

Số: /2026/QĐ-UBND

Hà Nội, ngày tháng năm 2026

QUYẾT ĐỊNH

Ban hành Quy định về bảo đảm an toàn, quản lý rủi ro và biện pháp ứng dụng trí tuệ nhân tạo trong quản lý nhà nước trên địa bàn thành phố Hà Nội

Căn cứ Luật Tổ chức chính quyền địa phương số 72/2025/QH15;

Căn cứ Luật Thủ đô số 02/2026/QH16;

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

Căn cứ Luật Dữ liệu số 60/2024/QH15;

Căn cứ Luật Bảo vệ dữ liệu cá nhân số 91/2025/QH15;

Căn cứ Luật An ninh mạng số 116/2025/QH15;

Căn cứ Luật Bảo vệ bí mật nhà nước số 117/2025/QH15;

Căn cứ Nghị định số 356/2025/NĐ-CP của Chính phủ quy định chi tiết một số điều và biện pháp thi hành Luật Bảo vệ dữ liệu cá nhân;

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 Thông tư số 05/2026/TT-BKHCN của Bộ Khoa học và Công nghệ về ban hành Khung đạo đức trí tuệ nhân tạo quốc gia;

Theo đề nghị của Giám đốc Sở Khoa học và Công nghệ tại Tờ trình số 5213/TTr-SKHCN ngày 16 tháng 7 năm 2026 về việc ban hành Quyết định ban hành Quy định về bảo đảm an toàn, quản lý rủi ro và biện pháp ứng dụng trí tuệ nhân tạo trong quản lý nhà nước trên địa bàn thành phố Hà Nội;

Ủy ban nhân dân thành phố Hà Nội ban hành Quyết định ban hành Quy định về bảo đảm an toàn, quản lý rủi ro và biện pháp ứng dụng trí tuệ nhân tạo trong quản lý nhà nước trên địa bàn thành phố Hà Nội.

Điều 1. Ban hành kèm theo Quyết định này “Quy định về bảo đảm an toàn, quản lý rủi ro và biện pháp ứng dụng trí tuệ nhân tạo trong quản lý nhà nước trên địa bàn thành phố Hà Nội”.

Điều 2. Quyết định này có hiệu lực thi hành kể từ ngày tháng năm 2026.

Điều 3. Chánh Văn phòng UBND Thành phố; Giám đốc các Sở, ban, ngành; Chủ tịch UBND xã, phường và các tổ chức, cá nhân có liên quan chịu trách nhiệm thi hành Quyết định này./.

Nơi nhận:

- Như Điều 3;
- Bộ Khoa học và Công nghệ;
- Thường trực Thành ủy;
- Thường trực HĐND Thành phố;
- Chủ tịch UBND Thành phố;
- Các Phó Chủ tịch UBND Thành phố;
- Cổng Thông tin điện tử Chính phủ;
- Cục KTVB&TCTHPL - Bộ Tư pháp;
- Các Sở, ban, ngành Thành phố;
- UBND xã, phường;
- VPUB: CVP, các PCVP, các phòng, ban, đơn vị trực thuộc;
- Lưu: VT, ĐMPT.

**TM. ỦY BAN NHÂN DÂN
CHỦ TỊCH**

Vũ Đại Thắng

QUY ĐỊNH

Về bảo đảm an toàn, quản lý rủi ro và biện pháp ứng dụng trí tuệ nhân tạo trong quản lý nhà nước trên địa bàn thành phố Hà Nội

(Ban hành kèm theo Quyết định số /2026/QĐ-UBND ngày / /2026 của Ủy ban nhân dân thành phố Hà Nội)

Chương I QUY ĐỊNH CHUNG

Điều 1. Phạm vi điều chỉnh

1. Quy định về bảo đảm an toàn, quản lý rủi ro và biện pháp ứng dụng trí tuệ nhân tạo trong quản lý nhà nước trên địa bàn thành phố Hà Nội (sau đây gọi tắt là Quy định) này quy định về công tác bảo đảm an toàn, quản lý rủi ro và ứng dụng trí tuệ nhân tạo phục vụ hoạt động quản lý nhà nước, cung cấp dịch vụ công và hoạt động công vụ trên địa bàn thành phố Hà Nội.

2. Hoạt động ứng dụng trí tuệ nhân tạo liên quan đến bí mật nhà nước thực hiện theo quy định của pháp luật về bảo vệ bí mật nhà nước, an ninh mạng và cơ yếu.

Điều 2. Đối tượng áp dụng

1. Quy định này áp dụng đối với các cơ quan, đơn vị thuộc UBND thành phố Hà Nội; các tổ chức, doanh nghiệp và cá nhân trực tiếp tham gia hoặc có liên quan đến hoạt động trí tuệ nhân tạo đối với các cơ quan, đơn vị thuộc UBND thành phố Hà Nội.

2. Trường hợp hệ thống trí tuệ nhân tạo dùng chung có liên quan đến cơ quan, tổ chức ngoài phạm vi quản lý của UBND Thành phố, khuyến khích áp dụng thực hiện theo Quy định này hoặc quy chế phối hợp hoặc văn bản của cơ quan có thẩm quyền.

Điều 3. Giải thích từ ngữ

1. Các thuật ngữ sử dụng trong Quy định này được hiểu theo Luật Trí tuệ nhân tạo, Luật An ninh mạng, Luật Dữ liệu, 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.

2. Trong Quy định này, các từ ngữ dưới đây được hiểu như sau:

a) Chủ quản hệ thống thông tin được quy định tại khoản 8 Điều 2 Luật An ninh mạng.

b) Chủ quản hệ thống trí tuệ nhân tạo là cơ quan, tổ chức, cá nhân có thẩm quyền quyết định đầu tư, quản lý hệ thống trí tuệ nhân tạo.

c) Đơn vị vận hành hệ thống trí tuệ nhân tạo là cơ quan, tổ chức được chủ quản hệ thống trí tuệ nhân tạo giao nhiệm vụ vận hành hệ thống trí tuệ nhân tạo. Trong trường hợp thuê dịch vụ công nghệ thông tin, đơn vị cung cấp dịch vụ theo quy định của pháp luật là đơn vị vận hành. Trường hợp hết thời hạn cung cấp dịch vụ, nếu hệ thống trí tuệ nhân tạo được thiết lập qua hình thức thuê dịch vụ vẫn tiếp tục duy trì hoạt động, đơn vị vận hành được xác định là đơn vị chủ trì thuê dịch vụ.

d) Trí tuệ nhân tạo mã nguồn mở (AI mã nguồn mở) là Hệ thống trí tuệ nhân tạo sử dụng mô hình, thành phần phần mềm, thư viện, bộ công cụ, mã nguồn hoặc thành phần kỹ thuật phục vụ trí tuệ nhân tạo được công bố để tổ chức, cá nhân có thể truy cập, sử dụng, sao chép, sửa đổi hoặc phân phối lại theo điều kiện của giấy phép nguồn mở hợp pháp.

đ) Tác nhân trí tuệ nhân tạo (Tác nhân AI hoặc AI Agent) là hệ thống trí tuệ nhân tạo hoặc chức năng của hệ thống trí tuệ nhân tạo có khả năng, trên cơ sở mục tiêu, nhiệm vụ hoặc ràng buộc do con người thiết lập, tự động thực hiện một hoặc nhiều bước xử lý, gọi công cụ, truy cập giao diện lập trình ứng dụng, tương tác với hệ thống thông tin, cơ sở dữ liệu, nền tảng số hoặc dịch vụ số khác trong phạm vi được phép.

e) Trí tuệ nhân tạo tạo sinh (Generative Artificial Intelligence hay còn gọi là Gen AI) là một lĩnh vực thuộc AI tập trung vào việc tạo ra dữ liệu mới, có thể là văn bản, hình ảnh, âm thanh, hoặc video, mã nguồn lập trình dựa trên dữ liệu đầu vào đã được huấn luyện trước đó.

Điều 4. Các hành vi bị nghiêm cấm

1. Các hành vi bị nghiêm cấm trong hoạt động trí tuệ nhân tạo thực hiện theo quy định của pháp luật về trí tuệ nhân tạo, an ninh mạng, dữ liệu và pháp luật có liên quan.

2. Trong hoạt động công vụ và trong quá trình cung cấp, triển khai, vận hành hệ thống trí tuệ nhân tạo cho cơ quan nhà nước thuộc Thành phố, các cơ quan, tổ chức, cá nhân không được thực hiện các hành vi sau đây:

a) Sử dụng hệ thống trí tuệ nhân tạo vượt quá mục đích, phạm vi, thẩm quyền được giao hoặc khi chưa có căn cứ pháp lý, quy trình nghiệp vụ được phê duyệt.

b) Đưa dữ liệu chưa được Chủ quản hệ thống trí tuệ nhân tạo phê duyệt để sử dụng trong hệ thống trí tuệ nhân tạo.

c) Can thiệp trái phép, xóa, sửa, làm sai lệch hoặc cố ý bỏ qua nhật ký hệ thống, cơ chế cảnh báo, cơ chế kiểm soát truy cập, cơ chế bảo mật hoặc luồng phê duyệt đã được thiết lập.

d) Để hệ thống trí tuệ nhân tạo tự động ký, ban hành, công bố, phát hành văn bản, quyết định hành chính, trả lời chính thức hoặc giao dịch có giá trị pháp lý.

đ) Che giấu việc sử dụng hệ thống trí tuệ nhân tạo trong hồ sơ hành chính, dịch vụ công, văn bản trả lời tổ chức, cá nhân hoặc giao diện tương tác với cá

nhân, doanh nghiệp trong trường hợp pháp luật yêu cầu thông báo, công khai, gắn nhãn hoặc giải trình.

e) Thiếu trách nhiệm trong việc kiểm tra, xác thực tính hợp pháp của kết quả đầu ra hệ thống trí tuệ nhân tạo; sử dụng kết quả đầu ra làm căn cứ tham mưu, xử lý hồ sơ, ban hành văn bản, quyết định hành chính hoặc thực hiện giao dịch công vụ gây ảnh hưởng, xâm phạm quyền và lợi ích hợp pháp của tổ chức, cá nhân, vi phạm quy định pháp luật.

g) Cản trở việc kiểm tra, giám sát, truy vết, xử lý sự cố; không cung cấp hoặc cung cấp không đầy đủ thông tin theo yêu cầu hợp pháp của cơ quan có thẩm quyền.

Điều 5. Nguyên tắc áp dụng

1. Việc áp dụng Quy định phải bảo đảm tính thống nhất, đồng bộ với các quy định pháp luật hiện hành về trí tuệ nhân tạo, an ninh mạng, dữ liệu, khung đạo đức trí tuệ nhân tạo quốc gia và các văn bản hướng dẫn chuyên ngành của Bộ Khoa học và Công nghệ.

2. Việc quản lý và sử dụng hệ thống trí tuệ nhân tạo trong hoạt động công vụ phải bảo đảm tuân thủ các nguyên tắc tại Điều 4 Luật Trí tuệ nhân tạo.

3. Tổ chức, doanh nghiệp, cá nhân cung cấp, phát triển, triển khai, tích hợp, vận hành, bảo trì, kiểm thử, đánh giá, xử lý dữ liệu nhằm cung cấp dịch vụ liên quan đến hệ thống trí tuệ nhân tạo cho cơ quan, đơn vị thuộc UBND thành phố Hà Nội có trách nhiệm nghiên cứu, quán triệt và tuân thủ các nguyên tắc an toàn, chuẩn mực quản lý và ứng dụng trí tuệ nhân tạo quy định tại Quy định này.

4. Việc quản lý và sử dụng hệ thống trí tuệ nhân tạo trong hoạt động công vụ phải bảo đảm các yêu cầu sau đây:

- a) Lấy con người làm trung tâm;
- b) Bảo đảm sự giám sát, kiểm soát và khả năng can thiệp có ý nghĩa của con người;
- c) Công khai, minh bạch, có trách nhiệm giải trình phù hợp với tính chất, phạm vi sử dụng và mức độ rủi ro của hệ thống;
- d) Quản lý rủi ro tương xứng với mức độ tác động của hệ thống đối với quyền, lợi ích hợp pháp của tổ chức, cá nhân và lợi ích công cộng;
- đ) Bảo đảm an ninh mạng, bảo vệ dữ liệu cá nhân, bí mật nhà nước, bí mật công tác, bí mật công nghệ và quyền sở hữu trí tuệ theo quy định của pháp luật;
- e) Phù hợp với Khung đạo đức trí tuệ nhân tạo quốc gia, các tiêu chuẩn kỹ thuật do cơ quan có thẩm quyền ban hành;
- g) Thúc đẩy phát triển, ứng dụng trí tuệ nhân tạo theo hướng an toàn, xanh, bao trùm, bền vững, sử dụng hiệu quả năng lượng, tiết kiệm tài nguyên và hạn chế tác động tiêu cực đến môi trường.

Chương II

BẢO ĐẢM AN TOÀN VÀ QUẢN LÝ RỦI RO ĐỐI VỚI HỆ THỐNG TRÍ TUỆ NHÂN TẠO

Điều 6 Phân loại rủi ro

1. Việc phân loại rủi ro, phân loại lại, thông báo kết quả phân loại và thực hiện nghĩa vụ tương ứng với từng mức rủi ro của hệ thống trí tuệ nhân tạo thực hiện theo quy định tại Điều 9, 10 Luật Trí tuệ nhân tạo, Chương II Nghị định số 142/2026/NĐ-CP ngày 30/4/2026 của Chính phủ và văn bản pháp luật có liên quan.

2. Cơ quan chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm sử dụng kết quả phân loại rủi ro làm căn cứ tổ chức quản lý, kiểm soát an toàn, bảo vệ dữ liệu, giám sát của con người, kiểm tra điều kiện vận hành và xử lý sự cố theo quy định của pháp luật.

3. Trường hợp sửa đổi, tích hợp, thay đổi chức năng, mục đích sử dụng, dữ liệu, mô hình, cấu hình, quyền truy cập hoặc môi trường vận hành có khả năng làm phát sinh rủi ro mới hoặc rủi ro cao hơn, cơ quan chủ quản hệ thống trí tuệ nhân tạo yêu cầu nhà cung cấp, bên triển khai hoặc đơn vị có liên quan phối hợp rà soát, phân loại lại theo quy định.

4. Mỗi hệ thống trí tuệ nhân tạo phải có hồ sơ quản lý bao gồm những thông tin tối thiểu như sau:

- a) Tên hệ thống, cơ quan chủ quản hệ thống trí tuệ nhân tạo;
- b) Mục đích sử dụng, phạm vi sử dụng;
- c) Loại dữ liệu đầu vào, nguồn dữ liệu;
- d) Mức độ rủi ro;
- đ) Thông tin về nhà cung cấp, mô hình, nền tảng sử dụng, kết quả thử nghiệm, tài liệu vận hành, API;
- e) Đánh giá hiệu năng, độ chính xác, độ tin cậy và chất lượng đầu ra;
- g) Cơ chế ghi nhật ký, truy vết, sao lưu, phục hồi;
- h) An ninh mạng và quy trình xử lý sự cố;
- i) Thời điểm đánh giá, rà soát lại;
- k) Các hồ sơ, giấy tờ khác được quy định tại Nghị định số 142/2026/NĐ-CP.

Điều 7. Kiến trúc kỹ thuật, tích hợp và khả năng kiểm soát hệ thống

1. Việc thiết kế, xây dựng, tích hợp hệ thống trí tuệ nhân tạo phải bảo đảm khả năng kết nối, liên thông, chia sẻ, sử dụng lại dữ liệu; khả năng giám sát, ghi nhật ký, truy vết, giải trình, kiểm soát thay đổi; khả năng tạm dừng, khôi phục, thay thế, chuyển đổi, chấm dứt sử dụng hoặc chuyển giao hệ thống, dữ liệu, cấu hình và tài liệu kỹ thuật khi cần thiết.

2. Cơ quan chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm bảo đảm việc tích hợp hệ thống trí tuệ nhân tạo với hệ thống thông tin chuyên ngành, hệ thống dùng chung của Thành phố trên cơ sở Khung kiến trúc số Thành phố, bảo đảm an ninh mạng và không làm phát sinh yêu cầu cung cấp lại dữ liệu trái quy định của pháp luật.

3. Khi tích hợp hệ thống trí tuệ nhân tạo vào quy trình nghiệp vụ đang vận hành, cơ quan chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm xác định tác động của việc tích hợp đối với quy trình xử lý công việc, thời hạn giải quyết công việc, trách nhiệm của cán bộ, công chức, viên chức, người lao động; quyền và lợi ích hợp pháp của tổ chức, cá nhân và khả năng chuyển sang quy trình xử lý thủ công khi hệ thống gặp sự cố.

4. Cơ quan chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm yêu cầu nhà cung cấp, nhà phát triển, đơn vị có liên quan bàn giao tài liệu kỹ thuật, tài liệu hướng dẫn sử dụng, quản lý, vận hành và các tài liệu khác liên quan đến hệ thống trí tuệ nhân tạo.

Điều 8. Sử dụng và kiểm soát rủi ro đối với trí tuệ nhân tạo mã nguồn mở

1. Thành phố khuyến khích việc nghiên cứu, thử nghiệm, phát triển, tích hợp và sử dụng AI mã nguồn mở trong hệ thống trí tuệ nhân tạo phục vụ quản lý nhà nước, cung cấp dịch vụ công và hoạt động công vụ phù hợp với quy định của pháp luật, yêu cầu bảo đảm an ninh, bảo vệ dữ liệu và quyền sở hữu trí tuệ.

2. Cơ quan chủ quản hệ thống trí tuệ nhân tạo xem xét sử dụng trí tuệ nhân tạo mã nguồn mở khi đáp ứng yêu cầu về mục tiêu sử dụng, hiệu quả kỹ thuật, an ninh mạng, bảo vệ dữ liệu, sở hữu trí tuệ, khả năng duy trì, vận hành, thay thế và các quy định của pháp luật có liên quan.

3. Việc lựa chọn, tích hợp, điều chỉnh, huấn luyện lại hoặc phân phối lại AI mã nguồn mở phải được rà soát về nguồn gốc, tính pháp lý, thành phần phụ thuộc, lỗ hổng bảo mật, dữ liệu huấn luyện, khả năng duy trì, khả năng thay thế, khả năng chuyển đổi và rủi ro vận hành trong hoạt động.

4. Các hệ thống trí tuệ nhân tạo mã nguồn mở khi triển khai phục vụ quản lý nhà nước và trong cung cấp dịch vụ công phải đánh giá tác động trước khi sử dụng. Các hệ thống trí tuệ nhân tạo xử lý dữ liệu quan trọng khuyến khích chạy thử nghiệm trong môi trường có kiểm soát (Sandbox) và được cơ quan có thẩm quyền đánh giá an ninh mạng trước khi vận hành chính thức.

Điều 9. Bảo đảm an ninh mạng, bảo vệ dữ liệu

1. Cơ quan chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm bảo đảm an ninh mạng, bảo vệ dữ liệu cá nhân, dữ liệu quan trọng, dữ liệu chuyên ngành, bí mật công tác và thông tin thuộc phạm vi hạn chế tiếp cận trong quá trình quản lý, triển khai, vận hành, cập nhật và chấm dứt sử dụng hệ thống trí tuệ nhân tạo theo quy định của pháp luật.

2. Hệ thống trí tuệ nhân tạo có kết nối, tích hợp hoặc xử lý dữ liệu từ hệ thống thông tin, cơ sở dữ liệu thuộc phạm vi quản lý của Thành phố phải được

kiểm soát truy cập, xác thực, phân quyền, sao lưu, phục hồi, giám sát, kiểm soát kết nối và tự động ghi nhật ký, mã hóa theo quy định của pháp luật, tiêu chuẩn, quy chuẩn kỹ thuật và hướng dẫn của cơ quan có thẩm quyền.

Nhật ký hệ thống phải lưu trữ đầy đủ các thông tin về thời gian truy cập, thông tin tài khoản truy cập, dữ liệu đầu vào, kết quả đầu ra và các tác vụ tự động được thực thi. Thời gian lưu trữ tối thiểu là 24 tháng kể từ thời điểm phát sinh dữ liệu. Định dạng dữ liệu nhật ký phải bảo đảm tính toàn vẹn, không thể chỉnh sửa bảo đảm phục vụ công tác kiểm tra, thanh tra, kiểm toán và xác minh trách nhiệm khi xảy ra sự cố.

3. Việc cấp quyền truy cập dữ liệu, hệ thống thông tin, giao diện lập trình ứng dụng (API), công cụ xử lý, tài khoản quản trị hoặc môi trường vận hành cho hệ thống trí tuệ nhân tạo, nhà cung cấp, bên triển khai, cá nhân hoặc đơn vị có liên quan phải bảo đảm đúng mục đích, đúng phạm vi, đúng thời hạn, quyền hạn, có khả năng thu hồi và truy vết.

4. Đối với hệ thống trí tuệ nhân tạo được triển khai trên hạ tầng, nền tảng, dịch vụ điện toán đám mây, dịch vụ phần mềm hoặc môi trường do bên thứ ba cung cấp, cơ quan chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm kiểm tra điều kiện pháp lý, điều kiện bảo mật, vị trí lưu trữ dữ liệu, cơ chế truy cập, chuyển giao, xóa, sao lưu, phục hồi dữ liệu và trách nhiệm của đơn vị cung cấp trước khi sử dụng.

5. Khi hệ thống trí tuệ nhân tạo có dấu hiệu bị tấn công, khai thác trái phép, rò rỉ dữ liệu, mất dữ liệu, mất quyền kiểm soát dữ liệu và mô hình, chiếm quyền điều khiển hoặc phát sinh hành vi bất thường, cơ quan chủ quản hệ thống trí tuệ nhân tạo kịp thời phối hợp với cơ quan có thẩm quyền để xử lý theo quy định của pháp luật.

Điều 10. Giám sát và xử lý sự cố

1. Cơ quan chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm thiết lập cơ chế giám sát, can thiệp và phân quyền phù hợp với mục đích sử dụng, mức độ rủi ro, mức độ tự động hóa, phạm vi tác động của hệ thống phù hợp quy định của pháp luật chuyên ngành.

2. Khi phát hiện hệ thống trí tuệ nhân tạo có nguy cơ gây thiệt hại đến quyền, lợi ích hợp pháp của tổ chức, cá nhân; vi phạm quy định pháp luật về trí tuệ nhân tạo, an ninh mạng, dữ liệu, cơ quan chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm kịp thời tạm dừng, hạn chế, cô lập, khắc phục hoặc chấm dứt sử dụng chức năng, mô hình, dữ liệu, kết nối hệ thống có liên quan.

3. Khi xảy ra sự cố liên quan đến hệ thống trí tuệ nhân tạo, cơ quan chủ quản hệ thống trí tuệ nhân tạo thực hiện việc ghi nhận, xác minh, phân loại sự cố; áp dụng biện pháp ngăn chặn bảo đảm hạn chế thiệt hại; bảo toàn nhật ký, dữ liệu, tài liệu; thông báo, báo cáo, phối hợp xử lý, khắc phục và phòng ngừa tái diễn theo quy định của pháp luật.

4. Trách nhiệm giải trình khi xảy ra rủi ro, sự cố hoặc kết quả đầu ra của hệ thống trí tuệ nhân tạo thiếu khách quan, sai lệch, xâm phạm quyền và lợi ích hợp pháp của tổ chức, cá nhân được phân định cụ thể như sau:

a) Chủ quản hệ thống trí tuệ nhân tạo chịu trách nhiệm trước UBND Thành phố và pháp luật đối với quyết định hành chính, hoạt động hành chính được ban hành dựa trên việc sử dụng kết quả đầu ra của hệ thống;

b) Đơn vị vận hành hệ thống trí tuệ nhân tạo chịu trách nhiệm trong phạm vi quản lý tài khoản truy cập, tuân thủ quy trình nghiệp vụ nhập dữ liệu, bảo mật hệ thống và thực hiện các biện pháp cô lập, ngăn chặn sự cố kịp thời theo quy trình và hướng dẫn kỹ thuật;

c) Tổ chức, doanh nghiệp cung cấp giải pháp công nghệ chịu trách nhiệm về tính chính xác kỹ thuật của thuật toán theo cam kết; chịu trách nhiệm bồi thường thiệt hại và liên đới chịu trách nhiệm trước pháp luật nếu rủi ro, sự cố xảy ra do lỗi mã nguồn, lỗ hổng bảo mật hoặc do sự sai lệch của mô hình công nghệ đã được xác định trong hợp đồng dịch vụ.

Chương III

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG QUẢN LÝ NHÀ NƯỚC

Điều 11. Nguyên tắc ứng dụng trí tuệ nhân tạo trong quản lý nhà nước

1. Việc ứng dụng hệ thống trí tuệ nhân tạo trong quản lý nhà nước, cung cấp dịch vụ công và hoạt động công vụ phải phục vụ mục tiêu nâng cao hiệu quả quản lý, chất lượng phục vụ người dân, doanh nghiệp, hiệu quả sử dụng nguồn lực công và năng lực chỉ đạo, điều hành của cơ quan nhà nước.

2. Hệ thống trí tuệ nhân tạo chỉ được ứng dụng khi xác định rõ mục tiêu, căn cứ pháp lý, phạm vi sử dụng, dữ liệu sử dụng, cơ quan chủ quản, đơn vị vận hành, người chịu trách nhiệm và biện pháp kiểm soát rủi ro tương ứng.

3. Việc ứng dụng hệ thống trí tuệ nhân tạo không làm thay đổi thẩm quyền, trách nhiệm của cơ quan trong hoạt động quản lý nhà nước, giải quyết thủ tục hành chính, cung cấp dịch vụ công, ban hành văn bản, quyết định hành chính và xử lý công việc. Kết quả đầu ra của hệ thống trí tuệ nhân tạo không được sử dụng là căn cứ pháp lý độc lập để ban hành quyết định hành chính, xử lý hồ sơ, kết luận thanh tra, kiểm tra, xử phạt vi phạm hành chính hoặc hạn chế quyền, lợi ích hợp pháp của tổ chức, cá nhân.

4. Chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm ưu tiên ứng dụng hệ thống có khả năng hỗ trợ tổng hợp, phân tích, khuyến nghị, dự báo, tối ưu hóa quy trình trong xử lý công việc.

Điều 12. Ứng dụng trí tuệ nhân tạo trong dịch vụ công và quy trình hành chính

1. Hệ thống trí tuệ nhân tạo được ứng dụng để hỗ trợ cải cách hành chính, giải quyết thủ tục hành chính, cung cấp dịch vụ công, tiếp nhận, phân loại, tổng

hợp, phân tích, trả lời hướng dẫn, cảnh báo rủi ro, phát hiện bất thường và nâng cao chất lượng phục vụ người dân, doanh nghiệp.

2. Hệ thống trí tuệ nhân tạo không được tự động từ chối, đình chỉ, thu hồi, hạn chế quyền tiếp cận dịch vụ công, hỗ trợ từ ngân sách nhà nước hoặc nguồn lực công.

3. Khi hệ thống trí tuệ nhân tạo được sử dụng để phân loại hồ sơ, xếp hạng rủi ro, chấm điểm, lập danh sách ưu tiên hoặc danh sách cần kiểm tra, cơ quan chủ quản hệ thống trí tuệ nhân tạo phải bảo đảm khả năng kiểm tra, truy vết, xem xét lại bởi con người theo quy định của pháp luật.

Chủ quản hệ thống trí tuệ nhân tạo có trách nhiệm tổ chức rà soát, đánh giá định kỳ tối thiểu 06 tháng một lần về độ chính xác, tính tin cậy và nguy cơ thiếu khách quan, phân biệt đối xử của thuật toán. Trên cơ sở phân tích đánh giá, Chủ quản hệ thống trí tuệ nhân tạo tạm dừng hoạt động của hệ thống để khắc phục nếu cần thiết.

4. Tổ chức, cá nhân chịu tác động bởi việc sử dụng hệ thống trí tuệ nhân tạo trong cung cấp dịch vụ công và quy trình hành chính được thực hiện quyền phản ánh, kiến nghị, khiếu nại, tố cáo hoặc yêu cầu xem xét lại theo quy định của pháp luật.

5. Việc thông báo, minh bạch, gán nhãn kết quả đầu ra cho người dùng thực hiện theo quy định tại Chương III Nghị định số 142/2026/NĐ-CP.

Điều 13. Ứng dụng trí tuệ nhân tạo tạo sinh (Gen AI) và tác nhân trí tuệ nhân tạo (AI Agent) trong hoạt động công vụ

1. Việc sử dụng trí tuệ nhân tạo tạo sinh và tác nhân trí tuệ nhân tạo trong hoạt động công vụ phải đúng mục đích, thẩm quyền, phạm vi nhiệm vụ; có kiểm tra, giám sát của con người; không sử dụng kết quả đầu ra của hệ thống làm căn cứ duy nhất để ban hành quyết định, kết luận hoặc thực hiện hành vi làm phát sinh, thay đổi, chấm dứt quyền, nghĩa vụ của tổ chức, cá nhân.

2. Việc cấp quyền cho tác nhân trí tuệ nhân tạo phải tuân thủ nguyên tắc phân quyền tối thiểu, đúng mục đích, phạm vi, đúng thời hạn, có khả năng thu hồi, có khả năng truy vết và có người chịu trách nhiệm phê duyệt.

3. Cơ quan, đơn vị thuộc Thành phố có trách nhiệm ban hành quy trình nội bộ về sử dụng trí tuệ nhân tạo tạo sinh, tác nhân trí tuệ nhân tạo trong hoạt động công vụ, trong đó xác định tối thiểu:

a) Nhóm công việc được sử dụng, giới hạn thực thi của trí tuệ nhân tạo và quyền phê duyệt;

b) Loại dữ liệu, tài liệu, thông tin được phép hoặc không được phép nhập vào hệ thống;

c) Giới hạn phạm vi của tác nhân trí tuệ nhân tạo bảo đảm không tự động tạo, sửa đổi, xóa, chuyển giao, chia sẻ hoặc công khai dữ liệu thuộc phạm vi quản lý của cơ quan nhà nước; tự mở rộng phạm vi nhiệm vụ, tự cấp thêm quyền, tự

kết nối hệ thống mới, tự thay đổi giới hạn hành động hoặc tự vô hiệu hóa cơ chế giám sát; tự thực hiện các hoạt động vượt quá phạm vi gây ảnh hưởng đến hệ thống thông tin;

d) Trách nhiệm kiểm chứng, chỉnh sửa, phê duyệt, lưu vết việc sử dụng kết quả đầu ra;

đ) Yêu cầu thông báo, gắn nhãn hoặc ghi nhận việc có sử dụng trí tuệ nhân tạo khi pháp luật có quy định;

e) Biện pháp bảo mật, bảo vệ dữ liệu, bảo vệ bí mật nhà nước, bảo vệ quyền sở hữu trí tuệ và xử lý sự cố;

g) Đối với các tác nhân trí tuệ nhân tạo tương tác trực tiếp với tổ chức, cá nhân, phải thiết lập cơ chế tự động chuyển hướng hỗ trợ sang cán bộ chuyên trách khi hệ thống không giải quyết đúng yêu cầu sau 02 lần tương tác, hoặc ngay khi có yêu cầu kết nối trực tiếp từ người dùng.

4. Khi tác nhân trí tuệ nhân tạo có dấu hiệu thực hiện hành động ngoài phạm vi nhiệm vụ, vượt giới hạn quyền, truy cập dữ liệu bất thường, tạo chuỗi hành động ngoài dự kiến, tiêu thụ tài nguyên bất thường, tạo đầu ra có nguy cơ gây thiệt hại hoặc có dấu hiệu bị tấn công, chiếm quyền điều khiển, cơ quan chủ quản hệ thống trí tuệ nhân tạo phải kịp thời tạm dừng, cô lập, thu hồi quyền truy cập, bảo toàn nhật ký và xử lý sự cố theo quy định tại Điều 10 Quy định này.

Điều 14. Hạ tầng dữ liệu, nền tảng dùng chung và bài toán ưu tiên của Thành phố

1. Thành phố khuyến khích ưu tiên sử dụng hạ tầng số, trung tâm dữ liệu, điện toán đám mây, nền tảng trí tuệ nhân tạo và nền tảng dữ liệu do doanh nghiệp Việt Nam cung cấp nếu đáp ứng yêu cầu pháp luật về trí tuệ nhân tạo, dữ liệu, an ninh mạng, cơ yếu và Khung kiến trúc số Thành phố.

2. Thành phố khuyến khích ứng dụng hệ thống trí tuệ nhân tạo trên cơ sở hạ tầng dữ liệu, nền tảng số dùng chung, Khung kiến trúc số Thành phố, hệ thống thông tin chuyên ngành và cơ sở dữ liệu thuộc phạm vi quản lý của Thành phố theo quy định của pháp luật.

3. Danh mục bài toán trí tuệ nhân tạo ưu tiên phải căn cứ vào nhu cầu quản lý, khả năng dữ liệu, điều kiện kỹ thuật, khả năng kiểm soát rủi ro, hiệu quả dự kiến, khả năng nhân rộng và khả năng bảo đảm nguồn lực.

4. Cơ quan, đơn vị chỉ đầu tư, triển khai hệ thống trí tuệ nhân tạo dùng riêng khi có nhu cầu đặc thù, không trùng lặp với hệ thống trí tuệ nhân tạo dùng chung của Thành phố. Chịu trách nhiệm trước pháp luật về việc đáp ứng yêu cầu về Khung kiến trúc số Thành phố, dữ liệu, an ninh mạng, bảo vệ bí mật nhà nước và khả năng duy trì vận hành, dùng sử dụng và tích hợp khi có yêu cầu.

5. Việc xây dựng, khai thác, sử dụng hạ tầng tính toán, nền tảng dữ liệu, môi trường kiểm thử, kho công cụ, bộ dữ liệu mẫu, tài liệu hướng dẫn và nền tảng dùng chung phục vụ trí tuệ nhân tạo của Thành phố phải tuân thủ quy định của

pháp luật về đầu tư công, quản lý tài sản công, dữ liệu, an ninh mạng và pháp luật có liên quan.

6. Thành phố khuyến khích hình thành mạng lưới chuyên gia, nhóm nghiên cứu, tư vấn, kiểm thử và đánh giá; khuyến khích hợp tác công tư và có kiểm soát đối với hệ thống trí tuệ nhân tạo phục vụ cơ quan nhà nước, cung cấp dịch vụ công, điều hành đô thị và đổi mới sáng tạo trong khu vực công.

Chương IV TỔ CHỨC THỰC HIỆN

Điều 15. Trách nhiệm của Sở Khoa học và Công nghệ

1. Là cơ quan đầu mối tham mưu UBND Thành phố tổ chức triển khai, hướng dẫn, theo dõi, đôn đốc, tổng hợp, đánh giá việc thực hiện Quy định này trong phạm vi quản lý nhà nước của Thành phố.

2. Chủ trì, phối hợp các cơ quan, đơn vị liên quan tổ chức kiểm tra hằng năm hoặc kiểm tra đột xuất công tác triển khai, tuân thủ Quy định tại các cơ quan, đơn vị thuộc Thành phố. Ưu tiên thực hiện bằng hình thức hậu kiểm.

3. Phối hợp, hỗ trợ các cơ quan, đơn vị ban hành quy trình nội bộ quy định tại khoản 3 Điều 13 Quy định này.

4. Phối hợp tổ chức đào tạo, bồi dưỡng, tập huấn, phổ biến kiến thức về quản trị, pháp lý, kỹ thuật, an ninh và rủi ro trí tuệ nhân tạo cho cán bộ, công chức, viên chức, người lao động trong cơ quan, tổ chức, đơn vị thuộc Thành phố.

5. Tổng hợp khó khăn, vướng mắc, kiến nghị của cơ quan, tổ chức, đơn vị trong quá trình thực hiện Quy định này; tham mưu, báo cáo UBND Thành phố xử lý theo thẩm quyền hoặc kiến nghị cơ quan nhà nước có thẩm quyền ở Trung ương xem xét, hướng dẫn, sửa đổi, bổ sung quy định có liên quan.

6. Tổng hợp đề xuất của các cơ quan, đơn vị gửi Sở Tài chính tham mưu UBND Thành phố bảo đảm kinh phí triển khai phù hợp quy định.

Điều 16. Trách nhiệm của các Sở, ban, ngành Thành phố; UBND xã, phường và tổ chức, doanh nghiệp, cá nhân có liên quan đến hệ thống trí tuệ nhân tạo

1. Công an thành phố Hà Nội

Chủ trì, phối hợp với Sở Khoa học và Công nghệ và các đơn vị có liên quan hướng dẫn, kiểm tra, cảnh báo, ứng cứu, xử lý sự cố an ninh mạng, bảo vệ bí mật nhà nước, bảo vệ dữ liệu đối với hệ thống trí tuệ nhân tạo theo quy định của pháp luật.

2. Sở Nội vụ

Chủ trì, phối hợp với Sở Khoa học và Công nghệ, các đơn vị có liên quan tham mưu, tổ chức hoặc lồng ghép nội dung về quản trị, sử dụng có trách nhiệm, đạo đức, pháp lý, an toàn và rủi ro trí tuệ nhân tạo vào chương trình đào tạo, bồi

dưỡng cán bộ, công chức, viên chức, người lao động thuộc Thành phố phù hợp với quy định của pháp luật

3. Sở Tài chính

Chủ trì, phối hợp với Sở Khoa học và Công nghệ, các đơn vị có liên quan tham mưu việc bố trí kinh phí, kinh phí do ngân sách Nhà nước bảo đảm theo phân cấp khi thực hiện nhiệm vụ quản lý, phát triển, triển khai, thử nghiệm, đào tạo, kiểm thử, đánh giá, giám sát và hoàn thiện hệ sinh thái trí tuệ nhân tạo của Thành phố theo quy định của pháp luật về ngân sách nhà nước và pháp luật có liên quan.

4. Các Sở, ban, ngành Thành phố, UBND xã, phường và tổ chức, doanh nghiệp, cá nhân có liên quan

a) Giám đốc, Thủ trưởng, cá nhân tại các cơ quan, đơn vị chịu trách nhiệm trước Chủ tịch UBND Thành phố và pháp luật nếu để xảy ra việc lộ bí mật nhà nước, mất an ninh mạng, gây hậu quả, thiệt hại nghiêm trọng tại đơn vị thuộc phạm vi quản lý.

b) Tổ chức quán triệt, thực hiện Quy định này trong phạm vi chức năng, nhiệm vụ, quyền hạn được giao.

c) Ban hành quy trình nội bộ về quản lý, sử dụng hệ thống trí tuệ nhân tạo phù hợp với tính chất, phạm vi hoạt động của cơ quan, tổ chức, đơn vị.

d) Bố trí nhân lực, kinh phí, điều kiện kỹ thuật, điều kiện bảo mật, điều kiện dữ liệu và thời gian cần thiết để thực hiện việc quản lý, giám sát, kiểm tra, đánh giá, hậu kiểm, xử lý sự cố và cập nhật hồ sơ quản lý hệ thống trí tuệ nhân tạo thuộc phạm vi quản lý theo quy định của pháp luật.

đ) Bảo đảm cán bộ, công chức, viên chức, người lao động được giao sử dụng, vận hành, giám sát hệ thống trí tuệ nhân tạo được hướng dẫn, tập huấn về chức năng, giới hạn, rủi ro, trách nhiệm và quy trình sử dụng hệ thống.

e) Kịp thời tạm dừng, hạn chế, cô lập, khắc phục, báo cáo và phối hợp xử lý theo quy định của pháp luật và Quy định này khi phát hiện hệ thống trí tuệ nhân tạo có rủi ro, sự cố, lỗi nghiêm trọng, có dấu hiệu vi phạm pháp luật, dấu hiệu mất an ninh mạng, rò rỉ dữ liệu, tác động bất lợi đến tổ chức, cá nhân hoặc nguy cơ ảnh hưởng đến hoạt động bình thường của cơ quan, tổ chức, đơn vị.

g) Có trách nhiệm phối hợp với Sở Khoa học và Công nghệ hướng dẫn cơ quan, tổ chức, đơn vị trong ngành, lĩnh vực quản lý thực hiện yêu cầu quản trị trí tuệ nhân tạo, bảo đảm phù hợp với quy định pháp luật chuyên ngành, đặc thù dữ liệu, đặc thù nghiệp vụ và mức độ rủi ro của hệ thống.

h) Có trách nhiệm lồng ghép báo cáo kết quả thực hiện Quy định này vào các báo cáo chuyển đổi số hằng năm.

i) Việc hợp tác với cơ sở giáo dục, cơ sở nghiên cứu, doanh nghiệp, tổ chức khoa học và công nghệ, chuyên gia, tổ chức quốc tế, đối tác nước ngoài trong nghiên cứu, phát triển, thử nghiệm, chuyển giao, ứng dụng, đào tạo, kiểm thử,

đánh giá hệ thống trí tuệ nhân tạo của Thành phố phải tuân thủ quy định của pháp luật về hợp tác quốc tế, khoa học và công nghệ, chuyển đổi số, trí tuệ nhân tạo, dữ liệu, an ninh mạng, bảo vệ bí mật nhà nước, đầu tư, đầu thầu và pháp luật có liên quan.

k) Trong quá trình thực hiện nếu có các vấn đề nảy sinh, không phù hợp hoặc chưa được quy định rõ, các cơ quan, tổ chức, đơn vị gửi kiến nghị, đề xuất về Sở Khoa học và Công nghệ để tổng hợp báo cáo UBND Thành phố kịp thời xem xét điều chỉnh, bổ sung phù hợp với tình hình thực tiễn.

Điều 17. Điều khoản chuyển tiếp

1. Đối với các hệ thống trí tuệ nhân tạo đã vận hành chính thức, đã ký hợp đồng cung cấp dịch vụ công nghệ thông tin, đang triển khai thử nghiệm hoặc đang trong quy trình lựa chọn nhà thầu trước ngày Quy định này có hiệu lực thi hành, chủ quản hệ thống có trách nhiệm tổ chức rà soát, phân loại mức độ rủi ro và hoàn thiện hồ sơ quản lý hệ thống theo quy định.

2. Thời hạn hoàn thành việc rà soát, hiệu chỉnh mô hình trí tuệ nhân tạo và bổ sung các biện pháp kiểm soát an toàn được quy định cụ thể như sau:

a) Không quá 06 tháng kể từ ngày Quy định này có hiệu lực thi hành đối với các hệ thống trí tuệ nhân tạo tham gia hỗ trợ trực tiếp quy trình hành chính, phân loại hồ sơ hoặc cung cấp dịch vụ công trực tuyến cho người dân, doanh nghiệp;

b) Không quá 12 tháng kể từ ngày Quy định này có hiệu lực thi hành đối với các hệ thống trí tuệ nhân tạo phục vụ hoạt động quản trị nội bộ hoặc các hệ thống ứng dụng chuyên ngành dùng riêng khác.

3. Việc thực hiện các thủ tục thông báo, báo cáo và công khai thông tin liên quan đến hệ thống trí tuệ nhân tạo trong giai đoạn công thông tin điện tử một cửa về trí tuệ nhân tạo chưa được Bộ Khoa học và Công nghệ công bố vận hành chính thức được thực hiện theo quy định tại Điều 46 Nghị định số 142/2026/NĐ-CP.

4. Trong quá trình thực hiện các biện pháp chuyển tiếp quy định tại Điều này, chủ quản hệ thống và đơn vị vận hành có trách nhiệm bảo đảm duy trì hoạt động ổn định của hệ thống; không làm gián đoạn hoặc ảnh hưởng đến tiến độ giải quyết thủ tục hành chính và cung cấp dịch vụ công cho người dân, doanh nghiệp.

5. Trong quá trình thực hiện, nếu phát sinh khó khăn, vướng mắc hoặc nội dung cần sửa đổi, bổ sung, các cơ quan, đơn vị gửi ý kiến về Sở Khoa học và Công nghệ để tổng hợp báo cáo UBND Thành phố xem xét, quyết định./.