

MỤC LỤC

Lời nói đầu	5
Lời cảm ơn	9
1 Giới thiệu: AI và cách mạng kĩ thuật số trong khoa học	11
2 AI trong việc quản lí tài liệu và bằng chứng khoa học	37
3 AI trong diễn giải dữ liệu	63
4 AI cho nghiên cứu có thể tái lập	89
5 Giới hạn của AI và các biện pháp chống thiên lệch	111
6 Kết luận: AI và tương lai của khám phá khoa học	135
Tài liệu tham khảo	153
Chỉ mục	180

1

GIỚI THIỆU

AI VÀ CÁCH MẠNG KỸ THUẬT SỐ
TRONG KHOA HỌC

Ít có ý tưởng nào trong thời đại của chúng ta lại khơi gợi trí tưởng tượng của đại chúng mạnh mẽ như trí tuệ nhân tạo: một viễn cảnh về những cỗ máy thông minh có thể hành động và suy nghĩ như con người. Những lời hứa hẹn và hiểm họa của các cỗ máy với tiềm năng trí tuệ luôn được hình dung, trình bày, tranh luận và tái định hình trên mọi kênh truyền thông và thông tin. Một số viễn cảnh trong đó hoàn toàn mang tính giả tưởng: chúng ta còn rất xa mới phải lo ngại đến chuyện những siêu robot sẽ thống trị thế giới hoặc khiến toàn bộ con người mất việc ngay trong thế hệ này. Tuy nhiên, các công nghệ trí tuệ nhân tạo – theo nghĩa thực tế (mà tôi sẽ phân tích chi tiết ngay sau đây) – vẫn đang dần thâm nhập vào nhiều lĩnh vực của đời sống con người. Gần như mỗi tuần, lại có một dòng tit lớn công bố một đột phá mới của công nghệ trí tuệ nhân tạo – từ việc chiến thắng trong các trò chơi

bài phức tạp, điều khiển xe tự lái, dự đoán cách protein gập lại cho đến việc giải thích các trò đùa, công nghệ AI dường như đang chinh phục mọi thách thức trên mọi mặt trận. Niềm hi vọng đặt vào công nghệ AI là nó sẽ cho phép tự động hóa, tăng tốc và mở rộng quy mô mọi hoạt động của con người, giúp chúng ta được giải phóng khỏi những công việc nhàm chán, lặp đi lặp lại hoặc quá khó khăn, đồng thời đẩy nhanh tốc độ tiến bộ trên mọi phương diện, đặc biệt là trong khám phá khoa học. Người ta kì vọng rằng các nhà khoa học sẽ hoàn toàn thoát khỏi cảnh vất vả và sự lặp lại, trong khi một “đội quân máy móc” sẽ hoạt động trong các phòng thí nghiệm: di chuyển và trộn vật liệu, xây dựng, điều khiển, lọc lựa và gợi ý các cách diễn giải – tất cả trong một cuộc đua ngày càng tăng tốc hướng tới những phát minh có thể làm thay đổi thế giới.

KHÁM PHÁ KHOA HỌC

KHÁM PHÁ KHOA HỌC LÀ GÌ?

Một trong những dấu hiệu hữu hình và cụ thể nhất của thời hiện đại là sự tiến bộ vững chắc của đổi mới khoa học và những tiến bộ công nghệ đi kèm. Trên nhiều lĩnh vực khác nhau, các câu hỏi khoa học được trả lời thông qua một loạt các phương pháp và thực hành phân tích đa dạng đáng kinh ngạc. Một vài khám phá tiêu biểu trong thế kỉ trước bao gồm thuyết tương đối rộng của Einstein, việc lập bản đồ bộ gen người, phát triển công nghệ vắc-xin mới để chống lại vi-rút SARS-CoV-2, những tiến bộ trong

việc tìm hiểu mối quan hệ phức tạp giữa các điều kiện kinh tế-xã hội và kết quả cuộc sống, và những dự đoán ngày càng chính xác về tác động của các hoạt động con người đối với khí hậu toàn cầu. Các khám phá khoa học củng cố sự tiến bộ của các tiêu chuẩn sống ở các nước phát triển và là tiền đề cho những hi vọng và mục tiêu phát triển toàn cầu trong tương lai.

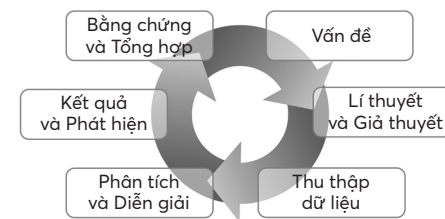
Khám phá khoa học được định nghĩa là “quá trình hoặc sản phẩm của một cuộc nghiên cứu khoa học thành công” (Schickore, 2018). Thật khó để xác định một cách chính xác “khám phá khoa học” nghĩa là gì, bởi khái niệm này có thể khác nhau về mặt lí thuyết lẫn thực tiễn, tùy thuộc vào loại câu hỏi, dạng thực thể, lĩnh vực và cách tiếp cận phương pháp luận khác nhau. Các chuyên ngành khác nhau, như vật lí, tâm lí học, hóa học, sinh học và y học, đã phát triển vô số phương pháp và cách thức để khám phá một cách chắc chắn, đáng tin cậy và hiệu quả hơn về mọi khía cạnh của thế giới nằm trong phạm vi quan tâm và ứng dụng của từng lĩnh vực.

Một khía cạnh của khám phá khoa học là tính mới. Từ *khám phá* đã bao gồm việc tìm ra điều gì đó chưa được biết đến trước đây, hoặc tạo ra một thứ gì đó thuộc loại chưa từng tồn tại trước đây. Một số ví dụ bao gồm cách sản xuất vật liệu mới cứng hơn và nhẹ hơn (Zeng và cộng sự, 2022), khám phá các loại vật chất kì lạ mới (Castelvecchi, 2021) và khám phá bằng chứng cho thấy con người đã sử dụng thuốc lá 12.300 năm trước, sớm hơn 9.000 năm so với thời điểm từng được cho là dấu vết sớm nhất của việc sử dụng thuốc lá (Duke và cộng sự, 2022).

Một khía cạnh khác của khám phá khoa học là *tính khoa học*. Dù những sắc thái chính xác của khái niệm này hay cách định nghĩa nó vẫn là chủ đề của nhiều cuộc tranh luận sâu rộng (Kampourakis, 2016), một định nghĩa từ điển điển hình lại đưa ra một cách hiểu có phần vòng vo: đó là “dựa trên hoặc mang đặc trưng là các phương pháp và nguyên lí của khoa học.” Điều này gợi ý rằng, để hiểu rõ hơn “khoa học” nghĩa là gì, chúng ta nên bắt đầu từ việc tìm hiểu phương pháp khoa học là gì và các nguyên tắc khoa học là gì.

Phương pháp khoa học có thể được mô tả như một quá trình chủ động dựa trên luồng thông tin, trong đó bằng chứng làm nảy sinh các vấn đề, từ đó dẫn đến các lí thuyết và giả thuyết. Những giả thuyết này sau đó được kiểm tra thông qua các thí nghiệm, từ đó tạo ra dữ liệu, rồi dữ liệu lại được diễn giải để rút ra phát hiện mới. Các phát hiện này tiếp tục tạo ra bằng chứng mới và từ đó gợi mở những vấn đề tiếp theo (Hình 1.1).

Trên thực tế, việc lặp lại các bước có thể diễn ra ở bất kì giai đoạn nào, và có rất nhiều biến thể khác nhau của



Hình 1.1 Phương pháp khoa học được khái niệm hóa như một quá trình định hướng theo chu kì.

quy trình khám phá khoa học, tùy thuộc vào bản chất cụ thể của câu hỏi mà khám phá khoa học đó đang theo đuổi. Ngoài ra, yếu tố tình cờ hay may mắn cũng đã được chứng minh là có vai trò nhất định (Gaughan, 2010). Khám phá khoa học tiến bộ một cách tích lũy, thông qua tổng hòa các hoạt động nghiên cứu khoa học, bao gồm nhiều quy trình có chủ đích như tổng hợp, kiểm nghiệm và báo cáo, với mục tiêu tạo ra những hiểu biết mới và tích lũy tri thức về thế giới và mọi thứ trong đó. Hơn nữa, nghiên cứu khoa học chiếm một phần đáng kể trong nền kinh tế của các quốc gia phát triển (Macilwain, 2010), nhờ vào nguồn đầu tư từ cả khu vực chính phủ lẫn phi chính phủ.

CÔNG NGHỆ VÀ SỐ HÓA TRONG KHÁM PHÁ KHOA HỌC

Công nghệ đóng vai trò quan trọng trong việc thúc đẩy nghiên cứu khoa học theo nhiều cách khác nhau. Trong nhiều lĩnh vực, loại câu hỏi mà các nhà khoa học đặt ra thường chỉ có thể được trả lời thông qua việc phát triển những công nghệ có khả năng thu thập được dữ liệu cần thiết để tìm ra lời giải. Ví dụ, công nghệ thường đóng vai trò trung gian trong đo lường và quan sát: Chúng ta không thể quan sát trực tiếp vi-rút bằng mắt thường, nhưng có thể nhìn thấy chúng qua kính hiển vi mạnh; cũng như chúng ta không thể nhìn thấy bức xạ tia gamma được tạo ra từ các vụ va chạm khổng lồ của những thiên thể xa xôi, nhưng có thể ghi lại chúng bằng kính thiên văn công suất lớn. Công nghệ đôi khi còn

cho phép chúng ta thực hiện những thí nghiệm mà nếu không có nó thì không thể làm được. Chẳng hạn, để kiểm nghiệm dự đoán trong thuyết tương đối hẹp của Einstein rằng thời gian sẽ nhanh lên hoặc chậm đi dưới tác động của gia tốc, người ta đã cho các đồng hồ nguyên tử bay vòng quanh thế giới ở tốc độ cao, và quả thực chúng đã chạy chậm lại hoặc nhanh lên đúng như lí thuyết đã dự báo (Hafele và Keating, 1972). Thông thường, những tiến bộ công nghệ còn giúp cho việc nghiên cứu của chúng ta trở nên quy mô hơn và toàn diện hơn so với trước đây.

Nghiên cứu khoa học, cùng với nhiều lĩnh vực khác, đang được số hóa ngày càng nhiều, tức là được hỗ trợ và thực hiện thông qua các công nghệ kĩ thuật số thông tin. Quá trình chuyển dịch này đang trở thành phương thức làm việc phổ biến trong thế giới hiện đại, gọi là *chuyển đổi kĩ thuật số*. Chuyển đổi kĩ thuật số đang được ca ngợi như một sự thay đổi mô hình trong cách nghiên cứu được tiến hành trên toàn bộ các lĩnh vực khoa học khác nhau – chẳng hạn như trong việc phát triển vật liệu mới (Kimmig, Zechel và Schubert, 2021), trong y học (Berisha và cộng sự, 2021), bao gồm cả sức khỏe tâm thần (Bucci, Schwannauer và Berry, 2019; Hastings, 2020), và trong khoa học hành vi (Hesse, 2020). Tất nhiên, chuyển đổi kĩ thuật số cũng đang ảnh hưởng đến thế giới bên ngoài nghiên cứu khoa học. Ở các nước phát triển, cuộc sống và công việc của chúng ta ngày càng được số hóa. Tốc độ phát triển công nghệ thông tin và tác động phổ biến của nó đối với các phương thức làm việc của chúng ta trên các lĩnh vực và khía cạnh khác nhau của cuộc sống đã rất ngoạn mục, một số người thậm chí coi nó ngang tầm với

cuộc cách mạng công nghiệp – một sự thay đổi toàn diện trong cách chúng ta sống, làm việc và kết nối với nhau.

Chuyển đổi kỹ thuật số có thể mang nghĩa là “*nhiều hơn*” theo nhiều cách. Ví dụ, chúng ta tiêu thụ nhiều thông tin hơn, nhanh hơn, với phạm vi địa lý rộng hơn. Khoa học cũng không phải là ngoại lệ. Đối với nghiên cứu khoa học, công nghệ kỹ thuật số cũng có nghĩa là “*nhiều hơn*”: nhiều dự án hơn với các nhóm lớn hơn, các thí nghiệm lớn hơn dẫn đến nhiều dữ liệu hơn, và cuối cùng là nhiều kết quả được công bố hơn trong nhiều ấn phẩm khoa học với sự đa dạng ngày càng tăng. Công nghệ thông tin đang thâm nhập sâu rộng, làm thay đổi cách đo lường khoa học, phân tích và diễn giải dữ liệu, cách công bố các phát hiện khoa học, cũng như quá trình tìm kiếm và tổng hợp các nghiên cứu hiện có.

Và trong số tất cả các công nghệ số, một trong những công nghệ có tiềm năng lớn nhất trong việc tác động đến cách chúng ta tiến hành nghiên cứu khoa học chính là trí tuệ nhân tạo: lời hứa về những hệ thống tự động có khả năng hỗ trợ quá trình điều tra khoa học bằng cách thực hiện các nhiệm vụ ở mức độ tương đương với con người.

TRÍ TUỆ NHÂN TẠO

TRÍ TUỆ NHÂN TẠO LÀ GÌ?

Bất kỳ công nghệ nào dù tiên tiến cũng chẳng khác gì phép thuật.

(CLARKE, 1968)

Trí tuệ nhân tạo trong mắt công chúng chắc hẳn mang theo một hào quang kì diệu, đúng nghĩa được phản ánh trong “định luật thứ ba” nổi tiếng của Arthur C. Clarke mà tôi vừa trích dẫn ở trên. Clarke đã viết về sự khó khăn trong việc dự đoán điều gì là khả thi và điều gì là bất khả thi, cũng như cách mà các công nghệ của tương lai thường vượt qua những ranh giới do trí tưởng tượng của ngay cả những nhà khoa học xuất sắc nhất đương thời đặt ra, khiến cho các đột phá công nghệ trông như phép màu khi được nhìn lại từ lăng kính của bất kì thời đại nào đã trôi qua. Tuy nhiên, cả trong trường hợp công nghệ nói chung và trí tuệ nhân tạo nói riêng, đây có thể là một con dao hai lưỡi. Đôi khi, những dự đoán của chúng ta về tương lai thực sự mang màu sắc huyền diệu, theo nghĩa là các công nghệ mà chúng đòi hỏi không chỉ chưa tồn tại, mà về nguyên lý còn không thể tồn tại được. Ngoài ra, cũng có một nét nghĩa ngược lại của điều này, đó là chúng ta thường không coi một thứ mà chúng ta hiểu cách hoạt động là một điều huyền diệu, mà chỉ là một công nghệ thông thường. Ý tưởng về trí tuệ nhân tạo vốn dĩ đã mang màu sắc huyền diệu, nhưng lại đi kèm một hệ quả: Chúng ta thường không coi các hệ thống số tự động là “trí tuệ nhân tạo” nếu đã hiểu rõ cách chúng vận hành. Nguy cơ ở đây là, mỗi đổi mới trong công nghệ số nhằm hỗ trợ suy luận, học tập và hành động tự động thì ban đầu thường được nhìn nhận như trí tuệ nhân tạo. Nhưng rồi khi trở nên phổ biến và được chấp nhận rộng rãi, nó lại bị đẩy lùi vào hậu trường và chỉ được gọi là hệ thống hoạt động theo quy tắc, suy luận hay học máy, tùy theo cách thức hoạt động. Chính vì vậy, cái nhãn công nghệ

“trí tuệ nhân tạo” luôn duy trì được sự bí ẩn, nhưng phải đánh đổi bằng việc trở thành một khái niệm thay đổi liên tục và khó nắm bắt. Chúng ta cần cảnh giác với xu hướng này và ghi nhớ rằng không phải công nghệ số nào mà ta chưa hiểu cách vận hành cũng đều có trí thông minh, và ngược lại, những gì ta đã hiểu rõ chưa chắc đã không thông minh. Khi xem xét ý nghĩa thực tiễn của “trí tuệ nhân tạo”, chúng ta cũng nên lưu ý đến mức độ mà các công nghệ số thực sự là “nhân tạo” hay không, theo nghĩa rằng chúng vẫn được vận hành nhờ vào các dữ liệu đầu vào và thuật toán do con người cẩn thận tạo ra. Cuối cùng, chúng ta phải nhớ rằng, bản thân khái niệm “trí khôn/trí thông minh” vốn đã là một ý niệm mơ hồ và luôn biến đổi khó xác định, ngay cả ở con người.

Việc thế nào là “có trí thông minh”, và trí thông minh liên quan như thế nào đến các năng lực và khả năng khác của con người, đã là chủ đề tranh luận và bất đồng suốt thế kỉ qua (Sternberg, 2019). Một số nhà nghiên cứu cho rằng chỉ tồn tại một chiều kích trí tuệ chung áp dụng cho mọi lĩnh vực và nhiệm vụ, trong khi những người khác lại xem các lĩnh vực, kĩ năng và năng lực là những chiều kích riêng biệt. Không thể nào lập ra một danh sách đầy đủ về những yếu tố cấu thành trí thông minh, nhưng nó thường được cho là bao gồm ít nhất các năng lực cốt lõi của con người như: hiểu thế giới và các bối cảnh cụ thể; giao tiếp bằng ngôn ngữ và kí hiệu; suy luận và rút ra kết luận, lên kế hoạch và thực hiện các hành động hiệu quả để đạt được mục tiêu; tự nhận thức; động lực; khả năng học hỏi; tiếp nhận thông tin và xây dựng các mô hình

cũng như sự khái quát hóa có thể được thao tác một cách logic thông qua tư duy phản biện, sáng tạo và giải quyết vấn đề. Nói rộng hơn, ta có thể mô tả trí thông minh là năng lực tổng quát để thích ứng và hành động phù hợp trong một môi trường hoặc bối cảnh cụ thể, ngay cả khi bối cảnh đó chưa từng được gặp trước đây. Và đối với khám phá khoa học, điều quan trọng là không được bỏ qua vai trò của sự sáng tạo – nhiều phát hiện mới là kết quả trực tiếp của việc suy nghĩ vượt ra ngoài những giới hạn quen thuộc, kết nối những yếu tố trước đó vốn chưa từng liên quan.

Trí tuệ nhân tạo có thể được định nghĩa theo nhiều cách khác nhau (Russell, Norvig và Davis, 2010; Marcus, 2020; Landgrebe và Smith, 2022), tùy thuộc vào mức độ và việc ta đang coi trọng chiều kích nào hơn: (1) việc máy móc có thể làm được gì hay (2) cách thức chúng làm điều đó. Định nghĩa phổ biến nhất hiện nay thường đề cập đến khả năng *thực hiện* những nhiệm vụ mà con người phải cần đến trí thông minh mới có thể làm được. “Phép thử Turing”, do nhà toán học người Anh và cũng là một trong những người tiên phong trong lĩnh vực trí tuệ nhân tạo là Alan Turing đề xuất, là một ví dụ cho định nghĩa này về một hệ thống trí tuệ nhân tạo. Một cỗ máy được coi là vượt qua phép thử Turing nếu con người không thể phân biệt được câu trả lời của cỗ máy với của một người khác, khi cả hai được đưa cho cùng một câu hỏi. Về mặt trực giác, điều này có vẻ hợp lí – còn gì thể hiện trí tuệ rõ hơn những hành vi, ví dụ như các câu trả lời đối thoại, không thể phân biệt được với con người? Tuy nhiên, cũng có thể