TẠP CHÍ NGHIÊN CỨU
PHẬT HỌC
CƠ QUAN NGÔN LUẬN CỦA TRUNG ƯƠNG GIÁO HỘI PHẬT GIÁO VIỆT NAM
ISSN: 2734-9195

AI và những tác động đối với đạo đức và tôn giáo

ISSN: 2734-9195 08:05 10/06/2026

Trong bối cảnh đó, chính niệm, từ bi và trí tuệ không chỉ là các giá trị tôn giáo, mà còn có thể được xem như những năng lực nhân bản thiết yếu giúp con người duy trì tự do nội tâm giữa thế giới ngày càng được vận hành bởi dữ liệu và thuật toán.

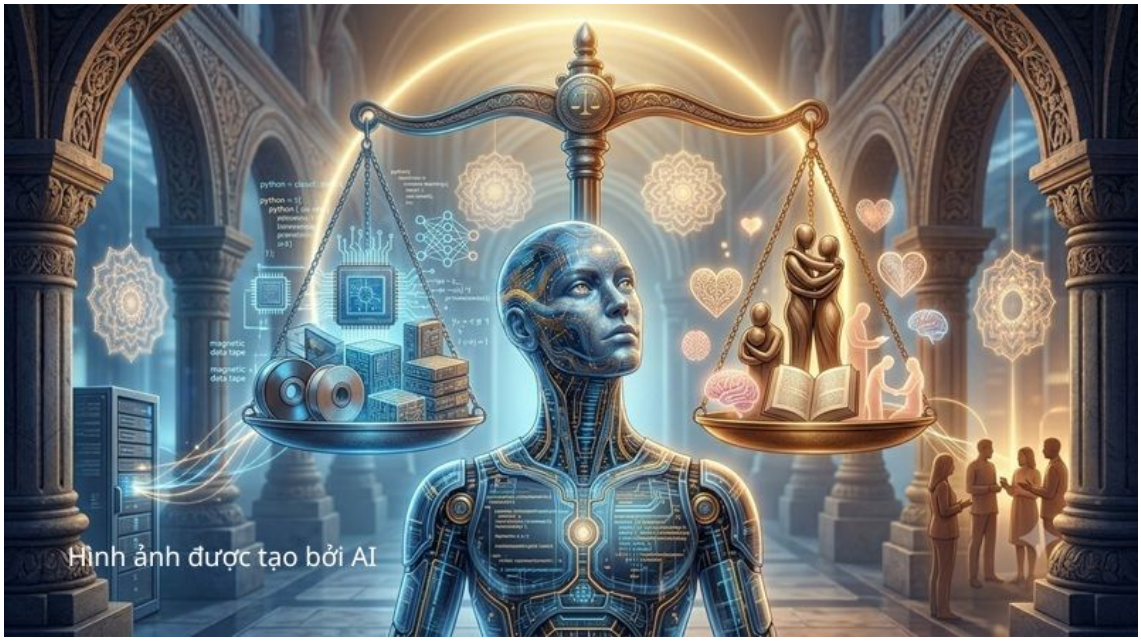
*[Tự chủ, đồng cảm và truy vấn đạo đức trong **thời đại AI** thuật toán].*

Mặc dù đang sống trong những xã hội được công nghệ hóa ở mức độ cao, chúng ta vẫn thường xuyên nghe thấy lời khẳng định: "Điều này sẽ thay đổi tất cả!". Hoặc có lẽ chính vì công nghệ mà chúng ta nghe câu nói ấy quá nhiều. Đây là một cách diễn đạt nhằm vượt qua sự nhiễu loạn thông tin để thu hút sự chú ý, trong khi sự chú ý ngày nay lại là một nguồn lực ngày càng khan hiếm.

Những phát triển mới nhất được cho là đang "thay đổi tất cả" bao gồm AI tạo sinh (Generative AI), tiêu biểu như **ChatGPT**. Những tiến bộ này tiếp nối hàng loạt cuộc chuyển đổi mang tính cách mạng trong thời gian gần đây: sự bùng nổ của các nền tảng mạng xã hội, các thiết bị thông minh, ứng dụng di động, điện thoại thông minh và xa hơn nữa là sự ra đời của Internet. Đó đều là những thay đổi thực sự với những tác động thực sự. Và chúng ta vẫn đang sống trong một thời kỳ biến đổi nhanh chóng.

Tôi không phải là chuyên gia về các **công nghệ** này. Tôi có sử dụng một số trong đó và có hiểu biết ở mức phổ thông về cách chúng vận hành, ai là người hưởng lợi từ chúng, cũng như ai là người thiết kế và kiểm soát chúng. Điều khiến tôi quan tâm dưới góc độ triết học là cách những công nghệ này định hình trải nghiệm của chúng ta về chính mình, về người khác và về tự nhiên.

Điều gì xảy ra đối với cảm nhận về quyền tự chủ của những người không chuyên khi họ bắt đầu tích hợp các hệ thống này vào thói quen hằng ngày? Khả năng đồng cảm với người khác sẽ thay đổi như thế nào? Và cách con người đặt câu hỏi, tìm hiểu và khám phá thế giới sẽ bị tác động ra sao?



AI và những câu hỏi đạo đức của nhân loại

Những câu hỏi ấy liên quan trực tiếp đến giáo dục, dân chủ và đạo đức.

Về cơ bản, điều khiến tôi quan tâm là các công nghệ vận hành dựa trên những **thuật toán** mạnh mẽ và AI tạo sinh, bởi chúng ảnh hưởng đến cách thức cũng như động cơ khiến chúng ta hành động. Trong bài viết này, tôi khảo sát những tác động tiềm tàng của chúng đối với quyền tự chủ trong việc ra quyết định, năng lực đồng cảm và phương thức truy vấn nhằm giải quyết vấn đề.

Thông qua năm phần nội dung và phần kết luận, bài viết tập trung vào các chủ đề sau:

1) Trong phần bối cảnh công nghệ, tôi phác họa một số dữ kiện nền tảng về thực trạng công nghệ mà chúng ta đang đối diện.

2) Khi bàn về lựa chọn, sự dẫn hướng (nudging) và siêu dẫn hướng (hypernudging), tôi phân tích AI tạo sinh nhằm làm rõ sự khác biệt quan trọng giữa hai khái niệm này.

3) Với câu hỏi “Liệu quyền tự chủ có đang bị đe dọa?”, tôi sử dụng các dữ kiện đã trình bày để nêu bật một mối quan tâm đạo đức: công nghệ có thể làm suy giảm khả năng tự quyết của con người như thế nào.

4) Tôi tiếp tục xem xét tác động của các công nghệ đương đại đối với khả năng kết nối với cảm xúc và quan điểm của người khác, thông qua câu hỏi: “Liệu sự đồng cảm có đang bị đe dọa?”

5) Cuối cùng, khi đặt câu hỏi “Liệu năng lực truy vấn có đang bị đe dọa?”, tôi phân tích mối liên hệ giữa siêu dẫn hướng, các công cụ thuật toán và quá trình

tìm kiếm tri thức. Bởi cách con người đặt câu hỏi và tìm kiếm câu trả lời mang tính cá nhân sâu sắc, tôi cho rằng cần phải hiểu rõ những ảnh hưởng tiềm tàng của các công cụ AI đối với tính độc đáo và cá tính trong hoạt động truy vấn. Bài viết kết thúc bằng phần tóm lược ngắn gọn cùng một số câu hỏi mở cho những suy tư tiếp theo.

Bối cảnh công nghệ: Chúng ta đang ở đâu?

Nhiều khía cạnh trong đời sống hằng ngày hiện nay được hỗ trợ và trung gian hóa bởi các công nghệ thông tin và kỹ thuật số. Chúng ta sử dụng thiết bị gia dụng, hệ thống báo động, điện thoại, đồng hồ theo dõi sức khỏe, bộ điều nhiệt thông minh, hệ thống định vị ô tô và vô số công cụ khác.

Sự phối hợp vận hành của các thiết bị này được quản lý bởi các chương trình trí tuệ nhân tạo, đồng thời được kết nối thông qua những mạng lưới điện toán đám mây phức tạp. Một số mạng lưới còn lưu trữ dữ liệu về lịch sử sử dụng trước đó, thường mang tính cá nhân hóa cao.

Việc công chúng chấp nhận và sử dụng những công nghệ này ngày càng phổ biến, trong khi các doanh nghiệp phát triển và kinh doanh chúng không ngừng quảng bá về tiềm năng cũng như lợi ích của chúng.

Hiểu được cấu trúc vận hành và các động lực kinh tế đứng sau những công nghệ này là điều thiết yếu nếu muốn nhận thức đúng đắn về những gì chúng đang mang lại hoặc tước đi khỏi các lợi ích của chúng ta.

Trong tác phẩm *The Age of Surveillance Capitalism* (Kỷ nguyên của chủ nghĩa tư bản giám sát), Shoshana Zuboff đã trình bày lịch sử phát triển của các tập đoàn công nghệ lớn hiện nay, đồng thời phân tích những cấu trúc kinh tế và tâm lý của cái mà bà gọi là “quyền lực công cụ hóa” (instrumentarian power).

Yếu tố cốt lõi tạo nên thành công của mô hình này nằm ở việc các thiết bị trực tuyến và thiết bị thông minh liên tục theo dõi, ghi nhận hành vi của người dùng, sau đó tổng hợp và phân tích lượng dữ liệu được “khai thác” từ những hoạt động ấy.

Hệ thống đó được Zuboff gọi là “kiến trúc khai thác dữ liệu” (extraction architecture). Đó không chỉ thu hút sự chú ý của chúng ta mà còn sản sinh ra “dữ liệu lớn” (big data).

Dữ liệu lớn có giá trị bởi giúp dự đoán, thậm chí trong một số trường hợp còn có thể tác động đến hành vi tương lai của con người. Như sẽ được trình bày ở phần sau, hiện tượng “siêu dẫn hướng” phụ thuộc rất nhiều vào dữ liệu lớn.

Kiến trúc khai thác dữ liệu ngày càng trở nên xâm nhập sâu hơn. Đó không còn giới hạn trong các hoạt động trực tuyến mà mở rộng sang cả những khía cạnh riêng tư của đời sống cá nhân.

“*Dưới góc nhìn Phật giáo, sự chú ý không chỉ là một nguồn lực kinh tế mà còn là nền tảng của chính niệm. Khi các nền tảng số cạnh tranh để chiếm giữ thời gian và sự chú tâm của người dùng, vấn đề không đơn thuần là tiêu thụ thông tin mà còn là khả năng làm chủ tâm thức. Nếu sự chú ý liên tục bị dẫn dắt bởi các thuật toán tối ưu hóa tương tác, con người có thể đánh mất năng lực nhận biết rõ ràng những gì đang diễn ra trong chính mình và trong thế giới xung quanh***”***.*

Theo Zuboff, quá trình khai thác dữ liệu ngày nay hướng tới cái gọi là “chiều sâu tâm lý”, tức nguồn “thặng dư hành vi” vô cùng béo bở được khai thác từ những khuôn mẫu thân mật nhất của bản ngã. Các hoạt động khai thác nhắm đến tính cách, tâm trạng, cảm xúc, những lời nói dối và cả những điểm yếu dễ tổn thương của con người.

Lợi ích mà các công ty công nghệ thu được từ việc thấu hiểu thói quen và trạng thái cảm xúc của người dùng nằm ở khả năng “dẫn hướng” hoặc tác động đến hành vi của họ.

Theo Zuboff, những người triển khai các công nghệ này đã nhận ra rằng cách chắc chắn nhất để dự đoán hành vi là can thiệp vào chính nguồn gốc tạo ra hành vi ấy. Các quy trình máy móc được thiết kế để can thiệp vào thực tại giữa con người và sự vật; chúng dẫn hướng, điều chỉnh, tập hợp, thao túng và sửa đổi hành vi theo những hướng cụ thể.

Trước khi bàn sâu hơn về ChatGPT, có lẽ cần làm rõ một số khái niệm cơ bản.

+ Trí tuệ nhân tạo (AI) là bất kỳ lý thuyết, hệ thống máy tính hoặc phần mềm nào được phát triển nhằm cho phép máy móc thực hiện những nhiệm vụ vốn đòi hỏi trí thông minh của con người.

+ Học máy (Machine Learning) là lĩnh vực phát triển và ứng dụng các thuật toán cùng mô hình thống kê để giúp hệ thống máy tính học hỏi và thích nghi mà không cần phải được lập trình bằng những chỉ dẫn cụ thể cho từng trường hợp.

Những định nghĩa đơn giản này giúp chúng ta hiểu rõ hơn điều gì đang thực hiện việc “dẫn dắt”, “điều chỉnh” và “gây ảnh hưởng” mà Zuboff đề cập.

Các công cụ khai thác dữ liệu ấy đồng thời gợi nhớ đến những băn khoăn mà triết gia John Dewey từng nêu ra từ đầu thế kỷ XX. Ông lo ngại rằng đời sống

hiện đại, bao gồm cả công nghệ, có thể làm cho trải nghiệm của con người trở nên vội vã hơn, bị giới hạn hơn và bị phân mảnh hơn.

Những định nghĩa trên cũng tạo tiền đề để hiểu rõ hai hiện tượng sẽ được bàn tiếp theo: dẫn hướng (nudging) và siêu dẫn hướng (hypernudging). Cả hai đều liên quan trực tiếp đến những vấn đề đạo đức được thảo luận trong bài viết.

Lựa chọn, dẫn hướng và siêu dẫn hướng

Như đã thấy, nhiều công nghệ đương đại đang "điều chỉnh" hoặc "dẫn dắt" hành vi trực tuyến nhằm khiến các kết quả trở nên dễ dự đoán hơn. Nói cách khác, hành vi của chúng ta đang chịu những tác động mang tính dẫn hướng.



AI đối thoại với tôn giáo và tâm linh

Tuy nhiên, một số học giả cho rằng bên cạnh dẫn hướng thông thường còn tồn tại một hiện tượng khác gọi là "siêu dẫn hướng" (hypernudging) và hiện tượng này đặt ra những vấn đề đạo đức nghiêm trọng hơn nhiều.

Để làm rõ điều đó, trước hết cần phân biệt giữa dẫn hướng và siêu dẫn hướng.

Dẫn hướng (Nudges)

Khái niệm dẫn hướng được phổ biến rộng rãi qua các bài viết và đặc biệt là cuốn sách bán chạy *Nudge* xuất bản năm 2008 của Richard Thaler và Cass Sunstein.

Theo hai tác giả, dẫn hướng là: *"Bất kỳ khía cạnh nào của kiến trúc lựa chọn có khả năng thay đổi hành vi của con người theo hướng có thể dự đoán được mà*

không cấm đoán các lựa chọn khác hoặc làm thay đổi đáng kể các động cơ kinh tế của họ. Để được xem là một sự dẫn hướng, tác động đó phải dễ dàng và ít tốn kém để tránh né. Dẫn hướng không phải là mệnh lệnh bắt buộc".

Một hình thức dẫn hướng phổ biến là thiết lập lựa chọn mặc định mà người tổ chức mong muốn người dùng chấp nhận. Nếu muốn lựa chọn khác, người dùng phải chủ động từ chối hoặc thay đổi cài đặt mặc định.

Một hình thức khác là sắp xếp không gian hoặc các phương án lựa chọn theo cách khiến một số lựa chọn trở nên dễ nhận thấy hoặc thuận tiện hơn.

Ngoài ra, áp lực xã hội cũng có thể được sử dụng như một hình thức dẫn hướng. Chẳng hạn, một công ty điện lực có thể thông báo rằng phần lớn hàng xóm của bạn đang sử dụng ít điện hơn và tiết kiệm được nhiều chi phí hơn.

Thaler và Sunstein gọi cách tiếp cận này là "chủ nghĩa gia trưởng tự do" (libertarian paternalism). Bởi các hình thức dẫn hướng thường minh bạch và có thể được nhận ra nếu người tham gia chú ý, chúng tránh được sự cưỡng ép và vẫn bảo vệ quyền tự chủ của người lựa chọn.

Siêu dẫn hướng (Hypernudges)

Siêu dẫn hướng là một hiện tượng hoàn toàn khác. Các nghiên cứu về chủ đề này hiện còn tương đối hạn chế. Một số học giả cho rằng siêu dẫn hướng chỉ đơn giản là dẫn hướng trong môi trường số, số khác lại cho rằng đây là một hiện tượng mới về bản chất.

Dù chưa thể giải quyết cuộc tranh luận đó ở đây, điều quan trọng là nhận diện những khác biệt cả về chất và lượng giữa hai hình thức này.

Theo Viktorija Morozovaite, siêu dẫn hướng là: *"Một trong những hình thức dẫn hướng kỹ thuật số tinh vi nhất, cho phép định hướng người dùng theo cách cá nhân hóa và linh hoạt, nhằm tiếp cận đúng người, với đúng thông điệp, bằng đúng phương tiện, vào đúng thời điểm và lặp lại nhiều lần khi cần thiết".*

Karen Yeung đưa ra một định nghĩa mang tính kỹ thuật hơn: *"Siêu dẫn hướng vận hành một cách linh hoạt, khó nhận biết nhưng có sức tác động rất mạnh, tạo ra cho chủ thể dữ liệu một môi trường lựa chọn được cá nhân hóa ở mức độ cao. Đó dựa trên những mối tương quan do thuật toán xác định và liên tục tái cấu hình bối cảnh thông tin của người dùng theo hướng có chủ đích nhằm ảnh hưởng đến các quyết định của họ".*

Điều cuối cùng cần lưu ý là siêu dẫn hướng không phải là một hành động đơn lẻ như một cú dẫn hướng thông thường.

Theo Stuart Mills, siêu dẫn hướng thực chất là "một chuỗi các dẫn hướng được kết nối với nhau dựa trên những lựa chọn hiện tại và các lựa chọn trong quá khứ của người ra quyết định". Chuỗi tác động này tiếp tục diễn ra cho đến khi người dùng chấp nhận kết quả mong muốn hoặc hoàn toàn rời bỏ hệ thống.

Một ví dụ minh họa

Hãy xem xét một ví dụ đơn giản để nhận diện sự khác biệt nổi bật giữa dẫn hướng và siêu dẫn hướng: so sánh giữa một gờ giảm tốc trên đường và việc sử dụng Google Maps để tìm đường.

Dẫn hướng (Nudge): Một gờ giảm tốc là yếu tố mà người tham gia giao thông có thể chấp nhận hoặc bỏ qua. Nó không được thiết kế riêng cho từng cá nhân và cũng không thay đổi theo hành vi của người sử dụng.

Siêu dẫn hướng (Hypernudge): Google Maps đề xuất một lộ trình mà người dùng có thể lựa chọn đi theo hoặc không. Tuy nhiên, nếu người dùng không đi theo chỉ dẫn, hệ thống sẽ ngay lập tức tính toán lại tuyến đường mới, đồng thời vẫn giữ nguyên mục tiêu cuối cùng là đưa người dùng đến đích. Nói cách khác, cách duy nhất để "thỏa mãn" chuỗi dẫn hướng liên tục của hệ thống là hoặc làm theo các chỉ dẫn, hoặc rời bỏ hoàn toàn ứng dụng.

Từ ví dụ này, có thể nhận diện một số đặc điểm nổi bật của siêu dẫn hướng. Đó là những hệ thống: Được vận hành bằng máy tính; Dựa trên các thuật toán; Có tính động và liên tục điều chỉnh; Mang tính dự đoán; Được cá nhân hóa dựa trên dữ liệu và lựa chọn của người dùng; Liên tục tác động vào bối cảnh ra quyết định của người dùng; Hướng tới những mục tiêu xác định trước (đặc biệt là các mục tiêu của bên thứ ba); Có cấu trúc phức tạp và thường không minh bạch đối với người sử dụng về cơ chế vận hành bên trong.

AI tạo sinh và ChatGPT

Khi áp dụng sự phân biệt trên vào các công nghệ đã đề cập, chúng ta có thể đặt câu hỏi: ChatGPT và các hệ thống AI tạo sinh đang thực hiện hoạt động dẫn hướng hay siêu dẫn hướng?

Thoạt nhìn, có vẻ như không có bất kỳ hình thức dẫn hướng nào diễn ra. Xét cho cùng, chính con người chủ động đưa ra câu hỏi, yêu cầu AI sáng tác thơ hoặc hỗ trợ giải quyết các vấn đề khác nhau.



Con người trước sự phát triển của AI

Tuy nhiên, khi quan sát kỹ hơn, chúng ta sẽ thấy nhiều đặc điểm trong cách vận hành của các hệ thống này có những điểm tương đồng đáng kể với siêu dẫn hướng.

Chẳng hạn, quá trình tương tác qua lại nhiều lần giữa người dùng và hệ thống là một đặc điểm phổ biến. Các nguồn dữ liệu và cơ chế tạo ra câu trả lời thường không minh bạch đối với người sử dụng. Nội dung phản hồi có thể được cá nhân hóa theo từng người dùng. Đồng thời, các câu trả lời được tạo ra dựa trên các mô hình ngôn ngữ dự đoán, vốn có xu hướng khuyến khích những cách diễn đạt và những kiểu tư duy phổ biến trong tập dữ liệu mà hệ thống được huấn luyện. Một số điểm tương đồng khác còn phụ thuộc vào từng phiên bản AI cụ thể.

Mặc dù người dùng chủ động khởi tạo cuộc trò chuyện với ChatGPT, khác với những hình thức siêu dẫn hướng diễn ra âm thầm trong môi trường số nhưng một số hệ thống AI có khả năng xây dựng hồ sơ người dùng qua nhiều lần tương tác. Hồ sơ này có thể âm thầm định hình các loại phản hồi được thiết kế nhằm làm hài lòng người sử dụng và củng cố những phản ứng tích cực.

Điều đó cho thấy mục tiêu của một phiên tương tác với AI không hoàn toàn do người dùng xác định. Trong nhiều trường hợp, vẫn tồn tại những mục tiêu khác được tích hợp từ phía nhà phát triển hoặc đơn vị vận hành hệ thống, chẳng hạn như duy trì mức độ gắn bó của người dùng với nền tảng.

Từ những đặc điểm trên, có cơ sở để cho rằng ChatGPT và các hệ thống AI tương tự có thể vô tình dẫn dắt người dùng theo những khuôn mẫu tư duy mang tính số đông hoặc duy trì hiện trạng, bởi các phản hồi của chúng thường được hình thành từ những mẫu câu trả lời phổ biến nhất trong dữ liệu huấn luyện.

Vì vậy, nhiều phê phán đạo đức vốn được đặt ra đối với hiện tượng siêu dẫn hướng cũng có thể được áp dụng để xem xét các công nghệ AI tạo sinh.

Quyền tự chủ có đang bị đe dọa?

Lựa chọn và quyền tự chủ: Sự đồng thuận dựa trên hiểu biết đầy đủ

Nhiều đặc điểm của siêu dẫn hướng khiến chúng trở nên khó nhận diện và khó đánh giá một cách lý tính. Điều này dẫn đến những vấn đề đáng quan ngại liên quan đến nguyên tắc "đồng thuận dựa trên hiểu biết đầy đủ" (informed consent).

Nếu một người không nhận thức được cách tư duy của mình đang bị định hướng hoặc dẫn dắt, có thể vì quá trình đó diễn ra quá nhanh hoặc quá kín đáo thì họ khó có thể đưa ra lựa chọn một cách thực sự tự nguyện: đồng thuận hay phản kháng trước những phương án được trình bày.

Trong trường hợp đó, người sử dụng thậm chí có thể không nhận ra rằng còn tồn tại những lựa chọn khác đã bị loại bỏ hoặc chưa từng được đưa vào phạm vi cân nhắc của mình.

"*Từ quan điểm Phật giáo, tự do không chỉ là khả năng lựa chọn giữa nhiều phương án mà còn là năng lực nhận diện những điều kiện đang chi phối quá trình lựa chọn ấy. Một quyết định chỉ thực sự tự chủ khi người đưa ra quyết định có khả năng quan sát các động cơ, cảm xúc, định kiến và tác động bên ngoài đang ảnh hưởng đến mình. Vì vậy, thách thức của thời đại AI không đơn thuần là bảo vệ quyền lựa chọn mà còn là duy trì sự tỉnh thức đối với những yếu tố vô hình đang định hướng hành vi và nhận thức của con người***"**.

Khi nào "dẫn hướng" trở thành "thống trị"?

Một số nhà nghiên cứu hiện tập trung phân tích cách thức các hệ thống này được thiết kế dựa trên thuật toán.

Chẳng hạn, John Danaher cho rằng siêu dẫn hướng có thể trở thành một hình thức "thao túng tinh vi". Theo ông, siêu dẫn hướng là:

"Một kỹ thuật thay đổi hành vi vận hành bên dưới ngưỡng nhận thức có ý thức, diễn ra một cách linh hoạt và được cá nhân hóa ở mức độ cao."

Nếu như các hình thức dẫn hướng truyền thống còn mang dáng vẻ tương đối vô hại, chẳng hạn việc sắp xếp quầy salad sao cho các món tráng miệng khó tiếp cận hơn thì trong môi trường số, điều đó đã thay đổi đáng kể.

Danaher giải thích: *"Các công nghệ được điều hành bằng thuật toán đã đẩy dẫn hướng đến mức cực đoan. Thay vì tạo ra một kiến trúc lựa chọn áp dụng chung cho mọi người và chỉ được điều chỉnh chậm rãi theo thời gian, giờ đây có thể xây dựng những kiến trúc lựa chọn mang tính cá nhân hóa cao, có khả năng học hỏi và thích nghi với từng người dùng. Điều này khiến việc nhận diện và từ chối các tác động dẫn hướng trở nên khó khăn hơn rất nhiều".*

Danaher cùng một số học giả trong lĩnh vực nghiên cứu khuyết tật cho rằng ảnh hưởng tích lũy của những cơ chế này đối với đời sống con người không còn đơn thuần là sự "gợi ý" mà đã tiến gần đến hình thức "định hướng hành vi".

Từ góc độ đó, siêu dẫn hướng có thể bị phê phán như một dạng "thống trị vi mô" (micro-domination).

Thoạt nghe, đây có vẻ là một kết luận cường điệu. Tuy nhiên, nếu xét đến lượng thời gian mà nhiều người dành cho môi trường trực tuyến, cũng như việc trải nghiệm trực tuyến ngày nay được đặt trong một hệ thống rộng lớn gồm khai thác dữ liệu, giám sát và dự đoán hành vi, thì những lo ngại về sự suy giảm quyền tự chủ không phải là điều thiếu cơ sở.

Daniel Susser cũng đưa ra nhận định tương tự: *"Hành động một cách tự chủ là hành động một cách độc lập, dựa trên những lý do của chính mình... Những ảnh hưởng vô hình đe dọa lý tưởng đó bằng cách khiến con người hành động vì những lý do mà họ không hiểu và do đó không thể thực sự chấp nhận".*

Từ những phân tích trên, có thể nói rằng chúng ta có đủ lý do để quan ngại về quyền tự chủ của con người trong bối cảnh họ ngày càng thường xuyên chịu tác động của các cơ chế siêu dẫn hướng.

Ngoài cách hiểu về bản ngã theo truyền thống Kant hoặc chủ nghĩa tự do, vốn nhấn mạnh tính độc lập và khả năng tự định hướng bằng lý trí, cách tiếp cận của chủ nghĩa thực dụng (pragmatism) còn đặt ra một mối quan tâm khác.

Đối với John Dewey, quyền tự chủ và bản sắc cá nhân không hình thành trong sự cô lập mà được nuôi dưỡng thông qua các tương tác xã hội sống động. Những trải nghiệm ấy phong phú hơn rất nhiều so với sự tương tác với các hệ thống tạo sinh ký hiệu như máy tính.

Từ góc nhìn này, siêu dẫn hướng không chỉ đe dọa quyền tự chủ vì nó là một hình thức ảnh hưởng vô hình, mà còn bởi nó có thể thu hẹp những tương tác mang tính thử nghiệm, khám phá và phát triển những yếu tố cấu thành quá trình trưởng thành của con người.

Liệu sự đồng cảm có đang bị đe dọa?

Vấn đề đạo đức thứ hai mà các công nghệ đương đại đặt ra liên quan đến năng lực đồng cảm.



Trí tuệ nhân tạo có thể thay thế trí tuệ đạo đức?

Cho đến đây, chúng ta chủ yếu tập trung vào cách các công nghệ thuật toán vận hành một cách âm thầm và khó nhận biết. Tuy nhiên, bên cạnh tính vô hình ấy, còn có một yếu tố quan trọng khác cần được xem xét: tốc độ.

Tốc độ tương tác ngày càng nhanh và lượng thông tin, kích thích liên tục được cung cấp đã thúc đẩy một nhịp sống ngày càng gấp gáp. Những khuôn mẫu sống ấy có thể kéo theo những hệ quả đáng suy ngẫm về mặt đạo đức.

Trải nghiệm bị gia tốc, không ma sát và thoáng qua

Khi nhìn lại những *biến* đổi công nghệ trong khoảng nửa thế kỷ qua, chúng ta thường nhắc đến sự xuất hiện của radio, truyền hình, chu kỳ tin tức 24 giờ và nhịp độ dồn dập của Internet.

Các nền tảng mạng xã hội vận hành bằng thuật toán cùng các thiết bị thông minh ngày nay đã đẩy quá trình ấy tiến thêm một bước nữa.

Theo Shannon Vallor, những trải nghiệm này được đặc trưng bởi hai yếu tố nổi bật: tốc độ và tính "không ma sát" (frictionless).

Bà nhận xét: *"So với các phương tiện truyền thông truyền thống, mạng xã hội mới tạo ra những hành vi giao tiếp nhanh hơn, linh hoạt hơn, ngắn gọn hơn, đa dạng hơn về phong cách và nội dung, đồng thời linh động hơn về giọng điệu, nhịp độ, phạm vi và cấu trúc. Tuy nhiên, chúng cũng cổ vũ cho một lý tưởng đáng nghi ngờ: những tương tác không ma sát, khéo léo né tránh sự nhàm chán, lúng túng, xung đột, sợ hãi, hiểu lầm, bực bội và cả những sự thân mật không thoải mái vốn thường xuất hiện trong các hình thức giao tiếp truyền thống, đặc biệt là trong các cuộc gặp gỡ trực tiếp ngoài đời thực".*

Khó có thể phủ nhận rằng chính chúng ta cũng góp phần tạo nên lối sống đầy áp lực này.

Chúng ta tự đẩy nhanh nhịp độ bên trong của mình, trở nên thiếu kiên nhẫn và cố gắng hoàn thành nhiều công việc trong thời gian ngắn nhất có thể.

Do đó, sự vội vã không chỉ xuất phát từ môi trường công nghệ xung quanh mà còn đến từ cách chúng ta tiếp nhận môi trường ấy và biến những phản ứng quen thuộc thành các thói quen được tôn vinh như phương thức sống tối ưu.

Chẳng hạn, nhiều người xem khả năng "đa nhiệm" là dấu hiệu của hiệu quả và năng suất.

Nói cách khác, chúng ta tạo ra các thói quen, rồi chính các thói quen ấy lại định hình chúng ta.

Khi các hệ thống thuật toán với những mục tiêu thiết kế và cơ chế siêu dẫn hướng của chúng trở thành một phần trong đời sống thường nhật, chúng cũng dần trở thành một phần cấu thành nên các thói quen ấy.

Tính cách và lối sống của chúng ta ngày càng phản ánh logic vận hành của các thuật toán.

Đó là lý do vì sao việc tự nhận thức bản thân trong thời đại ngày nay phải bao gồm cả việc xem xét các mục tiêu và cấu trúc thiết kế của công nghệ AI. Chỉ khi đó chúng ta mới có thể hiểu mình đang dần trở thành ai trong quá trình sử dụng chúng.

Đồng cảm và truy vấn đạo đức

Từ lối sống ngày càng tăng tốc ấy, Shannon Vallor rút ra một kết luận đạo đức quan trọng.

Theo bà, tốc độ, sự phân tâm và tính phân mảnh của đời sống số đang dần trở thành những thói quen mới, nhưng chính những thói quen đó lại làm suy yếu khả năng đồng cảm của con người.

Dựa trên các nghiên cứu của nhà tâm lý học Daniel Goleman, bà viết: *"Đối với những người thường xuyên đa nhiệm trên các phương tiện truyền thông, tình trạng dễ bị phân tâm có thể trở thành một trở ngại đáng kể đối với sự quan tâm đồng cảm. Các bằng chứng cho thấy những cơ chế thần kinh liên quan đến sự chú ý, vốn là nền tảng của mối quan tâm đạo đức mang tính đồng cảm cần nhiều thời gian hơn để được kích hoạt so với các cơ chế nhận thức khác".*

Nhận định này có lẽ khá dễ hiểu nếu chúng ta nghĩ đến cảm giác khi trò chuyện với một người liên tục nhìn vào điện thoại và không thể dành sự chú ý trọn vẹn cho cuộc đối thoại.

Như Vallor mô tả: *"Việc thỉnh thoảng ngẩng đầu lên trong vài giây để giao tiếp bằng mắt khó có thể đủ để nhận biết chính xác trạng thái cảm xúc của người đối diện hoặc kích hoạt các cơ chế đạo đức cấp cao liên quan đến sự đồng cảm. Hơn nữa, có bằng chứng cho thấy bên cạnh đồng cảm, những quá trình chú ý đạo đức mang tính tự nhiên còn hỗ trợ nhiều phẩm chất quan trọng khác như tính linh hoạt, sự quan tâm chăm sóc và khả năng nhìn nhận từ nhiều góc độ khác nhau".*

"*Trong Phật giáo, từ bi không chỉ là một cảm xúc nhất thời mà là khả năng hiện diện sâu sắc trước niềm vui và nỗi khổ của người khác. Sự đồng cảm đòi hỏi một chất lượng chú ý bền vững, khả năng lắng nghe và tiếp xúc trực tiếp với thực tại của tha nhân. Khi đời sống bị chi phối bởi nhịp độ tăng tốc, sự phân tâm liên tục và các tương tác ngắn hạn trên môi trường số, năng lực này có nguy cơ suy giảm. Vì vậy, việc nuôi dưỡng chánh niệm không chỉ mang ý nghĩa tâm linh mà còn có thể được xem như một điều kiện để bảo tồn khả năng đồng cảm trong xã hội công nghệ***"*.*

Nền tảng của đồng cảm và hành động đạo đức

Cốt lõi của đời sống đạo đức là khả năng thực sự nhìn sự việc từ góc nhìn của người khác. Chỉ khi đó, chúng ta mới có thể suy xét một cách nghiêm túc về điều nên làm.

Các nhà thực dụng luận cũng nhìn nhận mối liên hệ mật thiết giữa đồng cảm và hành động đạo đức. Chẳng hạn, John Dewey định nghĩa đồng cảm là khả năng "bước vào hoàn cảnh của người khác bằng trí tưởng tượng".

Theo Dewey, như Steven Fesmire nhận xét, "đồng cảm là điều kiện cần thiết cho sự suy tư đạo đức... bởi nó cung cấp bối cảnh cảm xúc nền tảng cho quá trình phản tư đạo đức; nếu thiếu điều đó, chúng ta thậm chí sẽ không bận tâm đến việc khám phá những khả thể đang hiện hữu trong một tình huống".

Điều đáng lưu ý là Dewey, khác với Shannon Vallor, không phải là một nhà đạo đức học đức hạnh theo truyền thống Aristotle. Ông bác bỏ quan niệm về những đức tính cố định, thay vào đó nhấn mạnh vai trò của quá trình truy vấn và thử nghiệm nhằm làm cho các thói quen ngày càng trở nên thông minh hơn.

Tuy khác biệt về lập trường triết học, cả Dewey và Vallor đều gặp nhau ở một điểm quan trọng: họ cùng khẳng định vai trò không thể thiếu của sự đồng cảm và cùng bác bỏ ý tưởng cho rằng truy vấn đạo đức chỉ là một quá trình lý tính hoặc tính toán thuần túy.

Theo cách hiểu đó, đạo đức không phải là thứ có thể đơn giản được "lập trình" vào AI.

Điều quan trọng hơn là chúng ta không thể tách rời bản thân mình khỏi những thói quen đang từng ngày được định hình bởi các công nghệ thuật toán.

Nguy cơ nằm ở chỗ các công nghệ này đang làm thay đổi nhịp điệu trải nghiệm của con người. Theo nghĩa đó, công nghệ không hề trung tính về giá trị.

Shannon Vallor viết: *"Công nghệ không chủ động quyết định cũng không đơn thuần phơi bày phẩm chất đạo đức của chúng ta; nó trung gian hóa đời sống đạo đức bằng cách vừa định hình vừa bị định hình bởi các thói quen và thực hành đạo đức của con người... Chúng ta cần nhận ra rằng không tồn tại một 'trí tuệ đạo đức tập thể' nào trên đám mây có thể thay thế cho sự chú ý chân thực và khả năng đồng cảm của con người... Vì vậy, dù làm gì đi nữa, chúng ta cũng không nên trao phó những năng lực đạo đức căn bản nhất của mình cho các biến động của thị trường hoặc cho ảo tưởng về một tâm trí tập thể. Năng lực chú ý là thành tố thiết yếu của nhân tính, cũng như chính cơ thể chúng ta".*

Liệu năng lực truy vấn có đang bị đe dọa?

Vấn đề cuối cùng liên quan đến ảnh hưởng của siêu dẫn hướng và AI đối với hoạt động truy vấn nói chung.

Sức mạnh của công nghệ đương đại nằm ở khả năng cung cấp khối lượng thông tin và dữ liệu khổng lồ được chọn lọc từ các sự kiện trong quá khứ.

Khi hồ sơ dữ liệu của từng người dùng ngày càng hoàn thiện, lượng thông tin được cung cấp cũng ngày càng được cá nhân hóa theo những gì máy móc dự đoán rằng chúng ta muốn hoặc cần.

Thoạt nhìn, đây dường như là một thiết kế lý tưởng. Xét cho cùng, tại sao tôi lại muốn nhìn thấy quảng cáo về một chiếc xe tải khi từ trước đến nay tôi chỉ sử dụng những chiếc xe cỡ nhỏ?

Nhưng sẽ ra sao nếu tôi mong muốn ít tính dự đoán hơn và nhiều cơ hội ngẫu nhiên hơn?

Điều gì xảy ra nếu tôi coi trọng sự mới mẻ, khả năng mở lòng trước những điều chưa biết hơn là cảm giác được thỏa mãn tức thời?

Bởi AI vận hành dựa trên dữ liệu và nhận diện khuôn mẫu, nó có xu hướng ưu tiên sự lặp lại hơn là đổi mới.

Điều đó dễ dàng khiến con người bị khóa chặt trong những lựa chọn của quá khứ.

Như Eli Pariser nhận xét: *“Các bộ lọc cá nhân hóa tạo ra một dạng tuyên truyền tự động vô hình, liên tục nhồi nhét chúng ta bằng chính các ý tưởng của mình, khuếch đại xu hướng tìm đến những điều quen thuộc và khiến chúng ta không nhận ra những nguy cơ đang ẩn mình trong vùng đất chưa biết”.*

Nói cách khác, các thuật toán dữ liệu hóa thường phản ánh và duy trì những thiên kiến hiện có.

Trong khi đó, bản sắc con người luôn vận động và biến đổi. Chính vì vậy, điều cốt yếu là quá trình truy vấn phải được tự do phát triển theo những hướng không bị các thuật toán áp đặt hoặc dẫn dắt.

Truy vấn không chỉ là công cụ giúp giải quyết vấn đề mà còn là động lực hình thành cá tính riêng của mỗi người.

“*Phật giáo luôn khuyến khích tinh thần quán chiếu và tự mình thể nghiệm chân lý. Trong Kinh Kalama, Đức Phật khuyên không nên chấp nhận một điều gì chỉ vì truyền thống, uy tín hay số đông tin tưởng, mà cần trực tiếp quan sát và kiểm chứng bằng kinh nghiệm. Từ góc nhìn này, nguy cơ lớn nhất của các hệ thống AI không phải là chúng cung cấp câu trả lời sai, mà là việc con người dần từ bỏ thói quen tự đặt câu hỏi, tự khảo sát và tự tìm hiểu thực tại***”.**

Sự phát triển của con người không thể chỉ là việc liên tục thỏa mãn những ham muốn quen thuộc bằng những hình thức thỏa mãn tương tự. Nó còn phải bao gồm khả năng đổi mới, thử nghiệm và chuyển hóa.

Theo Dewey, truy vấn bắt đầu từ việc nhận ra một khó khăn được cảm nhận trực tiếp trong kinh nghiệm sống. Từ đó, con người hình dung các giả thuyết khác nhau, dự phóng những hệ quả khả dĩ và kiểm nghiệm chúng trong hành động thực tiễn.

Siêu dẫn hướng không chỉ can thiệp vào những câu trả lời mà chúng ta nhận được; nó còn tác động đến chính quá trình hình thành vấn đề. Nó có khả năng định hình những gì chúng ta cảm nhận là đáng quan tâm hoặc đáng lo ngại.

Theo nghĩa đó, ba mối quan tâm được nêu ra trong bài viết – quyền tự chủ, sự đồng cảm và năng lực truy vấn – thực chất là các biểu hiện khác nhau của cùng một vấn đề sâu xa hơn: cách các hệ thống thuật toán có thể làm thu hẹp những điều kiện cần thiết cho sự trưởng thành về trí tuệ và đạo đức.

Do vậy, nguy cơ mà các thuật toán và những công cụ như ChatGPT đặt ra đối với hoạt động truy vấn không chỉ mang tính nhận thức luận hay đạo đức, mà còn mang tính hiện sinh.

Nếu dần hình thành thói quen chấp nhận các câu trả lời hoặc thậm chí chỉ những gợi ý được cung cấp bởi các hệ thống dự đoán khuôn mẫu, chúng ta có nguy cơ hợp lý hóa hiện trạng thay vì nghĩ ra những ý tưởng mới và nuôi dưỡng những giấc mơ mới.

Kết luận

Chúng ta đang sống trong một thời đại vừa phức tạp vừa dễ gây mất phương hướng.

Con người liên tục kết nối, liên tục tiếp nhận thông tin và liên tục bị cuốn vào nhịp độ tăng tốc của đời sống.

Các công cụ kỹ thuật số và hệ thống dựa trên dữ liệu hứa hẹn mang lại cho những con người bận rộn một cảm giác kiểm soát nhất định giữa dòng chảy biến động không ngừng của thế giới.

Tuy nhiên, bài viết này cho rằng chúng ta cần chậm lại đôi chút để suy tư một cách triết học về những đánh đổi đang âm thầm diễn ra.

Một trong những đánh đổi ấy là sự suy giảm quyền tự chủ - khả năng đưa ra những lựa chọn thực sự tự do. Điều này được liên hệ với hiện tượng siêu dẫn hướng.

Một đánh đổi khác là sự suy giảm khả năng đồng cảm, vốn chịu tác động của lối sống chuyển đổi liên tục và tình trạng phân tâm kéo dài.

Cuối cùng, bài viết đã xem xét ảnh hưởng của các công cụ thuật toán đối với hoạt động truy vấn, đồng thời chỉ ra rằng tính độc đáo và sáng tạo trong quá trình truy vấn đang đứng trước những nguy cơ ngày càng lớn; kéo theo đó là nguy cơ đối với chính tính cá nhân của con người.

Theo cách hiểu của Dewey, quyền tự chủ, sự đồng cảm và năng lực truy vấn là những điều kiện thiết yếu của sự phát triển, quá trình liên tục tái kiến tạo kinh nghiệm sống.

Những thói quen được hình thành dưới tác động của AI, dù thông qua siêu dẫn hướng hay các cơ chế tương tự, đều có thể cản trở sự phát triển ấy bằng cách khiến kinh nghiệm hiện tại bị chi phối bởi các khuôn mẫu quá khứ, thay vì mở ra những cuộc gặp gỡ thực sự mới mẻ với thế giới.

Những rủi ro này sẽ không biến mất. Ngược lại, cùng với sự phát triển của học máy, nhiều công nghệ đang được thảo luận hiện nay dường như còn có khả năng tự nhân rộng, tự sáng tạo và tự phát triển. Đây là thực tế mà nhân loại sẽ phải tiếp tục đối diện.

Tôi không phải là người theo thuyết quyết định luận, và John Dewey cũng vậy. Tuy nhiên, chúng ta vẫn nên nghiêm túc lắng nghe những cảnh báo của Jacques Ellul khi ông viết: *“Kỹ thuật đã trở nên tự trị; nó tạo dựng một thế giới khổng lồ vận hành theo các quy luật riêng và từ bỏ truyền thống. Kỹ thuật không còn dựa trên truyền thống nữa mà dựa trên chính các quy trình kỹ thuật có trước; và sự tiến hóa của nó diễn ra quá nhanh, quá mạnh mẽ đến mức không thể dung hợp với những truyền thống cũ”.*

Khác với Ellul, tôi tin rằng thay đổi vẫn là điều khả thi. Nhưng điều đó đòi hỏi những cuộc truy vấn tự do, có trách nhiệm và đầy tỉnh táo đối với những công cụ công nghệ đang ngày càng hiện diện trong đời sống của chúng ta.

❝*Từ góc nhìn Phật giáo, những vấn đề mà tác giả đặt ra về quyền tự chủ, sự đồng cảm và năng lực truy vấn không chỉ là những thách thức công nghệ mà còn là những thách thức của đời sống tâm thức. Nếu AI ngày càng có khả năng dự đoán và định hướng hành vi, thì câu hỏi quan trọng không phải là máy móc thông minh đến đâu, mà là con người có còn đủ tỉnh thức để nhận diện những*

*điều đang tác động đến mình hay không. Trong bối cảnh đó, chính niệm, từ bi và trí tuệ không chỉ là các giá trị tôn giáo, mà còn có thể được xem như những năng lực nhân bản thiết yếu giúp con người duy trì tự do nội tâm giữa thế giới ngày càng được vận hành bởi dữ liệu và thuật toán***”***.*

Tác giả: **David L. Hildebrand/**Chuyển ngữ và biên tập: **Thường Nguyên**

Nguồn: https://secularhumanism.org/2026/03/autonomy-empathy-and-moral-inquiry-amid-algorithmic-ai/

Chú thích:

1] Shoshana Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York, NY: Public Affairs, 2019.

2] Ibid., p. 199.

3] Ibid., p. 200.

4] “ChatGPT is a chatbot developed by OpenAI that uses generative AI and natural language processing to simulate human-like conversations in a chat window where the user can ask the bot to help with a variety of tasks, including drafting emails, essays, code, and more.” See Harvard Online, “The Benefits and Limitations of Generative AI: Harvard Experts Answer Your Questions,” April 19, 2023. Available online at https://www.harvardonline.harvard.edu/blog/benefits-limitations-generative-ai.

(Tạm dịch: “ChatGPT là một chatbot do OpenAI phát triển, sử dụng AI tạo sinh và công nghệ xử lý ngôn ngữ tự nhiên để mô phỏng các cuộc trò chuyện giống như con người trong giao diện hội thoại. Người dùng có thể yêu cầu hệ thống hỗ trợ nhiều công việc khác nhau như soạn thư điện tử, viết bài luận, lập trình và nhiều tác vụ khác”.

Xem Harvard Online, *The Benefits and Limitations of Generative AI: Harvard Experts Answer Your Questions*, ngày 19/4/2023).

5] Richard H. Thaler and Cass R. Sunstein, Nudge: The Final Edition. New York, NY: Penguin, 2021.

6] Viktorija Morozovaite, “Two Sides of the Digital Advertising Coin: Putting Hypernudging into Perspective.” Market and Competition Law Review vol. 5, no. 2 (2021), p. 117.

7] Karen Yeung, “Hypernudge: Big Data as a Mode of Regulation by Design.” Information, Communication & Society vol. 20, no. 1 (2017), p. 122.

8] Stuart Mills, “Finding the ‘Nudge’ in Hypernudge.” Technology in Society 71 (2022), pp. 102-117.

9] John Danaher, “Freedom in an Age of Algocracy.” In Shannon Vallor (ed.), The Oxford Handbook of Philosophy of Technology. Oxford, UK: Oxford University Press, 2022, p. 265.

10] Ibid,. p. 265

11] Daniel Susser, “Invisible Influence: Artificial Intelligence and the Ethics of Adaptive Choice Architectures.” Proceedings of the AIES’19. New York, NY: Association for Computing Machinery, 2019, p. 406.

12] Shannon Vallor, Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford, UK: Oxford University Press, 2016, p. 162.

13] Ibid., p. 171.

14] Ibid., p. 171.

15] John Dewey and James H. Tufts, “Ethics” (1908). In Jo Ann Boydston (ed.), The Middle Works of John Dewey, Volume 5, 1899-1924. Carbondale, IL: Southern Illinois University Press, 1932, p. 150.

16] Scott Fesmire, Dewey. London, UK: Routledge, 2015, p. 133.

17] Vallor, Technology and the Virtues, pp. 184, 173.

18. Eli Pariser, The Filter Bubble: What the Internet Is Hiding from You. New York, NY: Penguin Press, 2011, p. 15.

19] As Dewey put it, “Human beings are not like … mechanical products which are as much like each other as peas in the pod … When we get human individuals, we get the principle that each individual has something that is unique or irreplaceable. No one else will quite take his place in the world or do quite the same thing that he will do… The principle of individuality, then, is having a place and work in the world that no one else can quite do. This gives us a measure. The more mechanical, the more things are similar to each other; but as we rise to what is vital, rise in the sphere of life to what is spiritual, moral and intellectual, the principle of individuality counts for more and more. That is why the principle of individuality has such claims in education. It is the measure

of whatever is elevating in the rank of life in spiritual, moral and intellectual beings" (John Dewey, "Individuality in Education" (1923). In Jo Ann Boydston and Anne Sharpe (eds.), The Middle Works, Volume 15, 1899-1924. Carbondale, IL: Southern Illinois University Press, 1932, pp. 170-171).

(Tạm dịch: Như Dewey từng viết:

"Con người không giống như những sản phẩm cơ khí được chế tạo hàng loạt, giống nhau như những hạt đậu trong cùng một vỏ. Khi nói đến các cá nhân con người, chúng ta phải thừa nhận rằng mỗi người đều mang trong mình một điều gì đó độc đáo và không thể thay thế. Không ai khác có thể hoàn toàn thế chỗ của họ trong thế giới này hay thực hiện chính xác những gì họ có thể thực hiện.

Vì vậy, nguyên lý cá biệt tính nằm ở việc mỗi người có một vị trí và một công việc riêng trong đời sống mà không ai khác có thể đảm nhiệm theo cùng một cách. Đây cũng là thước đo để đánh giá sự phát triển của con người. Càng mang tính cơ giới, mọi thứ càng giống nhau; nhưng khi bước vào lĩnh vực của sự sống, rồi tiến lên các bình diện tinh thần, đạo đức và trí tuệ, nguyên lý cá biệt tính càng trở nên quan trọng hơn.

Đó là lý do vì sao nguyên lý cá biệt tính giữ vai trò đặc biệt trong giáo dục. Nó là tiêu chuẩn để nhận diện những gì nâng cao phẩm giá đời sống tinh thần, đạo đức và trí tuệ của con người".

(John Dewey, "Individuality in Education", 1923).

20] Jacques Ellul, The Technological Society. New York, NY: Vintage Books, 1964, p. 14.David L. Hildebrand