

Sắp có “ChatGPT made in Việt Nam 100%” cho người dùng cuối?

11:11 25/12/2023

Năm 2023, “cú nổ” ChatGPT làm rung chuyển thị trường công nghệ thế giới, mở ra cuộc đua chinh phục AI tạo sinh giữa các quốc gia và ông lớn trong lĩnh vực công nghệ. Việt Nam không nằm ngoài xu hướng này với hàng loạt các sản phẩm AI tạo sinh ra đời.



Năm 2023, “cú nổ” ChatGPT làm rung chuyển thị trường công nghệ thế giới, mở ra cuộc đua chinh phục AI tạo sinh giữa các quốc gia và ông lớn trong lĩnh vực công nghệ. Việt Nam không nằm ngoài xu hướng này với hàng loạt các sản phẩm AI tạo sinh ra đời.

Mô hình đằng sau AI tạo sinh

AI tạo sinh (Generative AI) được coi là thế hệ tiếp theo của trí tuệ nhân tạo, tập trung vào việc tạo ra nội dung, giải pháp hoặc ý tưởng mới, thay vì chỉ phân tích dữ liệu như hiện tại. Mô hình cơ bản đằng sau hoạt động của AI tạo sinh được

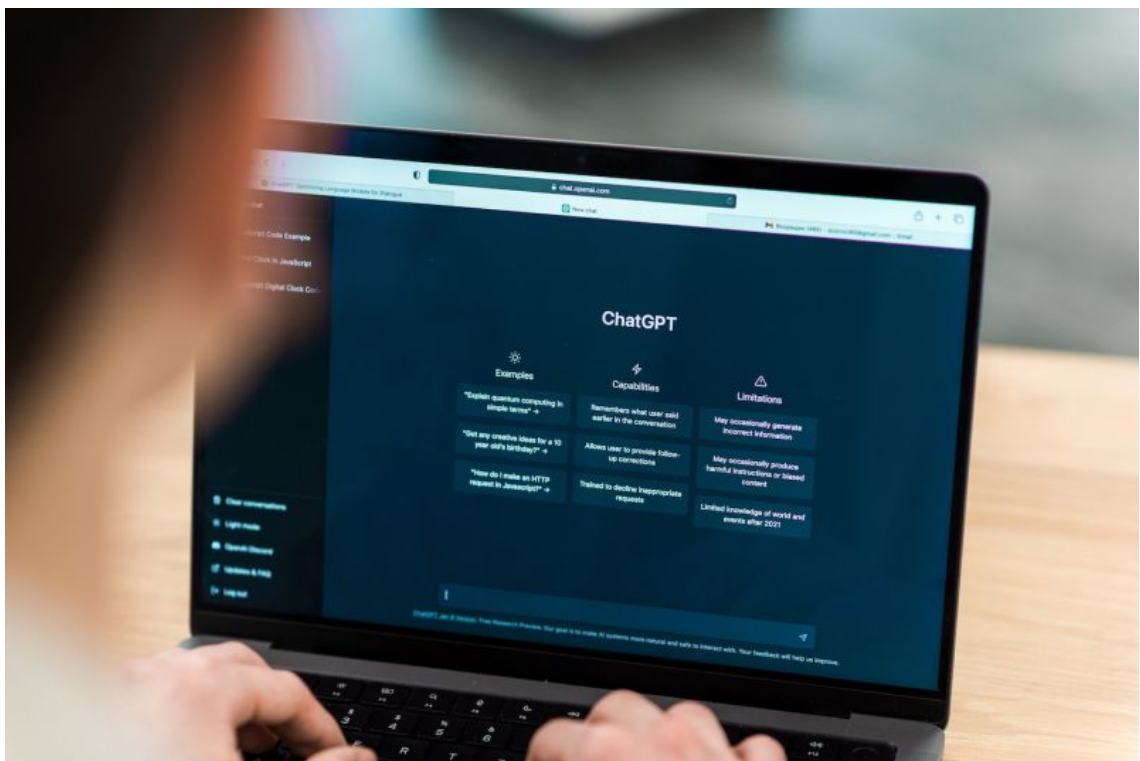
gọi là mô hình ngôn ngữ lớn (Large Language Model hay LLM).

Đây là một loại mô hình ngôn ngữ được đào tạo bằng cách sử dụng các kỹ thuật học sâu trên tập dữ liệu văn bản khổng lồ. Các mô hình này có khả năng tạo văn bản tương tự như con người và thực hiện nhiều tác vụ xử lý ngôn ngữ tự nhiên. Một số mô hình phổ biến có thể kể đến như GPT-4 do OpenAI phát triển, hay Bard của Google.

Do đó, để làm chủ AI tạo sinh hoàn toàn từ tầng lõi sâu nhất, các doanh nghiệp cần xây dựng và huấn luyện mô hình ngôn ngữ lớn từ những bước sơ khai. Cụ thể, ở giai đoạn nền tảng, mô hình có thể được đào tạo trên tập dữ liệu lớn để ghi nhớ các quy luật và cấu trúc ngôn ngữ. Qua giai đoạn đào tạo, mô hình bắt đầu rút ra mối quan hệ giữa các từ, cụm từ, câu và các khái niệm khác nhau và sau đó tiếp tục được tinh chỉnh qua nhiều bước để tạo ra nền tảng phát triển các ứng dụng thực tế của AI tạo sinh.

Thực tế ứng dụng AI tạo sinh tại Việt Nam

Hiện nay, các đơn vị nghiên cứu và phát triển AI tạo sinh trên thế giới thường cung cấp sản phẩm theo một số hình thức như mã nguồn mở (open source), mã nguồn đóng (closed source) hoặc mã nguồn đóng nhưng có sẵn qua API (APIs) và ứng dụng cho người dùng cuối (application).



Một số đơn vị đã cho ra mắt mô hình AI tạo sinh ở các định dạng khác nhau nhưng hiện nay chưa có ứng dụng 100% “made in Vietnam” cho người dùng cuối.

Với các mã nguồn mở như LLaMA (Meta), mô hình được công khai để các nhà phát triển/cộng đồng có thể tải xuống sử dụng, chỉnh sửa và tùy biến. Ngược lại, các mã nguồn đóng thường không công khai hoặc cung cấp qua giao thức kết nối (APIs) như GPT-4 (OpenAI) cho phép doanh nghiệp tích hợp phần lõi công nghệ này vào sản phẩm riêng. Ứng dụng như ChatGPT hay Bard là sản phẩm được hoàn thiện cho người dùng cuối có thể dễ dàng truy cập/cài đặt và sử dụng luôn các tác vụ cụ thể.

Tại Việt Nam, một số đơn vị đã cho ra mắt mô hình AI tạo sinh ở các định dạng khác nhau như nền tảng FPT GenAI dành cho doanh nghiệp, mô hình mã nguồn mở PhởGPT và gần đây là Zalo AI LLM. Tính đến thời điểm hiện tại, chưa có ứng dụng 100% “made in Vietnam” cho người dùng cuối được chính thức công bố tại thị trường Việt Nam.

Nhiều nguồn tin cho biết, VinBigdata đơn vị thuộc Tập đoàn Vingroup sẽ sớm ra mắt ứng dụng ViGPT - “ChatGPT phiên bản Việt” dành cho người dùng cuối và doanh nghiệp, tập trung vào một số nội dung đặc thù của Việt Nam như văn hoá, lịch sử, địa lý, danh nhân,... Theo thông tin chia sẻ, người dùng có thể sử dụng phiên bản trải nghiệm giới hạn của ViGPT bằng tiếng Việt trên giao diện web, tương tự như với ChatGPT. Có khả năng đơn vị này sẽ giới thiệu sản phẩm chính thức từ ngày 27/12/2023.

Bộ Thông tin & Truyền thông mới đây cũng ban hành kế hoạch, đặt mục tiêu đến năm 2025, Việt Nam có ít nhất một nền tảng công nghệ LLM tiếng Việt. Kế hoạch nhấn mạnh việc nghiên cứu, phát triển, đưa vào ứng dụng LLM tiếng Việt là một nhiệm vụ quan trọng, cần thiết và ý nghĩa. LLM tiếng Việt sử dụng tri thức, dữ liệu đào tạo đã được sàng lọc của Việt Nam, với chi phí thấp cho người dân, doanh nghiệp, tổ chức tại Việt Nam sử dụng để phát triển các ứng dụng mới.

Việt Nam đang ở những bước đầu trên hành trình chinh phục AI tạo sinh, việc ra mắt những ứng dụng tương tự ChatGPT của riêng người Việt là tín hiệu tốt cho thấy những nỗ lực của các đơn vị công nghệ trong nước nhằm xóa bỏ sự phụ thuộc vào những sản phẩm quốc tế, đảm bảo tính chính xác của thông tin và giảm thiểu dòng chảy dữ liệu ra thị trường nước ngoài.

Đức Anh