

# Mục lục

1	Dẫn nhập	5
2	AI và tri thức hằng ngày	21
3	AI và nghiên cứu	57
4	AI sinh dữ liệu và thực tiễn nghiên cứu	73
5	AI và truyền thông học thuật: Bình duyệt	93
6	Thư viện và kho lưu trữ	103
7	Đảm bảo lợi ích của AI được chia sẻ công bằng	117
8	Kết luận	131
	Tài liệu tham khảo	139
	Chỉ mục	150

bất kể tuổi tác, quốc tịch, niềm tin chính trị, thể chất hay tinh thần, bản dạng giới, di sản văn hóa, thu nhập, tình trạng di cư và tị nạn, tình trạng hôn nhân, nguồn gốc, tôn giáo hay khuynh hướng tính dục (IFLA, 2012). Nhưng để đảm bảo được cam kết đó thì cần thiết lập nhiều điều kiện tiên quyết. Một trong số đó là khả năng đọc viết cơ bản. Ngoài ra, một vấn đề cũng cần được đặc biệt quan tâm là trong bối cảnh thông tin ngày càng mang tính kỹ thuật số, thì càng phải để ý hơn tới những bất bình đẳng kỹ thuật số. Những bất bình đẳng kỹ thuật số này không chỉ là rào cản về khả năng tiếp cận công nghệ, băng thông và chi phí, mà còn là các vấn đề liên quan đến kỹ năng, sự tự tin, thời gian và động lực. Bất bình đẳng kỹ thuật số là kết quả từ việc thiếu tiếp cận giáo dục, việc làm lành nghề và tiền bạc, thường tạo ra một vòng luẩn quẩn của việc bị tước đi quyền lợi. Cũng có những bất bình đẳng kỹ thuật số mang tính quốc tế. Khi suy ngẫm về những hệ quả mà AI mang lại đối với tri thức, chúng ta cần tránh rơi vào cái bẫy chỉ tập trung vào trải nghiệm của những cư dân thành thị giàu có ở các nước phương Tây phát triển, mà phải ghi nhớ đến phần lớn nhân loại đang sống ở những khu vực ít đặc quyền hơn trên thế giới (ngoài phạm vi châu Âu, Mỹ và cái gọi là “phương Tây”). Điều này cũng bao gồm cả trải nghiệm của người cao tuổi và người khuyết tật – nhất là khi ta thừa nhận sự đa dạng trong bối cảnh của các quốc gia thuộc miền Nam toàn cầu (Global South) và trong những trải nghiệm sống với khuyết tật.

Câu hỏi đặt ra trong cuốn sách này cũng đề cập đến tính bền vững của các lợi ích mà AI mang lại. Nếu xét theo

cách giải thích về các Mục tiêu phát triển bền vững của Liên Hợp Quốc, vấn đề này còn liên quan đến những mối lo ngại về công bằng xã hội và môi trường toàn cầu. Có thể AI được sử dụng để giải quyết khủng hoảng khí hậu, nhưng bản thân nó cũng có tác động đáng kể đến môi trường.

Ngành Khoa học Thông tin coi quyền tiếp cận tri thức và quyền tự do biểu đạt là những giá trị cốt lõi. Tuy nhiên, cũng cần thừa nhận rằng trong một số hoàn cảnh nhất định, những quyền này không phải lúc nào cũng là tuyệt đối. Nhiều xã hội công nhận rằng các tác giả cá nhân, tổ chức và cộng đồng có thể có quyền sở hữu đối với các biểu đạt tri thức. Điều này bao gồm quyền tác giả và quyền sở hữu trí tuệ đối với các tác phẩm. Nếu không có những sự bảo vệ này, một trong những động lực chính thúc đẩy tác giả công bố ý tưởng của mình sẽ biến mất. Khoa học Thông tin cũng ủng hộ quyền của các cộng đồng trong việc kiểm soát các biểu đạt văn hóa truyền thống, như truyện kể, bài hát, nghi lễ, sản phẩm thủ công, hoa văn và nghệ thuật – những thứ mang trong mình tri thức. Các quyền này tồn tại trong một mâu thuẫn với niềm tin vào quyền tiếp cận tri thức mở cho tất cả mọi người. Ngoài ra, ngành cũng quan tâm đến quyền kiểm soát dữ liệu cá nhân, tức là những thông tin mô tả chính chúng ta. Khoa học Thông tin nhận thức rõ sự mâu thuẫn giữa nhu cầu minh bạch của các tổ chức và chính phủ với quyền được giữ bí mật trong một số quyết định. Tương tự, quyền riêng tư cá nhân cũng mâu thuẫn với việc cho phép truy cập tri thức một cách hoàn toàn tự do. Những giới hạn và cân nhắc này là rất quan trọng,

và cũng giống như quyền tiếp cận tri thức, chúng có tác động trực tiếp đến cách mà AI vận hành.

Dù tồn tại những giới hạn như đã nêu, cam kết của ngành Khoa học Thông tin đối với quyền tiếp cận và tạo lập tri thức cho tất cả mọi người vẫn là trung tâm trong cách cuốn sách này đặt và triển khai vấn đề. Tầm quan trọng của các quyền tiếp cận và sáng tạo tri thức phản ánh thực tế rằng đây là những quyền liên tục bị đe dọa. Nhiều chính phủ tìm cách ngăn chặn quyền tiếp cận tri thức thông qua việc giữ bí mật không chính đáng, kiểm duyệt, giám sát, tuyên truyền và thậm chí là sử dụng bạo lực. Các nhóm hoặc cá nhân ác ý – những kẻ thù của dân chủ và sự bao trùm – cũng là nguồn phát tán thông tin sai lệch nguy hiểm, tức những thông tin được chia sẻ một cách có chủ ý nhằm đánh lừa người khác. Bên cạnh đó là quyền lực của các tổ chức thương mại và cơ quan truyền thông – những chủ thể có thể làm ngập các kênh thông tin bằng nội dung tiếp thị, hoặc lựa chọn, định hướng lại hay kiểm duyệt thông tin vì động cơ lợi nhuận; đồng thời mở rộng giám sát thương mại đối với hành vi của chúng ta. Câu hỏi trọng tâm đối với ngành Khoa học Thông tin là: làm thế nào để nhận diện và vượt qua những mối đe dọa này – thông qua các giải pháp kỹ thuật, cơ chế quản trị, hoặc bằng cách trang bị cho con người năng lực tư duy phản biện để hiểu và đối phó một cách hiệu quả?

Do đó, cách AI đang tác động đến những mối đe dọa này là một khía cạnh quan trọng của câu hỏi được đặt ra trong cuốn sách này.

Khoa học Thông tin cũng chống lại sự bất cẩn của chính chúng ta đối với thông tin và các thói quen dễ thỏa mãn thông thường – xảy ra khi chúng ta không dành đủ thời gian để xem xét độ chính xác của thông tin, cảm thấy quá tải bởi lượng thông tin khổng lồ, hoặc thậm chí tự gây nguy hiểm cho quyền riêng tư và quyền sở hữu trí tuệ của mình khi bất cẩn chấp nhận các điều khoản và điều kiện trực tuyến. Khoa học Thông tin thúc đẩy các kỹ năng cần thiết để đánh giá thông tin, quản lý thông tin và dữ liệu, cũng như tạo ra kiến thức mới.

Vì vậy, cuốn sách này dựa trên góc nhìn đặc thù của Khoa học Thông tin, với lập trường chuẩn mực, có tính tự phản biện.

## MỘT KHÁI NIỆM VỀ AI

Để làm rõ câu hỏi mà cuốn sách này đặt ra, điều quan trọng là cần giải thích cách hiểu về thuật ngữ “AI” trong bối cảnh đang được bàn đến. Khi nghĩ về công nghệ, chúng ta thường chỉ nghĩ đến các chức năng mà nó thực hiện, chẳng hạn như lọc thư rác hay tóm tắt văn bản. Cách nhìn hẹp này phản ánh lối diễn ngôn phổ biến về AI, vốn xem nó như “chỉ là một công cụ”, hàm ý rằng bản thân nó là trung tính, còn lợi ích, chi phí hay rủi ro chỉ phụ thuộc vào cách con người lựa chọn sử dụng. Một số lối diễn ngôn phổ biến khác về các công nghệ như AI lại mang nặng tính tất định công nghệ (technological determinism) và chủ nghĩa giải pháp (solutionism):

tức giả định rằng công nghệ sẽ tất yếu dẫn đến những thay đổi không thể tránh khỏi, và rằng bản thân nó có thể tự giải quyết các vấn đề xã hội phức tạp. Chúng ta cũng thường nghe thấy những câu như “AI là tương lai”, như thể xã hội và cá nhân không thể ảnh hưởng đến bản chất hay phản kháng lại việc sử dụng công nghệ này.

Trái ngược với những diễn ngôn mang tính tất định hay chủ nghĩa giải pháp, ngành Khoa học Thông tin xem công nghệ như những hệ thống xã hội-kỹ thuật. Ngành này thừa nhận rằng công nghệ được tạo ra bởi các chủ thể xã hội, những người đưa vào đó lợi ích, giá trị và giả định của chính họ.

Công nghệ không chỉ là công cụ trung lập, mà còn mang theo các hệ quả xã hội, thường là củng cố hoặc làm dịch chuyển các cấu trúc quyền lực và bất bình đẳng. Chẳng hạn, các nhà tuyển dụng có thể sử dụng công nghệ để giám sát nhân viên và gia tăng mức độ kiểm soát đối với họ. Chính phủ cũng có thể vận dụng công nghệ như một công cụ kiểm soát xã hội.

Những bất bình đẳng quốc tế cũng có thể bị khoét sâu thêm bởi việc áp dụng các công nghệ được thiết kế cho một bối cảnh nhưng lại bị áp đặt vào một bối cảnh khác. Tuy vậy, cũng tồn tại khả năng các chủ thể xã hội có thể chống lại những áp lực đó và sử dụng công nghệ theo những cách bất ngờ, mang tính trao quyền. Bản thân công nghệ không quyết định sự thay đổi xã hội. Chức năng của các công cụ và hệ thống không thể tách rời khỏi bối cảnh sản sinh và triển khai chúng – chẳng hạn như vấn đề

sở hữu hay mô hình kinh doanh của các công ty công nghệ thông tin. Tương tự, các công nghệ nền tảng cũng không thể tách biệt khỏi những hệ thống quản trị đang tìm cách định hình thiết kế và cách chúng được triển khai. Cuốn sách này không xem AI chỉ đơn thuần là tập hợp các chức năng, mà còn đi sâu tìm hiểu quá trình hình thành xã hội của nó và những hệ quả mà nó mang lại.

Đặc biệt trong trường hợp của AI, điều quan trọng là phải nhận ra rằng ngành công nghiệp AI – nơi gương các công cụ tìm kiếm – đã cung cấp miễn phí hàng loạt công cụ như ChatGPT cho công chúng nhằm phục vụ các mục tiêu tìm kiếm lợi nhuận lâu dài. Có một quan điểm rất đáng lưu ý, cho rằng ngành công nghiệp AI về bản chất mang tính khai thác và bóc lột sâu sắc (Crawford, 2021). Mặc dù chúng ta có thể cho rằng việc được truy cập miễn phí vào nhiều công cụ AI là điều hiển nhiên, nhưng không được phép bỏ qua thực tế rằng nội dung tri thức và dữ liệu đã bị trích xuất để huấn luyện các hệ thống này, cùng với những tác động nghiêm trọng lên môi trường và kinh tế – vốn thường xảy ra trong bóng tối ở các quốc gia thiếu đặc quyền. Đối với một số người, những khía cạnh này bắt nguồn từ một ngành công nghiệp vốn không hề mang tính bao trùm trong thành phần nhân sự, phụ thuộc vào việc khai thác lao động bấp bênh để thực hiện các nhiệm vụ thiết yếu, và nơi mà các quyết định cuối cùng được định hướng bởi lợi nhuận chứ không phải lợi ích cộng đồng. Cuốn sách này sẽ xem xét cách mà góc nhìn đó làm thay đổi nhận thức về mức độ bao trùm, công bằng và bền vững của những lợi ích tri thức mà AI mang lại.

Bên cạnh đó, AI không phải là một công nghệ đơn lẻ hay đơn giản để hiểu hoặc định nghĩa. Nó vừa là một ý tưởng, vừa là một công nghệ cụ thể. Thuật ngữ “AI” có thể được truy về những năm 1950, nhưng khái niệm “cỗ máy biết suy nghĩ” còn xuất hiện sớm hơn nữa. Ý nghĩa của khái niệm này luôn thay đổi theo thời gian. AI thường được mô tả là công nghệ thực hiện những nhiệm vụ mà bình thường đòi hỏi trí tuệ con người, nhưng đó vẫn là một định nghĩa khá mơ hồ.

Một đặc điểm then chốt của một số, dù không phải tất cả, công nghệ AI là khả năng “học” từ dữ liệu. Điều này đồng nghĩa với việc các đầu ra mà AI tạo ra có thể vượt ngoài dự đoán, ngay cả đối với chính những người đã phát triển nó, và quá trình học này có thể bao hàm cả thiên kiến lẫn thông tin sai lệch. Khả năng học tập này cũng kéo theo một mức độ tự chủ nhất định.

AI hiện nay chủ yếu được dẫn dắt bởi dữ liệu, tức là cần huấn luyện các hệ thống AI trên một khối lượng dữ liệu khổng lồ.

Trong cách hiểu hiện tại, thuật ngữ “AI” bao gồm một loạt công nghệ đang không ngừng biến đổi và phát triển với tốc độ nhanh chóng. Một điểm khởi đầu hữu ích là định nghĩa sau đây:

... bất kì phần mềm công nghệ nào có ít nhất một trong các khả năng sau: nhận thức – bao gồm âm thanh, hình ảnh, văn bản và xúc giác (ví dụ: nhận dạng khuôn mặt), dự đoán (ví dụ: dự báo thời tiết),

ra quyết định (ví dụ: hệ thống chẩn đoán y tế), nhận dạng mẫu và tự động trích xuất kiến thức từ dữ liệu (ví dụ: phát hiện tin giả trên mạng xã hội), và lập luận logic (ví dụ: phát triển lý thuyết từ các tiền đề). Lĩnh vực mới này bao gồm nhiều phân ngành khác nhau, bao gồm học máy.

*Vinuesa và cộng sự, 2020*

Trích dẫn này hữu ích với mục đích tập trung vào những gì AI thực hiện, thay vì các công nghệ cơ bản tạo nên nó như mạng nơ-ron hoặc học máy. Định nghĩa này cũng thuận tiện trong việc nắm bắt các loại hình chức năng khác nhau mà AI hiện tại có thể thực hiện. Trích dẫn này cũng thể hiện tính đa dạng của AI. AI rất khó hiểu một phần vì nó đã trở thành một thuật ngữ chung bao hàm nhiều công nghệ khác nhau. Đặc biệt trong cuốn sách này, chúng ta sẽ tập trung vào một vài trong số những khả năng đã được đề cập trong trích dẫn. Các chức năng “nhận thức” của AI rất quan trọng đối với việc số hóa giọng nói, văn bản, hình ảnh, v.v... ngày càng tăng. Điều này có nghĩa là lời nói, văn bản hoặc bản thảo viết tay hoặc hình ảnh tĩnh và chuyển động có thể được máy tính hiểu như một dạng dữ liệu. Máy tính cũng có thể tạo ra các dạng đầu ra khác nhau, chẳng hạn như giọng nói.

Khái niệm được đề cập trong trích dẫn “nhận dạng mẫu và tự động trích xuất kiến thức từ dữ liệu” rất quan trọng để chúng ta có thể mô hình hóa ý nghĩa của dữ liệu đó và sử dụng nội dung này theo những cách khác

nhau. Chúng ta có thể gọi loại công nghệ này là AI mô tả (descriptive AI)<sup>(1)</sup>.

Tất nhiên, định nghĩa này có từ trước khi AI sinh dữ liệu ra đời, vì vậy cần phải được cập nhật để bao gồm các khả năng phức tạp của nó. Không chỉ củng cố hơn nữa sự hiểu biết của chúng ta về AI với vai trò là một công nghệ đa dạng và không ngừng phát triển, AI sinh dữ liệu như ChatGPT còn mở ra cánh cửa cho nhiều khả năng mới rất quan trọng để truy cập và tạo tri thức, chẳng hạn như tóm tắt nội dung hoặc viết bản nháp văn bản.

Trong cuốn sách này, chúng ta sẽ sử dụng thuật ngữ AI sinh dữ liệu, dù nhiều người vẫn gọi chúng là các mô hình ngôn ngữ lớn (LLM – Large Language Models). LLM là những biểu diễn toán học của ngôn ngữ dựa trên việc dự đoán xác suất của từ tiếp theo trong một chuỗi từ. Một ví dụ dễ hiểu là phiên bản sơ khai của công nghệ này – ứng dụng tự động gợi ý từ trên điện thoại giúp bạn có thể tiết kiệm thời gian khi gõ phím. AI sinh dữ liệu hiện tại dựa trên một loại mạng nơ-ron, được gọi là mô hình Transformer (tạm dịch là bộ biến đổi). Chúng được đào tạo trên một lượng lớn văn bản để có thể dự đoán văn bản thích hợp để phản hồi một yêu cầu đơn giản của người dùng.

(1) AI mô tả (*Descriptive AI*): Là một nhánh của trí tuệ nhân tạo (AI) tập trung vào việc giải thích và tóm tắt dữ liệu đã có. Mục tiêu chính của AI mô tả là trình bày những gì đã xảy ra hoặc đang xảy ra trong dữ liệu dưới dạng dễ hiểu, chẳng hạn như thông qua các báo cáo, bảng biểu, trực quan hóa hoặc các mẫu hình tóm tắt. AI mô tả chỉ nhằm mục đích mô tả thực tế của tập dữ liệu đang được phân tích.

Sự phát triển của AI trong vài năm tới có thể sẽ thay đổi đáng kể. Khi nói về AI, chúng ta thường nói về tương lai. AI hầu như luôn được xây dựng như một khái niệm về một tương lai có thể xảy ra. Cuốn sách này được đặt trong thời điểm hiện tại, khi chúng ta đang nghiên cứu những tác động của AI sinh dữ liệu, cùng với quá trình dữ liệu hóa và xu hướng ngày càng được ứng dụng nhiều của AI mô tả.

## TÓM LƯỢC

Để làm rõ và trả lời câu hỏi trung tâm mà cuốn sách đặt ra, các chương được sắp xếp như sau: Chương 2 phân tích cách AI đang làm thay đổi việc tiếp cận tri thức trong các bối cảnh đời sống hằng ngày, đồng thời chỉ ra những hạn chế và rủi ro đi kèm với những thay đổi đó. Các chương 3, 4 và 5 tập trung cụ thể hơn vào tác động của AI trong việc chuyển hóa quá trình tạo lập và sử dụng tri thức học thuật – và một lần nữa, cuốn sách cũng chú ý đến các nguy cơ và rào cản trong lĩnh vực này. Chương 6 xem xét việc AI có thể biến đổi hoạt động tiếp cận tri thức thông qua thư viện và lưu trữ như thế nào, bởi đây là những điểm truy cập tri thức then chốt. Chương 7 bàn về mô hình quản trị nào là cần thiết để bảo đảm các lợi ích mà AI mang lại là công bằng và bền vững. Chương này cũng đề cập đến các kỹ năng AI mà mỗi cá nhân cần có để có thể hoạt động hiệu quả trong một môi trường thông tin đang bị AI định hình sâu sắc.