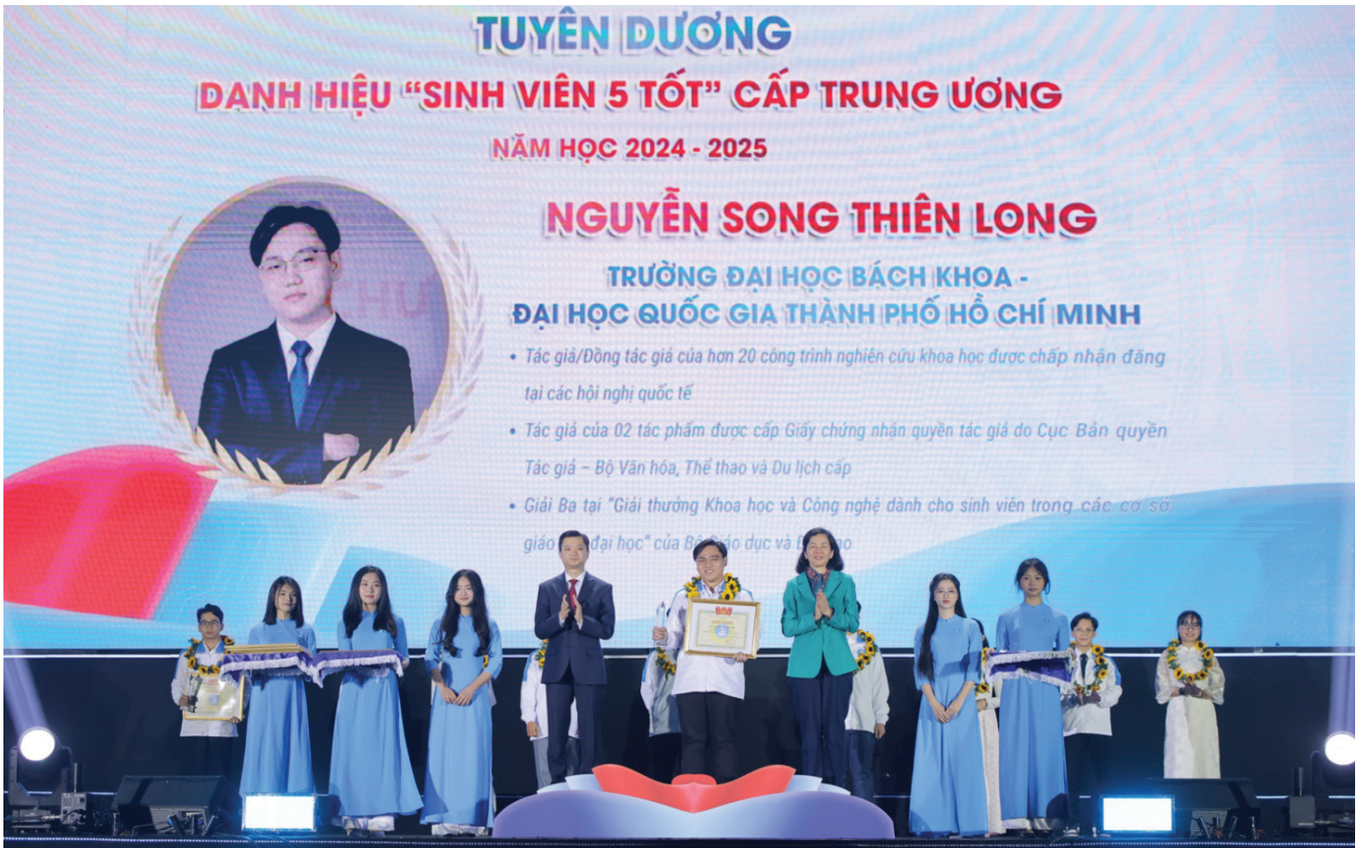

Nguyễn Song Thiên Long nhận danh hiệu "Sinh viên 5 tốt" tiêu biểu cấp Trung ương năm học 2024-2025. Ảnh: NVCC.

Người trẻ ứng dụng trí tuệ nhân tạo để bảo tồn văn hoá, ngôn ngữ

Nguyễn Song Thiên Long, sinh viên năm thứ ba ngành khoa học máy tính, Trường Đại học Bách khoa, Đại học Quốc gia TP. Hồ Chí Minh là gương mặt sinh viên tài năng trong lĩnh vực khoa học và công nghệ (KH&CN) với hơn 20 công trình nghiên cứu được công bố trong và ngoài nước. Những nỗ lực bền bỉ trong học tập và nghiên cứu đã giúp Thiên Long đạt danh hiệu "Sinh viên 5 tốt" tiêu biểu cấp Trung ương năm học 2024-2025, ghi dấu ấn của một người trí thức trẻ say mê sáng tạo và không ngừng theo đuổi tri thức.

Trí tuệ nhân tạo giúp "chạm" trực tiếp đến tri thức, ngôn ngữ và đời sống xã hội

Sinh năm 2005 tại Bến Tre, vào thời điểm máy tính vẫn còn là thiết bị xa xỉ, Thiên Long đã sớm có cơ hội tiếp xúc với công nghệ. Chia sẻ với phóng viên Tạp chí KH&CN Việt Nam, Thiên Long cho biết: "Em đến với nghiên cứu khoa học khá sớm, từ khi còn là học sinh phổ thông, nhưng phải đến khi vào đại học và được tham gia môi trường nghiên cứu bài bản thì mới thật sự hiểu nghiên cứu là một hành trình dài của việc học cách đặt câu hỏi, tìm lời giải và chấp nhận đi từng bước nhỏ để tạo ra tri thức mới".

"Trong quá trình đó, em chọn đi sâu vào nghiên cứu trí tuệ nhân tạo (AI), đặc biệt là xử lý ngôn ngữ tự nhiên, vì nhận ra đây là lĩnh vực mà công nghệ có thể "chạm" trực tiếp đến tri thức, ngôn ngữ và đời sống xã hội. Về lâu dài, em mong muốn theo đuổi những hướng nghiên cứu giúp hệ thống AI không chỉ mạnh hơn về mặt kỹ thuật, mà còn tin cậy hơn khi phục vụ các bài toán gắn với tri thức, ngôn ngữ và con người".

Thiên Long quan tâm nhiều hơn đến bài toán bảo tồn ngôn ngữ khi tham gia các dự án liên quan đến những ngôn ngữ ít tài nguyên như tiếng Ba Na hay tiếng Tày, thông qua một đề tài cấp Nhà nước mà nhóm nghiên cứu tham gia thực hiện có tên là: "Nghiên cứu xây dựng hệ thống dịch tự động văn bản tiếng Việt sang văn bản tiếng Ba Na và hệ thống tổng hợp tiếng Ba Na (bao gồm đủ các phương ngữ)".

"Khi làm việc với những ngôn ngữ này, em thấy rất rõ rằng nhiều ngôn ngữ dân tộc thiểu số hiện nay gần như chưa hiện diện nhiều trong môi trường số, trong khi ngôn ngữ lại là nơi lưu giữ ký ức, tri thức và bản sắc văn hóa của cả một cộng đồng. Nếu một ngôn ngữ không thể xuất hiện trên môi trường số, về lâu dài khoảng cách tiếp cận tri thức của cộng đồng sử dụng ngôn ngữ đó cũng sẽ lớn dần lên".

Theo Thiên Long, sự cấp thiết của việc nghiên cứu này nằm ở chỗ nếu không bắt đầu số hóa, thu thập dữ liệu và xây dựng công nghệ cho những ngôn ngữ này ngay từ bây giờ thì sau này sẽ càng khó hơn rất nhiều, vì dữ liệu ngôn ngữ vốn ít lại càng dễ "mai một" theo thời gian. AI có thể hỗ trợ từ dịch máy, nhận dạng tiếng nói, tổng hợp giọng nói cho đến xây dựng các hệ thống hỏi - đáp tri thức.

Quan trọng hơn, nếu xây dựng được các phương pháp đủ tốt cho một ngôn ngữ rất ít tài nguyên như tiếng Ba Na, thì về lâu dài hoàn toàn có thể mở rộng sang nhiều ngôn ngữ khác, đặc biệt là các ngôn ngữ có cùng họ ngôn ngữ hoặc có những đặc điểm ngôn ngữ gần nhau. Điều đó mở ra khả năng tạo ra tác động rộng hơn rất nhiều, không chỉ cho một cộng đồng riêng lẻ mà cho cả hệ sinh thái ngôn ngữ bản địa.

Thiên Long nhớ lại kỷ niệm cùng nhóm nghiên cứu đến tỉnh Bình Định (nay hợp nhất với Gia Lai thành tỉnh Gia Lai), nơi có cộng đồng người Ba Na sinh sống để thử nghiệm trực tiếp hệ thống dịch máy và tổng hợp giọng nói mà nhóm xây dựng. "Khi nhìn thấy người dân nghe thử tiếng nói được tổng hợp từ chính ngôn ngữ của họ, hoặc thử nhập câu để xem hệ thống dịch hoạt động, em cảm nhận những dòng lệnh (code) và mô hình được xây dựng trong phòng nghiên cứu thật sự có thể "chạm" đến đời sống thực tế. Với em, đó là lúc nghiên cứu trở nên có ý nghĩa rất rõ ràng, vì những gì trước đó còn nằm trên máy tính đã thật sự đi vào cuộc sống".

Thiên Long cho rằng, một công nghệ tốt không chỉ được đo bằng độ chính xác của mô hình, mà còn ở việc nó có thể khiến một cộng đồng cảm thấy ngôn ngữ của mình được lắng nghe và được hiện diện trong thời đại số.

Khi nghiên cứu là hành trình đi đến cùng

Một cột mốc đáng nhớ trong hành trình nghiên cứu cũng được Thiên Long "bật mí" là nghiên cứu bắt nguồn từ luận văn đại học của thầy và trò khoa Khoa học và Kỹ thuật Máy tính, Trường Đại học Bách khoa, Đại học Quốc gia TP. Hồ Chí Minh, sau nhiều vòng phản biện và chỉnh sửa liên tục, vừa mới đây đã được công bố trên tạp chí quốc tế hàng đầu về AI trong giáo dục là Computers and Education: Artificial Intelligence. Đây là tạp chí truy cập mở của Elsevier, hiện xếp hạng Q1, đứng đầu lĩnh vực AI, xếp hạng 2 ở lĩnh vực Giáo dục (Education), và đồng thời xếp hạng 3 ở lĩnh vực Khoa học Máy tính ứng dụng (Computer Science Applications), theo SCImago.

Nhóm tác giả bài báo "enhancing large language model performance for automatic zero-shot multiple-choice question answering via single-token logit prompting". Ảnh: ĐHBK.

Bài báo có tựa đề: "Enhancing large language model performance for automatic zero-shot multiple-choice question answering via single-token logit

prompting", tập trung vào việc đề xuất một kỹ thuật viết lời nhắc (prompting) mới có tên Single-Token Logit (STL), nhằm nâng cao độ chính xác và tính ổn định của các mô hình ngôn ngữ lớn (LLM) khi tự động trả lời câu hỏi trắc nghiệm (MCQ).

Công trình do nhóm nghiên cứu gồm giảng viên, cựu sinh viên và sinh viên của Trường Đại học Bách khoa thực hiện, dưới sự hướng dẫn của PGS.TS. Quản Thành Thơ, Trưởng khoa Khoa học và Kỹ thuật Máy tính. Công trình đã trải qua nhiều vòng phản biện kín hai chiều, với những yêu cầu chỉnh sửa liên tục về thí nghiệm, cách diễn giải, định hướng nghiên cứu và cấu trúc bản thảo.

Thiên Long cho biết: "Có những thời điểm chuyên gia phản biện chưa đồng tình với ý tưởng hoặc đặt ra những yêu cầu bổ sung rất sâu. Nhưng ở những vòng tiếp theo, sau khi nhóm đọc kỹ từng nhận xét, trả lời cẩn thận và tiếp tục chỉnh sửa bản thảo, quan điểm của phản biện dần thay đổi theo hướng tích cực hơn".

Nhìn lại quá trình thực hiện nghiên cứu, Thiên Long cho biết với em, nghiên cứu khoa học không chỉ là có một ý tưởng tốt, mà còn là quá trình kiên trì đi đến cùng với ý tưởng đó. Điều đáng nhớ nhất không chỉ nằm ở việc bài báo được công bố trên một tạp chí quốc tế uy tín, mà còn ở chặng đường thực hiện và rất dài phía sau kết quả đó. "Điều quan trọng hơn cả là quá trình ấy giúp những người trẻ làm nghiên cứu hiểu rằng tri thức khoa học luôn cần được kiểm chứng rất nghiêm túc trước khi được cộng đồng học thuật ghi nhận. Mỗi vòng phản biện không chỉ giúp công trình hoàn thiện hơn, mà còn rèn cho người làm nghiên cứu sự kiên nhẫn, khả năng tiếp nhận phản biện và tinh thần theo đuổi đến cùng một vấn đề mình tin là có giá trị. Với em, đó cũng là một cách để chứng minh sinh viên Việt Nam, nếu được đặt trong môi trường học thuật nghiêm túc và có sự dẫn dắt tốt, hoàn toàn có thể từng bước tham gia vào những công bố quốc tế có yêu cầu rất cao".

Công việc nghiên cứu khoa học đã thuận lợi hơn nhờ có trí tuệ nhân tạo

Với sự hỗ trợ của AI, công việc nghiên cứu khoa học hiện nay trở nên thuận lợi hơn ở rất nhiều khía cạnh, đặc biệt là trong việc xử lý lượng thông tin rất lớn. Hiện nay, các công cụ AI có thể hỗ trợ tìm kiếm tài liệu, tóm tắt bài báo, gợi ý tài liệu liên quan, hỗ trợ kiểm tra mã nguồn, phân tích dữ liệu hoặc thử nghiệm nhanh những ý tưởng ban đầu. Điều đó giúp người làm nghiên cứu tiết kiệm khá nhiều thời gian ở những công đoạn kỹ thuật.

Tuy nhiên, Thiên Long chia sẻ: "Điều cốt lõi vẫn nằm ở tư duy của con người. Nghiên cứu khoa học trước hết vẫn phải là quá trình con người tự đặt câu hỏi, tự suy nghĩ và tự chịu trách nhiệm với những kết luận mình đưa ra. AI có thể hỗ trợ tốc độ, nhưng không thể thay thế khả năng hiểu bản chất vấn đề, phát hiện điểm mới hay xây dựng một hướng nghiên cứu thật sự có giá trị. Có những lúc AI đưa ra câu trả lời rất nhanh, nhưng nếu người nghiên cứu không đủ nền tảng để kiểm chứng thì vẫn rất dễ đi sai hướng, thậm chí dẫn đến những kết luận thiếu chính xác hoặc bị ảnh hưởng bởi hiện tượng "ảo giác" của mô hình".

Theo Thiên Long, AI là một công cụ rất mạnh, nhưng công cụ chỉ thật sự có giá trị khi người sử dụng có đủ năng lực tư duy để điều khiển nó đúng cách. "Trong nghiên cứu, phần quan trọng nhất vẫn là trí tò mò, khả năng phản biện và việc dám suy nghĩ độc lập trước một vấn đề. Công nghệ có thể giúp mình đi nhanh hơn, nhưng hướng đi đúng vẫn phải do con người quyết định. Sau cùng, khoa học vẫn bắt đầu từ việc con người tự mình suy nghĩ nghiêm túc trước một câu hỏi chưa có lời giải".

Tuy nhiên, tốc độ phát triển rất nhanh của AI và xử lý ngôn ngữ tự nhiên hiện nay cũng là một thách thức. Gần như mỗi ngày đều xuất hiện những mô hình mới, những hướng tiếp cận mới hoặc những tiêu chuẩn mới trong cộng đồng nghiên cứu. "Điều đó buộc người làm nghiên cứu phải liên tục học hỏi nếu không muốn bị tụt lại quá xa trong "cuộc đua" tri thức này. Có nhiều giai đoạn em vừa phải đọc rất nhiều tài liệu mới, vừa cố gắng hiểu xem điều gì thật sự có giá trị và phù hợp với hướng mình đang theo đuổi".

Một khó khăn khác được Thiên Long tiếp tục chia sẻ là khoảng cách giữa ý tưởng và thực nghiệm. "Có những lúc một ý tưởng trong đầu nghe rất thuyết phục, em cảm thấy hướng đó có thể tạo ra kết quả tốt, nhưng khi triển khai thí nghiệm thì kết quả lại rất tệ hoặc không như mong đợi. Những lúc như vậy buộc em phải quay lại nhìn vấn đề kỹ hơn, kiểm tra từng giả định nhỏ và chấp nhận trong nghiên cứu, một ý tưởng hay chỉ mới là điểm bắt đầu, còn để biến nó thành kết quả khoa học thì cần rất nhiều kiểm chứng".

Nguyễn Song Thiên Long hướng dẫn trí tuệ nhân tạo cho học sinh THPT trong các chương trình kiến tập tại Trường Đại học Bách khoa. Ảnh: NVCC.

Học thuật là kết nối

Trong hành trình nghiên cứu về văn hoá, ngôn ngữ bản địa, những chuyến đi nghiên cứu mang lại nhiều ý nghĩa. Chỉ mới học năm ba đại học nhưng Thiên Long đã có cơ hội đi khá nhiều nơi, từ nhiều tỉnh, thành trong nước đến những hội nghị học thuật, cuộc thi và sự kiện quốc tế ở nước ngoài. Có những chuyến đi cùng nhóm nghiên cứu, có những lần một mình ra nước ngoài để tham dự một sự kiện học thuật, lúc đó phải tự chuẩn bị mọi thứ và học cách thích nghi một môi trường hoàn toàn mới.

"Mỗi chuyến đi đều để lại cho em nhiều cảm xúc, vì ở mỗi nơi lại gặp thêm những con người khác nhau: có người hoàn toàn mới, có người trước đó chỉ quen qua các bài báo, qua những lần trao đổi học thuật hoặc qua mạng xã hội, nhưng khi gặp trực tiếp lại có cảm giác rất gần gũi vì đã cùng chia sẻ một mối quan tâm về tri thức từ trước".

"Có lần đi dự hội nghị, em đã gặp một giáo sư mà trước đó chỉ quen nhau qua trao đổi trên mạng. Điều rất thú vị là giữa rất đông người, bằng một cách gần như rất tình cờ, em và vị giáo sư vẫn nhận ra nhau ngay từ lần gặp đầu tiên. Sau đó, hai anh em có dịp trao đổi sâu về nghiên cứu, và đến khi anh giáo sư trở lại nước Anh, hai anh em vẫn giữ liên lạc thường xuyên. Sau đó, hai anh em có cơ hội làm việc cùng trong một dự án nghiên cứu liên quan đến suy luận của mô hình ngôn ngữ lớn mà em làm chủ nhiệm", Thiên Long nhớ lại.

Một kỷ niệm khác cũng để lại nhiều ấn tượng là Hội nghị thường niên AAAI về AI lần thứ 40, một trong những hội nghị hàng đầu thế giới về AI, được tổ chức đầu năm nay tại Singapore.

"Em ấn tượng không chỉ là được trình bày nghiên cứu trong một môi trường học thuật lớn, mà còn được gặp nhiều người Việt Nam đang học tập và nghiên cứu ở nhiều nước trên thế giới. Sau thời gian hội nghị, mọi người tiếp tục cùng trao đổi về nghiên cứu, con đường học thuật, cuộc sống ở nhiều môi trường khác nhau. Với em, những khoảnh khắc như vậy vô cùng ý nghĩa, vì cho thấy tri thức không chỉ nằm trong bài báo hay phòng thí nghiệm, mà còn được tạo nên từ những kết nối rất con người như thế".

Những chuyến đi, những lần gặp gỡ và những trải nghiệm mang lại, đều được Thiên Long trân trọng. Thiên Long bày tỏ: "Em trân trọng những thầy cô, đồng nghiệp, bạn bè hay những người đi trước từng giúp đỡ và trao cơ hội, vì em hiểu rằng nhiều điều mình có được hôm nay bắt đầu từ sự tin tưởng chân thành của người khác. Làm việc trong lĩnh vực AI, em càng nhận thức rõ một điều: AI có thể xử lý rất nhanh, nhưng sự đồng cảm, lòng biết ơn và cách con người nâng đỡ nhau trên chặng đường dài vẫn là chỉ có con người mới có thể tạo ra. Có lẽ, điều đáng quý nhất không chỉ là mình đi được đến đâu, mà là mình vẫn giữ được sự tử tế với những người đã đồng hành cùng mình".

Nghiên cứu không chỉ là những con số, mà luôn gắn với con người và văn hoá. Từ những chuyến đi thực địa thu thập dữ liệu, trực tiếp trò chuyện cùng đồng bào dân tộc thiểu số, mỗi kết quả nghiên cứu dần trở thành những câu chuyện sống động về đời sống và bản sắc văn hoá của cộng đồng. Chính những trải nghiệm ấy đã giúp Thiên Long hình thành cách tiếp cận nghiên cứu nhân văn, đặt con người làm trung tâm và hướng các ứng dụng AI đến việc bảo tồn, phát huy các giá trị văn hóa Việt Nam ✍

Quỳnh Anh