

MỤC LỤC

Lời tựa	7
1 Giới thiệu	11
1.1. Giao tiếp phục vụ cho AI	13
1.1.1. Giao tiếp liên nhân	15
1.1.2. Ngôn ngữ	16
1.1.3. Hiểu sự giả tạo, hiểu AI	18
1.2. AI phục vụ cho giao tiếp	19
1.2.1. Dịch máy	20
1.2.2. Xử lí ngôn ngữ tự nhiên	20
1.2.3. Robot xã hội	21
1.2.4. Mô hình ngôn ngữ lớn	22
1.2.5. Sáng tạo của máy tính và AI sinh dữ liệu	23
1.3. Thuật ngữ	24
1.3.1. Giao tiếp	24
1.3.2. Trí tuệ nhân tạo	26
1.3.3. Gofai, hay phép biến đổi kí hiệu	27
1.3.4. Mạng nơ-ron nhân tạo và học máy	28
1.3.5. Robot	31
2 Dịch máy	33
2.1. Ý tưởng và lịch sử	34
2.2. Các phương pháp và hướng tiếp cận trong dịch máy	36
2.2.1. Dịch máy cổ điển	37

1

GIỚI THIỆU

AI VÀ GIAO TIẾP

Cuốn sách này kết nối những điểm giao nhau giữa trí tuệ nhân tạo (AI) và giao tiếp. Chúng ta sẽ cùng xác định và phân tích các cách mà những tiến bộ trong lĩnh vực AI – cụ thể là dịch máy (machine translation – MT), xử lý ngôn ngữ tự nhiên (natural language processing – NLP), mô hình ngôn ngữ lớn (large language model – LLM), robot xã hội và sáng tạo của máy tính – đang làm gián đoạn và đồng thời chuyển hóa cách con người giao tiếp. Cuốn sách được viết cho sinh viên, giảng viên và độc giả phổ thông muốn hiểu rõ AI giao thoa với giao tiếp như thế nào, cũng như những phát triển gần đây về kỹ thuật và công nghệ AI sẽ ảnh hưởng ra sao đến mọi khía cạnh của tương tác xã hội giữa con người với nhau.

Nhưng nếu đã đọc phần Lời tựa, có lẽ bạn đã biết tất cả những điều này. Thực tế, phần Lời tựa đã làm khá tốt – thậm chí có thể nói là tốt vượt mong đợi – trong việc mô tả cuốn sách này nói về điều gì và vì sao nó quan trọng.

Tuy nhiên, tôi không thể nhận công lao về phần đó, bởi Lời tựa không phải do tôi viết ra. Cũng không phải do một con người nào khác viết. Nó được tạo ra bằng cách nhập một chỉ dẫn (*prompt*) vào một mô hình ngôn ngữ lớn – cụ thể là ChatGPT của OpenAI – với nội dung sau:

Hãy viết lời tựa ngắn cho một cuốn sách của Giáo sư David J. Gunkel có tên “AI trong Giao tiếp”. Cuốn sách mô tả những cách khác nhau mà AI đang làm gián đoạn giao tiếp bằng cách buộc chúng ta phải suy nghĩ lại về việc ai hoặc cái gì có thể được coi là một chủ thể giao tiếp.

Điều quan trọng ở phần Lời tựa không chỉ nằm ở nội dung của những gì được viết ra, mà còn ở cách mà văn bản đó được tạo ra. Khác với ứng dụng xử lý văn bản mà tôi đang sử dụng để soạn phần này, ChatGPT không phải là một công cụ mà tôi điều khiển để viết ra Lời tựa. Chính bản thân AI đã tạo nên Lời tựa đó. Nhưng nếu vậy, chúng ta nên đọc và diễn giải một nội dung được máy móc tạo ra như thế nào? Thông thường, một văn bản viết ra được hiểu là sản phẩm của một tác giả, người có điều gì đó muốn nói. Vậy nếu chúng ta tiếp cận bằng những câu hỏi quen thuộc như: “Ai là người đang nói với chúng ta qua phần Lời tựa này?” và “Họ muốn truyền tải điều gì?”, thì những câu hỏi đó không chỉ khó trả lời, mà thậm chí có thể trở nên vô nghĩa, vì không có “ai đó”, ít nhất là không có một con người nào, chịu trách nhiệm cho những lời hiện ra trên trang giấy. Nhưng điều đó có khiến Lời tựa này vô nghĩa hay không có nội dung gì để truyền đạt? Đây chính là những câu hỏi cốt lõi của *AI trong Giao tiếp*,

và đó là những câu hỏi chạm đến tận bản chất của những gì chúng ta cho là giao tiếp.

Chương đầu tiên này nhằm khởi động hành trình khám phá. Cụ thể, nó sẽ (1) cho thấy rằng giao tiếp của con người, ngay từ ban đầu, đã là một điều kiện nền tảng để định nghĩa trí tuệ máy móc; (2) cung cấp cái nhìn tổng quan về các công nghệ và ứng dụng AI đang làm gián đoạn hoạt động giao tiếp, và sẽ được phân tích sâu hơn ở các chương tiếp theo; (3) giải nghĩa một số thuật ngữ quan trọng nhưng phức tạp sẽ được dùng xuyên suốt trong toàn bộ cuốn sách.

1.1. GIAO TIẾP PHỤC VỤ CHO AI

Giao tiếp – cụ thể là giao tiếp liên nhân ở cấp độ con người – từ lâu đã được xem là điều kiện nền tảng để định nghĩa trí tuệ của máy móc. Điều này thể hiện rõ ràng ngay trong bài viết đột phá của nhà toán học người Anh Alan Turing, mang tên “Computing Machinery and Intelligence” (Máy tính và trí tuệ), được công bố vào năm 1950. Turing (1999, tr. 37) mở đầu bài viết bằng mong muốn trả lời câu hỏi: “Liệu máy móc có thể suy nghĩ không?” Nhưng thay vì đưa ra một câu trả lời trực tiếp kiểu “có” hay “không”, ông ngay lập tức chỉ ra vấn đề nằm ở bản thân câu hỏi, và vấn đề đó liên quan đến ngôn ngữ.

Để trả lời được câu hỏi “Liệu máy móc có thể suy nghĩ không?”, trước hết ta cần hiểu các từ “máy móc” và “suy nghĩ” nghĩa là gì. Turing lập luận rằng ta có thể

khảo sát ý kiến của nhiều người để xem họ thường hiểu và sử dụng những từ này ra sao. Nhưng cách này, theo ông kết luận, sẽ không giúp ta tiến gần hơn đến một định nghĩa khách quan cho bất kì khái niệm nào. Vì vậy, Turing làm điều mà bất kì nhà nghiên cứu giỏi nào cũng sẽ làm khi gặp bế tắc: ông thay đổi câu hỏi. Thay vì đưa ra câu trả lời trực tiếp, ông đề xuất một phương pháp kiểm tra trí thông minh của máy móc – được gọi là “trò chơi mô phỏng” (the imitation game).

Trò chơi này – ngày nay thường được biết đến với tên gọi Bài kiểm tra Turing (Turing Test) – gồm ba người tham gia: một người đàn ông, một người phụ nữ và một người hỏi. Người hỏi ngồi một mình trong một phòng và gửi câu hỏi đến hai người còn lại, mỗi người cũng ngồi ở một phòng riêng. Mục tiêu của trò chơi là để người hỏi đoán giới tính của hai người kia chỉ dựa vào câu trả lời, mà không thấy mặt hay nghe giọng. Vì Turing hiểu rằng giọng nói có thể ảnh hưởng đến kết quả, nên ông yêu cầu mọi câu hỏi và trả lời đều phải được thực hiện bằng bút đàm⁽¹⁾ hoặc thông qua một hình thức trung gian như tin nhắn văn bản.

Với thiết lập tương tự như vậy, Turing (1999, tr. 38) đã đưa ra một biến thể: Điều gì sẽ xảy ra nếu một chiếc máy tính thay thế vị trí của người đàn ông hoặc người phụ nữ trong trò chơi này? Tức là ông hình dung một tình huống mà người hỏi trao đổi với một con người và một cỗ

(1) Trao đổi qua hình thức viết giấy chuyển tay. (Tất cả các chú thích trong sách này đều là của người dịch.)

máy, và mục tiêu mới của trò chơi là xem liệu người hỏi có phân biệt được đâu là con người thật, đâu là máy móc giả dạng con người qua cuộc trò chuyện bằng văn bản hay không. Hay nói cách khác: Liệu một cỗ máy có thể “đóng vai” một con người đủ thuyết phục trong một cuộc trò chuyện giao tiếp liên nhân, đến mức người đối thoại không thể phân biệt được đó là người hay chatbot⁽¹⁾ máy tính? Turing kết luận rằng: Nếu điều này xảy ra, thì chúng ta có lí do chính đáng để gọi cỗ máy đó là “thông minh”. Có khá nhiều điều cần lưu ý trong ví dụ này, vì vậy chúng ta sẽ tách nhỏ ra và tập trung vào ba điểm quan trọng trong phần tiếp theo.

1.1.1. GIAO TIẾP LIÊN NHÂN

Turing đã tái định nghĩa câu hỏi về trí tuệ máy móc trên cơ sở giao tiếp liên nhân. Do đó, như Simone Natale (2021, tr. 25) gợi ý trong cuốn sách *Deceitful Media (Truyện thông lừa dối)*, trò chơi mô phỏng hoàn toàn có thể được gọi là “trò chơi giao tiếp”. Lí do cho cách đặt tên này bắt nguồn từ một vấn đề mà các nhà triết học, tâm lí học và khoa học hành vi thường gọi là “vấn đề về những tâm trí khác” (the problem of other minds). Nhà thần kinh học, triết gia Paul Churchland (1999, tr. 67) đã có một mô tả nổi tiếng về vấn đề này như sau:

Làm sao để biết được một thực thể khác – như một sinh vật ngoài hành tinh, một robot tinh vi, một

(1) Chatbot là một chương trình máy tính được thiết kế để mô phỏng cuộc trò chuyện của con người, đặc biệt là qua tin nhắn văn bản hoặc giọng nói.

máy tính có đời sống xã hội, hay thậm chí là một con người khác – thực sự là một sinh vật có suy nghĩ, cảm xúc, ý thức; chứ không chỉ là một cỗ máy vô thức hành xử theo lập trình, không hề có trạng thái tâm trí đích thực?

Thông thường, chúng ta giải quyết vấn đề này bằng cách quan sát các hành vi có thể nhìn thấy được từ bên ngoài, vốn thường được coi là dấu hiệu hoặc biểu hiện của trí tuệ. Với Turing và nhiều người tiếp bước ông, trí tuệ là thứ vừa khó định nghĩa, lại vừa không thể trực tiếp quan sát. Vì vậy, cách duy nhất để phát hiện trí tuệ là thông qua các biểu hiện bên ngoài, đặc biệt là giao tiếp bằng lời ở cấp độ con người – vốn được xem là chỉ dấu của trí thông minh. Nói cách khác, vì suy nghĩ thông minh là thứ vô hình và không thể đo đạc trực tiếp, nên điều tốt nhất chúng ta có thể làm là xem xét các hình thức tương tác giao tiếp – thứ được giả định là sản phẩm của trí tuệ và có thể được quan sát, đo lường và đánh giá dựa trên thực nghiệm.

1.1.2. NGÔN NGỮ

Như bạn có thể đã nhận ra, Turing không sử dụng cụm từ “trí tuệ nhân tạo” (artificial intelligence). Đó là bởi vì, vào thời điểm ông viết bài luận của mình, thuật ngữ này chưa tồn tại. Chúng ta có được khái niệm “AI” từ nhà khoa học máy tính người Mỹ John McCarthy và một đề xuất cho hội thảo mùa hè mà ông cùng một vài

đồng nghiệp tổ chức tại Đại học Dartmouth vào năm 1955 – 1956. Trong bản đề xuất đó, McCarthy không định nghĩa AI một cách rõ ràng, mà thay vào đó liệt kê một loạt các năng lực hành vi để mô tả mục tiêu cần đạt được của lĩnh vực này: “Chúng tôi cố gắng tìm cách khiến máy móc có thể sử dụng ngôn ngữ, hình thành khái niệm và tư duy trừu tượng, giải quyết các loại vấn đề vốn dành cho con người, và tự cải thiện bản thân” (McCarthy và cộng sự, 1955). Đáng chú ý, mục đầu tiên trong danh sách này chính là năng lực giao tiếp, tức khả năng sử dụng ngôn ngữ.

Các định nghĩa về AI sau đó cũng đi theo cách tương tự. Trong cuốn *The Source Book of AI (Tuyển tập nguồn về AI)* – một giáo trình quan trọng xuất bản vào thập niên 1990 – biên tập viên Roger Schank, tiếp nối phương pháp của Turing, mở đầu bằng việc thừa nhận rằng rất khó để định nghĩa “trí tuệ”. Sau đó, ông làm theo McCarthy: liệt kê “những đặc điểm mà ta kì vọng một thực thể có trí tuệ phải có”. Và đặc điểm đầu tiên là giao tiếp: “Một thực thể có trí tuệ là thực thể mà ta có thể cùng giao tiếp. Chúng ta không thể nói chuyện với đá hay ra lệnh cho cây cối dù cố gắng đến đâu” (Schank, 1990, tr. 4). Giao tiếp – đặc biệt là giao tiếp bằng ngôn ngữ – không chỉ là một đặc tính của trí tuệ nhân tạo, mà là một điều kiện tiên quyết. Nếu một thực thể có thể diễn đạt và giải thích bản thân bằng một ngôn ngữ mà ta hiểu được, thì chúng ta có xu hướng xem nó vừa có thể hiểu được (intelligible), vừa có trí tuệ (intelligent).

1.1.3. HIỂU SỰ GIẢ TẠO, HIỂU AI

Trò chơi mô phỏng vận hành dựa trên sự giả tạo. Mục tiêu của máy móc trong trò chơi là giả làm một người tham gia đối thoại, nhằm khiến người hỏi tin rằng họ đang trò chuyện với một con người khác, dù thật ra chỉ là đang nói chuyện với chatbot. Đối với Turing, và ngay cả trong thực hành khoa học và kỹ thuật của ngành AI, kiểu giả tạo này không phải là một ngoại lệ, mà là quy tắc. Nhưng sự giả tạo, đặc biệt là trong các tương tác xã hội và trải nghiệm giao tiếp, không nhất thiết là điều tiêu cực. Khi bạn xem một bộ phim, các diễn viên đang đóng giả làm người khác để kể một câu chuyện. Xét về mặt kỹ thuật, đây là một hình thức giả tạo, nhưng đó lại là hành vi mà chúng ta không chỉ chấp nhận mà còn yêu thích vì nó mang tính giải trí.

Trong lĩnh vực AI, sự khác biệt này đã được Robert Sokolowski chỉ ra trong một bài viết không xoay quanh từ “trí tuệ” (mối bận tâm của Turing và McCarthy), mà là phần còn lại của cụm từ “trí tuệ nhân tạo” – tức là từ “nhân tạo”. Sokolowski (1988, tr. 45) nhận xét rằng từ “nhân tạo” (artificial) có nhiều nghĩa khác nhau và dường như mâu thuẫn với nhau. Nó có thể mang nghĩa là giả mạo, như trong “hoa giả” (artificial flowers). Nhưng nó cũng có thể chỉ một vật thể nhân tạo nhưng không hề giả, như “ánh sáng nhân tạo” (artificial light). Ánh sáng phát ra từ bóng đèn là ánh sáng nhân tạo so với ánh sáng tự nhiên của Mặt Trời, nhưng điều đó không có nghĩa nó

là “ánh sáng giả”. Đó vẫn là ánh sáng thật, chỉ là được tạo ra theo một cách khác.

Câu hỏi quan trọng đặt ra ở đây là: chúng ta đang dùng từ “nhân tạo” theo nghĩa nào khi nói về “trí tuệ nhân tạo”? Liệu AI là “trí tuệ giả”, chỉ là một mảnh khoe, một ảo ảnh đánh lừa người dùng? Hay AI là một dạng trí tuệ thật sự – mô phỏng hoặc tái tạo trí tuệ, được tạo ra theo cách khác? Không có gì ngạc nhiên khi câu hỏi này vẫn còn gây tranh cãi, thậm chí ngay trong nội bộ lĩnh vực AI.

1.2. AI PHỤC VỤ CHO GIAO TIẾP

Bởi giao tiếp vốn đã ăn sâu vào bản chất của khái niệm trí tuệ nhân tạo, nên không có gì ngạc nhiên khi các đổi mới trong lĩnh vực AI tác động trực tiếp đến hoạt động giao tiếp – không chỉ ảnh hưởng đến cách hiểu phổ thông của chúng ta về giao tiếp hằng ngày, mà còn đến các ngành học hàn lâm về giao tiếp, cũng như những ngành công nghiệp và thực hành chuyên môn liên quan đến truyền thông hiện nay. Dù tích cực hay tiêu cực, không có khía cạnh nào của giao tiếp và tương tác xã hội của con người nằm ngoài tầm ảnh hưởng của AI. Chính vì vậy, khi đi sâu vào nghiên cứu chủ đề này, chúng ta cần có sự chọn lọc để tập trung vào những phần tiêu biểu, thay vì cố gắng bao quát tất cả. Các chương tiếp theo của cuốn sách sẽ tập trung vào năm lĩnh vực cụ thể, trong đó ảnh hưởng theo cách gây gián đoạn giao tiếp của AI đã trở nên rõ rệt.

1.2.1. DỊCH MÁY

Chúng ta sẽ bắt đầu với dịch máy (machine translation – MT). Đúng như tên gọi, đây là việc sử dụng ứng dụng máy tính để xóa nhòa rào cản ngôn ngữ giữa các cộng đồng người. Nếu AI được hiểu là công nghệ giúp máy thực hiện những việc vốn cần đến trí thông minh của con người, thì dịch thuật chính là một ví dụ tiêu biểu.

Bạn có thể cảm nhận điều này từ chính trải nghiệm học ngoại ngữ của mình. Việc dịch giữa tiếng mẹ đẻ (ví dụ: tiếng Anh) và ngoại ngữ (ví dụ: tiếng Tây Ban Nha) đòi hỏi trí nhớ và năng lực nhận thức rất lớn. Vì vậy, việc lập trình máy tính để thực hiện công việc tương tự được xem là nhiệm vụ phù hợp với AI. Thực tế, dịch máy là một trong những ứng dụng đầu tiên của công nghệ AI. Tuy nhiên, khi đảm nhận việc chuyển ngữ giữa các ngôn ngữ tự nhiên, những công nghệ này không chỉ thay thế con người, cụ thể là các phiên dịch viên và biên dịch viên, mà còn buộc chúng ta đặt ra nhiều câu hỏi quan trọng về bản chất của ngôn ngữ loài người.

1.2.2. XỬ LÝ NGÔN NGỮ TỰ NHIÊN

Trong dịch máy, máy tính vẫn được xem là một “trung gian” giữa người gửi và người nhận. Nhưng với xử lý ngôn ngữ tự nhiên (natural language processing – NLP), tình thế đã thay đổi: AI không còn chỉ là công cụ trung gian, mà bắt đầu trở thành một chủ thể giao tiếp,