The Clean Paper

WHAT THE PAPER SAYS · WHAT IT DOESN'T · WHY IT MATTERS

Robot "foundation model" lớn có thật sự tốt hơn? Câu trả lời cẩn thận: có, vừa phải — và nhiều nghiên cứu khó phân biệt được

Toyota Research Institute huấn luyện "large behavior models" trên ~1.700 giờ thao tác robot đa dạng và so với chính sách một nhiệm vụ huấn luyện từ đầu bằng thử nghiệm mù, ngẫu nhiên hóa, cỡ mẫu lớn (~1.800 lượt thật, >47.000 mô phỏng) và thống kê đúng. Sau fine-tune theo nhiệm vụ, mô hình lớn tốt hơn trung bình, cần ít hơn khoảng 3–5× dữ liệu đặc thù và bền hơn khi điều kiện đổi; hiệu năng tăng trơn tru theo dữ liệu. Nhưng không fine-tune chúng không nhất quán thắng baseline, nhiều hiệu ứng rất nhỏ, và chuẩn hóa dữ liệu tầm thường còn quan trọng hơn kiến trúc. Đây là ủng hộ vừa phải cho hướng robot foundation model — không phải robot đa dụng, generalist zero-shot hay "cú nhảy emergent" — cùng cảnh báo rằng nhiều nghiên cứu robot có thể đang đo nhiễu.

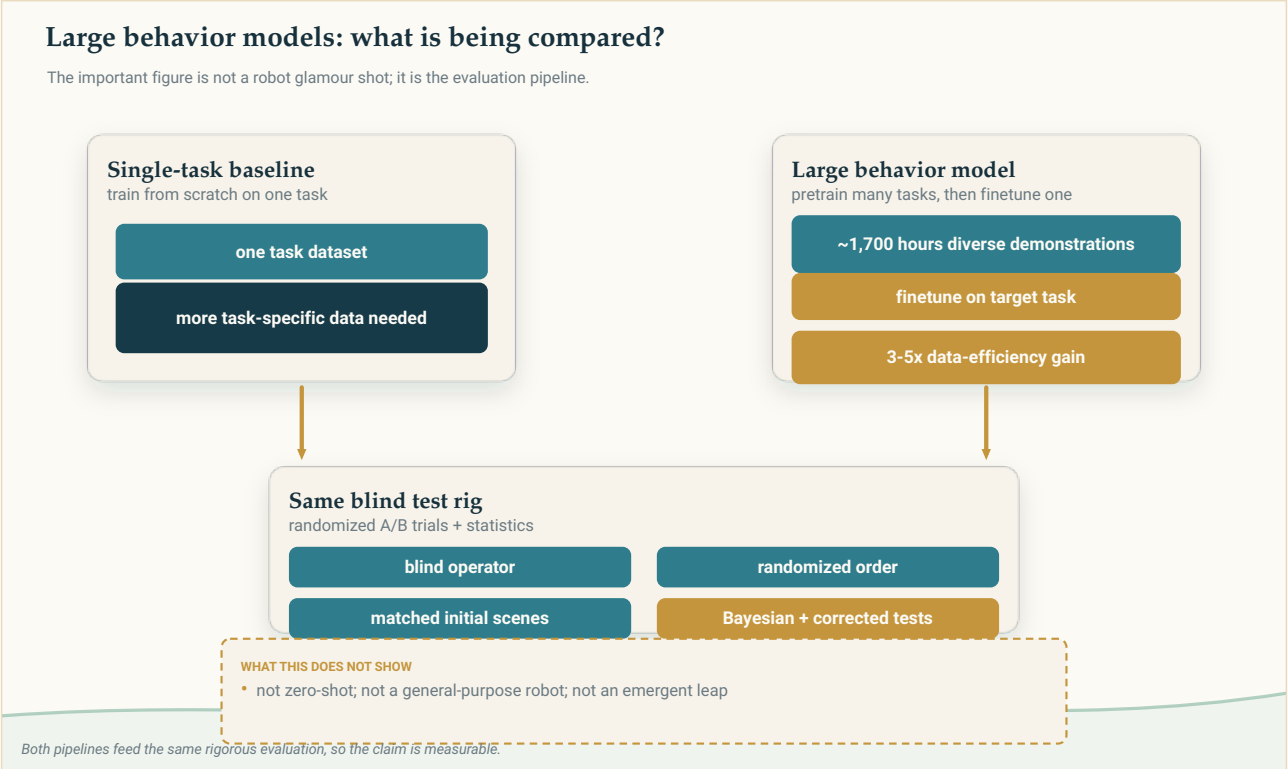
Written by Lucio Vaglio · Cross-checked by Laura Nesso · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

Paper: *A Careful Examination of Large Behavior Models for Multitask Dexterous Manipulation*, Toyota Research Institute Large Behavior Model Team — J. Barreiros, A. Beaulieu, et al.; senior authors incl. R. Ambrus, B. Burchfiel, S. Feng, H. Kress-Gazit (Cornell), R. Tedrake, Science Robotics (2026); preprint arXiv:2507.05331 · DOI: 10.1126/scirobotics.aea6201

Một bài báo robot mà chủ đề thật sự là sự trung thực

Robot học đang ở giữa một làn sóng "foundation model". Ý tưởng, mượn từ AI ngôn ngữ và hình ảnh, rất hấp dẫn: thay vì huấn luyện robot từng nhiệm vụ một, hãy huấn luyện một **"large behavior model" (LBM)** lớn trên một kho trình diễn khổng lồ và đa dạng, rồi thu được hệ thống có năng lực rộng và thích nghi nhanh. Sự hào hứng — cùng tiền đầu tư — là rất lớn. Tiêu đề tự viết: *bộ não robot đa dụng đã đến*.

Bài báo từ Toyota Research Institute này thú vị chính vì từ chối viết tiêu đề đó. Đóng góp thật sự không phải một robot hào nhoáng hơn. Nó là cái nhìn nghiêm túc vào câu hỏi tưởng đơn giản — *cách tiếp cận mô hình lớn có thật sự tốt hơn không, và làm sao chúng ta biết?* — với mức cẩn thận thống kê mà theo chính tác giả là hiếm trong lĩnh vực. Kết quả là một câu "có, nhưng" rất hữu ích, cùng cảnh báo rằng nhiều công trình robot có thể chỉ đang đo nhiễu.



Hai cách tiếp cận đều đi vào cùng một đánh giá mù và ngẫu nhiên hóa. Bài báo không tuyên bố robot foundation model là hệ tổng quát zero-shot, mà rằng tiền huấn luyện có thể cải thiện hiệu quả dữ liệu và độ bền khi được đo cẩn thận.

Các tác giả đã làm gì

Họ xây LBM theo nghĩa cụ thể: **chính sách thị giác-vận động dựa trên diffusion** (một Diffusion Transformer đọc ảnh camera, chỉ dẫn ngôn ngữ ngắn và vị trí khớp của robot, rồi phát ra các cụm lệnh vận động ngắn ở 10 Hz). Các mô hình được **tiền huấn luyện trên khoảng 1.700 giờ** trình diễn robot — hơn 500 nhiệm vụ khác nhau thu nội bộ cộng dữ liệu công khai — rồi **fine-tune** theo từng nhiệm vụ. Đối chứng xuyên suốt là **chính sách một nhiệm vụ huấn luyện từ đầu** chỉ trên dữ liệu của nhiệm vụ đó.

Nhưng trái tim bài báo là *đánh giá*, thứ tác giả xem như kết quả chính. Để tránh tự lừa mình, họ dùng:

- **Thử A/B mù, ngẫu nhiên hóa** ngoài đời thật — người vận hành robot không biết chính sách nào đang được thử, và thứ tự được xáo ngẫu nhiên.
- **Điều kiện ban đầu kiểm soát và lặp lại được** — trước mỗi lượt, người vận hành căn cảnh vật theo lớp ảnh phủ chuẩn.
- **Số lượt lớn** — 50 rollout thật cho mỗi nhiệm vụ, mỗi chính sách, mỗi điều kiện; 200 mỗi nhiệm vụ trong mô phỏng. Tổng cộng khoảng **1.800 rollout thật mù và hơn 47.000 rollout mô phỏng**.
- **Thống kê đúng nghĩa** — ước lượng Bayesian xác suất thành công và kiểm định giả thuyết từng cặp có hiệu chỉnh so sánh nhiều lần, thay vì nhìn bằng mắt các thanh sai số chồng nhau. Họ còn kiểm tra chất lượng một phần tư các trial do người chấm để đo sai số chấm.

[video pending]

So sánh cạnh nhau các mô hình chuẩn bị bàn ăn sáng: bên trái là baseline một nhiệm vụ, bên phải là LBM. Cả hai video chạy tốc độ 1×. Đây là một nhiệm vụ được đánh giá, không phải bằng chứng về tự chủ đa dụng.

Bộ máy đánh giá này chính là điểm cốt lõi. Toàn bài lập luận rằng nếu thiếu nó, ta không thể phân biệt cải thiện thật với may mắn.

Họ tìm thấy gì

Mô hình lớn sau fine-tune thắng mô hình một nhiệm vụ huấn luyện từ đầu — trung bình. Gộp qua các nhiệm vụ, LBM được tiền huấn luyện rồi fine-tune đáng tin cậy hơn chính sách huấn luyện từ đầu trên cùng dữ liệu nhiệm vụ, cả mô phỏng và thế giới thật, với khác biệt có ý nghĩa thống kê. Ở từng nhiệm vụ, LBM fine-tune bằng hoặc tốt hơn baseline theo thống kê trong gần mọi trường hợp (3/3 nhiệm vụ thật, 15/16 mô phỏng).

Lợi ích lớn và rõ nhất là hiệu quả dữ liệu. LBM fine-tune đạt hiệu năng tương đương huấn luyện từ đầu với **ít hơn khoảng 3–5 lần dữ liệu đặc thù nhiệm vụ**. Trong một nhiệm vụ thật — bày bàn ăn sáng — LBM fine-tune chỉ với **15%** số trình diễn vẫn thắng chính sách huấn luyện từ đầu bằng **100%** dữ liệu.

Tiền huấn luyện giúp nhất khi điều kiện thay đổi. Khi môi trường kiểm tra cố ý lệch khỏi điều kiện huấn luyện ("distribution shift"), lợi thế của LBM fine-tune *tăng*. Trong một bộ mô phỏng, nó thắng baseline có ý nghĩa ở 3/16 nhiệm vụ trong điều kiện bình thường nhưng 10/16 khi lệch phân phối. Vì triển khai thật luôn trôi khỏi điều kiện huấn luyện, độ bền này có thể là phát hiện thực dụng quan trọng nhất.

Thêm dữ liệu tiền huấn luyện giúp một cách trơn tru. Hiệu năng tăng đều khi thêm dữ liệu — **không có cú nhảy đột ngột hay "emergent leap"** ở quy mô đã thử. Hữu ích, dự đoán được và không kịch tính.

Nhưng câu chuyện "generalist không cần fine-tune" không đứng vững. LBM tiền huấn luyện dùng *zero-shot* — không fine-tune cho nhiệm vụ — *không* nhất quán thắng chính sách một nhiệm vụ. Một mạng có thể làm nhiều nhiệm vụ, nhưng giấc mơ "chỉ cần ra lệnh bằng prompt" không được chứng minh ở đây; tác giả cho rằng một phần do bộ mã hóa ngôn ngữ nhỏ của họ còn giòn.

Và lợi ích nhỏ đến mức rất dễ bỏ lỡ — hoặc vô tình tạo giả. Nhiều hiệu ứng chỉ hiện ra với cỡ mẫu lớn hơn thông thường và kiểm định cẩn thận. Các tác giả nói thẳng rằng với kích thước hiệu ứng và mức nhiễu, **có rủi ro đáng kể nhiều bài robot đang đo nhiễu thống kê.** Họ cũng thấy một lựa chọn rất tầm thường — cách *chuẩn hóa dữ liệu* — ảnh hưởng kết quả nhiều hơn thay đổi kiến trúc, và một **bug chuẩn hóa** trong tiền huấn luyện chỉ lộ ra *sau khi* đánh giá đã xong.

Điều này có thể có ý nghĩa gì

Cách đọc có thể bảo vệ được: tiền huấn luyện quy mô lớn trên dữ liệu robot đa dạng là một thành phần thật và đáng giá — nó giảm dữ liệu cần cho mỗi nhiệm vụ mới và khiến chính sách vững hơn khi thế giới không giống tập huấn luyện. Điều đó thực sự ủng hộ hướng mà lĩnh vực đang đặt cược. Nhưng lợi ích *vừa phải và có điều kiện* (chủ yếu sau fine-tune, rõ nhất khi gộp và dưới stress), không phải sự xuất hiện của robot tổng quát "cắm vào là chạy".

Ý nghĩa yên tĩnh nhưng quan trọng hơn là phương pháp. Bài báo về thực chất là một thước đo: cho thấy cần bao nhiêu bằng chứng để đưa ra tuyên bố đáng tin về chính sách robot, và ngầm nói rằng rất nhiều hào hứng đã xuất bản dựa trên quá ít dữ liệu. Lĩnh vực cần sự chỉnh sửa này hơn thêm một model mới.

Điều này *không* chứng minh gì

- Đây **không phải robot đa dụng**. Các lợi ích được chứng minh cho một kiến trúc cụ thể (diffusion policies), fine-tune từng nhiệm vụ, trong điều kiện kiểm soát và từ trình diễn teleoperation — không phải robot tự trị nhận mọi việc mới theo lệnh.
- Nó **không** xác nhận dùng **zero-shot**. Không fine-tune, mô hình lớn không nhất quán thắng

baseline một nhiệm vụ.
- Đây **không** phải bằng chứng về **"cú nhảy emergent"**. Scaling cải thiện trơn tru; không có điểm gián đoạn hỗ trợ câu chuyện "rồi bỗng nhiên nó biết làm".
- Các con số **tương đối và bị giới hạn trong lab**. Tỷ lệ thành công tuyệt đối được cố ý điều chỉnh quanh ~50% để nhạy với so sánh; chúng không đo độ tin cậy triển khai thật, và đây là một kiến trúc từ một lab.
- Nó **không** giải thích **vì sao** từng chính sách thành công hay thất bại; một số nhiệm vụ cụ thể nơi mô hình lớn làm *tệ hơn* được báo nhưng chưa giải thích.
- Nó **không nói gì về an toàn, tự chủ hay triển khai** ngoài giàn đánh giá.

Bằng chứng mạnh đến đâu?

Với các khẳng định so sánh trung tâm — LBM fine-tune thắng baseline từ đầu khi gộp, cần ít dữ liệu hơn nhiều và bền hơn dưới distribution shift — bằng chứng mạnh *và* được kiểm soát bất thường tốt: mù, ngẫu nhiên hóa, cỡ mẫu lớn, kiểm định thống kê, kèm QA chấm điểm. Đây là trường hợp hiếm phương pháp đủ vững để lấy kết luận chính gần với giá trị bề mặt.

Các cảnh báo trung thực chính là những thứ tác giả tự nêu. Thanh sai số của họ bắt độ ngẫu nhiên của *đánh giá* nhưng **không** bắt độ ngẫu nhiên của *huấn luyện* — huấn luyện cùng model hai lần có thể cho chính sách khác đáng kể, và biến thiên ấy không nằm trong thống kê. Nhiệm vụ thật có 50 trial, đủ bắt hiệu ứng trung bình nhưng có thể bỏ lỡ hiệu ứng nhỏ. Điều kiện ngôn ngữ dùng encoder khiêm tốn, nên tuyên bố kiểu "chỉ nói robot làm gì" có thể khác ở hệ lớn hơn. Và còn việc công khai một bug chuẩn hóa phát hiện hậu nghiệm. Không điểm nào đánh chìm kết quả chính; ngược lại, đó chính là loại vấn đề bài báo nói lĩnh vực thường bỏ qua.

Một ghi chú nguồn: bài giải thích này dựa trên preprint của tác giả. Chúng tôi không lấy được phiên bản đã xuất bản ở tạp chí, nên chưa kiểm tra có thay đổi gì giữa preprint và bản xuất bản hay không.

Vì sao điều này quan trọng

"Robot foundation model" là cụm từ sinh ra để bị thổi phồng, và nghiên cứu như vậy dễ bị đọc lệch theo cả hai hướng — *"nó hoạt động!"* đầy chiến thắng hoặc *"toàn hype"* đầy bác bỏ. Cách đọc chính xác hữu ích hơn: tiền huấn luyện trên dữ liệu đa dạng mang lại lợi ích thật, đo được nhưng vừa phải — chủ yếu ít dữ liệu mỗi nhiệm vụ hơn và bền hơn — và cải thiện có thể dự đoán theo quy mô.

Lý do sâu hơn là bài báo quay sự nghiêm khắc vào chính lĩnh vực của mình. Bằng cách cho thấy hiệu ứng thật nhỏ đến mức biến mất dưới đánh giá cẩu thả, và một lựa chọn nhàm chán như chuẩn hóa dữ liệu có thể quan trọng hơn kiến trúc thông minh, nó lập luận rằng nhiều tiến bộ robot-learning cần phép đo chắc hơn trước khi đáng tin. Một bài báo dùng uy tín của mình để canh ranh giới giữa kết quả và mong ước đang làm điều hiếm và có giá trị hơn việc

đứng đầu leaderboard.

Tóm tắt gọn

Các nhà nghiên cứu tại Toyota Research Institute huấn luyện "large behavior models" — chính sách robot diffusion được tiền huấn luyện trên ~1.700 giờ dữ liệu thao tác đa dạng — rồi so với chính sách một nhiệm vụ huấn luyện từ đầu bằng giao thức bất thường nghiêm ngặt: mù, ngẫu nhiên hóa, cỡ mẫu lớn (≈1.800 trial thật và >47.000 mô phỏng), có thống kê đúng. Sau fine-tune từng nhiệm vụ, mô hình lớn đáng tin cậy tốt hơn khi gộp, cần **ít hơn khoảng 3–5× dữ liệu nhiệm vụ** và **bền hơn khi điều kiện thay đổi**, còn hiệu năng tăng trơn tru khi dữ liệu tiền huấn luyện tăng. Nhưng **không fine-tune** chúng *không* nhất quán thắng mô hình một nhiệm vụ, nhiều hiệu ứng nhỏ đến mức chỉ cỡ mẫu lớn mới thấy, và lựa chọn chuẩn hóa dữ liệu tầm thường còn quan trọng hơn kiến trúc. Đây là ủng hộ đo lường cẩn thận cho hướng robot foundation model — **không** phải robot đa dụng, không phải generalist zero-shot, không phải "emergent leap" — cùng cảnh báo thẳng rằng nhiều bài robot có thể đang đo nhiễu.

Đánh giá thẳng thắn

Bài báo cho thấy gì: Với đánh giá nghiêm ngặt, mù và đủ lực thống kê (≈1.800 rollout thật + >47.000 mô phỏng), diffusion policies được tiền huấn luyện đa nhiệm rồi fine-tune (LBM) tốt hơn chính sách một nhiệm vụ từ đầu khi gộp, đạt hiệu năng tương đương với ~3–5× ít dữ liệu nhiệm vụ hơn, bền hơn dưới distribution shift; hiệu năng tăng trơn tru theo dữ liệu tiền huấn luyện.

Điều hợp lý nhưng chưa được chứng minh: Các lợi ích này chuyển sang vision-language-action model lớn hơn nhiều (encoder ngôn ngữ ở đây nhỏ); scaling trơn tru tiếp tục ngoài phạm vi dữ liệu đã thử.

Điều nó không cho thấy: Robot đa dụng hay zero-shot (không fine-tune → không có lợi thế nhất quán); bất kỳ cú nhảy "emergent" nào; lời giải thích cho thất bại ở từng nhiệm vụ; độ tin cậy tuyệt đối ngoài đời (tỷ lệ thành công được điều chỉnh gần 50% để tăng độ nhạy); bất cứ điều gì về an toàn hay triển khai tự trị.

Hạn chế chính: Thống kê phản ánh ngẫu nhiên của đánh giá nhưng không phải của từng lần huấn luyện; 50 trial thật mỗi nhiệm vụ có thể bỏ lỡ hiệu ứng nhỏ; một kiến trúc và một lab; encoder ngôn ngữ khiêm tốn; bug chuẩn hóa dữ liệu chỉ phát hiện sau đánh giá; phân tích dựa trên preprint (chưa kiểm bản xuất bản).

Một độc giả phổ thông nên tự tin đến mức nào? Tự tin cao rằng tiền huấn luyện đa nhiệm + fine-tune mang lợi ích thật nhưng vừa phải — đặc biệt về hiệu quả dữ liệu và độ bền — và được đo bất thường cẩn thận. Tự tin cao rằng đây **không** phải robot đa dụng/zero-shot và **không** phải cú nhảy emergent. Tự tin vừa về mức lợi ích khi scale lên model lớn hơn. Và đáng nghiêm

túc với cảnh báo của chính tác giả: hiệu ứng trong lĩnh vực nhỏ đến mức nghiên cứu thiếu lực thống kê có thể đang báo cáo nhiễu. Thái độ phù hợp: lạc quan có chừng mực với hướng tiếp cận và hoài nghi lành mạnh với kết quả robot-AI thiếu nền thống kê kiểu này.

Nguồn

arXiv:2507.05331
https://arxiv.org/abs/2507.05331

Peer-reviewed version (not retrieved) — Science Robotics, 10.1126/scirobotics.aea6201
https://doi.org/10.1126/scirobotics.aea6201

Project page
https://toyotaresearchinstitute.github.io/lbm1/