Ký bởi: CÔNG BÁO NƯỚC CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Cơ quan: VĂN PHÒNG CHÍNH PHỦ
Ngày ký: 2026-05-21 08:06:02 +07:00

THỦ TƯỚNG CHÍNH PHỦ

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

Số: 804/QĐ-TTg

Hà Nội, ngày 06 tháng 5 năm 2026

QUYẾT ĐỊNH
Ban hành Danh mục bộ dữ liệu phục vụ phát triển trí tuệ nhân tạo trong các lĩnh vực thiết yếu

THỦ TƯỚNG CHÍNH PHỦ

Căn cứ Luật Tổ chức Chính phủ ngày 18 tháng 02 năm 2025;

Căn cứ Luật Trí tuệ nhân tạo ngày 10 tháng 12 năm 2025;

Theo đề nghị của Bộ trưởng Bộ Khoa học và Công nghệ.

QUYẾT ĐỊNH:

Điều 1. Danh mục bộ dữ liệu phục vụ phát triển trí tuệ nhân tạo trong các lĩnh vực thiết yếu

1. Ban hành kèm theo Quyết định này "Danh mục bộ dữ liệu phục vụ phát triển trí tuệ nhân tạo trong các lĩnh vực thiết yếu", gồm:

a) Danh mục bộ dữ liệu phục vụ phát triển trí tuệ nhân tạo trong các lĩnh vực thiết yếu (Phụ lục I kèm theo Quyết định này).

b) Danh mục dữ liệu ưu tiên triển khai phục vụ phát triển trí tuệ nhân tạo (Phụ lục II kèm theo Quyết định này).

2. Việc tổ chức thu thập, tạo lập, chia sẻ và xử lý, sử dụng, khai thác các bộ dữ liệu thuộc Danh mục ban hành kèm theo Quyết định này phải tuân thủ các quy định của pháp luật về dữ liệu, bảo vệ dữ liệu cá nhân, bảo vệ bí mật nhà nước, sở hữu trí tuệ và các pháp luật khác có liên quan.

Điều 2. Hiệu lực thi hành

Quyết định này có hiệu lực thi hành từ ngày ký.

Điều 3. Trách nhiệm thi hành

1. Bộ Công an có trách nhiệm:

a) Làm đầu mối chủ trì, đôn đốc, giám sát và tổng hợp báo cáo tình hình tổ chức triển khai xây dựng các bộ dữ liệu thuộc Danh mục ban hành kèm theo Quyết định này tại các bộ, cơ quan ngang bộ và Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương.

b) Định kỳ hoặc theo nhu cầu thực tế tiến hành rà soát, tiếp nhận đề xuất, tổng hợp, trình Thủ tướng Chính phủ sửa đổi, bổ sung Danh mục ban hành kèm theo Quyết định này.

2. Các bộ, cơ quan ngang bộ, Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương được giao chủ trì tại Phụ lục I kèm theo Quyết định này có trách nhiệm:

a) Chủ trì, phối hợp với các cơ quan liên quan tổ chức xây dựng kế hoạch và triển khai thu thập, tạo lập, bảo đảm chất lượng dữ liệu theo quy định có liên quan, gán nhãn, chú thích, chuẩn hóa đối với các bộ dữ liệu thuộc phạm vi quản lý.

b) Phối hợp với Bộ Công an để thực hiện kết nối, đồng bộ các bộ dữ liệu này với cơ sở dữ liệu quốc gia về trí tuệ nhân tạo.

c) Đối với các nhóm dữ liệu tại Mục IX, XIII, XIV Phụ lục I có liên quan đến yếu tố quốc phòng, an ninh, phối hợp với Bộ Công an và Bộ Quốc phòng trong quá trình thực hiện quy định tại điểm a khoản này.

3. Bộ trưởng, Thủ trưởng cơ quan ngang bộ, Chủ tịch Ủy ban nhân dân các tỉnh, thành phố trực thuộc trung ương và các tổ chức, cá nhân có liên quan chịu trách nhiệm thi hành Quyết định này./.

KT. THỦ TƯỚNG
PHÓ THỦ TƯỚNG

Hồ Quốc Dũng

Phụ lục I

DANH MỤC BỘ DỮ LIỆU PHỤC VỤ PHÁT TRIỂN TRÍ TUỆ NHÂN TẠO TRONG CÁC LĨNH VỰC THIẾT YẾU

(Kèm theo Quyết định số 804/QĐ-TTg ngày 06 tháng 5 năm 2026 của Thủ tướng Chính phủ)

Các bộ dữ liệu quy định tại Phụ lục này được triển khai thực hiện trong giai đoạn 2026 – 2030; trong đó, ưu tiên tập trung cho các danh mục dữ liệu quy định tại Phụ lục II kèm theo Quyết định này.

STT	DANH MỤC BỘ DỮ LIỆU	MÔ TẢ KHÁI QUÁT	CƠ QUAN CHỦ TRÌ XÂY DỰNG
I	Bộ dữ liệu ngôn ngữ tiếng Việt và tiếng dân tộc thiểu số	Bao gồm dữ liệu văn bản, tiếng nói, chữ viết tay, ngữ liệu đa phương tiện, dữ liệu song ngữ, hội thoại, ngữ cảnh và ngôn ngữ ký hiệu phục vụ huấn luyện cốt lõi cho mô hình ngôn ngữ lớn.	Bộ Khoa học và Công nghệ, Bộ Văn hóa, Thể thao và Du lịch, Bộ Giáo dục và Đào tạo, các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
II	Bộ dữ liệu tri thức quốc gia	Bao gồm các dữ liệu tri thức tổng hợp và đã được chuẩn hóa của quốc gia, như sách (không bao gồm sách giáo khoa), từ điển, bách khoa, tài liệu tham khảo và các nguồn tri thức nền tảng phản ánh hiểu biết chung về ngôn ngữ, lịch sử, văn hóa và xã hội.	Bộ Khoa học và Công nghệ, Bộ Giáo dục và Đào tạo, các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
III	Bộ dữ liệu văn bản pháp luật và văn bản hành chính (không bao gồm văn bản mật)	Bao gồm các dữ liệu văn bản quy phạm pháp luật, án lệ, điều ước quốc tế mà nước Cộng hòa xã hội chủ nghĩa Việt Nam là thành viên và các văn bản hành chính (không bao gồm văn bản mật).	Bộ Tư pháp và các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
IV	Bộ dữ liệu khoa học, công nghệ và đổi mới sáng tạo	Bao gồm dữ liệu công bố khoa học, sáng chế, nhiệm vụ khoa học đã nghiệm thu, tiêu chuẩn, quy chuẩn kỹ thuật quốc gia và dữ liệu hệ sinh thái khởi nghiệp, nguồn lực khoa	Bộ Khoa học và Công nghệ và các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.

STT	DANH MỤC BỘ DỮ LIỆU	MÔ TẢ KHÁI QUÁT	CƠ QUAN CHỦ TRÌ XÂY DỰNG
		học công nghệ, dữ liệu về công nghệ, dữ liệu về doanh nghiệp khoa học công nghệ.	
V	Bộ dữ liệu dịch vụ công và thủ tục hành chính	Bao gồm danh mục, quy trình, hồ sơ giải quyết thủ tục hành chính, biểu mẫu, dữ liệu cung cấp dịch vụ công từ cơ quan nhà nước.	Bộ Công an, Bộ Tư pháp và các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
VI	Bộ dữ liệu y tế và chăm sóc sức khỏe	Bao gồm dữ liệu lâm sàng, cận lâm sàng, hình ảnh; Phòng bệnh và sức khỏe cộng đồng; Dược, thiết bị vật tư, an toàn thực phẩm; Tri thức chuyên ngành, hướng dẫn và nghiên cứu thử nghiệm; Quản trị ngành, nhân lực và cơ sở dữ liệu.	Bộ Y tế và các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
VII	Bộ dữ liệu giáo dục và đào tạo	Bao gồm sách giáo khoa, học liệu số, luận văn, ngân hàng câu hỏi, bài tập kiểm tra, đánh giá và quá trình học tập, kiểm tra, đánh giá của người học (đã khử định danh) và dữ liệu quản lý mạng lưới giáo dục.	Bộ Giáo dục và Đào tạo và các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
VIII	Bộ dữ liệu nông nghiệp	Bao gồm dữ liệu về thổ nhưỡng, giống cây trồng, vật nuôi, thủy sản, lâm nghiệp, thông tin phòng chống dịch bệnh và hệ thống truy xuất nguồn gốc chuỗi cung ứng nông sản.	Bộ Nông nghiệp và Môi trường, các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
IX	Bộ dữ liệu giao thông và đô thị	Bao gồm dữ liệu hạ tầng giao thông, điều hành đô thị, dữ liệu cảm biến vạn vật kết nối và bản đồ độ phân giải cao phục vụ phương tiện tự hành.	Bộ Xây dựng và các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.

STT	DANH MỤC BỘ DỮ LIỆU	MÔ TẢ KHÁI QUÁT	CƠ QUAN CHỦ TRÌ XÂY DỰNG
X	Bộ dữ liệu tài nguyên và môi trường	Bao gồm dữ liệu dự báo khí tượng, thủy văn, biến đổi khí hậu, các chỉ số quan trắc đa dạng sinh học và dữ liệu quản lý tài nguyên.	Bộ Nông nghiệp và Môi trường, các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
XI	Bộ dữ liệu kinh tế và thị trường	Bao gồm các chỉ số thống kê kinh tế vĩ mô, dữ liệu đăng ký doanh nghiệp, thương mại, xuất nhập khẩu, phân tích tài chính, tỷ giá và hoạt động của thị trường chứng khoán.	Bộ Tài chính, Bộ Công Thương và các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
XII	Bộ dữ liệu văn hóa, di sản và du lịch	Bao gồm dữ liệu di tích, bảo tàng, di vật, bảo vật quốc gia, di sản tư liệu, văn hóa phi vật thể, thiết chế văn hóa nghệ thuật truyền thống, tác phẩm văn học tiêu biểu và dữ liệu hoạt động du lịch.	Bộ Văn hóa, Thể thao và Du lịch, các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
XIII	Bộ dữ liệu bản đồ và không gian địa lý quốc gia	Bao gồm dữ liệu nền tảng theo hệ tọa độ chuẩn, bản đồ chuyên đề và hệ thống dữ liệu không gian địa lý đô thị.	Bộ Nông nghiệp và Môi trường, các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
XIV	Bộ dữ liệu viễn thông và hạ tầng số	Bao gồm dữ liệu tọa độ trạm thu phát sóng di động, quy hoạch tần số, mạng internet không dây công cộng, dữ liệu doanh nghiệp công nghệ số và hạ tầng số dùng chung.	Bộ Khoa học và Công nghệ, các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.
XV	Dữ liệu đa ngôn ngữ quốc tế về Việt Nam	Bao gồm các nguồn ngữ liệu bằng ngôn ngữ nước ngoài phản ánh các lĩnh vực văn hóa, lịch sử, chính trị, xã hội liên quan đến Việt Nam.	Bộ Ngoại giao và các bộ, ngành, địa phương có liên quan theo chức năng nhiệm vụ được giao.

Phụ lục II

DANH MỤC DỮ LIỆU ƯU TIÊN TRIỂN KHAI PHỤC VỤ PHÁT TRIỂN TRÍ TUỆ NHÂN TẠO

(Kèm theo Quyết định số 804/QĐ-TTg ngày 06 tháng 5 năm 2026 của Thủ tướng Chính phủ)

Danh mục dữ liệu quy định tại Phụ lục II được xây dựng trên cơ sở các bộ dữ liệu quy định tại Phụ lục I.

STT	DANH MỤC DỮ LIỆU	MÔ TẢ	DANH MỤC TƯƠNG ỨNG STT PHỤ LỤC I
I	**Dữ liệu phục vụ phát triển mô hình ngôn ngữ lớn tiếng Việt**		
1	Dữ liệu ngôn ngữ tiếng Việt phổ thông	Dữ liệu văn bản tiếng Việt chuẩn, phản ánh ngữ pháp, từ vựng và cách diễn đạt phổ biến trong đời sống và hành chính.	I
2	Dữ liệu hội thoại và tương tác tiếng Việt	Dữ liệu hội thoại, trao đổi và tương tác ngôn ngữ trong các ngữ cảnh giao tiếp xã hội khác nhau, để hiểu nội dung giao tiếp.	
3	Dữ liệu tiếng nói tiếng Việt	Dữ liệu âm thanh và giọng nói tiếng Việt đa vùng miền, phục vụ nhận dạng và tổng hợp tiếng nói, để hiểu và tạo âm thanh giọng nói vùng miền.	
4	Dữ liệu ngôn ngữ các dân tộc thiểu số	Dữ liệu ngôn ngữ, tiếng nói và văn bản của các dân tộc thiểu số tại Việt Nam.	
5	Dữ liệu hướng dẫn và căn chỉnh mô hình	Dữ liệu chỉ dẫn, phản hồi và chuẩn hóa hành vi nhằm tinh chỉnh và căn chỉnh mô hình trí tuệ nhân tạo.	
6	Dữ liệu tổng hợp nhân tạo	Dữ liệu được tạo sinh nhân tạo nhằm bổ sung, mở rộng và cân bằng tập dữ liệu huấn luyện.	

STT	DANH MỤC DỮ LIỆU	MÔ TẢ	DANH MỤC TƯƠNG ỨNG STT PHỤ LỤC I
7	Dữ liệu báo chí và truyền thông	Dữ liệu nội dung báo chí, truyền thông đa phương tiện phản ánh thông tin thời sự và xã hội.	II
8	Dữ liệu tri thức và thuật ngữ chuyên ngành	Dữ liệu tri thức chuyên sâu và hệ thống thuật ngữ trong các lĩnh vực chuyên ngành.	
9	Dữ liệu pháp luật và văn bản quản lý hành chính	Dữ liệu văn bản quy phạm pháp luật, án lệ và các văn bản hành chính.	III
10	Dữ liệu khoa học, học thuật, quy chuẩn kỹ thuật	Dữ liệu nhiệm vụ nghiên cứu khoa học đã nghiệm thu, tài liệu học thuật và tiêu chuẩn kỹ thuật, quy chuẩn kỹ thuật quốc gia.	IV
11	Dữ liệu mã nguồn và lập trình	Dữ liệu mã nguồn, tài liệu kỹ thuật và nội dung liên quan đến lập trình.	
12	Dữ liệu toán học và suy luận khoa học	Dữ liệu bài toán, công thức và nội dung phục vụ suy luận logic và khoa học.	IV, VII
13	Dữ liệu văn hóa và lịch sử Việt Nam	Dữ liệu phản ánh giá trị văn hóa, lịch sử và bản sắc dân tộc Việt Nam.	XII
14	Dữ liệu từ nguồn mở trên Internet và môi trường số (không bao gồm dữ liệu báo chí và truyền thông tại mục 7)	Dữ liệu công khai từ Internet và các nền tảng số đã được xử lý và chuẩn hóa.	II, XIV
15	Dữ liệu song ngữ và đa ngôn ngữ	Dữ liệu chứa nội dung đa ngôn ngữ phục vụ dịch máy và liên thông ngôn ngữ.	XV
II	**Dữ liệu phục vụ phát triển trí tuệ nhân tạo thị giác máy tính**		

STT	DANH MỤC DỮ LIỆU	MÔ TẢ	DANH MỤC TƯƠNG ỨNG STT PHỤ LỤC I
16	Dữ liệu đa phương tiện phục vụ hiểu ngữ cảnh hình ảnh và video	Dữ liệu hình ảnh và video kèm thông tin mô tả, nhãn hoặc phụ đề, phản ánh ngữ cảnh và nội dung liên quan, phục vụ phát triển các mô hình trí tuệ nhân tạo nhận diện, phân tích và hiểu ngữ nghĩa thị giác.	II
17	Dữ liệu hình ảnh y tế	Dữ liệu hình ảnh trong lĩnh vực y tế phục vụ chẩn đoán và phân tích.	VI
18	Dữ liệu hình ảnh nông nghiệp	Dữ liệu hình ảnh phục vụ giám sát, phân tích và quản lý nông nghiệp.	VIII
19	Dữ liệu hình ảnh giao thông	Dữ liệu hình ảnh phục vụ giám sát và điều hành giao thông.	IX
20	Dữ liệu hình ảnh đô thị và hạ tầng	Dữ liệu hình ảnh phản ánh hạ tầng kỹ thuật và không gian đô thị.	IX, XIV
21	Dữ liệu ảnh vệ tinh và viễn thám	Dữ liệu ảnh từ vệ tinh và công nghệ viễn thám phục vụ phân tích không gian.	X, XIII
III	**Dữ liệu phục vụ kiểm thử và đánh giá hệ thống trí tuệ nhân tạo**		
22	Dữ liệu đánh giá năng lực hiểu tiếng Việt	Dữ liệu kiểm tra khả năng hiểu ngôn ngữ tiếng Việt của mô hình trí tuệ nhân tạo.	I
23	Dữ liệu đánh giá năng lực hội thoại tiếng Việt	Dữ liệu đánh giá khả năng hội thoại và tương tác ngôn ngữ của trí tuệ nhân tạo.	
24	Dữ liệu đánh giá hiểu biết pháp luật Việt Nam	Dữ liệu kiểm tra khả năng hiểu và áp dụng pháp luật Việt Nam.	III
25	Dữ liệu đánh giá năng lực lập trình và kỹ thuật	Dữ liệu đánh giá khả năng lập trình và xử lý kỹ thuật của trí tuệ nhân tạo.	IV

STT	DANH MỤC DỮ LIỆU	MÔ TẢ	DANH MỤC TƯƠNG ỨNG STT PHỤ LỤC I
26	Dữ liệu đánh giá năng lực suy luận và tri thức	Dữ liệu đánh giá khả năng suy luận logic và tri thức tổng hợp.	IV, VII
27	Dữ liệu đánh giá hiểu biết văn hóa và xã hội	Dữ liệu đánh giá mức độ hiểu biết về văn hóa và xã hội Việt Nam.	XII
IV	**Dữ liệu phục vụ phát triển trí tuệ nhân tạo trong các lĩnh vực thiết yếu**		
28	Dữ liệu thủ tục hành chính	Dữ liệu về quy trình, hồ sơ, biểu mẫu giải quyết thủ tục hành chính của cơ quan nhà nước.	V
29	Dữ liệu y tế và chăm sóc sức khỏe	Dữ liệu phục vụ quản lý, chẩn đoán và chăm sóc sức khỏe.	VI
30	Dữ liệu giáo dục	Dữ liệu phục vụ giảng dạy, học tập, kiểm tra và quản lý giáo dục.	VII
31	Dữ liệu môi trường và khí hậu	Dữ liệu phục vụ giám sát, dự báo và quản lý môi trường, khí hậu.	X
32	Dữ liệu năng lượng và tài nguyên	Dữ liệu phục vụ quản lý, khai thác và sử dụng tài nguyên, năng lượng.	X
33	Dữ liệu tài chính và kinh tế	Dữ liệu phục vụ phân tích, điều hành và dự báo kinh tế – tài chính.	XI
34	Dữ liệu bản đồ số và địa lý quốc gia	Dữ liệu không gian địa lý và bản đồ phục vụ quản lý lãnh thổ.	XIII
35	Dữ liệu hạ tầng viễn thông và mạng số	Dữ liệu về hạ tầng mạng, kết nối và hệ thống viễn thông.	XIV
V	**Dữ liệu phục vụ phát triển trí tuệ nhân tạo phục vụ đánh giá an toàn và tin cậy**		
36	Dữ liệu phát hiện tin giả và thông tin sai lệch	Dữ liệu phục vụ nhận diện và xử lý thông tin sai lệch trên môi trường số.	II, III

STT	DANH MỤC DỮ LIỆU	MÔ TẢ	DANH MỤC TƯƠNG ỨNG STT PHỤ LỤC I
37	Dữ liệu nhận diện sản phẩm đa phương tiện do trí tuệ nhân tạo tạo ra	Dữ liệu phục vụ phát hiện nội dung giả mạo do trí tuệ nhân tạo tạo ra.	II
38	Dữ liệu kịch bản tấn công và kiểm thử đạo đức trí tuệ nhân tạo	Dữ liệu mô phỏng tấn công và kiểm thử độ an toàn, đạo đức của trí tuệ nhân tạo.	IV
39	Dữ liệu phát hiện mã độc và mối đe dọa an ninh mạng.	Dữ liệu phục vụ phát hiện và phòng chống các mối đe dọa an ninh mạng.	III, XIV