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

Khi thuật toán bước vào chiến trường: AI có thể "giữ giới"?

ISSN: 2734-9195 11:05 09/03/2026

Bởi nếu trí tuệ không đi kèm từ bi, đó có thể trở thành công cụ tinh vi tái chế khổ đau. Và khi thuật toán bước vào chiến trường, câu hỏi về giới luật không còn là vấn đề của tu viện, mà là vấn đề của toàn nhân loại.

Những ngày gần đây, một thông tin gây nhiều chú ý trong giới công nghệ và an ninh quốc phòng: hệ thống AI Claude của công ty Anthropic được cho là đã và đang được sử dụng trong một số hoạt động hoạch định, mô phỏng chiến trường và phân tích mục tiêu của quân đội Hoa Kỳ. Tuy nhiên, chỉ ít lâu sau đó, chính quyền của Tổng thống Donald Trump lại tuyên bố chấm dứt hợp tác với Anthropic, với lý do liên quan đến kiểm soát và quyền quyết định trong việc sử dụng AI cho mục đích quân sự.

Căng thẳng bùng phát khi Anthropic công khai khẳng định, họ từ chối cho phép công nghệ của mình được sử dụng trong hai lĩnh vực: *triển khai vũ khí tự động hoàn toàn* và giám sát đại chúng trong nước. Về phía chính quyền và Lầu Năm Góc, quan điểm được đưa ra rất rõ ràng: việc sử dụng công nghệ cho mục tiêu an ninh quốc gia thuộc thẩm quyền của nhà nước.

Đây không chỉ là một bất đồng thương mại. Đó là một câu hỏi lớn của thời đại: Khi thuật toán bước vào chiến trường, vai trò lương tâm đặt ở đâu?

Trí tuệ không đồng nghĩa với đạo đức

AI có thể xử lý dữ liệu với tốc độ vượt xa con người, có thể dự đoán quỹ đạo tên lửa, xác suất tổn thất, khả năng phản công và tối ưu hóa chiến thuật. Nhưng tất cả những điều đó chỉ là năng lực tính toán.

Trong Phật giáo, “trí tuệ” (prajñā) không đơn thuần là khả năng phân tích. Trí tuệ chân chính luôn đi cùng từ bi (karuṇā). Nếu thiếu từ bi, trí tuệ có thể trở thành công cụ phục vụ tham vọng, sợ hãi hoặc quyền lực.

Một hệ thống AI có thể nhận diện mục tiêu quân sự với độ chính xác cao. Nhưng AI không biết sợ hãi là gì, không cảm nhận được mất mát, không ý thức được đau khổ. AI không có kinh nghiệm về khổ (dukkha), nên cũng không có động cơ nội tại để tránh gây khổ.

Vì vậy, câu hỏi không phải là: AI có thông minh không?

Mà là: AI có khả năng tự hạn chế mình trước đau khổ của tha nhân không?

Giới thứ nhất trong thời đại vũ khí tự động

Giới đầu tiên trong Phật giáo là không sát sinh. Đó không chỉ là cấm đoán hành vi, mà là lời nhắc về thái độ căn bản đối với sự sống: tôn trọng và bảo hộ.

Khi AI được tích hợp vào hệ thống vũ khí tự động, một tình huống mới xuất hiện: quyết định tấn công có thể được đưa ra mà không cần sự can thiệp trực tiếp của con người trong từng khoảnh khắc.

Anthropic lập luận rằng các mô hình AI hiện nay chưa đủ đáng tin cậy để được trao quyền trong những quyết định sinh tử. Lập luận ấy, xét dưới góc nhìn Phật học, không chỉ là vấn đề kỹ thuật, mà là vấn đề **giới hạn đạo đức.**

Nếu một cỗ máy không có khả năng cảm nhận trách nhiệm, liệu nó có thể "giữ giới"?

Hay nói đúng hơn: giới có thể được lập trình?

Giới luật trong Phật giáo không phải là thuật toán *điều kiện - kết quả*. Giới bắt nguồn từ tâm ý, từ sự tỉnh thức và lòng từ. Một hệ thống máy học có thể được gắn rào chắn kỹ thuật, nhưng đó không phải là "giữ giới" theo nghĩa nội tâm.

Nghiệp thuộc về ai?

Một trong những vấn đề phức tạp nhất là câu hỏi về trách nhiệm.

Nếu một hệ thống AI đề xuất mục tiêu, một sĩ quan phê duyệt và một vũ khí tự động thực thi, thì ai là chủ thể tạo nghiệp?

+ Nhà lập trình?

+ Công ty công nghệ?

+ Nhà hoạch định chiến lược?

+ Quốc gia?

+ Hay cá nhân bấm nút cuối cùng?

Trong giáo lý về nghiệp (karma), yếu tố quyết định không chỉ là hành vi mà là ý định (cetanā). AI không có ý định theo nghĩa tâm linh. Nhưng con người đứng sau việc thiết kế, triển khai và cho phép sử dụng nó thì có.

Điều đó cho thấy: dù công nghệ có phát triển đến đâu, trách nhiệm đạo đức không thể được "chuyển giao" hoàn toàn cho máy móc.

Quyền lực và nỗi sợ

Trong bối cảnh địa chính trị căng thẳng, nhiều quốc gia cho rằng họ cần mọi công cụ có thể để bảo vệ lợi ích của mình. AI trở thành lợi thế chiến lược.

Nhưng nếu động lực cốt lõi là sợ hãi: sợ bị tấn công, sợ mất ưu thế, sợ bị tụt hậu, thì công nghệ chỉ là phương tiện khuếch đại tâm lý ấy.

Phật giáo không phủ nhận nhu cầu tự vệ. Nhưng tự vệ dựa trên tỉnh thức khác với tự vệ dựa trên hoảng loạn. Một thế giới nơi các thuật toán tối ưu hóa chiến tranh có thể vận hành nhanh hơn khả năng suy ngẫm đạo đức của con người là một thế giới tiềm ẩn rủi ro lớn.

AI có thể "giữ giới" không?

Câu trả lời trung thực có lẽ là: AI không thể giữ giới nhưng con người có thể đặt giới hạn cho AI.

Giới luật không phải là sản phẩm của mạch điện. Đó là kết quả của tiến trình tu tập. Nếu con người không có nội lực đạo đức đủ vững, thì dù công nghệ có được ràng buộc bằng bao nhiêu nguyên tắc, vẫn có thể bị sử dụng theo hướng gây hại.

Cuộc tranh luận giữa Anthropic và chính phủ Hoa Kỳ cho thấy một điểm đáng suy ngẫm: trong kỷ nguyên AI, xung đột không chỉ diễn ra giữa các quốc gia, mà còn giữa các quan niệm về trách nhiệm và giới hạn đạo đức.

Có lẽ câu hỏi quan trọng nhất không phải là: Liệu máy móc, hay trí tuệ nhân tạo có đạo đức không?

Mà là: Khi trao ngày càng nhiều quyền lực cho thuật toán, con người có còn đủ tỉnh thức để tự giữ giới cho chính mình?

Trong thời đại mà chiến trường có thể được mô phỏng bằng dữ liệu và mục tiêu được xác định bằng mô hình thuật toán công nghệ cao, lời dạy về chính niệm và từ bi không trở nên lỗi thời. Ngược lại, càng trở nên cấp thiết.

Bởi nếu trí tuệ không đi kèm từ bi, đó có thể trở thành công cụ tinh vi tái chế khổ đau. Và khi thuật toán bước vào chiến trường, câu hỏi về giới luật không còn là vấn đề của tu viện, mà là vấn đề của toàn nhân loại.

Nguồn: **The Morning Dispatch**

hello@newsletter.thedispatch.com

Chuyển ngữ và biên tập: **Thường Nguyên**

Nội dung gốc tiếng Anh:

Tiêu đề ảnh: Model Behavior (tạm dịch: Hành vi của mô hình). Giám đốc điều hành kiêm đồng sáng lập Anthropic, Dario Amodei và Bộ trưởng Quốc phòng Pete Hegseth. (Ảnh: Michael M. Santiago/Getty Images và Chip Somodevilla/Getty Images).

Claude, Anthropic's AI system, is likely playing a key role in America's ongoing war with Iran. U.S. Central Command, which oversees the Middle East, uses Claude for planning, operations, battle simulations, and target identification, the Wall Street Journal confirmed Saturday. But the Trump administration has also just labeled Claude a threat to U.S. national security.

The U.S. attacks on Iran came just hours after President Donald Trump directed all U.S. government agencies with a six-month transition period for the Department of Defense to stop using Anthropic's tools. Trump's announcement capped a roughly weeklong saga in which a dispute between the Pentagon and Anthropic over how AI technology can be used in weapons systems and surveillance boiled over, ending Anthropic's $200 million contract with the Department of Defense. Defense Secretary Pete Hegseth then designated the company a supply-chain risk, declaring that no military contractor, supplier, or partner may conduct any commercial activity with Anthropic.

The breakdown marks the first major collision between AI safety policy and government interests. But what led up to this? Do existing laws actually protect against the uses Anthropic tried to block? And what does the fallout mean for Anthropic and for the Pentagon's relationship with Silicon Valley?

Anthropic, which has cultivated a reputation as the frontier AI company most committed to safety research and the ethical dimensions of its products, said negotiations with the government failed over a matter of principle: It refuses to allow its models to be used for deploying autonomous weapons or for mass surveillance of U.S. citizens. “We support all lawful uses of AI for national security aside from the two narrow exceptions above”, the company said in a press release on Friday, adding “we do not believe that today’s frontier AI models are reliable enough to be used in fully autonomous weapons” and “we believe that mass domestic surveillance of Americans constitutes a violation of fundamental rights”.

The White House and Pentagon maintain they have the sole authority to decide how to use AI tools. Jeremy Lewin, the acting undersecretary of state for foreign assistance, humanitarian affairs, and religious freedom, tweeted, "This isn’t about Anthropic or the specific conditions at issue. It’s about the broader premise that technology deeply embedded in our military must be under the exclusive control of our duly elected/appointed leaders”.