

# SỰ GIA TĂNG CỦA ỨNG DỤNG **AI** TẠO SINH: **NHỮNG RỦI RO TIỀM ẨN CHO XÃ HỘI VÀ CON NGƯỜI**



HOÀNG TÙNG DƯƠNG, ĐẶNG TUẤN DŨNG, TS. HỒ MẠNH TÙNG  
*Viện Triết học, Viện Hàn lâm Khoa học Xã hội Việt Nam*

AI tạo sinh là một trong những thành tựu công nghệ mới nhất của con người trong thập niên 20 của thế kỷ XXI. Cho đến nay, sự ứng dụng của AI tạo sinh đã tạo ra nhiều cuộc tranh luận quan trọng trong các nghiên cứu xã hội, đặc biệt là trong lĩnh vực triết học. AI tạo sinh đã thách thức nhiều khái niệm và định kiến của chúng ta về bản thân mình, đặc biệt là về cách chúng ta hiểu về tư duy và bản chất của tư duy con người. Trong bài viết này, chúng tôi sẽ tổng hợp và tập trung làm rõ một số vấn đề đã được đưa ra về rủi ro tiềm ẩn cho xã hội và con người khi AI tạo sinh được ứng dụng nhiều hơn vào trong cuộc sống. Trong đó, hai vấn đề lớn nhất được chúng tôi đưa ra là AI tạo sinh sẽ giải quyết thách thức về bất công xã hội như thế nào khi nó góp phần gia tăng những định kiến ngấm trong xã hội, cũng như khiến cho những nhóm người với đặc trưng xã hội thiểu số bị bỏ lại phía sau và sự bất bình đẳng giữa các ngôn ngữ. Vấn đề thứ hai là quá trình phổ cập khoa học và chính sách khi AI tạo sinh có thể sử dụng nguồn thông tin không được xác định và làm cho các cuộc tranh luận trở nên phức tạp.

**Tóm tắt:**

- AI tạo sinh là công nghệ mới nhất với khả năng tạo ra nội dung mới, được ứng dụng rộng rãi nhưng gây tranh cãi.
- Rủi ro về bất công xã hội:
  - + Định kiến ngầm: AI tạo sinh có thể làm gia tăng các định kiến xã hội hiện có, gây ra bất công trong các mô hình ngôn ngữ lớn (LLMs).
  - + Nhóm thiểu số bị bỏ lại: Những nhóm người với di sản văn hóa ít phổ biến có thể bị AI tạo sinh bỏ qua, dẫn đến mất mát văn hóa.
  - + Bất bình đẳng ngôn ngữ: AI tạo sinh ưu tiên ngôn ngữ phổ biến, tạo ra sự bất bình đẳng về chất lượng kết quả giữa các ngôn ngữ.
- Phổ cập khoa học và chính sách:
  - + Tri thức khoa học: AI tạo sinh có thể phổ biến khái niệm khoa học nhưng dễ dẫn đến thông tin sai lệch, ảnh hưởng đến nhận thức xã hội.
  - + Truyền thông chính sách: Sự phát triển của AI tạo sinh có thể làm phức tạp quá trình truyền thông chính sách, dễ bị thao túng và gây rối loạn thông tin.
- Kết luận: AI tạo sinh sẽ có ảnh hưởng sâu rộng đến đời sống xã hội, đòi hỏi phải có sự đánh giá chính xác và phát triển các quy tắc ứng xử và giá trị đạo đức mới.





## Mở đầu

AI tạo sinh (Generative AI) là một công cụ trí tuệ nhân tạo mới xuất hiện trong thời gian gần đây. Điểm khác biệt của công cụ này với các công cụ AI trước đó là nó có khả năng sáng tạo nên những nội dung mới (bao gồm cả chữ viết, hình ảnh, âm thanh). Sự xuất hiện của AI tạo sinh đã được chào đón nồng nhiệt trong nhiều ngành nghề, trong đó dễ thấy nhất là giáo dục và truyền thông bởi đây là nơi mà khả năng trả lời câu hỏi và sáng tạo nội dung của AI tạo sinh có thể được sử dụng một cách hữu hiệu nhất. Báo chí cũng là một lĩnh vực mà AI tạo sinh đang dần được chấp nhận và sử dụng rộng rãi. Một ví dụ cho việc này là ngay tại Việt Nam, các công cụ AI tạo sinh đã được một số tòa soạn báo đưa vào sử dụng khá sớm [1].

Tuy nhiên, không phải lúc nào việc sử dụng AI tạo sinh cũng đem lại kết quả tốt. Sự lạm dụng AI tạo sinh có thể kéo theo nhiều hệ lụy mà đến nay chúng ta chưa thể biết được, nhất là khi chúng được sử dụng trong những lĩnh vực yêu cầu độ chính xác cao như trong lĩnh vực y tế, điều này đã được chỉ ra bởi Siontis và các cộng sự [2]. Ngay cả trong giáo dục, một trong những khuyết điểm của AI tạo sinh là việc chúng có xu hướng tự chế tạo ra những thông tin không có thật (được các chuyên gia trong ngành gọi là “ảo giác AI” [3]) cũng sẽ ảnh hưởng đến việc sử dụng các công cụ AI tạo sinh này. Điều này khiến nhiều người đề xuất ra những giải pháp để ứng dụng AI tạo sinh vào công tác giáo dục phù hợp hơn. Một trong những giải pháp đáng chú ý gần đây là của Kasneci và các cộng sự, họ cho rằng việc sử dụng AI tạo sinh trong giáo dục cần phải đi đôi với chiến lược sư phạm rõ ràng cùng với điều kiện là người dạy và người học phải có sự hiểu biết về AI tạo sinh nếu muốn chúng thực sự đem lại hiệu quả [4].

Trong khuôn khổ bài viết này, nhóm tác giả sẽ tổng hợp và trình bày một số vấn đề mà xã hội sẽ gặp phải khi mà AI tạo sinh ngày càng len lỏi vào cuộc sống con người. Bài viết sẽ tổng hợp và

phân tích các rủi ro tiềm ẩn của AI tạo sinh dựa trên cơ sở của một tiểu luận gần đây tương phản nỗi lo về sự căn cổ hóa tâm hồn và sự phong phú ngôn ngữ của AI tạo sinh, được công bố trên tạp chí khoa học AI & Society của NXB Springer Nature [5]. Cụ thể là bài viết tập trung vào rủi ro tiềm ẩn đối với công bằng xã hội và vấn đề phổ cập tri thức khoa học cũng như truyền thông chính sách.

## AI tạo sinh sẽ giải quyết thách thức về bất công xã hội ra sao?

Trước hết, việc sử dụng AI tạo sinh có thể tạo ra những vấn đề nằm trong phạm vi của sự bất công xã hội [6]. Điều này có thể quan sát được trong thực tế qua hai điểm sau đây.

### *Định kiến ngầm*

Thứ nhất, bất công xã hội có thể đến từ việc ưu tiên lựa chọn nguồn thông tin đang phổ biến hơn trong xã hội. Trong thực tế thì xã hội không phải chỉ có những thông tin mà tất cả mọi người đều đồng ý là nó đúng, luôn tồn tại nhiều loại định kiến, khuôn mẫu và thành kiến sẵn có đã được chuyển hóa thành những sơ đồ tiên nghiệm mà ta gọi là “văn hóa”, chúng là thứ kiến tạo nên những quan điểm và góc nhìn của chúng ta về các vấn đề xã hội một cách định kiến.

Trước tiên là vấn đề về định kiến ngầm trong các mô hình ngôn ngữ lớn và các AI tạo sinh khác. Hiện nay, mô hình ngôn ngữ lớn (Large Language Models (LLMs)) – một sản phẩm cốt lõi từ AI tạo sinh – đã được lập trình có chủ đích để tránh những vi phạm trắng trợn những quy tắc bất thành văn trong văn hóa này. Trong một nghiên cứu vào đầu năm 2024, nhóm tác giả Hofmann và cộng sự đã cho thấy bằng chứng sự phân biệt chủng tộc ngầm của LLMs đối với phương ngữ của người Mỹ gốc Phi [7]. Trong thí nghiệm này, nhóm tác giả đã cho AI chatbot đánh giá năng lực của người được tuyển dụng và hành vi phạm tội dựa trên văn bản có chứa tiếng địa phương. Và



kết quả là tiếng địa phương của người Mỹ gốc Phi bị đánh giá kém hơn trong việc tuyển dụng cũng như trong phạm tội. Hiển nhiên, các từ “lóng” xuất hiện trên không gian mạng, hay các câu nói mang ẩn ý, các ngôn ngữ cơ thể đặc trưng của một số nhóm người sẽ khó trở thành khuyến hướng xác suất chủ đạo của công cụ AI tạo sinh. Hơn nữa, ngôn ngữ luôn thay đổi và nhiều nhận thức về xã hội chỉ tồn tại ở chiều kích ngầm định, hoặc có nguồn thông tin về chúng chỉ tồn tại với số lượng nhỏ nên huấn luyện thế nào để LLMs “hiểu” về những cảm nhận xã hội và đạo đức luôn thay đổi này là một bài toán khó.

#### ***Những nhóm người sẽ bị bỏ lại phía sau?***

Thứ hai, trong quá trình đưa hệ thống trí tuệ nhân tạo vào xã hội, luôn có một số nhóm cảm thấy bị bỏ lại phía sau và quyền lợi của họ bị lãng quên. Đó là các nhóm người có những kiến thức bản địa của mình chỉ được lưu truyền dưới

dạng truyền miệng, hoặc có di sản văn hóa gắn liền với những kỹ năng ít phổ biến, ít được lưu truyền dưới dạng số hóa, văn bản hóa. Do đó các kiến thức về văn hóa này của họ sẽ ít được các công cụ AI tạo sinh tiếp cận đến để làm nguồn cho các câu trả lời của mình, gây nên việc những nhóm dân số nhỏ có nguy cơ bị thu nhỏ hơn nữa về mặt văn hóa. Những lo ngại về việc các công cụ AI có thể được sử dụng nhằm tạo ra lợi nhuận cho các tập đoàn công nghệ [8], hay tệ hơn nữa là dẫn đến việc thuộc địa hóa các nền văn hóa yếu hơn và làm các nền văn hóa này phụ thuộc vào phương Tây đã được nêu lên trước đây [9]. Với sự xuất hiện của các công cụ AI tạo sinh, những lo ngại này có thể sẽ tiếp tục xuất hiện ở một mức độ mạnh hơn nữa.

#### ***Bất bình đẳng giữa các ngôn ngữ***

Một rủi ro bất công xã hội tiềm ẩn đến từ hiệu ứng ưu tiên một ngôn ngữ với độ sử dụng phổ

biến hơn. Hiện tượng này thường xuất hiện khi ta nói đến những ngôn ngữ có độ phổ biến cao hơn như tiếng Anh, tiếng Trung trong mối tương quan với những ngôn ngữ khác. Có thể thấy rằng các phiên bản của các AI tạo sinh hoạt động ở các ngôn ngữ phổ biến hơn sẽ đưa ra các kết quả tốt hơn các phiên bản của bản thân chúng nhưng được hoạt động trên các ngôn ngữ ít phổ biến hơn, điều này thể hiện ở việc chúng sẽ đưa ra các kết quả chi tiết hơn, rõ ràng hơn.

Ta có thể lấy ví dụ cho vấn đề này qua ưu thế của tiếng Anh: tiếng Anh là một ngôn ngữ phổ biến, được sử dụng bởi nhiều quốc gia trên thế giới và nhiều chuyên gia hàng đầu từ các nền văn hóa đa dạng làm việc bằng tiếng Anh. Tiếng Anh thậm chí đã trở thành một *lingua franca* - một thứ ngôn ngữ chung trong các nghiên cứu khoa học. Như vậy, có thể suy ra rằng thành quả của việc tích hợp các công cụ AI tạo sinh vào đời sống, dù mang hiệu quả cao, sẽ có thể càng khuếch đại ưu thế có sẵn cho những tầng lớp dùng được, hoặc có điều kiện để tiếp cận ngoại ngữ, nhất là các ngôn ngữ “mạnh” như tiếng Anh, tiếng Trung. Khác biệt về ưu thế này có thể dẫn tới phân hóa thu nhập giữa các nhóm nhanh chóng: nhóm có ưu thế ban đầu sẽ được hưởng lợi nhiều và nhanh hơn các nhóm yếu thế hơn và từ đó khoảng cách ngày càng xa dần (hiệu ứng runaway). Đây là một vấn đề cần tính tới để đảm bảo công cụ AI tạo sinh không làm xấu đi tình hình sự phân hóa xã hội, vốn đang là vấn đề quan trọng rất được các cấp có thẩm quyền ở nước ta quan tâm và tìm giải pháp kiểm soát [10, 11].

Ví dụ như tại Việt Nam, báo chí đã ghi nhận việc có khả năng sử dụng tiếng Anh vốn đã là một lợi thế cạnh tranh lớn trong thị trường lao động, mở rộng khả năng lựa chọn công việc, tạo điều kiện cho việc nâng cao thu nhập [12], thì giờ đây với sự mở rộng của công cụ AI tạo sinh, lợi thế cạnh tranh này có khả năng ngày càng được nới rộng khi mà những người thành thạo tiếng Anh có thể tìm thấy các câu trả lời cho các vấn đề

họ gặp phải thông qua các công cụ này một cách toàn diện hơn, nhanh hơn so với những người chỉ có thể sử dụng tiếng Việt trong công việc.

Mặt khác, cũng có thể bàn rộng hơn rằng vấn đề này thực chất là một phần của việc làm thế nào để phát triển và bảo vệ tiếng Việt trong sự phát triển của các công cụ AI tạo sinh? Các công cụ AI tạo sinh bằng tiếng Việt hiện nay là không đủ mạnh so với các phiên bản AI tạo sinh của các thứ tiếng phổ biến trên thế giới. Điều này sẽ làm người sử dụng tiếng Việt tiếp tục mất đi lợi thế so với các ngôn ngữ lớn trên thế giới không chỉ trong cạnh tranh việc làm như vừa trình bày ở trên, mà còn trong tuyên truyền văn hóa, chính sách, học tập, nghiên cứu khoa học.

Thực tế thì việc huấn luyện LLMs bằng các ngôn ngữ bản địa, dù có nhiều thách thức, nhưng đã có nhiều quốc gia trên thế giới thực hiện (như Trung Quốc, Nhật Bản, v.v... [13, 14]). Bên cạnh việc cạnh tranh về kinh tế và công nghệ, huấn luyện LLMs bằng ngôn ngữ bản địa đem tới hai lợi ích có thể nhìn thấy được. Thứ nhất là AI tạo sinh được huấn luyện trên ngôn ngữ bản địa sẽ có thể được tinh chỉnh để chú trọng hơn tới các khía cạnh văn hóa đặc trưng của nước chủ thể qua đó phục vụ tốt hơn tới công việc truyền thông cũng như kinh doanh tại nước đó. Ví dụ là một AI trợ lý ảo dùng trong công sở của Nhật Bản cần tuân thủ các chuẩn mực về ngôn ngữ như một nhân viên tại nước này [15]. Lợi ích thứ hai là ý nghĩa quan trọng, lâu dài đối với vấn đề an ninh thông tin, không chỉ ở do tính bảo mật của các phiên bản AI tạo sinh hiện tại vẫn còn là một vấn đề nan giải [16, 17].

## **AI tạo sinh và quá trình phổ cập khoa học, chính sách**

Việc sử dụng các công cụ AI tạo sinh một cách đại trà cũng gây nên một số lo ngại trong việc phổ cập các khái niệm khoa học và chính sách.

### ***Vấn đề phổ cập các tri thức khoa học***

Thứ nhất là trong việc phổ cập các kiến thức, khái niệm khoa học. Các công cụ AI tạo sinh



sử dụng LLMs sẽ giúp phổ biến các khái niệm khoa học, nhưng bên cạnh đó là sự phát sinh câu chuyện về sự tôn trọng những nội hàm, ý nghĩa của các khái niệm đó. Đây là một nguyên tắc quan trọng, nền tảng trong đạo đức nghiên cứu khoa học. Quả thực, LLMs có thể trở thành một công cụ để giúp các nhà nghiên cứu ở các nước đang phát triển chỉnh sửa và cải thiện đáng kể các bản thảo của họ, qua đó tiệm cận trình độ nghiên cứu khoa học và truyền thông khoa học ở trình độ thế giới. Và đây chính là đóng góp tích cực cho cộng đồng khoa học.

Tuy nhiên thì những tác động của AI tạo sinh đối với khoa học và thực hành khoa học còn chưa rõ ràng, đặc biệt là trong hoạt động phổ biến các khái niệm khoa học xã hội và nhân văn. Trong những lĩnh vực này, sức mạnh và giá trị của một khái niệm khoa học tới từ những nỗ lực trí tuệ của nhiều nhà tư tưởng, thậm chí nhiều thế hệ nhà tư tưởng. Sự phát triển của tri thức khoa học thường dẫn đến sự chỉnh sửa, bổ sung, lược bỏ cái chưa đúng, hoàn thiện cái đã đúng bằng những bằng chứng mới, hoặc đơn giản là mở rộng nội hàm của khái niệm đó ra cho nhiều trường hợp hơn. Bất kể hoạt động ấy là gì thì nó cũng đã làm cho khái niệm ấy có nhiều nội hàm khác nhau, tùy theo những giai đoạn khác nhau trong lịch sử.

Điều lo ngại với AI tạo sinh là, chúng có thể sản xuất số lượng từ lớn một cách nhanh chóng, nhưng chưa có gì đảm bảo chúng có thể truyền đạt các khái niệm, tri thức, hay cách hiểu mới từ cộng đồng khoa học một cách công bằng, sát nghĩa, và đảm bảo được các ngụ ý. Với sự phát triển ồ ạt của ngành công nghiệp nội dung (content industry), vấn đề nguy khoa học, thông tin khoa học sai lệch, thiếu chính xác là một rủi ro lớn đối với nhiều mặt của cuộc sống [18] trong đó là truyền thông chính sách về kinh tế, xã hội, giáo dục, v.v....

Thực chất thì điều này đến từ hai hiện tượng đã tồn tại trong xuyên suốt lịch sử phát triển của tư

duy con người: tính độc quyền và tính tương đối của tri thức, đặc biệt là tri thức khoa học. Tình trạng độc quyền của tri thức là một xu hướng tất yếu trên nhu cầu phát triển xã hội của con người. Sự phát triển của phân công lao động xã hội đã tạo nên yêu cầu cấp thiết phải xây dựng một nhóm đối tượng với nhiệm vụ chuyên môn là tìm hiểu về thế giới một cách khoa học, từ đó đưa ra những tri thức cần thiết cho sự phát triển của xã hội. Hiện tượng này được gọi là “tính độc quyền của tri thức”. Tính độc quyền ấy thể hiện ở chỗ, do đặc thù của lĩnh vực nghiên cứu khoa học, nên chỉ có những nhà khoa học với chuyên môn nhất định thì mới có thể hiểu được bản chất của vấn đề mà họ đang nói đến. Điều này khiến cho bất cứ một thực thể nào nhân danh khoa học để áp đặt các tuyên ngôn của mình chỉ bởi vì họ có thể biết những điều mà người thường không được biết. Và bởi vì nhu cầu của sự tìm hiểu về khoa học thì không bao giờ ngừng, nên người ta dễ dàng nghe theo bất cứ một cá nhân nào “trông có vẻ là người hiểu chuyện” mà chưa từng kiểm tra xem điều người đó nói có đúng không [19]. Nếu như ngày xưa các nhà khoa học và trí thức là những thực thể có thẩm quyền ấy, thì ngày nay chúng ta có mạng Internet, Google và giờ là AI tạo sinh, những gì chúng ta đã nói thực chất là khi tính độc quyền của tri thức đã trở nên đa diện hơn với sự xuất hiện của AI tạo sinh.

Tính tương đối của tri thức cũng là một vấn đề phát sinh từ vấn đề tính độc quyền của tri thức. Nhìn chung thì hầu như các tri thức khoa học đều mang một ý nghĩa tương đối. Tương đối ở đây là khi nó được áp dụng cho một phạm vi nhất định, với một đối tượng nhất định, vậy nên trong phạm vi của mình thì những tri thức ấy là đúng đắn, và nếu nằm ngoài phạm vi đó thì nó sẽ không còn đúng nữa. Và nhìn chung thì cách truyền tải tri thức của xã hội đều hoạt động dựa trên nguyên tắc này. Các tri thức phổ thông, dễ học mà ta thường được biết trong sách giáo khoa không phải là những tri thức đúng nhất, nhưng

nó là những tri thức phù hợp với phạm vi sinh tồn và mức độ cần thiết của chúng ta trong hoàn cảnh sống cụ thể, khi mà đại đa số người dân không phải là nhà khoa học cần phải nắm tất thảy mọi kiến thức đó.

Tương tự như vậy, những tri thức với mức độ phức tạp hơn, cần đến sự sâu sắc hơn thì chủ yếu chỉ nằm ở trong phạm vi các nhà khoa học, bởi những tri thức này đều có những cấu trúc logic của riêng nó mà không phải ai cũng hiểu được nếu không có nền tảng khoa học chuyên môn. Điều này thể hiện rõ tính độc quyền trong tri thức như chúng ta đã thấy bên trên [19]. Nhưng bởi vì cách tiếp cận tri thức của đại đa số chúng ta đều phù hợp với hoàn cảnh sống của chúng ta, nó sẽ dẫn đến một thiên kiến xác nhận rằng những tri thức mà ta có là đúng nhất, mà không nhận ra rằng nó chỉ đúng bởi nó đã được chọn lọc để đúng với nhu cầu của chúng ta trước đó. Tương tự với trường hợp tính độc quyền tri thức, LLMs sẽ củng cố những nhận thức có sẵn của con người và, với sự bổ sung về thẩm quyền của LLMs như một “nhà khoa học” mà chúng ta đã thấy bên trên, LLMs sẽ làm cho các luồng kiến thức với nền tảng khác nhau trực tiếp đối đầu nhau.

### ***Về công tác truyền thông chính sách***

Song song với những rủi ro đến từ việc phổ cập các khái niệm khoa học một cách sai lầm trong công việc và trong giáo dục thì một mối nguy khác là việc các công cụ AI tạo sinh có thể gây ra các tác động tiêu cực đến công tác truyền thông chính sách của các Chính phủ. Việc truyền thông chính sách cho các công dân hiểu được các chính sách này đóng một vai trò quan trọng trong việc thực thi chúng của các quốc gia: điều này sẽ giúp đưa thông tin về chính sách đến người dân, lấy ý kiến người dân về chính sách, tạo ra sự đồng thuận của người dân đối với chính sách được ban hành. Tuy nhiên, quá trình này có thể bị thao túng bởi các thế lực chính trị đối lập [20], nhất là trong điều kiện các công cụ AI tạo sinh hiện

đang có thể tạo ra các thông điệp tuyên truyền tương đối thuyết phục [21]. Hơn nữa, với việc ngày càng nhiều người sử dụng trực tiếp hoặc gián tiếp thông tin được đưa ra bởi các công cụ AI, việc tuyên truyền các chính sách này hoàn toàn có thể bị cản trở, do AI tạo sinh có khả năng đưa ra các thông tin sai lầm, ảo giác như đã được bàn ở trên.

Có thể xảy ra những trường hợp như những người có ảnh hưởng trên mạng xã hội có thể sử dụng LLMs để tạo ra nội dung nghe có vẻ khoa học, gây ra sự nhầm lẫn và những tác hại đối với những chính sách hay định hướng mới. Ở đây, rủi ro thứ nhất là về sự gia tăng rối loạn trong hệ thống thông tin và truyền thông. Nhưng xa hơn nữa là việc phản ứng thái quá của dư luận có thể khiến nhiều chính sách hay khó đi vào cuộc sống hơn. Hoặc sẽ có những tình huống mà các nhà hoạch định chiến lược, chính sách sử dụng LLMs gặp phải những diễn giải sai lệch về một số khái niệm và phải chịu hậu quả như là sự lãng phí thời gian, nhưng tệ nhất có thể là sự chệch hướng trong sự nghiệp của họ, thậm chí cả của đồng nghiệp của họ. Điều này là đặc biệt nguy hiểm nếu các công cụ AI tạo sinh này được sử dụng trong những công việc yêu cầu độ chính xác cao, hay gây nguy hiểm đến con người, ví dụ như y tế, dược liệu [22].

## **Kết luận**

Rõ ràng, trong thời kỳ trước mắt, AI tạo sinh sẽ dần có ảnh hưởng rõ nét hơn tới đời sống xã hội và trong những công việc cụ thể như giáo dục, truyền thông, khoa học, v.v... Việc đánh giá chính xác và thích ứng với các rủi ro tiềm ẩn [23, 24] đối với xã hội trong công nghệ AI nói chung và AI tạo sinh nói riêng sẽ là một bài toán liên quan trực tiếp tới sự hưng thịnh của mỗi cá nhân, gia đình, các công ty, các cơ quan giáo dục [25], và hệ thống chính trị. Đây là vấn đề mang tính khoa học đòi hỏi sự phối hợp liên ngành của



hiều chuyên môn khác nhau để đi đến những đánh giá có thực chứng khoa học và rút ra những hàm ý chính sách tốt nhất. Trong dài hạn, có nhiều công việc và tương tác xã hội có thể hoàn toàn bị thay thế bởi công cụ AI, cho thấy một nhu cầu cấp bách về việc phát triển một bộ giá trị, đạo đức và quy tắc ứng xử khi tương tác và sử dụng AI để đảm bảo yếu tố nhân văn, xã hội, phẩm giá con người không bị tổn hại bởi sự lạm dụng AI. ■ THÔNG TIN & TRUYỀN THÔNG

#### Tài liệu tham khảo:

- Lê Quốc Minh. *Tri tuệ nhân tạo và báo chí: đối đầu hay hợp tác*. Báo Nhân dân điện tử 2024 [cited 2024 August 27]; Available from: <https://special.nhandan.vn/tri-tue-nhan-cao-va-bao-chi/index.html>.
- Siontis, K.C., et al., *ChatGPT hallucinating: can it get any more humanlike?* *European Heart Journal*, 2024. 45(5): p. 321-323.
- VinFuture. *Chuyên gia: Công cụ ai hiệu quả, nhưng đừng rơi vào bẫy "ảo giác"*. 2023 [cited 2024 June 6]; Available from: <https://vinfutureprize.org/vi/tin-tuc/chuyen-gia-cong-cu-ai-hieu-qua-nhung-dung-roi-vao-bay-ao-giac/>.
- Kasneji, E., et al., *ChatGPT for good? On opportunities and challenges of large language models for education*. *Learning and Individual Differences*, 2023. 103: p. 102274.
- Vuong, Q.-H. and M.-T. Ho, *Abundance of words versus poverty of mind: the hidden human costs co-created with LLMs*. *AI & SOCIETY*, 2024.
- Ho, M.-T. and Q.-H. Vuong, *Five premises to understand human-computer interactions as AI is changing the world*. *AI & SOCIETY*, 2024.
- Hofmann, V., et al., *Dialect prejudice predicts AI decisions about people's character, employability, and criminality* Preprint on arXiv, 2024.
- Hutchinson, C.S., *Potential abuses of dominance by big tech through their use of Big Data and AI*. *Journal of Antitrust Enforcement*, 2022. 10(3): p. 443-468.
- Birhane, A., *Algorithmic Colonization of Africa* Abeba Birhane, in *Imagining AI: How the World Sees Intelligent Machines*, S. Cave and K. Dihal, Editors. 2023, Oxford University Press. p. 247-259.
- Đỗ Văn Quân and Nguyễn Ngọc Huy, *Kiểm soát phân hóa giàu nghèo ở Việt Nam hiện nay*. *Tạp chí Tuyên Giáo*, 2024.
- Phạm Minh Chính and Vương Quân Hoàng, *Kinh tế Việt Nam: Thăng trầm và đột phá*. 2009: Nxb Chính trị quốc gia-Sự thật.
- Nguyen Thi Mai Suong, *Đào tạo tiếng Anh góp phần đa dạng hóa nguồn thu nhập cho lao động trẻ Việt Nam trong bối cảnh hội nhập kinh tế*. *Kinh tế Châu Á- Thái Bình Dương*, 2023. 6/2023: p. 86-88.
- Hornyak, T., *Why Japan is building its own version of ChatGPT*. *Nature*, 2023.
- Yang, Z. *Chinese AI chatbots want to be your emotional support*. *MIT Technology Review* 2023 [cited 2023 September 27]; Available from: <https://www.technologyreview.com/2023/09/06/1079026/chinese-ai-chatbots-emotional-support/>.
- Ho, M.-T. and M.-T. Ho, *Bridging divides: Empathy-augmenting technologies and cultural soul-searching*. *AI & SOCIETY*, 2024.
- Vương, Q.H., et al., *Một số vấn đề an ninh thông tin trong yếu tố trong kỷ nguyên AI*. *Trang Thông tin Điện tử Hội đồng Lý luận Trung ương*, 2024(3/2024).
- Mantello, P. and M.-T. Ho, *Losing the information war to adversarial AI*. *AI & SOCIETY*, 2023.
- Ho, M.-T. and T.H.-K. Nguyen, *Bài toán về hệ thống hóa quá trình sàng lọc thông tin trước sự trỗi dậy của thuật toán thông minh* *Tạp chí Thông Tin & Truyền Thông*, 2024(March/2024).
- Ilencov, E., *Logic học Biện chứng*. 2002, Hà Nội: Nxb. Văn hóa-Thông tin.
- Hà Thị Thu Hương, *Vai trò của truyền thông chính sách đối với hoạt động của chính phủ ở các nước*. *Tạp chí Tổ chức nhà nước*, 2021.
- Goldstein, J.A., et al., *How persuasive is AI-generated propaganda?* *PNAS Nexus*, 2024. 3(2): p. pgae034.
- Sallam, M. *ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns*. in *Healthcare*. 2023. MDPI.
- Vuong, Q.H. and M.-H. Nguyen, *Better Economics for the Earth: A Lesson from Quantum and Information Theories*. 2024: AISDL.
- Ho, M.-T., P. Mantello, and M.-T. Ho, *An analytical framework for studying attitude towards emotional AI: The three-pronged approach*. *Methods X*, 2023. 10.
- Vuong, Q.-H. and M.-T. Ho, *The disruptive AlphaGeometry: is it the beginning of the end of mathematics education?* *AI & SOCIETY*, 2024.