The Clean Paper

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

क्याबड़े robot "foundation models" सचमुच बेहतर कम करते हैं? सावधानीसे मापाजवाब: थोड़ाहाँ— और अधकिांश अध्ययन शायद बताहीनहीपाते

Toyota Research Institute ने "large behavior models" — लगभग 1,700 घंटे के ववविध manipulation data पर pretraining पाए robot policies — कोunusual rigor के साथ शुरुआत-से प्रशक्षिति single-task policies से तुलनाकी blind, randomized, बड़े नमूने वाले trials (लगभग 1,800 real-world और 47,000+ simulation runs) और वास्तवकि statistics। प्रतत-task fine-tuning के बाद बड़े models औसतन बेहतर रहे, लगभग 3–5× कम task-specific data चाहए थाऔर परस्थितियोंबदलने पर वे अधकि robust थे; अधकि pretraining data के साथ performance धीरे-धीरे बढ़ी। लेकनि fine-tuning के बनिावे लगातार single-task models कोनहींहराते; कई effects इतने छोटे थे कि बड़े नमूनोंके बनिादखिते हीनहीं, और सावधान data-normalisation choice architecture से अधकि महत्वपूर्ण नकिली। यह robot-foundation-model दशिाके पक्ष में मापाहुआ evidence है — general-purpose robot याzero-shot generalist नहीं, और कोई "emergent leap" भीनहीं— साथ हीचेतावनीकि robotics में बहुत कुछ signal के बजाय noise मापाजासकताहै।

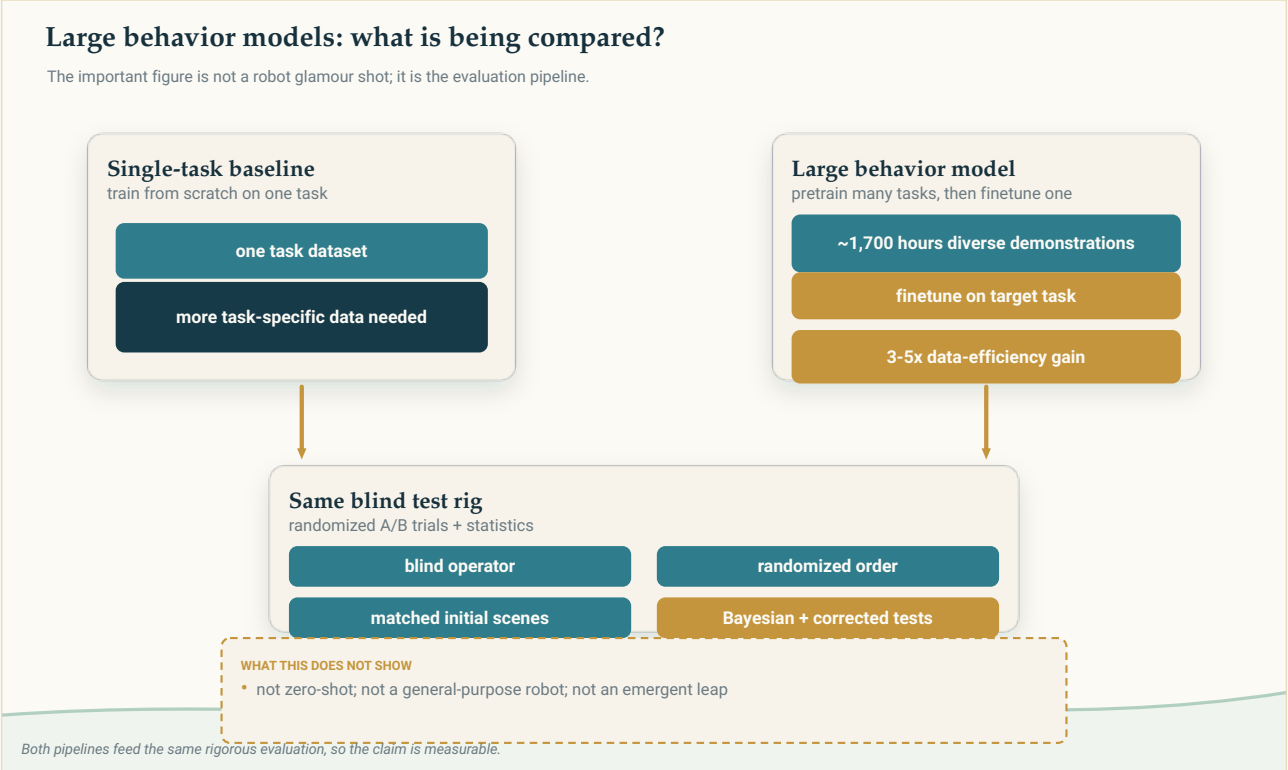
Written by Lucio Vaglio · Cross-checked by Laura Nesso · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

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

ऐसा रोबोटिक्स शोधपत्र जिसका असली विषय ईमानदारी है

रोबोटिक्स इस समय “foundation model” की दौड़ के बीच है। भाषा और छवि AI से लिया गया विचार आकर्षक है: हर काम के लिए रोबोट को अलग से प्रशिक्षित करने के बजाय, विविध प्रदर्शनों के बहुत बड़े संग्रह पर एक बड़ा **“large behavior model” (LBM)** प्रशिक्षित किया जाए और ऐसी प्रणाली मिले जो कई तरह के काम कर सके तथा नए कामों के लिए जल्दी अनुकूलित हो जाए। उत्साह — और निवेश — दोनों बहुत बड़े हैं। सुर्खी लगभग खुद लिख जाती है: *समान्य-उद्देश्य वाले रोबोट मस्तिष्क आ गए हैं।*

Toyota Research Institute का यह शोधपत्र ठीक इसलिए दिलचस्प है क्योंकि वह यह सुर्खी लिखने से इनकार करता है। इसका वास्तविक योगदान कोई अधिक चमकदार रोबोट नहीं, बल्कि एक भ्रमक रूप से सरल सवाल की कठोर जाँच है: *क्या बड़ा मॉडल सचमुच बेहतर काम करता है, और हम यह विश्वसनीय ढंग से कैसे जानेंगे?* लेखकों के अनुसार इस क्षेत्र में असमान्य स्तर की सांख्यिकीय सावधानी के साथ दिया गया उत्तर उपयोगी “हाँ, लेकिन…” है — और साथ ही चेतावनी कि रोबोटिक्स के बहुत-से अध्ययन शायद वास्तविक प्रभाव के बजाय शोर माप रहे हों।



दोनों तरफ़ को समान, blind और randomized मूल्यांकन से गुज़ारा गया। शोधपत्र का दावा यह नहीं कि robot foundation models बिना अतिरिक्त प्रशिक्षण के समान्य-उद्देश्य विशेषज्ञ हैं, बल्कि यह कि सावधानी से मापने पर pretraining डेटा-दक्षता और robustness सुधार सकता है।

लेखकों ने क्या किया

उन्होंने LBM को एक खास, ठोस अर्थ में बनाया: **diffusion-based visuomotor policies** — Diffusion Transformer जो कैमरे की छवियाँ, छोटा भाषा-निर्देश और रोबोट के अपने joints की स्थिति पढ़ता है और 10 Hz पर मोटर commands के छोटे समूह निकालता है। इन मॉडलों को लगभग **1,700 घंटे** के रोबोट प्रदर्शनों पर पहले से प्रशिक्षित किया गया — संस्थान के भीतर एकत्र 500 से अधिक अलग काम, साथ ही सार्वजनिक datasets — और फिर हर विशेष काम के लिए **fine-tune** किया गया। पूरे शोधपत्र में तुलना उस **single-task policy** से है जिसे उसी एक काम के डेटा पर शुरू से प्रशिक्षित किया गया।

लेकिन शोधपत्र का केंद्र वास्तव में *मूल्यांकन* है, जिसे लेखक मुख्य परिणाम जितना महत्व देते हैं। खुद को धोखा देने से बचने के लिए उन्होंने इस्तेमाल किया:

- वास्तविक दुनिया में **blind, randomized A/B testing** — रोबोट चलाने वाले इंसान को नहीं पता था कि कौन-सी policy परखी जा रही है और क्रम यादृच्छिक था।
- **नियंत्रित और दोहराई जा सकने वाली शुरुआती स्थितियाँ** — हर परीक्षण से पहले operator दृश्य को image overlay से मिलाता था।
- **बहुत बड़ी परीक्षण-संख्या** — हर काम, policy और परिस्थिति के लिए 50 वास्तविक-world rollouts; simulation में हर काम के लिए 200। कुल मिलाकर लगभग **1,800 blind वास्तविक-world rollouts और 47,000 से अधिक simulation rollouts**।
- **उचित सांख्यिकी** — सफलता की संभावना के Bayesian अनुमान और overlapping error bars को देखकर अंदाज़ लगाने के बजाय multiple-comparison corrections वाले pairwise hypothesis tests। इंसानों द्वारा आंकित परीक्षणों में scoring error मापने के लिए एक चौथाई trials पर quality-assurance pass भी किया गया।

[video pending]

नाश्ते की मेज़ लगाने वाले मॉडलों की साथ-साथ तुलना: बाएँ single-task baseline, दाएँ LBM। दोनों वीडियो 1x गति पर हैं। यह मूल्यांकन किया गया एक काम है, समस्या-उद्देश्य स्वयं तटस्थता का प्रमाण नहीं।

यही पूरी व्यवस्था असली मुद्दा है। शोधपत्र का तर्क है कि इसके बिना आप वास्तविक सुधार और किस्मत के बीच भरोसेमंद अंतर नहीं कर सकते।

उन्हें क्या मिला

Fine-tune किए गए बड़े मॉडल औसतन शुरू से प्रशिक्षित single-task मॉडलों से बेहतर रहे। सभी कामों को जोड़कर देखें तो पहले बड़े विविध dataset पर प्रशिक्षित और फिर खास काम के लिए fine-tune किया गया LBM उसी काम पर शुरू से प्रशिक्षित policy से simulation और वास्तविक दुनिया दोनों में विश्वसनीय रूप से बेहतर रहा, और अंतर सांख्यिकीय रूप से महत्वपूर्ण था। अलग-अलग कामों पर fine-tuned LBM लगभग हर मामले में सांख्यिकीय रूप से बराबर या बेहतर था — वास्तविक दुनिया में 3/3 काम और simulation में 15/16।

सबसे बड़ा और सबसे साफ़ फ़ायदा डेटा-दक्षता का था। Fine-tuned LBM ने शुरू से प्रशिक्षित मॉडल के बराबर प्रदर्शन लगभग **3–5 गुना कम task-specific डेटा** से हासिल किया। वास्तविक दुनिया के एक काम — नाश्ते की मेज़ लगाने — में केवल **15%** demonstrations पर fine-tune किया गया LBM, **100%** demonstrations पर शुरू से प्रशिक्षित policy से बेहतर रहा।

परिस्थितियाँ बदलने पर pretraining का फ़ायदा सबसे अधिक दिखा। जब परीक्षण-पर्यावरण को जानबूझकर प्रशिक्षण की स्थितियों से बदला गया — यानी "distribution shift" बनाया गया — fine-tuned LBM की बढ़त बड़ी हो गई। एक simulation

समूह में समान्य स्थितियोंमें वह 16 में से 3 कामोंपर शुरू से प्रशिक्षित मॉडल से संख्यकीय रूप से बेहतर था, लेकिन distribution shift में 16 में से 10 पर। वास्तविक तैनातीमें परिस्थितियाँप्रशिक्षण से कभीन कभीअलग होतीहीहैं, इसलिए व्यावहारिक दृष्टि से यह robustness शायद सबसे महत्वपूर्ण नतीजाहै।

अधिक pretraining डेटाने धीरे-धीरे मदद की। डेटाबढ़ने के साथ प्रदर्शन लगातार सुधरा— पर जाँचे गए पैमाने पर **कोई अचानक छलाँग या "emergent" क्षमतानहींदिखाई दी।** उपयोगी, अनुमान लगाने योग्य, लेकिन नाटकीय नहीं।

लेकिन बिनाfine-tuning वालाgeneralist दावानहींटिका। पहले से प्रशिक्षित LBM को*zero-shot* — यानीकिसीखास काम के लिए अतिरिक्त fine-tuning के बिना— इस्तेमाल करने पर उसने single-task policies कोलगातार नहींहराया। एक ही network कई काम कर सकताथा, लेकिन "बस निर्देश दोऔर काम होजाएगा" वालासपनायहाँसाकार नहींहुआ। लेखक इसका कुछ हिस्साअपने छोटे language encoder कीसीमाओं से जोड़ते हैं।

और सुधार इतने छोटे थे कि उन्हें न देख पाना— याझूठासुधार समझ लेना— आसान था। कई प्रभाव केवल समान्य से बड़े sample sizes और सावधान statistical tests के कारण दिखाई दिए। लेखक साफ़ लिखते हैं कि प्रभावोंके आकार और प्रयोगात्मक शोर कोदेखते हुए **इस बात कामहत्वपूर्ण जोखिम है कि रोबोटिक्स के कई शोधपत्र संख्यकीय शोर कोवास्तविक परिणाम समझ रहे हों।** उन्होंने यह भीपायाकि एक साधारण-साचुनाव — डेटाकोकिस तरह *normalise* कियाजाताहै — कुछ architectural बदलावोंसे अधिक असर डाल सकताहै। और pretraining में normalisation कीएक **bug** मूल्यांकन पूराहोने के *बाद* सामने आई।

इसकाशायद क्यामतलब है

सबसे बचाव योग्य व्याख्यायह है: विविध रोबोट डेटापर बड़े पैमाने कापretraining सचमुच उपयोगीहै — नए काम के लिए कम डेटाचाहिए और जब वास्तविक दुनियाप्रशिक्षण से मेल नहींखातीतोpolicy अधिक मजबूत रहतीहै। यह उस दिशा कोवास्तविक समर्थन देताहै जिस पर क्षेत्र दाँव लगारहाहै। लेकिन फ़ायदा*मध्यम और शर्तोंके साथ* है: वह मुख्यतः fine-tuning के बाद दिखताहै और सभीकामोंकोजोड़कर याकठिन परिस्थितियोंमें सबसे साफ़ होताहै। यह तैयार-उपयोग सामान्य रोबोट काआगमन नहींहै।

शांत लेकिन शायद अधिक महत्वपूर्ण अर्थ पद्धतिगत है। यह शोधपत्र व्यवहार में एक मापदंड है: भरोसेमंद robot-policy दावाकरने के लिए वास्तव में कितने सक्ष्य चाहिए, यह दिखाताहै — और संकेत देताहै कि प्रकाशित उत्साह का अच्छाहिस्साबहुत कम सक्ष्य पर टिकाहोसकताहै। इस क्षेत्र कोएक और मॉडल से अधिक ऐसे सुधारात्मक अध्ययन कीजरूरत है।

यह क्यासाबित नहींकरता

- यह **सामान्य-उद्देश्य रोबोट नहीं**है। बढ़त एक खास architecture — diffusion policies — के लिए, हर काम पर fine-tuning, नियंत्रित परिस्थितियोंऔर teleoperated demonstrations के साथ दिखाई गई है; ऐसास्वायत्त रोबोट नहींजो आदेश मिलते हीकोई भीनयाकाम कर दे।
- यह **zero-shot** उपयोग कोप्रमाणित **नहीं**करता। Fine-tuning के बिनाबड़ामॉडल single-task baselines कोलगातार नहीं हरापाया।
- यह किसी**"emergent छलाँग"** कासाक्ष्य **नहीं**है। Scaling से प्रदर्शन धीरे-धीरे सुधरा, यहाँऐसाdiscontinuity नहींहै जो "और फिर अचानक क्षमताआ गई" वालीकहानीकासमर्थन करे।
- संख्याएँ **सीमेक्ष और प्रयोगशालासीमित** हैं। तुलनाकोसंवेदनशील बनाने के लिए absolute success rates जानबूझकर

लगभग 50% के आसपास रखे गए; वे वास्तविक-world reliability का माप नहीं हैं। और यह एक लैब की एक architecture है।
- यह यह **नहीं** बताता कि कोई policy खास काम पर क्यों सफल या असफल होती है। कई ऐसे काम जहाँ बड़ा मॉडल *खराब* रहा, रिपोर्ट किए गए लेकिन समझाए नहीं गए।
- मूल्यांकन व्यवस्था के बाहर **सुरक्षा, स्वायत्तता या तैनाती** के बारे में यह कुछ नहीं बताता।

सक्ष्य कितना मजबूत है?

मुख्य तुलनात्मक दावों के लिए — कि fine-tuned LBMs सभी काम को जोड़ने पर शुरू से प्रशिक्षित baselines से बेहतर हैं, कई गुना कम डेटा चाहते हैं और distribution shift में अधिक मजबूत हैं — साक्ष्य मजबूत और असामान्य रूप से अच्छी तरह नियंत्रित है: blind, randomized, बड़े sample वाला सांख्यिकीय रूप से जाँचा गया और scoring पर QA pass के साथ। यह उन दुर्लभ मामलों में से है जहाँ पद्धति इतनी मजबूत है कि मुख्य निष्कर्षों को लगभग सीधे स्वीकार किया जा सकता है।

ईमानदार सीमाएँ वही हैं जिन्हें लेखक खुद उठाते हैं। Error bars *मूल्यांकन* की randomness को पकड़ते हैं, लेकिन **training-run randomness को नहीं** — एक ही मॉडल को दोबारा प्रशिक्षित करने पर अर्थपूर्ण रूप से अलग policy मिल सकती है और वह variation इन आँकड़ों में नहीं है। वास्तविक दुनिया के हर काम में 50 trials थे, जो मध्यम प्रभाव पकड़ने के लिए पर्याप्त हैं लेकिन छोटे प्रभाव छूट सकते हैं। भाषा-conditioning के लिए अपेक्षाकृत छोटा encoder था, इसलिए "रोबोट को बस बता दो क्या करना है" वाला दावा बड़े systems में अलग निकल सकता है। और evaluation के बाद मिली normalisation bug की साफ़ disclosure भी है। इनमें से कोई मुख्य निष्कर्ष को गिरफ़्तार नहीं, लेकिन यही वे बातें हैं जिन्हें यह शोधपत्र कहता है कि क्षेत्र अक्सर नज़रअंदाज़ कर देता है।

उसी भावना में स्रोत संबंधी एक टिप्पणी: यह व्याख्या लेखक के preprint पर आधारित है। हमें journal में प्रकाशित संस्करण नहीं मिला, इसलिए preprint और प्रकाशित पाठ के बीच संभावित बदलाव जाँचे नहीं गए।

यह क्यों मायने रखता है

"Robot foundation models" ऐसा वाक्यांश है जिसमें बढ़ा-चढ़ाकर दावा करना आसान है, और इस अध्ययन को दो दिशाओं में गलत पढ़ा जा सकता है — विजयी *"यह काम करता है!"* या dismissive *"सब hype है।"* सही व्याख्या दोनों से अधिक उपयोगी है: विविध डेटा पर pretraining वास्तविक, मापने योग्य लेकिन मध्यम लाभ देता है — मुख्यतः हर काम के लिए कम डेटा और अधिक robustness — और डेटा के पैमाने के साथ प्रदर्शन अनुमान लगाने योग्य ढंग से सुधरता है।

इसका गहरा महत्व यह है कि शोधपत्र अपनी कठोरता अपने ही क्षेत्र पर लागू करता है। यह दिखाकर कि वास्तविक प्रभाव इतने छोटे हैं कि कमजोर मूल्यांकन में गायब हो सकते हैं, और data normalisation जैसा उबाऊ चुनाव किसी चतुर नई architecture से अधिक असर डाल सकता है, वह यह मामला बनाता है कि robot-learning की बहुत-सी कथित प्रगति पर विश्वास करने से पहले बेहतर मापन चाहिए। परिणाम और इच्छा के बीच फर्क की पहरेदारी पर अपनी विश्वसनीयता खर्च करने वाला शोधपत्र लीडरबोर्ड में पहला स्थान पाने से अधिक दुर्लभ और अधिक मूल्यवान काम करता है।

संक्षेप में

Toyota Research Institute के शोधकर्ताओं ने “large behavior models” बनाए — diffusion-based robot policies जिन्हें लगभग 1,700 घंटे के विविध manipulation डेटा पर पहले से प्रशिक्षित किया गया — और उन्हें असामान्य रूप से कठोर protocol में शुरू से प्रशिक्षित single-task policies से तुलना की। blind, randomized, बड़े sample वाला मूल्यांकन, जिसमें लगभग **1,800 वास्तविक-world और 47,000 से अधिक simulation trials** तथा उचित सांख्यिकीय मदद थी। हर काम के लिए fine-tuning के बाद बड़े मॉडल सभी को जोड़कर विश्वसनीय रूप से बेहतर रहे, लगभग **3–5 गुना कम task-specific डेटा** से समान प्रदर्शन तक पहुँचे और **परिस्थितियाँ बदलने पर अधिक मजबूत** रहे। Pretraining डेटा बढ़ने के साथ प्रदर्शन धीरे-धीरे सुधरा। लेकिन **fine-tuning के बिना** उन्होंने single-task मॉडलों को लगातार नहीं हराया; कई प्रभाव इतने छोटे थे कि केवल बड़े sample ने उन्हें उजागर किया और साधारण data-normalisation का चुनाव architecture से अधिक महत्वपूर्ण निकला। यह robot-foundation-model दिशा के लिए ठोस लेकिन मिला-जुला समर्थन है — **सामान्य-उद्देश्य रोबोट नहीं**, zero-shot generalist नहीं और कोई “emergent छलाँग” नहीं — साथ ही तीखी चेतावनी कि रोबोटिक्स का अच्छा हिस्सा शायद शोर मात्र रहा हो।

बिना बढ़ा-चढ़ाकर जाँच

शोधपत्र क्या दिखाता है: कठोर, blind और पर्याप्त statistical power वाले मूल्यांकन — लगभग 1,800 वास्तविक-world और 47,000+ simulation rollouts — में multitask pretraining के बाद fine-tune की गई diffusion policies (LBMs) सभी को जोड़कर शुरू से प्रशिक्षित single-task policies से बेहतर हैं, लगभग 3–5 गुना कम task-specific डेटा से समान प्रदर्शन तक पहुँचती हैं और distribution shift में अधिक मजबूत हैं। Pretraining डेटा बढ़ने पर प्रदर्शन धीरे-धीरे सुधरता है।

क्या संभव है, पर साबित नहीं किए लाभ बहुत बड़े vision-language-action मॉडलों तक भी उसी तरह पहुँचेंगे — इस अध्ययन का language encoder छोटा था — और कि जाँची गई सीमा से बहुत आगे भी यही smooth scaling जारी रहेगी।

यह क्या नहीं दिखाता सामान्य-उद्देश्य या zero-shot रोबोट — fine-tuning नहीं तो लगातार बढ़त नहीं; कोई “emergent” क्षमता छलाँग; खास task-level failures की व्याख्या; absolute वास्तविक-world reliability — सफलता दरें तुलना को संवेदनशील बनाने के लिए लगभग 50% के पास रखी गई; या सुरक्षा और स्वायत्त तैनाती के बारे में कुछ भी।

मुख्य सीमाएँ: सांख्यिकी evaluation randomness को पकड़ती है, training-run randomness को नहीं; हर वास्तविक-world काम पर 50 trials छोटे प्रभाव चूक सकते हैं; एक architecture और एक लैब; अपेक्षाकृत छोटा language encoder; evaluation के बाद मिली data-normalisation bug; और विश्लेषण preprint पर आधारित है, प्रकाशित संस्करण की जाँच नहीं हुई।

सामान्य पाठक को कितना भरोसा रखना चाहिए? इस बात पर उच्च कि multitask pretraining + fine-tuning वास्तविक लेकिन मध्यम लाभ देता है — खासकर डेटा-दक्षता और robustness में — और इन्हें असामान्य सावधानी से मापा गया। इस बात पर भी उच्च कि यह **सामान्य-उद्देश्य या zero-shot रोबोट नहीं** और **emergent छलाँग नहीं** है। बहुत बड़े मॉडलों तक लाभ कितना बढ़ेगा, इस पर मध्यम भरोसा रखें। और लेख को की चेतावनी को गंभीरता से लें: इस क्षेत्र में प्रभाव इतने छोटे हो सकते हैं कि कम शक्ति वाले अध्ययन शोर को परिणाम समझ लें। उचित रुख है — इस दिशा पर मिला-जुला आशावाद, और ऐसे robot-AI दावों पर स्वस्थ संदेह जिनके पीछे इस स्तर की सांख्यिकीय जाँच नहीं है।

स्रोत

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/