

Sáng kiến và giải pháp phân loại, gắn nhãn dữ liệu phục vụ nghiên cứu, ứng dụng trí tuệ nhân tạo

Sáng ngày 16/08/2024, tại Trường ĐH Nguyễn Tất Thành, cơ sở quận 4 đã diễn ra hội thảo khoa học “Sáng kiến và giải pháp phân loại, gắn nhãn dữ liệu phục vụ nghiên cứu, ứng dụng trí tuệ nhân tạo” thu hút rất đông các chuyên gia từ các trường đại học, viện nghiên cứu tham gia.



Ông Đào Mạnh Thắng - Phó Cục trưởng Cục Thông tin Khoa học và Công nghệ Quốc gia, phát biểu khai mạc hội thảo

Hội thảo do Cục Thông tin khoa học và công nghệ quốc gia - Bộ Khoa học và Công nghệ phối hợp với Trường ĐH Nguyễn Tất Thành tổ chức nhằm mục đích tạo ra một diễn đàn để các nhà khoa học, nhà nghiên cứu, và sinh viên có cơ hội thảo luận, trao đổi ý tưởng về các phương pháp phân loại và gắn nhãn dữ liệu, từ đó đóng góp vào sự phát triển của trí tuệ nhân tạo tại Việt Nam. Bên cạnh đó tạo điều kiện kết nối các trường đại học, viện nghiên cứu, và các cơ quan quản lý Nhà nước cùng đẩy mạnh hợp tác nghiên cứu và ứng dụng AI, xử lý tạo lập dữ liệu học máy, thảo luận về các ứng dụng tiềm năng của AI cũng như đề xuất các giải pháp cụ thể để triển khai những ứng dụng này tại Việt Nam.

Tham dự hội thảo có ông Đào Mạnh Thắng - Phó Cục trưởng Cục Thông tin khoa học và công nghệ quốc gia. Đại diện Trường ĐH Nguyễn Tất Thành có: TS. Trần Ái Cẩm - Ủy viên Hội đồng trường, Hiệu trưởng Nhà trường; PGS.TS. Trần Thị Hồng - Phó hiệu trưởng; PGS.TS. Bạch Long Giang - Trưởng phòng Khoa học Công nghệ; cùng đông đảo quý thầy cô là trưởng phó các đơn vị phòng ban và sinh viên trong toàn trường quan tâm đến tham dự.

Với sự phát triển mạnh mẽ của khoa học công nghệ, chuyển đổi số và trí tuệ nhân tạo đã được ứng dụng ngày càng sâu rộng, tạo nên sự thay đổi ngoạn mục trong nhiều lĩnh vực đời sống xã hội. Do đó, việc tổ chức hội thảo “Sáng kiến và giải pháp phân loại, gắn nhãn dữ liệu phục vụ nghiên cứu, ứng dụng trí tuệ nhân tạo” càng trở nên ý nghĩa với định hướng xây dựng môi trường cho các nhà khoa học, chuyên gia trao đổi, chia sẻ các “sản phẩm” AI mới nhất và làm thế nào để tạo ra các bộ dữ liệu được gắn nhãn chính xác nhất phục vụ huấn luyện AI, ngoài ra hội thảo còn góp phần tuyên truyền phổ biến tri thức về AI, giúp nâng cao chất lượng giáo dục, thúc đẩy phát triển nghiên cứu khoa học, phát triển các giải pháp AI có thể đáp ứng nhu cầu phát triển bền vững đất nước.

Phát biểu khai mạc hội thảo, ông Đào Mạnh Thắng - Phó Cục trưởng Cục Thông tin khoa học và công nghệ quốc gia, cho biết hội thảo lần này được tổ chức nhằm thúc đẩy, chia sẻ kinh nghiệm và cơ hội để các trường đại học, viện nghiên cứu và các trung tâm thư viện có thêm thông tin, cơ hội tiếp cận để cùng khai thác các nguồn tin khoa học công nghệ và xây kho dữ liệu nghiên cứu mở trong nước và quốc tế phục vụ đào tạo và nghiên cứu khoa học. Hy vọng hội thảo lần này sẽ là cầu nối cho các giảng viên, nhóm nghiên cứu trong cả nước tiếp tục đẩy mạnh hợp tác, thúc đẩy nâng cao năng lực nghiên cứu, đóng góp vào sự phát triển của nền khoa học công nghệ nước nhà.

Tại hội thảo, TS. Trần Ái Cẩm - Hiệu trưởng Trường ĐH Nguyễn Tất Thành, cũng đã thay mặt Ban lãnh đạo nhà trường gửi lời cảm ơn đến tất cả quý vị lãnh đạo, quý diễn giả đã đến tham dự chương trình, đặc biệt là phía Cục Thông tin Khoa học và Công nghệ Quốc gia đã tin tưởng lựa chọn Trường ĐH Nguyễn Tất Thành để phối hợp tổ chức một hội thảo ý nghĩa như “Sáng kiến và giải pháp phân loại, gắn nhãn dữ liệu phục vụ nghiên cứu, ứng dụng trí tuệ nhân tạo” trong bối cảnh mà các ngành, các lĩnh vực trong đời sống xã hội Việt Nam đang đẩy mạnh chuyển đổi số.

Trường ĐH Nguyễn Tất Thành với mục tiêu khuyến khích sinh viên, giảng viên và nghiên cứu viên về trí tuệ nhân tạo và khoa học dữ liệu, thời gian qua đã tổ chức thành công nhiều hoạt động, nhiều cuộc thi về trí tuệ nhân tạo và khoa học dữ liệu, thu hút nhiều dự án nghiên cứu của sinh viên tham gia, phê duyệt được nhiều đề tài nghiên cứu có liên quan đến lĩnh vực trí tuệ nhân tạo và khoa học dữ liệu.



TS. Trần Ái Cẩm - Hiệu trưởng Trường ĐH Nguyễn Tất Thành, chia sẻ tại chương trình

Trong khuôn khổ của hội thảo, các đại biểu tham dự cũng được nghe các chuyên gia cung cấp nhiều thông tin về nhận diện xu thế phát triển trí tuệ nhân tạo, khai thác các nguồn tin, gán nhãn dữ liệu với 6 nội dung, chuyên đề hữu ích được trình bày tại hội thảo lần này:

1/ “*Khai thác dữ liệu khoa học trong nước và quốc tế hỗ trợ hoạt động nghiên cứu tại các trường đại học*” (bà Trần Thị Hải Yến - Phó giám đốc Thư viện KH&CN Quốc gia);

2/ “*Tổ chức dữ liệu hỗ trợ hoạt động đổi mới sáng tạo và khởi nghiệp tại các trường đại học*” (bà Trần Thị Thúy Kiều - Giám đốc Thư viện Trường Đại học Nguyễn Tất Thành);

3/ “*Dữ liệu nghiên cứu mở và vai trò của gán nhãn dữ liệu trong việc xây dựng bộ dữ liệu phục vụ huấn luyện AI*” (ông Phùng Công Định - Giám đốc Trung tâm Dữ liệu Quốc gia, Cục Thông tin KH&CN Quốc gia);

4/ “*Một số phương pháp gán nhãn dữ liệu trong tin sinh, Giới thiệu giải pháp, công cụ tối ưu cho gán nhãn và phân tích dữ liệu Gen quy mô lớn*” (ông Dương Chí Vinh - chuyên gia nghiên cứu, Viện Nghiên cứu dữ liệu lớn VinBigdata);

5/ “*Nâng cao hiệu quả dự án AI qua gán nhãn dữ liệu: Phương pháp, thách thức và giải pháp*” (TS. Cao Văn Kiên - Phó trưởng khoa Công nghệ Thông tin Trường ĐH Nguyễn Tất Thành);

6/ “*Giải pháp tối ưu hóa quá trình gán nhãn dữ liệu trong trí tuệ nhân tạo*” (TS. Nguyễn Kim Quốc - Trưởng khoa Công nghệ Thông tin; ThS. Đặng Như Phú - Phó trưởng khoa Công nghệ Thông tin Trường ĐH Nguyễn Tất Thành);







Hình ảnh các báo cáo viên tham gia trình bày báo cáo tại hội thảo

Hội thảo trở thành điểm đến của trí thức, diễn đàn khoa học mở để các chuyên gia, các nhà khoa học cùng chia sẻ, trao đổi ý tưởng về các phương pháp phân loại, gắn nhãn dữ liệu, phục vụ cho nghiên cứu, ứng dụng và phát triển trí tuệ nhân tạo tại Việt Nam. Cũng trong khuôn khổ Hội thảo, các đại biểu tham dự cùng ban tổ chức cũng đã thảo luận và đưa ra một số các nội dung có tính gợi mở liên quan đến chủ đề này:

- Cách thức tạo ra dữ liệu huấn luyện chất lượng: Bộ dữ liệu mẫu được phân loại và gắn nhãn đúng cách là yếu tố cơ bản quyết định cho việc huấn luyện mô hình AI. Các nhà nghiên cứu cần có dữ liệu đáng tin cậy và đa dạng để xây dựng và đào tạo các mô hình AI hiệu quả.
- Cách thức nghiên cứu và phát triển mô hình AI: Bằng cách có sẵn các bộ dữ liệu mẫu phân loại và gắn nhãn, các nhà nghiên cứu có thể tiến hành nghiên cứu và phát triển các mô hình AI mới và cải tiến hiện có trong các lĩnh vực khác nhau. Điều này giúp tăng cường hiểu biết và khả năng áp dụng của trí tuệ nhân tạo trong các ứng dụng thực tế.
- Thảo luận về phương pháp kiểm tra và đánh giá mô hình AI: Các bộ dữ liệu mẫu được phân loại và gắn nhãn sử dụng để kiểm tra và đánh giá hiệu suất của các mô hình AI. Bằng cách sử dụng các bộ dữ liệu kiểm tra có chất lượng, các nhà nghiên cứu, phát triển có thể đảm bảo rằng mô hình AI hoạt động đúng mục đích và tin cậy trong các tình huống thực tế.
- Thảo luận về việc ứng dụng phân loại, gắn nhãn dữ liệu trong các lĩnh vực cụ thể: Các bộ dữ liệu mẫu được phân loại và gắn nhãn có thể được sử dụng để áp dụng trí tuệ nhân tạo trong các lĩnh vực cụ thể như y học, tự động hóa, công nghệ, nông nghiệp, môi trường, và nhiều lĩnh vực khác. Điều này giúp tạo ra các giải pháp hiệu quả, tiết kiệm thời gian và công sức trong nghiên cứu và tạo ra các ứng dụng thực tiễn.



Đại biểu tham gia thảo luận



Toàn bộ dữ liệu sau khi phân loại, gắn nhãn sẽ được đưa vào hệ thống Openscience.vn. Đây là nền tảng cho phép thu thập, lưu trữ quản trị, chia sẻ dữ liệu nghiên cứu KH&CN, tích hợp các công cụ phát triển, thực thi mô hình phân tích dữ liệu, học máy. Đối tượng sử dụng là các nhà khoa học dữ liệu, các cán bộ nghiên cứu, giảng viên, học viên, sinh viên các trường đại học, các cán bộ quản lý trong lĩnh vực khoa học dữ liệu, học máy, trí tuệ nhân tạo (AI). Nền tảng cũng là môi trường cho phép cộng đồng đóng góp, khai thác, chia sẻ dữ liệu và xây dựng, phát triển các mô hình học máy, phân tích dữ liệu, góp phần quan trọng vào việc thực hiện thắng lợi Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030.

NASATI

Nguồn: Cục Thông tin Khoa học và Công nghệ Quốc gia.