ các hoạt động quản trị, điều hành nội bộ; các hệ thống trí tuệ nhân tạo đang được triển khai trong khuôn khổ cơ chế thử nghiệm có kiểm soát theo quy định của pháp luật về trí tuệ nhân tạo và pháp luật có liên quan;

c) Các dịch vụ cung cấp bởi hệ thống trí tuệ nhân tạo của bên thứ ba được sử dụng như công cụ hỗ trợ cá nhân mà không được tích hợp vào quy trình hoạt động nghiệp vụ hoặc hệ thống thông tin của đối tượng áp dụng quy định tại khoản 2 Điều này thì không thuộc phạm vi điều chỉnh của Thông tư này.

2. Đối tượng áp dụng

Thông tư này áp dụng đối với các tổ chức tín dụng, chi nhánh ngân hàng nước ngoài, tổ chức cung ứng dịch vụ trung gian thanh toán, tổ chức cung ứng dịch vụ Tiền di động, công ty thông tin tín dụng, Công ty Quản lý tài sản của các tổ chức tín dụng Việt Nam, Nhà máy in tiền quốc gia, Bảo hiểm tiền gửi Việt Nam (sau đây gọi chung là đơn vị).

Điều 2. Giải thích từ ngữ và thuật ngữ

Trong Thông tư này, các từ ngữ dưới đây được hiểu như sau:

1. *Vòng đời hệ thống trí tuệ nhân tạo* là chuỗi các giai đoạn liên tiếp và lặp lại trong quá trình tồn tại của một hệ thống trí tuệ nhân tạo, bao gồm các giai đoạn: thiết kế, phát triển, kiểm thử, triển khai, vận hành và ngừng sử dụng hệ thống.

2. *Tấn công trích xuất mô hình (Model extraction*) là hành vi khai thác giao diện, kết quả đầu ra hoặc cơ chế tương tác của hệ thống trí tuệ nhân tạo để suy luận, sao chép trái phép cấu trúc, tham số hoặc logic hoạt động của mô hình hệ thống trí tuệ nhân tạo (sau đây gọi tắt là mô hình) nhằm tạo ra mô hình mới giống mô hình gốc.

3. *Tấn công né tránh (Evasion attack)* là hành vi cố ý làm nhiễu dữ liệu đầu vào trong giai đoạn vận hành nhằm đánh lừa hệ thống trí tuệ nhân tạo phân loại

sai bản chất của dữ liệu, từ đó đưa ra kết quả sai lệch, không chính xác hoặc không phù hợp với mục đích thiết kế ban đầu.

4. *Tấn công tiêm lệnh (Prompt injection)* là hành vi chèn, thao túng nội dung đầu vào đối với hệ thống trí tuệ nhân tạo sử dụng mô hình ngôn ngữ lớn nhằm vượt qua các ràng buộc kiểm soát an toàn để làm sai lệch hành vi hoặc kết quả đầu ra của hệ thống.

5. *Tấn công vi phạm quyền riêng tư dữ liệu (Privacy leakage)* là hành vi lợi dụng các điểm yếu trong hệ thống trí tuệ nhân tạo nhằm trích xuất, thu thập trái phép hoặc gây rò rỉ, lộ lọt dữ liệu của tổ chức, cá nhân.

6. *Trôi mô hình (Model drift)* là hiện tượng hiệu năng của mô hình hệ thống trí tuệ nhân tạo bị suy giảm theo thời gian do sự thay đổi trong mối quan hệ giữa các biến dữ liệu đầu vào và kết quả đầu ra dự báo so với giai đoạn huấn luyện ban đầu hoặc do trôi dữ liệu.

7. *Trôi dữ liệu (Data drift)* là hiện tượng phân phối thống kê hoặc đặc tính của dữ liệu đầu vào, dữ liệu vận hành thực tế thay đổi so với dữ liệu đã được sử dụng để huấn luyện hoặc kiểm thử mô hình, dẫn đến nguy cơ sai lệch trong kết quả xử lý.

8. *Khả năng giải thích (Explainability)* là khả năng cung cấp các thông tin, lý luận hoặc bằng chứng giúp con người hiểu được nguyên nhân hoặc các yếu tố ảnh hưởng dẫn đến kết quả hoặc quyết định cụ thể của hệ thống trí tuệ nhân tạo.

9. *Thay đổi đáng kể của hệ thống trí tuệ nhân tạo* là thay đổi thuộc các trường hợp phải đánh giá lại sự phù hợp theo quy định tại khoản 1 Điều 13 Nghị định số 142/2026/NĐ-CP ngày 30/4/2026 của Chính phủ quy định chi tiết một số điều và biện pháp để tổ chức, hướng dẫn thi hành Luật Trí tuệ nhân tạo.

10. *Cấp có thẩm quyền của đơn vị* là Hội đồng quản trị, Hội đồng thành viên hoặc tổ chức, cá nhân có thẩm quyền tương đương *(sau đây gọi chung là Hội đồng quản lý cấp cao)* hoặc người đại diện theo pháp luật, người đại diện theo ủy quyền của đơn vị *(sau đây gọi chung là người đại điện hợp pháp)* hoặc người được Hội đồng quản lý cấp cao, người đại diện hợp pháp phân cấp, ủy quyền bằng văn bản để thực hiện một hoặc một số chức năng, nhiệm vụ của đơn vị.

11. *Người đại diện* là người đại diện theo pháp luật, người giám hộ của cá nhân mở tài khoản thanh toán, thẻ, ví điện tử, tài khoản Tiền di động theo quy định của pháp luật có liên quan.

12. *Bên thứ ba* là các cá nhân, tổ chức (không bao gồm tổ chức tín dụng nước ngoài và các thành viên thuộc tổ chức tín dụng nước ngoài trong trường hợp tổ chức là đơn vị phụ thuộc, hiện diện thương mại tại Việt Nam của tổ chức tín dụng

nước ngoài) có thỏa thuận bằng văn bản với đơn vị nhằm cung cấp dịch vụ công nghệ số.

13. *Thuê ngoài* là việc đơn vị thỏa thuận bằng văn bản về việc thuê bên thứ ba để thực hiện một phần hoặc toàn bộ hoạt động ứng dụng hệ thống trí tuệ nhân tạo.

Điều 3. Nguyên tắc ứng dụng hệ thống trí tuệ nhân tạo trong ngành Ngân hàng

1. Việc ứng dụng hệ thống trí tuệ nhân tạo phải tuân thủ quy định của pháp luật về trí tuệ nhân tạo, pháp luật về an ninh mạng, pháp luật về dữ liệu, pháp luật về bảo vệ dữ liệu cá nhân và các quy định về quản lý rủi ro của Ngân hàng Nhà nước.

2. Việc ứng dụng hệ thống trí tuệ nhân tạo phải được quản lý phù hợp với mức độ rủi ro và trên cơ sở kết quả đánh giá mức độ ảnh hưởng đối với các hoạt động nghiệp vụ, quyền lợi của khách hàng và hoạt động an toàn của các hệ thống thông tin.

3. Trí tuệ nhân tạo không thay thế thẩm quyền và trách nhiệm của con người. Đơn vị chịu trách nhiệm trước pháp luật và khách hàng đối với mọi quyết định, hoặc hỗ trợ ra quyết định do hệ thống trí tuệ nhân tạo đưa ra mà đơn vị dựa vào đó để cung cấp dịch vụ cho khách hàng. Đơn vị không được viện dẫn các lý do liên quan đến đặc tính kỹ thuật, tính tự chủ của hệ thống trí tuệ nhân tạo hoặc kết quả do bên thứ ba cung cấp để từ chối, miễn trừ hoặc giảm nhẹ trách nhiệm pháp lý, trách nhiệm quản lý và trách nhiệm bồi thường thiệt hại, ngoại trừ trường hợp được miễn trừ theo quy định của pháp luật.

4. Hệ thống trí tuệ nhân tạo phải được thiết kế, triển khai và vận hành an toàn, tin cậy, bảo đảm an toàn hệ thống thông tin, an toàn dữ liệu và tính liên tục trong hoạt động ngân hàng.

5. Hệ thống trí tuệ nhân tạo phải được thiết kế bảo đảm khả năng giám sát và can thiệp của con người đối với hệ thống trong quá trình sử dụng, phù hợp với mức độ rủi ro của hệ thống trí tuệ nhân tạo và bảo đảm hệ thống hoạt động chính xác theo mục tiêu của đơn vị.

Điều 4. Phân loại hệ thống trí tuệ nhân tạo theo mức độ rủi ro

1. Trước khi triển khai hệ thống trí tuệ nhân tạo, đơn vị phải thực hiện phân loại hệ thống trí tuệ nhân tạo theo mức độ rủi ro cao, trung bình, thấp theo quy định của pháp luật về trí tuệ nhân tạo và quy định tại khoản 2 Điều này.

2. Đơn vị xác định hệ thống trí tuệ nhân tạo có rủi ro cao trên cơ sở Danh mục hệ thống trí tuệ nhân tạo có rủi ro cao do Thủ tướng Chính phủ ban hành,

trong đó hệ thống trí tuệ nhân tạo có rủi ro cao trong lĩnh vực ngân hàng, được xác định như sau:

a) Đối với hệ thống trí tuệ nhân tạo tự động thực hiện giao dịch điện tử trong hoạt động ngân hàng, được xác định là hệ thống trí tuệ nhân tạo có rủi ro cao nếu có thực hiện tối thiểu một trong các loại giao dịch sau:

(i) Giao dịch thanh toán của khách hàng cá nhân có giá trị lớn hơn 10 triệu đồng hoặc có tổng giá trị giao dịch trong một ngày đối với một khách hàng do hệ thống trí tuệ nhân tạo tự động thực hiện lớn hơn 20 triệu đồng;

(ii) Giao dịch thanh toán của khách hàng tổ chức có giá trị lớn hơn 50 triệu đồng hoặc có tổng giá trị giao dịch trong một ngày đối với một khách hàng do hệ thống trí tuệ nhân tạo tự động thực hiện lớn hơn 100 triệu đồng;

b) Đối với hệ thống trí tuệ nhân tạo tự động quyết định cấp tín dụng, được xác định là hệ thống trí tuệ nhân tạo có rủi ro cao nếu có xử lý khoản cấp tín dụng có giá trị từ 50 triệu đồng trở lên.

3. Kết quả phân loại hệ thống trí tuệ nhân tạo phải được phê duyệt bởi cấp có thẩm quyền của đơn vị. Đối với chi nhánh ngân hàng nước ngoài, thẩm quyền phê duyệt thực hiện theo quy định của ngân hàng mẹ.

4. Đơn vị được kế thừa kết quả phân loại của nhà cung cấp theo quy định của pháp luật về trí tuệ nhân tạo. Trường hợp sửa đổi, tích hợp hoặc thay đổi chức năng của hệ thống trí tuệ nhân tạo làm phát sinh rủi ro mới hoặc rủi ro cao hơn, đơn vị phối hợp nhà cung cấp thực hiện phân loại lại.

5. Đơn vị phải lập danh mục hệ thống trí tuệ nhân tạo và rà soát, cập nhật danh mục tối thiểu hằng năm. Danh mục tối thiểu bao gồm các trường thông tin: tên hệ thống trí tuệ nhân tạo, mức độ rủi ro, mục đích sử dụng, nhà cung cấp, phiên bản, văn bản phê duyệt mức độ rủi ro.

Điều 5. Điều kiện tự động hóa và kiểm soát của con người đối với hệ thống trí tuệ nhân tạo

1. Việc tự động hóa thực hiện quyết định do hệ thống trí tuệ nhân tạo đề xuất chỉ được thực hiện khi đáp ứng các điều kiện sau:

a) Hoạt động nghiệp vụ được ứng dụng hệ thống trí tuệ nhân tạo không thuộc trường hợp pháp luật quy định phải có sự xem xét, quyết định hoặc phê duyệt của con người;

b) Mô hình đã được kiểm tra, đánh giá đạt ngưỡng chất lượng theo quy định nội bộ của đơn vị;

c) Đơn vị đã xác định rõ ngưỡng rủi ro, tiêu chí cảnh báo và quy trình giám sát hoạt động của hệ thống trí tuệ nhân tạo. Đối với hệ thống trí tuệ nhân tạo có

rủi ro cao, đơn vị phải thiết lập công cụ giám sát liên tục hoạt động của hệ thống trí tuệ nhân tạo;

d) Bảo đảm khả năng can thiệp của con người theo quy định tại khoản 2 Điều này;

đ) Kết quả vận hành của hệ thống phải được rà soát và báo cáo theo quy định tại khoản 6 Điều 12 Thông tư này;

e) Bảo đảm khả năng xác định được giao dịch do hệ thống trí tuệ nhân tạo tự động thực hiện hay do con người thực hiện;

g) Trường hợp triển khai hệ thống trí tuệ nhân tạo tự động thực hiện giao dịch thanh toán, đơn vị phải triển khai đáp ứng các yêu cầu:

(i) Khách hàng và đơn vị phải xác lập thỏa thuận về việc cho phép hệ thống trí tuệ nhân tạo tự động thực hiện giao dịch thanh toán. Khi xác lập thỏa thuận đơn vị có trách nhiệm kiểm tra, đối chiếu, xác minh thông tin nhận biết khách hàng, bảo đảm thực hiện bởi chính chủ tài khoản thanh toán, thẻ, ví điện tử, tài khoản Tiền di động hoặc người đại diện của khách hàng cá nhân hoặc người đại diện theo pháp luật hoặc người đại diện theo ủy quyền của khách hàng tổ chức; trong trường hợp thực hiện bằng phương tiện điện tử phải kiểm tra tối thiểu bằng khớp đúng thông tin sinh trắc học kết hợp với một hình thức xác nhận khác quy định tại Thông tư số 50/2024/TT-NHNN của Thống đốc Ngân hàng Nhà nước Việt Nam quy định về an toàn, bảo mật cho việc cung cấp dịch vụ trực tuyến trong ngành Ngân hàng được sửa đổi, bổ sung bởi Thông tư số 77/2025/TT-NHNN *(sau đây gọi tắt là Thông tư số 50/2024/TT-NHNN).* Thỏa thuận phải xác định rõ: phạm vi giao dịch được tự động thực hiện; hạn mức theo từng giao dịch và theo ngày; thời hạn hiệu lực; điều kiện tạm dừng; quyền của khách hàng chấm dứt thỏa thuận bất kỳ thời điểm nào và trách nhiệm của đơn vị thực hiện ngay yêu cầu chấm dứt;

(ii) Tổng giá trị giao dịch do hệ thống trí tuệ nhân tạo tự động thực hiện trong một ngày của một khách hàng không vượt quá hạn mức theo quy định nội bộ của đơn vị, nhưng không vượt quá 100 triệu đồng đối với khách hàng cá nhân, không vượt quá 500 triệu đồng đối với khách hàng tổ chức;

(iii) Có công cụ cho phép khách hàng chủ động tạm dừng hoặc hủy quyền cho phép hệ thống trí tuệ nhân tạo tự động thực hiện giao dịch thanh toán;

(iv) Trong phạm vi thỏa thuận đã được xác nhận, các giao dịch do hệ thống trí tuệ nhân tạo tự động thực hiện không phải áp dụng xác nhận giao dịch theo điểm a, điểm c khoản 1 Điều 10 Thông tư số 50/2024/TT-NHNN và giá trị giao dịch không tính vào T, T_{ksth} theo quy định tại Phụ lục 01, Phụ lục 03 ban hành kèm theo Thông tư số 50/2024/TT-NHNN;

h) Trường hợp triển khai hệ thống trí tuệ nhân tạo tự động quyết định cấp tín dụng, đơn vị phải tuân thủ quy định về cho vay bằng phương tiện điện tử theo quy định của pháp luật.

2. Đơn vị phải thiết lập cơ chế giám sát và bảo đảm khả năng can thiệp của con người phù hợp với mức độ rủi ro của hệ thống trí tuệ nhân tạo, trong đó:

a) Bộ phận thực hiện giám sát hệ thống trí tuệ nhân tạo phải có đầy đủ quyền hạn và năng lực để:

(i) Ghi đè hoặc điều chỉnh, tạm dừng quyết định của hệ thống trí tuệ nhân tạo hoặc kích hoạt cơ chế ngắt khẩn cấp để dừng hoặc cô lập hệ thống trong trường hợp cần thiết;

(ii) Can thiệp trong các tình huống cần thiết nhằm bảo đảm an toàn hoạt động nghiệp vụ, quyền lợi của khách hàng và hoạt động của các hệ thống thông tin;

b) Mọi hoạt động can thiệp, điều chỉnh hoặc ghi đè quyết định của hệ thống trí tuệ nhân tạo phải được ghi nhận đầy đủ, tối thiểu bao gồm: nội dung thực hiện, thời gian thực hiện, người thực hiện, lý do can thiệp.

3. Nhân sự tham gia giám sát, kiểm soát và can thiệp đối với hệ thống trí tuệ nhân tạo phải được đào tạo và có năng lực phù hợp để giám sát, phát hiện sai lệch, thực hiện hoặc kích hoạt các biện pháp can thiệp, xử lý kịp thời khi hệ thống trí tuệ nhân tạo hoạt động sai lệch.

CHƯƠNG II
YÊU CẦU BẢO ĐẢM AN TOÀN TRONG ỨNG DỤNG HỆ THỐNG TRÍ TUỆ NHÂN TẠO

Điều 6. Bảo đảm an toàn hệ thống trí tuệ nhân tạo

1. Hệ thống trí tuệ nhân tạo phải được kiểm tra, đánh giá an ninh mạng trước khi đưa vào vận hành chính thức và định kỳ hoặc ngay khi có thay đổi đáng kể của hệ thống trí tuệ nhân tạo, tối thiểu bao gồm:

a) Kiểm tra, đánh giá việc phòng, chống các kịch bản tấn công, rủi ro đặc thù đối với hệ thống trí tuệ nhân tạo, tối thiểu bao gồm:

(i) Tấn công vi phạm quyền riêng tư dữ liệu;

(ii) Tấn công đầu độc dữ liệu, tấn công đầu độc mô hình, tấn công né tránh và tấn công trích xuất mô hình (bắt buộc đối với hệ thống trí tuệ nhân tạo có rủi ro cao; khuyến khích áp dụng đối với các hệ thống trí tuệ nhân tạo khác);

(iii) Tấn công tiêm lệnh, rủi ro tạo sinh nội dung sai lệch hoặc không có căn cứ thực tế, rủi ro thiên lệch trong kết quả đầu ra (đối với hệ thống trí tuệ nhân tạo tạo sinh);

b) Đối với các hệ thống trí tuệ nhân tạo có rủi ro cao hoặc hệ thống trí tuệ nhân tạo có rủi ro trung bình ra quyết định trong hoạt động nghiệp vụ, áp dụng các biện pháp phòng, chống các rủi ro, lỗ hổng, tấn công mạng theo các tiêu chuẩn và thông lệ quốc tế như OWASP (OWASP Machine Learning Security Top Ten đối với hệ thống học máy, OWASP Top 10 for Large Language Model Applications đối với hệ thống sử dụng mô hình ngôn ngữ lớn; Phiên bản OWASP áp dụng là phiên bản mới nhất hoặc phiên bản gần nhất với phiên bản được ban hành trong vòng 06 tháng) hoặc tương đương;

c) Thời gian định kỳ kiểm tra, đánh giá an ninh mạng phù hợp với cấp độ an ninh mạng của hệ thống trí tuệ nhân tạo theo quy định của pháp luật về an ninh mạng;

d) Kết quả kiểm tra, đánh giá an ninh mạng và các biện pháp khắc phục phải được lập thành hồ sơ, lưu trữ phục vụ công tác quản lý, kiểm tra, thanh tra theo quy định.

2. Đơn vị phải thiết lập cơ chế giám sát hiệu năng và hành vi của hệ thống trí tuệ nhân tạo, tối thiểu bao gồm: việc phát hiện các dấu hiệu bất thường như sai lệch đầu ra, suy giảm chất lượng, thiên lệch, trôi mô hình, trôi dữ liệu hoặc hành vi bất thường khác cho thấy hệ thống hoạt động không đúng với mục đích thiết kế hoặc phạm vi sử dụng đã được phê duyệt.

Điều 7. Quản lý an toàn dữ liệu

1. Đơn vị phải tuân thủ đầy đủ quy định của pháp luật về dữ liệu, bảo vệ dữ liệu cá nhân, sở hữu trí tuệ và các quy định của Ngân hàng Nhà nước có liên quan trong quá trình xử lý dữ liệu phục vụ hệ thống trí tuệ nhân tạo.

2. Dữ liệu đơn vị sử dụng để huấn luyện, kiểm thử và vận hành hệ thống trí tuệ nhân tạo phải bảo đảm các yêu cầu:

a) Có nguồn gốc rõ ràng, hợp pháp và được lưu thông tin về nguồn gốc. Trường hợp sử dụng dữ liệu từ bên thứ ba, đơn vị phải yêu cầu bên thứ ba cam kết về nguồn gốc hợp pháp, phạm vi và quyền chia sẻ dữ liệu; cung cấp thông tin hoặc bằng chứng phù hợp để chứng minh quyền chia sẻ dữ liệu hoặc mô tả nguồn gốc dữ liệu ở mức phù hợp nhưng không bắt buộc tiết lộ bí mật thương mại, trừ trường hợp sử dụng dữ liệu mở được công khai hợp pháp;

b) Trường hợp sử dụng dữ liệu nhân tạo mô phỏng các đặc điểm của dữ liệu thực tế, đơn vị phải thực hiện đánh giá và có biện pháp phòng ngừa rủi ro suy giảm chất lượng mô hình và sai lệch kết quả theo thời gian;

c) Lưu nhật ký quá trình xử lý dữ liệu đầu vào trước khi đưa dữ liệu vào hệ thống trí tuệ nhân tạo nhằm bảo đảm khả năng truy vết dữ liệu khi cần thiết.

3. Đơn vị phải thiết lập quy trình, chính sách kiểm soát chất lượng dữ liệu sử dụng cho huấn luyện, kiểm thử và vận hành hệ thống trí tuệ nhân tạo, tối thiểu bao gồm các nội dung sau:

a) Dữ liệu huấn luyện, kiểm thử phải bảo đảm:

(i) Có tính đại diện và phù hợp với mục đích sử dụng, đối tượng và phạm vi áp dụng của hệ thống trí tuệ nhân tạo;

(ii) Được cập nhật phù hợp với đặc điểm nghiệp vụ và chu kỳ thay đổi của dữ liệu trong môi trường vận hành thực tế;

b) Thiết lập tiêu chí đo lường và đánh giá chất lượng đối với các tập dữ liệu sử dụng cho huấn luyện, kiểm thử và vận hành hệ thống trí tuệ nhân tạo;

c) Phát hiện và xử lý dữ liệu lỗi, dữ liệu bị sai lệch, bị thao túng hoặc bị đầu độc;

d) Giám sát trôi dữ liệu và thực hiện các biện pháp cập nhật dữ liệu, hiệu chỉnh hoặc huấn luyện lại mô hình nếu cần thiết.

4. Đơn vị phải thiết lập cơ chế bảo vệ dữ liệu khỏi nguy cơ rò rỉ từ hệ thống trí tuệ nhân tạo, tối thiểu bao gồm:

a) Hạn chế và kiểm soát quyền truy cập của hệ thống trí tuệ nhân tạo vào dữ liệu của đơn vị theo nguyên tắc chỉ cấp quyền tối thiểu, khi cần thiết, phù hợp với đối tượng, mục đích sử dụng và phạm vi xử lý dữ liệu đã được phê duyệt;

b) Kiểm soát quyền truy cập đối với dữ liệu huấn luyện, kiểm thử, vận hành của hệ thống trí tuệ nhân tạo và thông tin về cấu trúc mô hình nhằm phòng ngừa nguy cơ rò rỉ, lộ lọt dữ liệu.

Điều 8. Quản lý thay đổi hệ thống trí tuệ nhân tạo

1. Đơn vị phải tuân thủ đầy đủ quy định về quản lý thay đổi hệ thống thông tin theo quy định của Ngân hàng Nhà nước.

2. Đơn vị phải quản lý phiên bản đối với mô hình, tối thiểu bao gồm: thông tin phiên bản mô hình; thời điểm áp dụng, thay đổi hoặc ngừng sử dụng; phạm vi ảnh hưởng; cấu hình liên quan; nhật ký thay đổi; kết quả kiểm tra, đánh giá mô hình.

3. Quản lý thay đổi phiên bản mô hình:

a) Trường hợp mô hình do đơn vị trực tiếp quản lý, đơn vị phải có phương án phục hồi trong trường hợp thực hiện thay đổi không thành công hoặc gặp các sự cố không có khả năng dự tính trước;

b) Đối với các hệ thống trí tuệ nhân tạo có rủi ro cao hoặc hệ thống trí tuệ nhân tạo có rủi ro trung bình ra quyết định trong hoạt động nghiệp vụ sử dụng mô hình do bên thứ ba cung cấp thông qua dịch vụ nền tảng hoặc giao diện lập trình ứng dụng (API), đơn vị phải thực hiện đánh giá tác động và có phương án dự phòng khi nhà cung cấp cập nhật, thay đổi mô hình.

4. Mọi thay đổi đáng kể của hệ thống trí tuệ nhân tạo đều phải được quản lý thông qua quy trình quản lý sự thay đổi, bao gồm: việc đánh giá lại tác động của việc triển khai đến các hệ thống hiện có; cập nhật tài liệu hướng dẫn, kịch bản ứng phó và phương án dự phòng trước khi áp dụng.

Điều 9. Quản lý và xử lý sự cố đối với hệ thống trí tuệ nhân tạo

1. Đơn vị xây dựng và ban hành quy trình xử lý sự cố hệ thống trí tuệ nhân tạo, bao gồm các loại sự cố đặc thù của hệ thống trí tuệ nhân tạo và phương án xử lý nghiệp vụ khi hệ thống trí tuệ nhân tạo gặp sự cố.

2. Khi xảy ra sự cố nghiêm trọng đối với hệ thống trí tuệ nhân tạo, đơn vị phải:

a) Kích hoạt ngay cơ chế ngắt khẩn cấp, chuyển sang vận hành thủ công trong trường hợp cần thiết;

b) Triển khai ngay biện pháp khắc phục, hạn chế thiệt hại;

c) Báo cáo sự cố cho cơ quan quản lý nhà nước có thẩm quyền và Ngân hàng Nhà nước Việt Nam (Cục Công nghệ thông tin) theo quy định của pháp luật về trí tuệ nhân tạo;

d) Thông báo cho khách hàng bị ảnh hưởng theo quy định của pháp luật về bảo vệ dữ liệu cá nhân (nếu có);

đ) Lập báo cáo hoàn thành khắc phục sự cố tối thiểu bao gồm nguyên nhân, thiệt hại và biện pháp khắc phục.

3. Đơn vị phải lưu trữ đầy đủ hồ sơ, tài liệu liên quan đến sự cố để phục vụ kiểm tra, điều tra khi cần thiết.

4. Sau khi xử lý sự cố nghiêm trọng, đơn vị phải tổ chức đánh giá lại mức độ rủi ro của hệ thống trí tuệ nhân tạo; trường hợp cần thiết, thực hiện phân loại lại rủi ro, điều chỉnh biện pháp kiểm soát hoặc tạm dừng đưa hệ thống vào sử dụng để ngăn ngừa tái diễn sự cố.

Điều 10. Ghi nhật ký vận hành hệ thống trí tuệ nhân tạo

1. Hệ thống trí tuệ nhân tạo phải được cấu hình để ghi nhật ký các sự kiện trong quá trình vận hành, tối thiểu bao gồm:

a) Thời gian, hệ thống, người dùng thực hiện các giao dịch hoặc tác vụ;

b) Thông tin về dữ liệu đầu vào và kết quả đầu ra của hệ thống, kèm theo thông tin về phiên bản mô hình, cấu hình hoặc câu lệnh (nếu có) bảo đảm đủ để truy xuất, kiểm tra và tái lập quá trình xử lý khi cần thiết;

c) Các trường hợp kết quả nằm ngoài phạm vi hoạt động dự định hoặc có độ tin cậy thấp;

d) Các sự kiện liên quan đến lỗi của hệ thống hoặc cảnh báo sự cố an ninh mạng.

2. Thời gian lưu trữ nhật ký hoạt động của hệ thống trí tuệ nhân tạo và nhật ký xử lý dữ liệu quy định tại điểm c khoản 2 Điều 7 Thông tư này phải phù hợp với mức độ rủi ro, mục đích sử dụng của hệ thống, tối thiểu đáp ứng quy định về lưu trữ nhật ký của hệ thống thông tin theo cấp độ an ninh mạng.

Điều 12. Cung cấp thông tin và hỗ trợ người sử dụng nội bộ

1. Đơn vị cung cấp đầy đủ, rõ ràng thông tin cho người quản lý, người điều hành và nhân viên sử dụng hệ thống trí tuệ nhân tạo (sau đây gọi chung là người sử dụng nội bộ), bao gồm:

a) Mục đích sử dụng, phạm vi áp dụng, vai trò của hệ thống trong việc hỗ trợ hoặc đưa ra quyết định, khả năng và các giới hạn của hệ thống trí tuệ nhân tạo;

b) Hướng dẫn sử dụng hệ thống và các khuyến nghị về việc diễn giải, sử dụng kết quả đầu ra của hệ thống, bao gồm cảnh báo về nguy cơ sai lệch hoặc sử dụng ngoài mục đích thiết kế;

c) Quy trình và cách thức yêu cầu sự can thiệp của con người khi phát hiện bất thường hoặc rủi ro hoặc sự cố trong quá trình vận hành.

2. Đơn vị phải thiết lập kênh tiếp nhận và xử lý phản hồi từ người sử dụng nội bộ và các bộ phận liên quan về các sự cố, hành vi bất thường, tác động bất lợi hoặc kết quả sai lệch phát sinh trong quá trình vận hành hệ thống trí tuệ nhân tạo, làm cơ sở cho việc điều chỉnh mô hình, quy trình hoặc biện pháp kiểm soát rủi ro.

CHƯƠNG III
QUẢN LÝ RỦI RO ĐỐI VỚI VIỆC ỨNG DỤNG HỆ THỐNG TRÍ TUỆ NHÂN TẠO

Điều 12. Trách nhiệm quản lý rủi ro trong ứng dụng hệ thống trí tuệ nhân tạo

1. Đơn vị xác định rõ vai trò, trách nhiệm của các cá nhân, bộ phận chức năng trong quản lý rủi ro ứng dụng hệ thống trí tuệ nhân tạo và thiết lập cơ chế phối hợp giữa các bộ phận liên quan.

2. Hội đồng quản lý cấp cao chịu trách nhiệm cao nhất về việc ứng dụng hệ thống trí tuệ nhân tạo tại đơn vị, bao gồm:

a) Phê duyệt định hướng phát triển và ứng dụng hệ thống trí tuệ nhân tạo phù hợp với chiến lược kinh doanh và chính sách quản lý rủi ro của đơn vị;

b) Phê duyệt khung quản lý rủi ro đối với ứng dụng hệ thống trí tuệ nhân tạo và các chính sách quan trọng có liên quan;

c) Phân bổ nguồn lực đầy đủ cho việc triển khai, vận hành và quản lý rủi ro đối với việc ứng dụng hệ thống trí tuệ nhân tạo;

d) Tổ chức triển khai giám sát việc thực hiện quản lý rủi ro đối với việc ứng dụng hệ thống trí tuệ nhân tạo.

3. Đối với hệ thống trí tuệ nhân tạo có rủi ro cao, Hội đồng quản lý cấp cao phải phê duyệt chủ trương triển khai và quyết định đưa hệ thống vào vận hành chính thức.

4. Việc phê duyệt chi tiết về kỹ thuật, triển khai và vận hành hệ thống trí tuệ nhân tạo được thực hiện theo phân cấp, ủy quyền nội bộ của đơn vị.

5. Thẩm quyền phê duyệt các nội dung quy định tại khoản 1, 2, 3 và 4 Điều này đối với ứng dụng hệ thống trí tuệ nhân tạo của chi nhánh ngân hàng nước ngoài thực hiện theo quy định của ngân hàng mẹ.

6. Đơn vị phải báo cáo Hội đồng quản lý cấp cao, ngân hàng mẹ về tình hình vận hành, rủi ro, sự cố nghiêm trọng, khiếu nại và xu hướng sai lệch của hệ thống định kỳ tối thiểu 6 tháng đối với hệ thống trí tuệ nhân tạo có rủi ro cao và rủi ro trung bình hoặc khi có thay đổi đáng kể của hệ thống trí tuệ nhân tạo có rủi ro cao.

Điều 13. Đánh giá tác động đối với hệ thống trí tuệ nhân tạo có rủi ro cao

1. Đối với hệ thống trí tuệ nhân tạo có rủi ro cao, đơn vị phải thực hiện đánh giá tác động của hệ thống trí tuệ nhân tạo trước khi triển khai, tối thiểu bao gồm các nội dung sau:

a) Mô tả mục đích, phạm vi và nguyên lý hoạt động của hệ thống; mô tả cơ chế bảo đảm khả năng giám sát, can thiệp của con người;

b) Phân tích rủi ro tiềm ẩn đối với hoạt động nghiệp vụ; rủi ro an ninh mạng đối với hệ thống thông tin; quyền lợi của khách hàng và các bên liên quan (trong

đó đánh giá cụ thể tác động về tính công bằng, sự minh bạch, khả năng giải thích và nguy cơ phân biệt đối xử đối với khách hàng);

c) Đánh giá tác động xử lý dữ liệu cá nhân theo quy định của pháp luật (nếu có);

d) Xác định các biện pháp kiểm soát và giảm thiểu rủi ro;

đ) Xây dựng kế hoạch giám sát và phương án ứng phó sự cố trong quá trình vận hành.

2. Kết quả đánh giá tác động phải được lập thành hồ sơ, phê duyệt bởi cấp có thẩm quyền của đơn vị; được rà soát, cập nhật khi có sự thay đổi về chức năng, phạm vi xử lý dữ liệu hoặc điều kiện vận hành của hệ thống trí tuệ nhân tạo; và lưu trữ hồ sơ trong suốt thời gian hoạt động của hệ thống.

Điều 14. Khung quản lý rủi ro đối với ứng dụng hệ thống trí tuệ nhân tạo

1. Đơn vị phải xây dựng, ban hành và tổ chức thực hiện khung quản lý rủi ro đối với ứng dụng hệ thống trí tuệ nhân tạo, bảo đảm phù hợp, đồng bộ với chính sách quản lý rủi ro tổng thể và không tách rời khỏi hệ thống kiểm soát nội bộ chung của đơn vị.

2. Đơn vị tổ chức quản lý rủi ro đối với việc ứng dụng hệ thống trí tuệ nhân tạo theo 03 tuyến bảo vệ độc lập, trong đó:

a) Tuyến bảo vệ thứ nhất là bộ phận hoặc cá nhân thực hiện chức năng phát triển, triển khai, vận hành, sử dụng hệ thống trí tuệ nhân tạo; nhận dạng, thực hiện các biện pháp kiểm soát, theo dõi, giảm thiểu rủi ro;

b) Tuyến bảo vệ thứ hai là bộ phận hoặc cá nhân thực hiện chức năng kiểm tra, đánh giá hệ thống trí tuệ nhân tạo; xây dựng chính sách quản lý rủi ro, quy định nội bộ về quản lý rủi ro; theo dõi, kiểm soát rủi ro và tuân thủ quy định của pháp luật;

c) Tuyến bảo vệ thứ ba là bộ phận hoặc cá nhân thực hiện chức năng kiểm toán nội bộ;

d) Đơn vị được phép thuê ngoài để thực hiện một số chức năng của các tuyến bảo vệ, nhưng phải bảo đảm tính độc lập giữa các tuyến.

3. Các loại rủi ro đối với ứng dụng hệ thống trí tuệ nhân tạo được quản lý tối thiểu bao gồm:

a) Rủi ro mô hình;

b) Rủi ro quản trị dữ liệu;

c) Rủi ro an ninh mạng;

d) Rủi ro tuân thủ pháp luật;

đ) Rủi ro đạo đức;

e) Rủi ro từ bên thứ ba;

g) Rủi ro chuỗi cung ứng;

h) Rủi ro danh tiếng (đối với ngân hàng thương mại và chi nhánh ngân hàng nước ngoài).

4. Đơn vị xác định nội hàm, tiêu chí nhận dạng, phương pháp đo lường, ngưỡng chấp nhận và biện pháp kiểm soát đối với từng loại rủi ro quy định tại khoản 3 Điều này.

Điều 15. Quản lý vòng đời hệ thống trí tuệ nhân tạo

1. Đơn vị phải thiết lập, thực hiện và duy trì quản lý rủi ro đối với hệ thống trí tuệ nhân tạo trong suốt vòng đời của hệ thống trong phạm vi trách nhiệm và khả năng kiểm soát của đơn vị, bao gồm từ giai đoạn thiết kế, phát triển, kiểm thử, triển khai, vận hành đến khi ngừng sử dụng được quy định tại khoản 2, 3, 4 và 5 của Điều này.

2. Trong giai đoạn thiết kế và phát triển hệ thống trí tuệ nhân tạo, đơn vị phải:

a) Nhận dạng, phân tích các rủi ro đã biết và rủi ro có thể lường trước có thể ảnh hưởng đến hoạt động nghiệp vụ, quyền lợi của khách hàng và an ninh mạng;

b) Thực hiện quản trị dữ liệu huấn luyện theo quy định tại điểm a khoản 3 Điều 7 Thông tư này;

c) Tích hợp các yêu cầu về an ninh mạng, cơ chế giám sát, kiểm soát và khả năng can thiệp của con người ngay từ khâu thiết kế hệ thống.

3. Trong giai đoạn kiểm thử, đơn vị phải:

a) Kiểm tra, đánh giá an ninh mạng hệ thống trí tuệ nhân tạo trước khi đưa vào sử dụng, bảo đảm đáp ứng các yêu cầu quy định tại khoản 1 Điều 6 Thông tư này;

b) Kiểm tra, đánh giá mô hình trước khi đưa vào sử dụng, bảo đảm đáp ứng các yêu cầu quy định tại Điều 16 Thông tư này.

4. Trong giai đoạn triển khai và vận hành, đơn vị phải:

a) Thực hiện giám sát hiệu năng và hành vi của hệ thống trí tuệ nhân tạo theo quy định tại khoản 2 Điều 6 Thông tư này;

b) Xây dựng quy trình báo cáo và xử lý các sự cố nghiêm trọng;

c) Xem xét việc thực hiện lại quy trình đánh giá rủi ro; kiểm tra, đánh giá an ninh mạng tương tự như triển khai mới để bảo đảm bản cập nhật không gây ra các tác động tiêu cực khi có thay đổi đáng kể của hệ thống trí tuệ nhân tạo.

5. Giai đoạn ngừng hoạt động:

a) Đơn vị phải lưu trữ hồ sơ, tài liệu cần thiết nhằm phục vụ công tác kiểm toán theo quy định của pháp luật, tối thiểu bao gồm: tài liệu thiết kế; hồ sơ liên quan đến điều kiện trước khi đưa hệ thống trí tuệ nhân tạo vào vận hành quy định tại Điều 22 Thông tư này (và tài liệu cập nhật trong quá trình vận hành nếu có); hồ sơ kỹ thuật của hệ thống trí tuệ nhân tạo có rủi ro cao theo quy định của pháp luật về trí tuệ nhân tạo;

b) Khi hệ thống trí tuệ nhân tạo không còn sử dụng, đơn vị phải thực hiện quy trình ngừng hoạt động an toàn, bao gồm việc xóa (sau khi thực hiện việc lưu trữ dữ liệu theo quy định của pháp luật) hoặc vô hiệu hóa an toàn hoặc cô lập dữ liệu, mô hình, cấu hình và các thành phần liên quan thuộc phạm vi quản lý, kiểm soát của đơn vị nhằm ngăn chặn rò rỉ dữ liệu hoặc khôi phục trái phép;

c) Đối với dữ liệu huấn luyện và tham số mô hình do bên thứ ba quản lý, đơn vị phải có cơ chế phối hợp, yêu cầu và lưu trữ hồ sơ, tài liệu về việc xử lý phù hợp theo thỏa thuận hợp đồng.

Điều 16. Quản lý rủi ro mô hình

1. Đơn vị thực hiện quản lý rủi ro mô hình theo quy định tại Điều này, tối thiểu bao gồm các mô hình của hệ thống trí tuệ nhân tạo có rủi ro cao và mô hình của hệ thống trí tuệ nhân tạo có rủi ro trung bình ra quyết định trong hoạt động nghiệp vụ. Đối với hệ thống trí tuệ nhân tạo khác, đơn vị thực hiện theo quy định nội bộ của đơn vị.

2. Đơn vị thực hiện kiểm tra, đánh giá mô hình trong các trường hợp sau:

a) Trước khi đưa vào vận hành lần đầu;

b) Định kỳ tối thiểu hằng năm trong quá trình vận hành;

c) Khi xảy ra sự cố nghiêm trọng; khi có thay đổi đáng kể của hệ thống trí tuệ nhân tạo; khi phát hiện bất thường, sai lệch nghiêm trọng trong quá trình vận hành; hoặc khi mức độ rủi ro của hệ thống thay đổi.

3. Việc kiểm tra, đánh giá mô hình phải tối thiểu bao gồm các nội dung sau:

a) Đối với mô hình do đơn vị trực tiếp quản lý:

(i) Sự phù hợp của dữ liệu huấn luyện và dữ liệu kiểm thử;

(ii) Tính hợp lý của thuật toán và kiến trúc mô hình;

(iii) Hiệu năng và độ chính xác của mô hình;

(iv) Khả năng giải thích về cơ chế ra quyết định và mức độ ảnh hưởng của các yếu tố đầu vào chính đối với kết quả của mô hình;

(v) Tính công bằng và kiểm soát nguy cơ thiên lệch;

(vi) Độ ổn định và khả năng chống chịu của mô hình trong các điều kiện vận hành;

b) Đối với mô hình do bên thứ ba cung cấp dưới dạng dịch vụ thông qua nền tảng hoặc giao diện lập trình ứng dụng:

(i) Mục đích sử dụng, phạm vi triển khai và mức độ rủi ro của hệ thống;

(ii) Kết quả kiểm thử hộp đen thực tế trên tập dữ liệu của đơn vị;

(iii) Kết quả đầu ra của mô hình trong điều kiện vận hành thực tế, trừ trường hợp quy định tại điểm a khoản 2 Điều này;

(iv) Tài liệu, thông tin, báo cáo hoặc cam kết do nhà cung cấp công bố;

(v) Các biện pháp kiểm soát, giám sát và can thiệp của con người trong quá trình sử dụng.

4. Việc kiểm tra, đánh giá mô hình được thực hiện bởi bộ phận độc lập với bộ phận phát triển, triển khai, vận hành hệ thống trí tuệ nhân tạo và nhà cung cấp mô hình.

5. Đơn vị phải thiết lập cơ chế giám sát hiệu năng của mô hình, bao gồm:

a) Phát hiện kịp thời các dấu hiệu suy giảm chất lượng mô hình;

b) So sánh định kỳ giữa kết quả dự đoán của mô hình và kết quả thực tế;

c) Thiết lập ngưỡng cảnh báo và biện pháp can thiệp khi phát hiện bất thường.

6. Kết quả kiểm tra, đánh giá và giám sát phải được báo cáo cấp có thẩm quyền của đơn vị.

Điều 17. Quản lý rủi ro đạo đức

1. Đơn vị phải áp dụng các biện pháp kiểm soát phù hợp nhằm phát hiện và giảm thiểu thiên lệch dữ liệu, thiên lệch mô hình và thiên lệch vận hành; bảo đảm xem xét đầy đủ tác động đến nhóm dễ bị tổn thương bao gồm trẻ em, người cao tuổi, người khuyết tật, nhóm yếu thế.

2. Đơn vị phải đánh giá cụ thể tác động về tính công bằng, sự minh bạch, khả năng giải thích và nguy cơ phân biệt đối xử đối với khách hàng theo quy định tại điểm b khoản 1 Điều 13 Thông tư này đối với hệ thống trí tuệ nhân tạo có rủi ro cao.

3. Khuyến khích áp dụng Khung đạo đức trí tuệ nhân tạo quốc gia do Bộ Khoa học và Công nghệ ban hành.

Điều 18. Quản lý rủi ro đối với việc thuê ngoài

1. Phạm vi và nguyên tắc chung về thuê ngoài

a) Đơn vị được phép thuê ngoài việc phát triển, triển khai, vận hành hệ thống trí tuệ nhân tạo hoặc sử dụng giải pháp trí tuệ nhân tạo do bên thứ ba cung cấp theo quy định của pháp luật và quy định tại Thông tư này;

b) Việc thuê ngoài không làm thay đổi trách nhiệm của đơn vị đối với việc bảo đảm an toàn hoạt động nghiệp vụ, bảo vệ quyền lợi của khách hàng, việc tuân thủ pháp luật về an ninh mạng, bảo vệ dữ liệu cá nhân và quản lý rủi ro.

2. Yêu cầu đối với việc thuê ngoài hệ thống trí tuệ nhân tạo:

a) Khi sử dụng giải pháp trí tuệ nhân tạo do bên thứ ba cung cấp, đơn vị phải thực hiện đánh giá rủi ro khi sử dụng dịch vụ bên thứ ba theo quy định của pháp luật và xác định các tiêu chí lựa chọn bên thứ ba tối thiểu bao gồm:

(i) Năng lực kỹ thuật, kinh nghiệm và uy tín của bên thứ ba;

(ii) Khả năng tuân thủ các quy định của pháp luật Việt Nam về an ninh mạng và bảo vệ dữ liệu cá nhân;

(iii) Phương án dự phòng và năng lực hỗ trợ khi có sự cố;

(iv) Các chứng nhận về an ninh mạng và bảo vệ dữ liệu cá nhân trong trường hợp triển khai hệ thống trí tuệ nhân tạo có rủi ro cao;

b) Đơn vị phải xây dựng và ban hành hướng dẫn thực hiện quản lý rủi ro đối với việc thuê ngoài hệ thống trí tuệ nhân tạo, trong đó xác định rõ:

(i) Phạm vi công việc được thuê ngoài;

(ii) Yêu cầu về an toàn, bảo mật, chất lượng và kiểm soát rủi ro;

(iii) Cơ chế phối hợp, giám sát và báo cáo.

3. Yêu cầu về hợp đồng thuê ngoài hệ thống trí tuệ nhân tạo tối thiểu bao gồm:

a) Tuân thủ yêu cầu về hợp đồng sử dụng dịch vụ công nghệ thông tin với bên thứ ba theo quy định của Thống đốc Ngân hàng Nhà nước;

b) Bên thứ ba phải cung cấp báo cáo kiểm toán tuân thủ công nghệ thông tin do tổ chức kiểm toán độc lập thực hiện hằng năm trong thời gian thực hiện hợp đồng hoặc các chứng chỉ bảo mật, quản lý rủi ro quốc tế tương đương (như SOC 2, ISO/IEC 27001, ISO/IEC 42001) còn hiệu lực. Trường hợp bên thứ ba không cung cấp được tài liệu này, đơn vị thực hiện đánh giá hoặc thuê tổ chức độc lập đánh giá tuân thủ công nghệ thông tin đối với bên thứ ba;

c) Bên thứ ba phải cung cấp thông tin hoặc báo cáo kết quả kiểm tra, đánh giá chất lượng mô hình hằng năm hoặc theo từng phiên bản được phát hành;

d) Bên thứ ba phải cung cấp công cụ hoặc giao diện lập trình ứng dụng (API) hoặc dữ liệu nhật ký (logs) hoặc các phương tiện kỹ thuật cần thiết khác để đơn vị giám sát hiệu năng của mô hình;

đ) Bên thứ ba có trách nhiệm cung cấp thông tin về đặc tính kỹ thuật, giới hạn mô hình, cơ chế cập nhật và các biện pháp bảo mật liên quan để đơn vị thực hiện đánh giá rủi ro và quản trị vận hành;

e) Bên thứ ba cam kết bảo đảm tính liên tục trong hoạt động cung cấp dịch vụ, có phương án dự phòng và giai đoạn chuyển tiếp phù hợp khi ngừng hỗ trợ hoặc có thay đổi đáng kể về mô hình;

g) Bên thứ ba phải phối hợp, cung cấp thông tin cho cơ quan nhà nước có thẩm quyền trong kiểm tra, đánh giá, hậu kiểm và khắc phục sự cố liên quan đến hệ thống theo quy định của pháp luật về trí tuệ nhân tạo;

h) Bên thứ ba cam kết không sử dụng dữ liệu của đơn vị để huấn luyện, tinh chỉnh hoặc cải thiện mô hình của bên thứ ba mà không có sự đồng ý bằng văn bản của đơn vị.

4. Giám sát việc thuê ngoài

a) Đơn vị phải thiết lập và duy trì cơ chế giám sát thường xuyên đối với hoạt động của bên nhận thuê ngoài trong suốt thời gian hợp tác, bảo đảm việc phát triển, triển khai và vận hành hệ thống trí tuệ nhân tạo phù hợp với chính sách, quy trình và hợp đồng đã ký kết;

b) Đối với các hệ thống trí tuệ nhân tạo có rủi ro cao, đơn vị phải:

(i) Xem xét, phê duyệt kế hoạch phát triển, triển khai và vận hành hệ thống;

(ii) Yêu cầu bên thứ ba cung cấp đầy đủ thông tin, tài liệu phục vụ công tác kiểm tra, đánh giá, giám sát và kiểm toán;

(iii) Thiết lập quy trình và biện pháp xử lý khi phát hiện sai sót, vi phạm hoặc sự cố;

c) Đơn vị phải thực hiện kiểm tra, đánh giá toàn diện việc thuê ngoài hệ thống trí tuệ nhân tạo theo định kỳ hoặc khi có dấu hiệu rủi ro gia tăng;

d) Đơn vị và bên thứ ba phải phối hợp thiết lập các biện pháp phòng ngừa nhằm hạn chế rủi ro phát sinh từ việc vận hành hệ thống trí tuệ nhân tạo;

đ) Đơn vị phải đánh giá và quản lý rủi ro phát sinh từ việc phụ thuộc vào một nền tảng trí tuệ nhân tạo khi nhà cung cấp gặp sự cố;

e) Trường hợp hệ thống trí tuệ nhân tạo do thuê ngoài gây thiệt hại cho khách hàng hoặc bên liên quan, đơn vị phải:

(i) Kịp thời thực hiện các biện pháp khắc phục, bảo vệ quyền lợi của khách hàng;

(ii) Tổ chức tiếp nhận, xử lý khiếu nại và bồi thường thiệt hại theo quy định của pháp luật và thỏa thuận hợp đồng.

Điều 19. Quản lý rủi ro chuỗi cung ứng hệ thống trí tuệ nhân tạo

1. Đơn vị phải thiết lập quy trình kiểm tra, đánh giá an ninh mạng tối thiểu đối với các thư viện phần mềm, bộ dữ liệu và các mô hình được huấn luyện sẵn có nguồn gốc từ bên thứ ba hoặc mã nguồn mở trước khi tích hợp và trong suốt vòng đời sử dụng hệ thống, bảo đảm các thành phần này không chứa mã độc hoặc lỗ hổng bảo mật.

2. Đơn vị phải xây dựng và duy trì danh mục chi tiết các thành phần phần mềm của hệ thống trí tuệ nhân tạo, tối thiểu bao gồm: Tên thành phần (thư viện, mô hình, tập dữ liệu huấn luyện và các thành phần phụ thuộc); tên nhà cung cấp hoặc nhà phát triển thành phần; phiên bản đang sử dụng; giấy phép sử dụng; mã kiểm tra tính toàn vẹn và xác thực nguồn gốc; thời điểm tạo lập danh mục để phục vụ việc truy vết và ứng phó kịp thời đối với các rủi ro, sự cố an ninh mạng.

3. Đối với các mô hình được chuyển giao từ bên thứ ba hoặc tải về từ các kho mã nguồn mở, đơn vị phải áp dụng các biện pháp kỹ thuật để xác minh tính toàn vẹn, nguồn gốc bảo đảm mô hình không bị can thiệp hoặc bị thay đổi trái phép trong quá trình chuyển giao.

4. Đơn vị phải cập nhật thường xuyên thông tin về các lỗ hổng bảo mật, cảnh báo an ninh và khuyến nghị từ nhà cung cấp liên quan đến các nền tảng, thư viện trí tuệ nhân tạo đang sử dụng; thực hiện đánh giá rủi ro và áp dụng kịp thời các biện pháp khắc phục phù hợp với mức độ rủi ro.

5. Đơn vị phải thực hiện đánh giá định kỳ quản lý rủi ro chuỗi cung ứng hệ thống trí tuệ nhân tạo phù hợp với mức độ rủi ro của hệ thống, tối thiểu hằng năm với hệ thống trí tuệ nhân tạo có rủi ro cao.

CHƯƠNG IV
ĐIỀU KIỆN TRIỂN KHAI ỨNG DỤNG HỆ THỐNG TRÍ TUỆ NHÂN TẠO

Điều 20. Điều kiện triển khai đối với hệ thống trí tuệ nhân tạo ra quyết định trong các hoạt động nghiệp vụ

1. Hệ thống trí tuệ nhân tạo ra quyết định trong các hoạt động nghiệp vụ phải đáp ứng các điều kiện trước khi đưa vào vận hành quy định tại Điều 22 Thông tư này và các yêu cầu quy định tại Điều này.

2. Đơn vị có trách nhiệm cung cấp thông tin hướng dẫn sử dụng và nêu rõ các giới hạn của hệ thống trí tuệ nhân tạo để khách hàng biết trước khi thoả thuận sử dụng dịch vụ của đơn vị.

3. Đối với các hệ thống trí tuệ nhân tạo đóng vai trò thay thế con người trong việc giao tiếp trực tiếp với khách hàng, đơn vị phải có biện pháp thông báo cho khách hàng biết việc đang tương tác với hệ thống trí tuệ nhân tạo ngay từ khi bắt đầu phiên làm việc hoặc hội thoại.

4. Đơn vị phải thiết lập bộ phận và quy trình tiếp nhận, xử lý khiếu nại liên quan đến việc sử dụng hệ thống trí tuệ nhân tạo; trong đó bảo đảm việc rà soát, đánh giá lại bởi con người đối với các quyết định tự động bị khiếu nại và sử dụng kết quả rà soát của con người làm căn cứ cuối cùng để giải quyết khiếu nại.

5. Đối với các hệ thống trí tuệ nhân tạo tự động ra quyết định có ảnh hưởng trực tiếp đến trách nhiệm pháp lý, nghĩa vụ tài chính hoặc việc tiếp cận, sử dụng sản phẩm, dịch vụ của khách hàng, đơn vị phải có biện pháp lưu trữ thông tin về cơ sở và các yếu tố chính dẫn đến quyết định đó để phục vụ việc cung cấp, giải thích khi khách hàng có yêu cầu. Việc cung cấp thông tin giải thích không được làm lộ mã nguồn, thuật toán chi tiết, bộ tham số hoặc thông tin thuộc bí mật kinh doanh, bí mật công nghệ. Trường hợp quyết định của hệ thống trí tuệ nhân tạo nhằm mục đích phòng chống gian lận, lừa đảo, rửa tiền, tài trợ khủng bố hoặc bảo đảm an ninh mạng, đơn vị chỉ cần thông báo lý do phù hợp với việc rà soát rủi ro, bảo đảm an toàn tài sản cho khách hàng.

Điều 21. Yêu cầu về nguồn nhân lực triển khai và quản lý hệ thống trí tuệ nhân tạo

1. Trường hợp hệ thống trí tuệ nhân tạo do đơn vị trực tiếp phát triển, triển khai, vận hành, đơn vị phải bố trí nguồn lực có trình độ chuyên môn và kinh nghiệm phù hợp để thực hiện các nhiệm vụ sau:

a) Phát triển, triển khai, vận hành, giám sát và bảo trì hệ thống trí tuệ nhân tạo;

b) Quản lý rủi ro, kiểm tra, đánh giá mô hình;

c) Kiểm toán hệ thống trí tuệ nhân tạo.

2. Trường hợp đơn vị thuê ngoài phát triển, triển khai hoặc vận hành hệ thống trí tuệ nhân tạo, đơn vị có trách nhiệm quản lý, giám sát và đánh giá năng lực của bên cung cấp dịch vụ bảo đảm triển khai các nhiệm vụ tại khoản 1 Điều này.

3. Đơn vị phải xây dựng và triển khai chương trình đào tạo, nâng cao năng lực cho cán bộ, nhân viên liên quan về:

a) Kiến thức cơ bản và ứng dụng của trí tuệ nhân tạo trong hoạt động ngân hàng;

b) Quản lý rủi ro đối với việc ứng dụng hệ thống trí tuệ nhân tạo và các phương pháp đánh giá, kiểm soát rủi ro;

c) Đạo đức trí tuệ nhân tạo, bảo vệ dữ liệu cá nhân và quyền riêng tư;

d) An ninh mạng và các quy định của pháp luật liên quan.

4. Định kỳ tối thiểu hằng năm, đơn vị phải đánh giá và cập nhật năng lực của đội ngũ nhân sự quy định tại khoản 1 Điều này.

Điều 22. Điều kiện trước khi đưa hệ thống trí tuệ nhân tạo vào vận hành

1. Đơn vị lập hồ sơ liên quan đến hệ thống trí tuệ nhân tạo theo khoản 2 Điều này và báo cáo cấp có thẩm quyền của đơn vị quyết định việc đưa hệ thống trí tuệ nhân tạo vào vận hành chính thức.

2. Hồ sơ bao gồm:

a) Hồ sơ phân loại rủi ro theo quy định của pháp luật về trí tuệ nhân tạo;

b) Kết quả kiểm tra, đánh giá an ninh mạng;

c) Kết quả kiểm tra, đánh giá mô hình;

d) Kết quả đánh giá tác động của hệ thống trí tuệ nhân tạo có rủi ro cao theo quy định tại Điều 13 Thông tư này;

đ) Kết quả đánh giá sự phù hợp của hệ thống trí tuệ nhân tạo có rủi ro cao theo quy định của pháp luật về trí tuệ nhân tạo;

e) Kế hoạch vận hành an toàn, bao gồm giám sát, bảo đảm hoạt động liên tục và ứng phó sự cố;

g) Kết quả đánh giá đáp ứng điều kiện triển khai đối với hệ thống trí tuệ nhân tạo ra quyết định trong các hoạt động nghiệp vụ theo quy định tại Điều 20 Thông tư này;

h) Báo cáo về nguồn nhân lực triển khai và quản lý hệ thống trí tuệ nhân tạo theo quy định tại Điều 21 Thông tư này;

i) Điều kiện dừng và khôi phục hệ thống.

3. Đơn vị phải thiết lập các tiêu chí cụ thể cần đáp ứng trước khi đưa hệ thống trí tuệ nhân tạo vào vận hành, tối thiểu bao gồm:

a) Chỉ số về độ chính xác, độ tin cậy của mô hình;

b) Chỉ số về hiệu năng của hệ thống;

c) Kết quả kiểm tra, đánh giá an ninh mạng.

4. Đơn vị phải xây dựng kế hoạch triển khai chi tiết, bao gồm lộ trình, phương án chuyển đổi dữ liệu, phương án phục hồi khi gặp sự cố và đánh giá tác động của việc triển khai đến các hệ thống hiện có.

CHƯƠNG V
ĐIỀU KHOẢN THI HÀNH

Điều 23. Hiệu lực thi hành

1. Thông tư này có hiệu lực thi hành kể từ ngày 15 tháng 11 năm 2026, trừ quy định tại khoản 2, 3 và 4 Điều này.

2. Quy định tại Điều 14 Thông tư này có hiệu lực thi hành từ ngày 01 tháng 01 năm 2027.

3. Quy định tại điểm b khoản 1 Điều 6 Thông tư này có hiệu lực thi hành kể từ ngày 01 tháng 09 năm 2027.

4. Quy định tại Điều 21 Thông tư này có hiệu lực thi hành kể từ ngày 01 tháng 09 năm 2027, ngoại trừ trường hợp đơn vị có hệ thống trí tuệ nhân tạo có rủi ro cao.

Điều 24. Điều khoản chuyển tiếp

Các hệ thống trí tuệ nhân tạo được triển khai đưa vào vận hành chính thức trước ngày Thông tư này có hiệu lực thi hành thì phải tuân thủ các quy định về bảo đảm an toàn, quản lý rủi ro và điều kiện triển khai quy định tại Chương II, Chương III, Chương IV Thông tư này kể từ ngày 01 tháng 03 năm 2027.

Điều 25. Trách nhiệm tổ chức thực hiện

1. Cục Công nghệ thông tin có trách nhiệm theo dõi, kiểm tra và phối hợp với các đơn vị liên quan để xử lý những vướng mắc phát sinh trong quá trình thực hiện Thông tư này.

2. Thủ trưởng các đơn vị thuộc Ngân hàng Nhà nước, Giám đốc Ngân hàng

Nhà nước chi nhánh tại các khu vực, các tổ chức tín dụng, chi nhánh ngân hàng nước ngoài, các tổ chức cung ứng dịch vụ trung gian thanh toán, công ty thông tin tín dụng, tổ chức cung ứng dịch vụ Tiền di động, Công ty Cổ phần Thanh toán Quốc gia Việt Nam, Công ty Quản lý tài sản của các tổ chức tín dụng Việt Nam, Nhà máy in tiền quốc gia, Bảo hiểm tiền gửi Việt Nam chịu trách nhiệm tổ chức thực hiện Thông tư này./.

KT. THỐNG ĐỐC
PHÓ THỐNG ĐỐC

Phạm Tiến Dũng