



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