Tâm sơ khởi - kiến tạo tác ý tập thể trong kỷ nguyên AI

ISSN: 2734-9195 16:25 11/05/2026

AI cũng nên được mô hình hóa tương tự, nơi không có dữ liệu hay quan điểm nào giữ vị thế ưu tiên tuyệt đối, người dùng, chuyên gia lĩnh vực, kỹ sư và nhà quản lý đều đóng vai trò như những "ong trinh sát", đóng góp tri thức riêng mình.

Beginner's Mind (tạm dịch: Tâm sơ khởi) là một dự án đặc biệt, tập hợp các bài luận sâu sắc do sinh viên đại học tại Hoa Kỳ viết sau khi tham gia các khóa học trải nghiệm liên quan đến **Phật giáo**.

Một số tác giả tự nhận mình là phật tử, trong khi với những người khác, đây là lần đầu họ tiếp cận *Buddhadharma* (phật pháp). Tất cả đều chia sẻ những suy tư và ấn tượng về những gì họ đã học, cách điều đó tác động đến đời sống của họ và cách họ có thể tiếp tục gắn bó với giáo pháp.



Carter Kawaguchi đã viết bài luận cho khóa học mùa Xuân năm 2026 về *Buddhist Economics* (Kinh tế học Phật giáo) tại Đại học Nam California (University of Southern California - USC). Carter là sinh viên năm cuối, theo học các ngành *Computational Linguistics* (Ngôn ngữ học tính toán), Kinh doanh và Thiết kế. Anh quan tâm đến quản lý sản phẩm, thiết kế, và việc kết nối ngôn

ngữ, công nghệ, sáng tạo và cộng đồng.

Nuôi dưỡng tác ý tập thể thông qua việc khử thiên lệch trong AI

Đối với chủ đề bài luận này, tôi lựa chọn tập trung vào một lực lượng cụ thể, mang tính dẫn dắt đang tái định hình nền kinh tế hiện đại: trí tuệ nhân tạo (*artificial intelligence - AI*). Sự hiện diện của AI trong xã hội hiện nay là điều không thể đảo ngược, và chúng ta sẽ còn tiếp tục chứng kiến những tác động kinh tế to lớn từ sự phát triển này.

Tuy nhiên, việc *de-biasing* (khử thiên lệch) trong AI đang nổi lên như một thực hành ngày càng quan trọng. Nhiều vấn đề triết học và đạo đức đã xuất hiện khi các *LLMs - large language models* (mô hình ngôn ngữ lớn) kéo theo những hệ quả sâu rộng đối với nhận thức về chân lý và sự lan truyền tri thức, đồng thời được ứng dụng trong hầu hết các ngành nghề. Việc khử thiên lệch mở ra một cơ hội đặc biệt để áp dụng triết học tiến trình (*processual philosophy* - triết học nhấn mạnh tính vận hành và biến đổi liên tục) của Phật giáo vào công cụ hiện đã trở nên phổ biến này.

Những quan điểm phổ biến hiện nay xem thiên lệch như một lỗi kỹ thuật, thay vì nhìn nhận đó như một yếu tố tạo ra *precarity* (tình trạng bấp bênh, bất ổn) và kiểm soát **nhận thức** về chân lý. Tôi đề xuất rằng thực hành khử thiên lệch trong AI có thể được diễn giải lại như việc nuôi dưỡng *cetanā* tập thể (*Skt. cetanā*: ý chí, tác ý). Bằng cách tiếp cận việc khử thiên lệch trong AI với tư duy triết học tiến trình, chúng ta có thể góp phần hình thành *cetanā* tập thể thông qua các vòng đánh giá lặp lại (*iterative* - tuần hoàn cải tiến), mang tính liên chủ thể (*intersubjective* - giữa nhiều chủ thể nhận thức) từ các bên liên quan đa dạng, nhằm giảm thiểu bất ổn và kiến tạo một xã hội vận hành trong trạng thái liên kết, tương thuộc sâu sắc hơn, từ đó tạo ra chuyển hóa.



AI hiện diện trong gần như mọi ngành nghề và không gian của nền kinh tế - xã hội hiện đại, khiến AI trở thành yếu tố trung tâm trong việc kiến tạo một nền kinh tế quan tâm (*economy of care*) và giảm thiểu bất ổn. Việc khử thiên lệch trong AI là thiết yếu cho tiến trình này, khi các hệ thống **thuật toán** ngày càng đóng vai trò trung gian trong việc tiếp cận việc làm, tài nguyên, sự công nhận và tri thức. Điều này có thể thấy rõ qua sự trỗi dậy của các công ty chuyên áp dụng AI vào các quy trình hiện có như chấm điểm tín dụng (*credit scoring*), tuyển dụng (*hiring*) và xét duyệt phúc lợi xã hội (*welfare eligibility*).

Tri thức mà các mô hình này lan truyền mang những hệ quả thực tế đối với hàng triệu người trên toàn thế giới, có khả năng làm gia tăng khổ đau và *duhkha* (Pāli/Skt. *dukkha*: khổ, bất toại nguyện) một cách không công bằng. Khi *Cakkavattisutta (DN 26)* (Kinh Chuyển Luân Thánh Vương) xác định nghèo đói và thiếu thốn vật chất là căn nguyên của tội phạm, bạo lực và suy đồi đạo đức, ta có thể thấy rõ sai lệch trong các hệ thống AI có thể nhanh chóng khuếch đại tác động tiêu cực như thế nào.

Việc khử thiên lệch trong AI trở nên dễ hình dung hơn như một dạng *cetanā* tập thể khi được phân tích song song với *Honeybee Democracy* (Nền dân chủ của loài ong mật) và chủ nghĩa hoài nghi Pyrrhonian-Buddhist (*Pyrrhonian skepticism kết hợp với Phật học*), qua đó cho thấy ý định và phán đoán nảy sinh từ các tiến trình phân tán thay vì từ một tâm thức đơn lẻ.

Trong *Honeybee Democracy* (Seeley, 2010), chúng ta quan sát thấy không có con ong nào "có ý định" **kiểm soát** địa điểm làm tổ mới một cách riêng lẻ. Đây là một hành động tập thể, nơi các ong trinh sát khám phá những địa điểm tiềm

năng, biểu đạt đánh giá chất lượng thông qua "vũ điệu lắc" (*waggle dance*) và cả đàn dần hình thành sự đồng thuận (*quorum* - đạt ngưỡng đồng thuận) cho vị trí tốt nhất. Đàn ong giúp chúng ta tái định nghĩa "ý định" không phải là thứ nằm trong từng cá thể, mà là cách một mạng lưới nghiêng về kết quả nhất định dựa trên phản hồi và kinh nghiệm tập thể. AI cũng nên được mô hình hóa tương tự, nơi không có dữ liệu hay quan điểm nào giữ vị thế ưu tiên tuyệt đối, người dùng, chuyên gia lĩnh vực, kỹ sư và nhà quản lý đều đóng vai trò như những "ong trinh sát", đóng góp tri thức riêng mình.

Tiếp đó, trong phân tích của Jonathan Gold về Phật giáo của Pyrrho, chúng ta có thể rút ra một góc nhìn mới về đầu ra của các mô hình này (Gold, 2023). Văn bản cho thấy lời khuyên không dựa trên những quan điểm cố định, mà dựa trên việc thử nghiệm các cách sống và trải nghiệm hệ quả của chúng. Trong AI, chúng ta có thể áp dụng cách tiếp cận này bằng cách xem các đầu ra của mô hình là tạm thời (*provisional* - chưa cố định), **minh bạch** về mức độ bất định và xây dựng các thể chế nhằm liên tục điều chỉnh mô hình, thước đo (*metrics*) và chính sách để giảm thiểu thiên lệch và tác hại. Hai ví dụ này cùng giúp ta hiểu rõ hơn về *cetanā* tập thể trong việc khử thiên lệch AI như một hình thái phân phối cấu trúc của quyền lực và ảnh hưởng trong các mạng lưới xã hội - công nghệ của AI, đồng thời nuôi dưỡng thực hành khiêm tốn tri thức (*epistemic humility* - ý thức về giới hạn hiểu biết) và sự điều chỉnh liên tục.

Giờ đây, khi đã hiểu việc nhìn nhận khử thiên lệch AI như hành vi nuôi dưỡng *cetanā* tập thể, chúng ta có thể đề xuất ba chiến lược có thể triển khai ngay nhằm xây dựng một nền kinh tế tương thuộc hơn. Những chiến lược này chuyển hóa AI từ sản phẩm kỹ thuật biệt lập thành một tiến trình xã hội - kỹ thuật tích hợp, có định hướng rõ ràng nhằm giảm bất ổn và thúc đẩy phúc lợi chung.

Thứ nhất, cần triển khai kiểm toán xã hội - kỹ thuật (*socio-technical auditing*) trên diện rộng trong toàn bộ các giai đoạn phát triển sản phẩm AI. Từ thu thập dữ liệu, thiết kế tính năng (*feature design*), mục tiêu huấn luyện (*training targets*), bối cảnh triển khai (*deployment context*), đến giao diện người dùng (*UI/UX - user interface/user experience*), các mô hình và sản phẩm AI cần được thiết kế xoay quanh việc giảm thiểu thiên lệch. Việc áp dụng quy trình kiểm toán ở mỗi giai đoạn sẽ giúp giảm thiên lệch theo hai cách: tiếp xúc với các góc nhìn đa dạng và đảm bảo phân tích tiến trình nhất quán.

Thứ hai, cần mã hóa các thước đo công bằng lấy phúc lợi làm trung tâm (*welfare-centered fairness metrics*) nhằm thể hiện cam kết giảm thiểu bất ổn. Các thước đo hiện nay thường tập trung vào các yếu tố kỹ thuật, mang tính lợi nhuận như tốc độ, giọng điệu (*tone*), tỷ lệ rời bỏ (*churn*),... Việc chuyển hướng

sang các thước đo không chỉ tối ưu độ chính xác hay cân bằng (*parity*), mà còn hướng đến phúc lợi xã hội, sẽ thể hiện rõ ý định của thể chế.

Thứ ba, cần thiết lập cơ chế **quản trị** có sự tham gia liên tục (*ongoing participatory governance*) bao gồm tất cả các bên liên quan và người sử dụng AI, nhằm kiểm định chất lượng đầu ra và giám sát, điều chỉnh các tính năng AI có tác động lớn. Các thiết chế tham gia này gồm người dân, chuyên gia và quan chức cần có không gian thảo luận dựa trên thông tin xác thực và sự điều phối ôn hòa. Trên thực tế, điều này thể chế hóa một dạng "*sangha*" (tăng đoàn - cộng đồng tu tập) nhằm hình thành cộng đồng thực hành cùng nhau điều chỉnh các công cụ AI.

Không cần phải nói thêm, việc khử thiên lệch AI sẽ ngày càng đóng vai trò quan trọng trong xã hội và nền kinh tế. Như đã phân tích, tiến trình này không nên chỉ được hiểu như một giải pháp kỹ thuật, mà là thực hành nuôi dưỡng *cetanā* tập thể. Điều thiết yếu là tất cả cá nhân chịu tác động bởi AI đều có "ý phần" trong các mô hình này. Nhà phát triển, thể chế, người dùng và nhà hoạch định chính sách cần được đưa vào tiến trình ra quyết định tập thể của AI.

Tiến bộ hướng đến ngành công nghiệp AI đặt trọng tâm vào *cetanā* tập thể có thể đạt được thông qua cả quy định pháp lý lẫn ý hướng thể chế nhằm nâng cao phúc lợi và giảm bất ổn, thay vì chỉ ưu tiên lợi nhuận và hiệu suất. Câu hỏi trung tâm của AI trong nền kinh tế và thế giới hôm nay không còn đơn thuần là AI có thể làm gì, mà là làm thế nào để đảm bảo những công cụ này luôn là tương liên, hành động tập thể nhằm cải thiện đời sống cho tất cả mọi người.

Suy ngẫm về việc sử dụng AI

Quá trình viết một bài luận với việc sử dụng AI minh bạch là trải nghiệm mới mẻ. Đặc biệt với những thuật ngữ mà tôi gặp khó khăn khi nắm bắt trong khóa học, AI đã hỗ trợ tôi ôn lại kiến thức và áp dụng các khái niệm vào bài viết. Cách tôi sử dụng AI khá có cấu trúc và phần lớn các câu lệnh (*prompts*) chi tiết. Mỗi câu lệnh đều bắt nguồn từ suy nghĩ và ý định ban đầu của tôi và AI được dùng để hỗ trợ cấu trúc hóa cũng như cung cấp các góc nhìn thay thế trong quá trình viết.

Đặc biệt, do tôi không muốn hiểu sai các văn bản hay khái niệm đang sử dụng, sự hỗ trợ của AI rất hữu ích. Khi bài luận của tôi xoay quanh chính AI và mối tương tác của AI với thiên lệch và tri thức, việc suy tư trong quá trình viết càng trở nên thú vị, tôi gần như đang hỏi AI làm thế nào để tự sửa chữa chính nó. Khả năng của AI trong việc trình bày góc nhìn thiểu số về hướng phát triển của ngành mang lại phần nào sự yên tâm. Tuy nhiên, tôi mong muốn các chính sách

được đề xuất trong bài thực sự được triển khai, vì tôi tin rằng những hành động như vậy sẽ góp phần tạo ra công nghệ công bằng hơn và hướng thiện cho xã hội.

Tác giả: **Carter Kawaguchi**/Chuyển ngữ và biên tập: **Thường Nguyên**

Nguồn:**https://www.buddhistdoor.net/features/beginners-mind-cultivating-collective-cetana-through-ai-debiasing/**

***

Tài liệu tham khảo:

Seeley, Thomas D. 2010. *Honeybee Democracy*. Princeton, NJ: Princeton University Press.

Gold, Jonathan. “Pyrrho’s Buddha on Duḥkha and the Liberation from Views”. *Journal of the American Academy of Religion* 91, số 3 (tháng 9/2023). https://academic.oup.com/jaar/article/91/3/655/7606269