

Trí tuệ nhân tạo và tương lai của lưu trữ tài liệu khoa học và công nghệ tại Việt Nam

Nguyễn Thị Vân Nga, Chu Thị Ngân

Trung tâm Dữ liệu và Thông tin khoa học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam



Lưu trữ tài liệu khoa học và công nghệ (KH&CN) có vai trò thiết yếu trong việc bảo tồn và phát triển tri thức quốc gia. Sự bùng nổ của công nghệ trí tuệ nhân tạo (AI) đã mở ra những hướng đi mới nhằm tối ưu hóa công tác lưu trữ, quản lý và khai thác dữ liệu KH&CN. Bài viết phân tích tiềm năng ứng dụng AI trong lưu trữ tài liệu KH&CN tại Việt Nam, đánh giá thực trạng hiện nay, đồng thời đề xuất mô hình tích hợp AI, hướng tới xây dựng hệ thống lưu trữ hiện đại, thông minh, phục vụ hiệu quả hoạt động nghiên cứu và đổi mới sáng tạo.



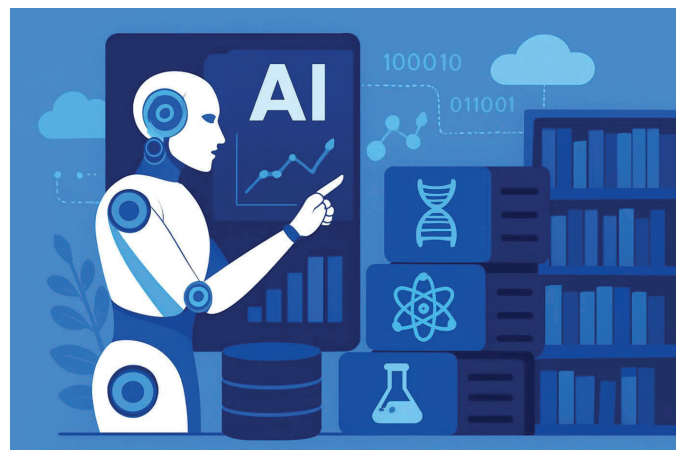
Xu hướng ứng dụng trí tuệ nhân tạo trong lưu trữ tài liệu

Lưu trữ tài liệu KH&CN đóng vai trò quan trọng trong việc bảo vệ và phát huy giá trị của nguồn tri thức khoa học, góp phần vào sự phát triển bền vững của đất nước. Các tài liệu KH&CN không chỉ mang giá trị cho nghiên cứu khoa học mà còn là công cụ hiệu quả trong hoạch định chính sách, phát triển công nghệ và đổi mới sáng tạo. Chính vì vậy, hệ thống lưu trữ tài liệu KH&CN phải đảm bảo tính bảo mật, khả năng truy xuất dễ dàng và chia sẻ hiệu quả. Trong bối cảnh chuyển đổi số quốc gia, sự chuyển đổi từ lưu trữ truyền thống sang lưu trữ điện tử giúp nâng cao hiệu quả quản lý và khai thác tài liệu KH&CN.

Đặc biệt, việc ứng dụng AI trong lưu trữ tài liệu KH&CN đang mở ra những cơ hội mới. Nhờ khả năng tự động phân loại, gắn nhãn, trích xuất thông tin, nhận dạng nội dung và phát hiện các mối liên hệ giữa tài liệu mà con người có thể bỏ sót nên AI được ứng dụng vào quá trình quản lý, tìm kiếm và khai thác dữ liệu khoa học để cải thiện độ chính xác và hiệu quả. Bên cạnh đó, AI còn hỗ trợ phân tích xu hướng nghiên cứu, dự báo công nghệ, đánh giá tác động của KH&CN và cung cấp các khuyến nghị có giá trị cho việc hoạch định chính sách. Việc tích hợp AI vào hệ

thống lưu trữ sẽ góp phần xây dựng hệ sinh thái dữ liệu mở và thông minh, đáp ứng yêu cầu phát triển bền vững và chuyển đổi số quốc gia.

Trong những năm gần đây, công nghệ AI không chỉ giúp tự động hóa các quy trình lưu trữ mà còn cải thiện đáng kể khả năng truy xuất quản lý và phân tích dữ liệu. Cụ thể, học máy (machine learning) là công nghệ AI giúp máy tính cải thiện hiệu suất mà không cần lập trình lại. Công cụ này hỗ trợ phân tích, phân loại dữ liệu, dự đoán nhu cầu sử dụng dựa trên hành vi người dùng, giúp nâng cao hiệu quả tìm kiếm và truy xuất thông tin. Học sâu (deep learning) là một nhánh



Hình ảnh minh họa về vai trò của trí tuệ nhân tạo trong lưu trữ tài liệu khoa học và công nghệ.

của học máy, dùng mạng nơ-ron để xử lý các tài liệu phức tạp như văn bản không cấu trúc hoặc hình ảnh, nhằm cải thiện độ chính xác trong phân tích và tìm kiếm. Xử lý ngôn ngữ tự nhiên (NLP) cho phép máy tính hiểu và xử lý ngôn ngữ con người. Trong lưu trữ, NLP giúp trích xuất thông tin, phân loại theo chủ đề và xây dựng hệ thống tìm kiếm thông minh có khả năng hiểu ngữ cảnh truy vấn. Công nghệ hỗ trợ chuyển đổi văn bản hình ảnh hoặc tài liệu quét thành văn bản số (Optical Character Recognition - OCR), ứng dụng hiệu quả trong hoạt động số hóa và tìm kiếm tài liệu. Đáng chú ý, chatbot AI đóng vai trò là trợ lý ảo trong việc hỗ trợ người dùng tìm kiếm tài liệu, nhờ tích hợp NLP để hiểu và phản hồi các câu hỏi tự động và chính xác.

Các quốc gia phát triển như Hoa Kỳ, Nhật Bản và các nước châu Âu đã tích cực ứng dụng AI trong hệ thống lưu trữ tài liệu KH&CN, mang lại những bước tiến quan trọng trong việc tối ưu hóa quản lý dữ liệu và thúc đẩy chia sẻ tri thức.

Tại Hoa Kỳ, các tổ chức nghiên cứu lớn như Cơ quan Hàng không và Vũ trụ Hoa Kỳ và Viện Y tế Quốc gia Hoa Kỳ đã ứng dụng AI để tự động hóa quy trình lưu trữ, từ phân loại, gắn mã, đến phân tích dữ liệu và dự báo nhu cầu người dùng. Các nền tảng như Google Scholar và JSTOR sử dụng công nghệ học máy và xử lý NLP để hiểu ngữ nghĩa tài liệu, tối ưu hóa kết quả tìm kiếm và đề xuất tài liệu liên quan theo ngữ cảnh.

Tại Nhật Bản, Viện nghiên cứu RIKEN đã triển khai hệ thống số hóa tài liệu sử dụng công nghệ OCR kết hợp AI, cho phép tự động nhận diện, phân loại và lưu trữ tài liệu KH&CN, đặc biệt hiệu quả trong việc xử lý những tài liệu giấy đã cũ và xuống cấp. Tương tự, các quốc gia châu Âu như Đức và Anh cũng đã đầu tư lớn cho AI trong lưu trữ tài liệu KH&CN; công nghệ deep learning được sử dụng để xử lý các tài liệu phi cấu trúc như hình ảnh, âm thanh, video nhằm mở rộng phạm vi lưu trữ và nâng cao giá trị sử dụng của dữ liệu đa phương tiện trong nghiên cứu. Ngoài ra, chatbot AI đang trở thành công cụ hỗ trợ phổ biến trong các thư viện số quốc tế. Với khả năng hiểu ngôn ngữ, chatbot giúp người dùng tìm kiếm tài liệu, trả lời các câu hỏi và hướng dẫn tra cứu, giúp giảm tải cho cán bộ thủ thư.

Tất cả các mô hình này đều thể hiện rõ hiệu quả của việc ứng dụng AI trong lưu trữ tài liệu KH&CN, từ tự động hóa quy trình, nâng cao khả năng truy xuất, đến dự báo hành vi người dùng. Việc nghiên cứu, chọn lọc và triển khai phù hợp các mô hình này ứng dụng tại Việt Nam là cần thiết, nhằm nâng cao năng lực khai thác tri thức và kết nối cộng đồng nghiên cứu quốc tế.

Thực trạng và triển vọng ứng dụng tại Việt Nam

Tại Việt Nam, khối lượng tài liệu KH&CN ngày càng gia tăng nhanh chóng, cùng với sự phát triển mạnh mẽ của công nghệ, việc ứng dụng AI vào hệ thống lưu trữ tài liệu là một xu hướng tất yếu. Một mô hình ứng dụng nổi bật là nền tảng Lưu trữ số quốc gia do Cục Văn thư và Lưu trữ Nhà nước phối hợp với Công ty TNHH Hệ thống thông tin FPT (FPT IS) phát triển. Nền tảng này tích hợp các công nghệ AI như nhận dạng ký tự OCR và học máy để tự động hóa quy trình số hóa, phân loại và truy xuất tài liệu, đảm bảo tuân thủ tiêu chuẩn quốc tế OAIS (ISO 14721:2012). Ngoài ra, các thư viện lớn tại Việt Nam cũng đang tích cực triển khai các giải pháp ứng dụng AI trong lưu trữ và quản lý tài liệu. Thư viện Quốc gia Việt Nam đã thí điểm công nghệ AI trong hoạt động phân loại và biên mục tài liệu, giúp xử lý nhanh và đồng bộ kho tài liệu số với độ chính xác cao hơn. Học viện Cảnh sát Nhân dân đã áp dụng AI để quản lý tài liệu và cung cấp các tiện ích như tìm kiếm thông minh và chatbot hỗ trợ tra cứu. Về mặt pháp lý và chính sách, việc ban hành Luật Lưu trữ (sửa đổi) năm 2024 với quy định rõ hơn về lưu trữ số đang được kỳ vọng sẽ tạo nền tảng pháp lý vững chắc cho các cơ quan triển khai hệ thống lưu trữ ứng dụng công nghệ AI trong tương lai. Tuy nhiên, phần lớn các hệ thống mới chỉ dừng lại ở mức thí điểm hoặc ứng dụng đơn lẻ, chưa có sự tích hợp đồng bộ và quy mô lớn. Việc thiếu cơ sở hạ tầng dữ liệu dùng chung, hạn chế về nhân lực có chuyên môn sâu trong cả AI và lưu trữ, cũng như vấn đề bảo mật, bản quyền và chia sẻ dữ liệu là những rào cản lớn cần vượt qua.

Để xây dựng hệ thống lưu trữ tài liệu KH&CN hiện đại, thông minh và phù hợp với bối cảnh chuyển đổi số quốc gia, cần những định hướng chiến lược rõ ràng trong việc ứng dụng AI. Trước hết, Việt Nam cần xác



Cán bộ Trung tâm Lưu trữ nhập dữ liệu để số hóa hồ sơ.

định lưu trữ số là hạ tầng chiến lược phục vụ nghiên cứu khoa học và đổi mới sáng tạo. Việc ứng dụng AI không chỉ là giải pháp công nghệ mà còn là bước đi tất yếu để đảm bảo hiệu quả, an toàn và bền vững trong quản trị tài liệu KH&CN.

Một trong những định hướng quan trọng là phát triển hệ thống lưu trữ tài liệu tích hợp AI theo hướng mở và tiêu chuẩn hóa, có khả năng xử lý dữ liệu lớn, học hỏi và thích nghi với các mô hình dữ liệu đa dạng. Hệ thống này nên kết hợp các công nghệ như nhận dạng NLP để tự động phân loại, gắn nhãn và phân tích ngữ cảnh nội dung tài liệu; học máy để gợi ý liên kết tài liệu có liên quan và AI hỗ trợ tìm kiếm thông minh, nâng cao trải nghiệm người dùng.

Tiếp theo, cần xây dựng nền tảng dữ liệu lưu trữ dùng chung cấp quốc gia, cho phép chia sẻ tài nguyên giữa các viện nghiên cứu, trường đại học và cơ quan quản lý khoa học. Nền tảng này nên tích hợp cơ chế AI học sâu để tối ưu hóa việc khai thác, truy xuất và kết nối dữ liệu theo ngữ nghĩa, hướng tới phát triển hệ sinh thái dữ liệu khoa học mở.

Việc hoàn thiện chính sách và quy định pháp lý phù hợp là cần thiết khi triển khai AI trong lưu trữ KH&CN. Trong đó, cần có quy định rõ ràng về tiêu chuẩn dữ liệu lưu trữ số, quy trình kiểm định công nghệ AI áp dụng trong lĩnh vực lưu trữ, cơ chế bảo vệ quyền sở hữu trí tuệ và chia sẻ dữ liệu. Chính phủ sẽ đóng vai trò kiến tạo trong việc ban hành các chiến lược AI quốc gia, đồng thời cần có chính sách khuyến khích khu vực tư nhân cùng tham gia phát triển nền tảng hạ tầng số.

Bên cạnh đó, phát triển nguồn nhân lực kép về lưu trữ học và trí tuệ nhân tạo là yếu tố then chốt. Xây dựng các chương trình đào tạo chuyên biệt cho cán bộ lưu trữ tại các viện nghiên cứu, thư viện và trung tâm dữ liệu, phải kết hợp giữa kiến thức chuyên môn và kỹ năng công nghệ mới. Đây là điều kiện tiên quyết để vận hành và khai thác hiệu quả các hệ thống lưu trữ dữ liệu tích hợp AI.

Điều quan trọng cần tính đến là việc triển khai mô hình thí điểm quy mô vừa và nhỏ tại các đơn vị có nền tảng dữ liệu tốt như thư viện quốc gia, thư viện các trường đại học trọng điểm hoặc các trung tâm quản lý dữ liệu KH&CN cấp bộ/ngành. Qua đó, đánh giá hiệu quả ứng dụng AI, điều chỉnh quy trình và công nghệ trước khi nhân rộng trong toàn quốc.

*
* *

Việc ứng dụng AI vào hệ thống lưu trữ tài liệu KH&CN là một hướng đi tiềm năng, giúp tối ưu hóa việc quản lý và chia sẻ tri thức. Tuy nhiên, để triển khai AI tại Việt Nam trong lĩnh vực này cần thận trọng trong từng bước đi, ngay từ khâu xây dựng nền tảng hạ tầng công nghệ thông tin phù hợp và đồng bộ. Hiện nay, Việt Nam vẫn đang trong giai đoạn khởi động và đang phải đối mặt với nhiều thách thức về cơ sở hạ tầng và nguồn lực. Để đạt được mục tiêu tích hợp AI vào lưu trữ KH&CN, cần xây dựng hệ thống hạ tầng dữ liệu hiện đại, phát triển các nền tảng dữ liệu số dùng chung và nâng cao chất lượng nguồn nhân lực trong cả lĩnh vực lưu trữ tài liệu KH&CN và công nghệ AI. Chính phủ và các cơ quan chức năng nên có chính sách hỗ trợ và khuyến khích đầu tư vào công nghệ mới, đồng thời tạo ra môi trường pháp lý rõ ràng để đảm bảo quyền lợi sở hữu trí tuệ và chia sẻ dữ liệu phù hợp. Mặc dù còn nhiều khó khăn, nhưng nếu có chiến lược phát triển rõ ràng và đồng bộ, Việt Nam có thể khai thác sức mạnh của AI để nâng cao hiệu quả trong công tác lưu trữ tài liệu KH&CN. Đây sẽ là bước tiến quan trọng trong quá trình chuyển đổi số và thúc đẩy sự phát triển của ngành KH&CN, góp phần nâng cao khả năng cạnh tranh và phát triển kinh tế tri thức của đất nước trong tương lai.