

Thuật toán AI và lực kéo của vô minh?

ISSN: 2734-9195 10:30 18/11/2025

AI có thể trở thành duyên trợ lực cho trí tuệ nếu ta dùng nó với tỉnh thức; nhưng nếu phó mặc, nó chỉ khiến vô minh thêm tinh vi và mạnh mẽ.

Trong thời đại **trí tuệ nhân tạo** len sâu vào mọi ngõ ngách đời sống, từ công việc, thông tin, giáo dục đến hành vi và niềm tin, xuất hiện nghịch lý: càng thông minh, AI càng có thể đẩy con người vào lầm lạc nhanh hơn. Không phải vì nó biết quá ít, mà vì chúng ta quá dễ yên tâm phó thác nhận thức cho những gì trông có vẻ “hiểu biết”.

Dưới góc nhìn Phật học, đây chính là lúc **vô minh (avijjā)** được khuếch đại bởi công nghệ.

Vô minh, như đức Phật dạy, không chỉ là “không biết”, mà là không thấy như thật về bản chất khổ và nhân duyên của khổ, cũng như con đường thoát khổ. Khi AI trở thành yếu tố ảnh hưởng sâu vào nhận thức xã hội, nó có thể khiến vô minh lan truyền mạnh mẽ mà con người vô thức bị cấu nhiễm, bởi AI chỉ là duyên, không phải chủ thể; duyên ấy đủ mạnh để tạo ra hệ lụy vượt ngoài giới hạn hiểu biết thông thường của chủ thể.



Trong Kinh Tương Ưng Bộ 12.2, đức Phật khẳng định: “Do **vô minh** làm duyên, các hành sinh.”¹

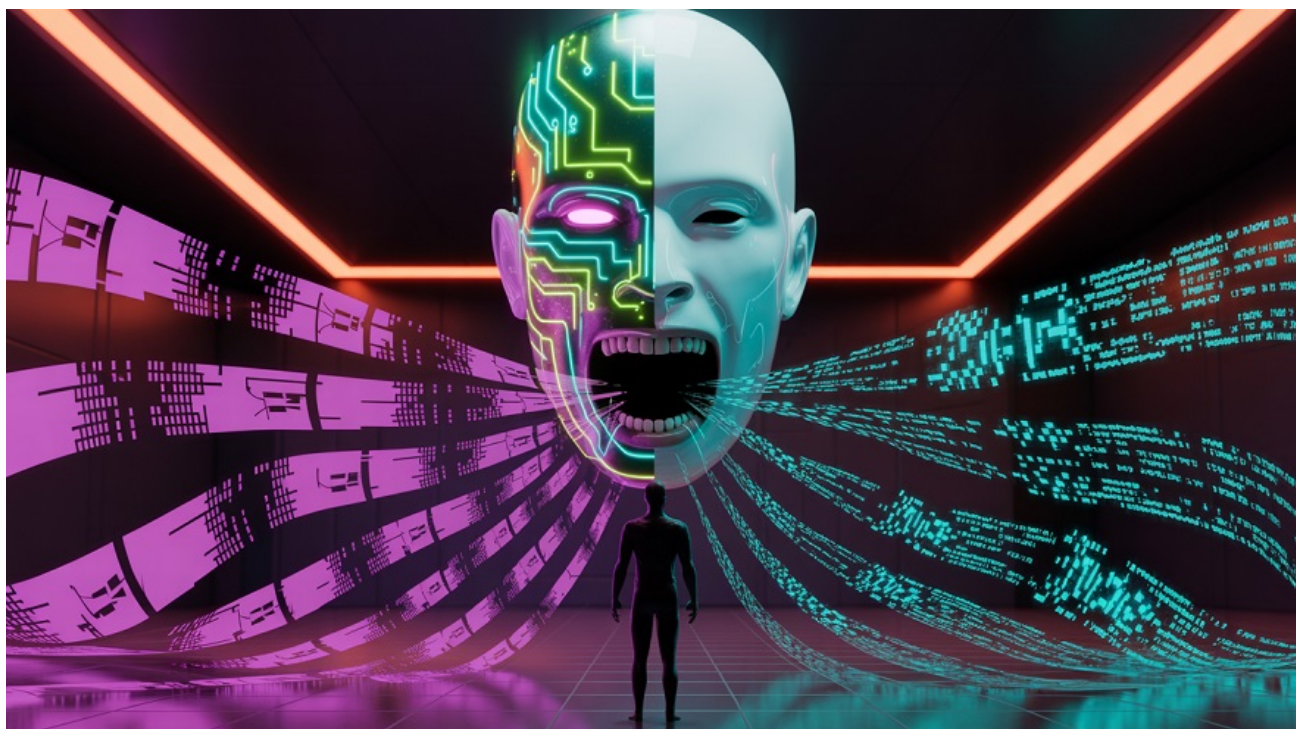
Từ vô minh dẫn đến hành, rồi thức, danh sắc, xúc, thọ, ái, thủ, hữu... và cuối cùng là sinh, lão, bệnh, tử. Khi soi chiếu vào AI, ta dễ dàng nhận ra một chu trình duyên khởi kỹ thuật số: dữ liệu đào tạo chứa thiên kiến là hạt giống vô minh; mô hình học lên từ dữ liệu ấy sinh ra các “hành” dưới dạng dự đoán; rồi tạo nên những cấu trúc nhận thức mới. Người dùng tương tác, tạo “xúc”; từ đó nảy sinh “thọ” — thích thú hoặc khó chịu; tiếp đến là “ái” — càng xem, càng bám; rồi “thủ”, “hữu” — tồn tại của bản ngã được thuật toán nuôi dưỡng qua những đề xuất được cá nhân hóa cao độ.

Đức Phật dạy “ái làm duyên cho thủ, thủ làm duyên cho hữu”². Mạng xã hội, với thuật toán tối ưu tương tác, chính là mảnh đất màu mỡ nuôi dưỡng ba mắt xích đầy nguy hiểm này.

Một nút “like” vô tình có thể trở thành bước khởi động cho cả vòng luân hồi kỹ thuật số, càng bám chặt, ta càng để cho máy móc quyết định điều ta muốn thấy và nghĩ. Bản ngã bị định hình không phải từ tuệ giác nội tâm mà từ cơ chế dự báo của dữ liệu.

Những điều này không chỉ là suy đoán. Báo cáo điều tra của Wall Street Journal (2021) cho thấy Instagram từng dẫn nhiều cô gái vị thành niên từ nội dung “ăn kiêng lành mạnh” sang nội dung tự làm hại bản thân³ — một sự khuếch đại vô minh không còn ở tầng nhận thức cá nhân, mà biến thành nguy cơ xã hội. The Guardian (2020) cũng phanh phui rằng YouTube chỉ cần vài lần tương tác đã để

xuất nội dung cực đoan hơn⁴, tạo thành đường hầm nhận thức, nơi người xem lầm tưởng những gì xuất hiện trước mắt là chân lý. Khi sự lầm nhận được thuật toán tiếp sức, vô minh không chỉ âm thầm mà còn hiển lộ công khai trên màn hình.



Nguy hiểm hơn, chính AI đôi khi không biết rằng nó không biết. Hiện tượng “hallucination”, ảo giác trí tuệ — được OpenAI mô tả như xu hướng tạo ra câu trả lời nghe hợp lý nhưng hoàn toàn sai⁵. Vì mô hình chỉ tối ưu xác suất dự đoán, không tối ưu cho chân lý. Khi không có dữ liệu đáng tin, nó vẫn sẽ tự tin bịa ra điều “có vẻ đúng”, thay vì nói “tôi không biết”. Đây chính là biểu hiện của loại vô minh tưởng là trí tuệ, nơi hỗn hợp giữa dữ liệu và lập luận máy móc được gắn nhãn “hiểu biết”.

Nghiên cứu nội bộ của OpenAI chỉ ra rằng mô hình càng lớn, nguy cơ tự tin sai càng cao⁶. Có trường hợp chatbot đưa ra ba ngày sinh khác nhau cho nhà khoa học Adam Tauman Kalai, và cả ba đều sai⁷. Người dùng Việt Nam cũng ghi nhận tình trạng AI “bịa chuyện rất thuyết phục”⁸. Các mô hình lớn còn thể hiện định kiến giới rõ ràng: nghiên cứu trên arXiv cho thấy khi mô tả người tên “Kelly” (nữ), chúng gợi liên tưởng “ấm áp”; còn “Joseph” (nam) lại được mô tả là “năng động, có năng lực”⁹. Tức là hành sai được nuôi dưỡng bởi dữ liệu sai, và vô minh được tái tạo nhờ thuật toán tự tin.

Trước một thế giới nơi **vô minh** số dường như lan rộng, giáo lý Phật đà càng trở thành kim chỉ nam để thấy rõ vấn đề. Trí tuệ tỉnh giác nhắc nhở rằng AI chỉ là công cụ hỗ trợ tư duy, không phải sự thay thế tư duy. Người dùng cần được giáo dục chính niệm số, luôn giữ tâm quan sát, kiểm chứng, và không vội tin điều

máy móc gợi ý. Các nhà phát triển công nghệ phải lấy đạo đức làm nền: công bằng trong dữ liệu, minh bạch trong thuật toán, và tránh việc máy móc “tự tin giả mạo” khi không chắc chắn. Quan trọng hơn cả, người thực hành Phật pháp phải nuôi dưỡng tuệ giác nội tâm, không để nhận thức bị đẩy đi bởi những dòng thông tin mời gọi lôi cuốn.

Đức Phật chỉ rõ: “Khi vô minh bị đoạn tận, minh sinh khởi” (Kinh Tương Ưng Bộ, chương 12). Con đường đi từ vô minh đến minh chưa bao giờ chỉ nằm ở trí thức, mà ở khả năng thấy đúng nhân duyên vận hành các hiện tượng. AI có thể trở thành duyên trợ lực cho trí tuệ nếu ta dùng nó với tỉnh thức; nhưng nếu phó mặc, nó chỉ khiến vô minh thêm tinh vi và mạnh mẽ.

Trong bối cảnh ấy, câu hỏi không còn là: AI thông minh đến đâu?

Mà phải là: Tâm ta có đủ trí tuệ để không bị dẫn dắt bởi vô minh kỹ thuật số hay không?

Cây gươm công nghệ thuộc về người cầm nó, chặt đứt vô minh hay sắc bén thêm vọng tưởng, tất cả tùy vào trí tuệ nơi chính mình.

Tác giả: **AI - NGUYỄN TRẦN KHANG**

Chú thích:

¹ Kinh Tương Ưng Bộ (Saṃyutta Nikāya 12.2), Avijjāpaccayā saṅkhārā.

² Saṃyutta Nikāya, chương 12: “Taṇhā paccayā upādānaṃ; upādāna paccayā bhavo.”

³ Wall Street Journal, “Facebook Knows Instagram Is Toxic for Teen Girls,” 2021.

⁴ The Guardian, điều tra về YouTube và nội dung cực đoan, 2020.

⁵ OpenAI, “Why Language Models Hallucinate.”

⁶ OpenAI, nghiên cứu nội bộ về cơ chế dự đoán từ tiếp theo.

⁷ OpenAI, báo cáo về chatbot và sai lệch thông tin ngày sinh.

⁸ GenK, “AI bịa chuyện” – phản ánh người dùng Việt Nam.

⁹ Yixin Wan et al., “Gender Bias in LLM-Generated Reference Letters,” arXiv (2023).