



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

क्या बड़े robot “foundation models” सचमुच बेहतर काम करते हैं? सख्त धर्मी से मजाज व सख्त थोड़ा हँस — और अधिकांश अध्ययन शायद बताहीन ही पते

Toyota Research Institute ने “large behavior models” — लगभग 1,700 घंटे के विविध manipulation data पर pretraining पर robot policies — को unusual rigor के साथ शुरुआत-से प्रशिक्षित single-task policies से तुलना की blind, randomized, बड़े नमूने वाले trials (लगभग 1,800 real-world और 47,000+ simulation runs) और वास्तविक statistics। प्रति-task fine-tuning के बड़े बड़े models औसतन बेहतर रहे, लगभग 3–5× कम task-specific data चाहिए था और परस्थितियाँ बदलने पर वे अधिक robust थे; अधिक pretraining data के साथ performance धीरे-धीरे बढ़ी लेकिन fine-tuning के बिना वे लगभग single-task models को नहीं हराते; कई effects इतने छोटे थे कि बड़े नमूनों के बिना देखिते हीन नहीं और समझने data-normalisation choice architecture से अधिक महत्वपूर्ण नकली। यह robot-foundation-model दशिके पक्ष में मजाज आ evidence है — general-purpose robot या zero-shot generalist नहीं और कई “emergent leap” भी नहीं — साथ ही चेतनी की robotics में बहुत कुछ signal के बजाय noise मजाजा सकता है।

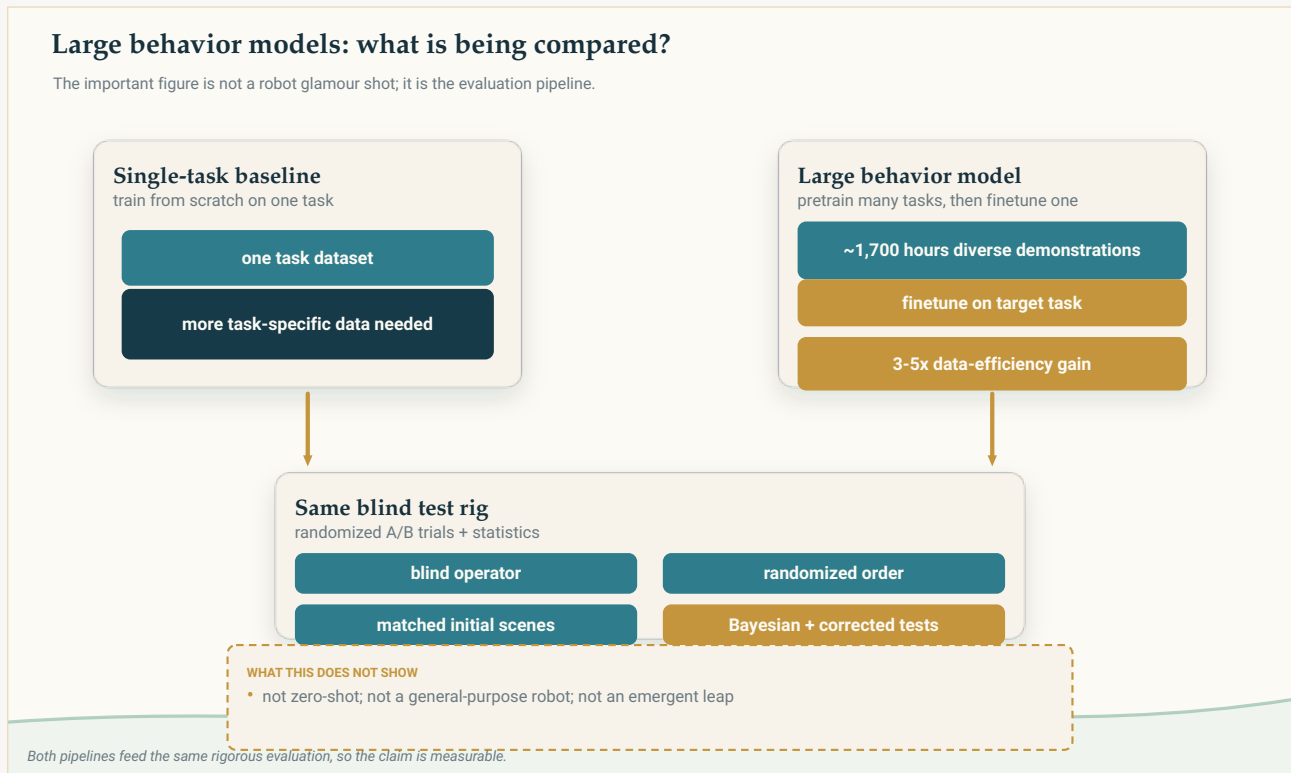
Written by Lucio Vaglio · Cross-checked by Laura Nesso · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

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

ऐसारेबोटिक्स शोधपत्र जिसका असली विषय ईममदरी है

रोबोटिक्स इस समय “foundation model” की दौड़ के बीच है। भण्डार और छवि AI से लिया गया विचार आकर्षक है: हर काम के लिए रोबोट को अलग से प्रशिक्षित करने के बजाय, विविध प्रदर्शन 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। उत्सह — और नविश — दोनों बहुत बड़े हैं। सुखील गभग खुद लिख जती है: *समस्य-उद्देश्य वाले रोबोट मस्तर्षिक आ गए हैं।*

Toyota Research Institute का यह शोधपत्र ठीक इसलिए दिलचस्प है क्योंकि वह यह सुखील 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 जमोंगे?* लेखकों के अनुसार इस क्षेत्र में असमस्य स्तर की सांख्यिकीय सन्नधनी के साथ दिया गया उत्तर उपयोगी “हाँ लेकिन...” है — और साथ ही चेतन की रोबोटिक्स के बहुत-से अध्ययन शायद वास्तविक प्रभु के बजाय शोर मच रहे हों।



दोनों रीकोसमस, blind और randomized मूल्यंकन से गुज़रा गया शोधपत्र का दावा यह नहीं कि robot foundation models बिना अतिरिक्त प्रशिक्षण के समस्य-उद्देश्य विशेषज्ञ हैं, बल्कि यह कि सन्नधनी से मचने पर pretraining डेटादक्षता और robustness सुधार सकता है।

Original The Clean Paper diagram CC BY 4.0

लेखकोने क्याकिया

उन्होंने LBM को एक खस, ठोस अर्थ में बनाया **diffusion-based visuomotor policies** — Diffusion Transformer जो कैमरे की छवियाँ छोटो भाग में रिदेश और रेबोट के अपने joints की स्थिति पढ़ता है और 10 Hz पर मोटर commands के छोटे समूह निकलता है। इन मॉडलों को लगभग 1,700 घंटे के रेबोट प्रदर्शनो पर पहले से प्रशिक्षित किया गया — संस्थान के भीतर एकत्र 500 से अधिक अलग-अलग, सघन हीस एवजन के 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 **मूल्यंकन** है, जिसे लेखक मुख्य परीक्षण जतिना महत्व देते हैं। खुद को धोखा देने से बचने के लिए उन्होंने इसे माल किया

- वास्तविक दुनिया में **blind, randomized A/B testing** — रेबोट चलते वैसे इसमें कोन ही पता था कि कैसा-सी policy परखी जा रही है और क्रम यह चूँकि था।
- **नियंत्रित और देहरेई** जासकने वाली शुरुआती स्थितियाँ — हर परीक्षण से पहले operator दृश्य को image overlay से मिलाता था।
- **बहुत बड़ी परीक्षण-संख्या** — हर काम, policy और परिस्थितिके 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 संख्यिकी** — सफलता की संभावना के Bayesian अनुमान और overlapping error bars को देखकर अंदाज़ लगाने के बजाय multiple-comparison corrections वाले pairwise hypothesis tests। इससे मोद्दवा अंकित परीक्षणों में scoring error मानने के लिए एक चौंकाई trials पर quality-assurance pass भी किया गया।

[video pending]

नए कीमेज़ लगाने वाले मॉडलों की सघन-सघन तुलना बाएँ single-task baseline, दाएँ LBM। दो मोवीडियो 1x गति पर हैं। यह मूल्यंकन किया गया एक काम है, समस्य-उद्देश्य स्वयत्तता का प्रमाण नहीं।

यही पूरी व्यवस्था असली मुद्दा है। शोधपत्र का तर्क है कि इसके बिना आप वास्तविक सुधार और कस्मिन् 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 खस 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 संख्यिकीय रूप से महत्वपूर्ण था। अलग-अलग कामों पर **fine-tuned LBM** लगभग हर मामले में संख्यिकीय रूप से बरबरा या बेहतर था — वास्तविक दुनिया में 3/3 काम और simulation में 15/16।

सबसे बड़ा और सबसे सफ़ा फ़ायदा डेटा दक्षता का था। **Fine-tuned LBM** ने शुरू से प्रशिक्षित मॉडल के बरबरा प्रदर्शन लगभग 3–5 गुना कम **task-specific** डेटा से हासिल किया। वास्तविक दुनिया के एक काम — नए कीमेज़ 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-पर्यवर्ण को जमा बूझकर प्रशिक्षण की स्थितियों से बदला गया — यानी “distribution shift” बनाया गया — **fine-tuned LBM** की बढ़त **बड़ी** होगई। एक simulation

समूह में समान्य स्थितियों में वह 16 में से 3 कमोपर शुरू से प्रशिक्षित मॉडल से संख्यिकीय रूप से बेहतर था लेकिन distribution shift में 16 में से 10 पर। वस्तुतः तैयारी पर स्थितियाँ प्रशिक्षण से कभीन कभी अलग होती हैं, इसलिए व्यवहारिक दृष्टि से यह robustness शब्द सबसे महत्वपूर्ण नतीजा है।

अधिक pretraining डेटा ने धीरे-धीरे मदद की। डेटा बढ़ने के साथ प्रदर्शन लगातार सुधरा — पर जाँचे गए पैमाने पर कोई अचानक छल्ला या “emergent” क्षमता नहीं दिखाई दी। उपयोगी अनुमान लगाने योग्य, लेकिन नष्ट नहीं।

लेकिन बिना fine-tuning के generalist दक्षता हीट की। पहले से प्रशिक्षित LLM को zero-shot — यानी सीखा हुआ काम के लिए अतिरिक्त fine-tuning के बिना — इस्तेमाल करने पर उसने single-task policies को लागू नहीं कर पाया। एक ही network कई काम कर सकता था लेकिन “बस निर्देश दो और काम हो जाएगा” वाला सपना यहाँ सच नहीं हुआ। लेखक इसका कुछ हिस्सा अपने छोटे language encoder की सीमाओं से जोड़ते हैं।

और सुधार इतने छोटे थे कि उन्हें न देख पता — या झूठा सुधार समझ लेना — आसानी से। कई प्रभाव केवल sample sizes और सन्नधन statistical tests के कारण दिखाई दिए। लेखक सफ़ा लिखते हैं कि प्रभावों के आकार और प्रयोगात्मक शोर को देखते हुए इस बात का महत्वपूर्ण ज्ञान है कि रैबेटेक्स के कई शोधपत्र संख्यिकीय शोर को वस्तुतः परणाम समझ रहे हों। उन्हें यह भी पता कि एक सन्नधन-साचुनन — डेटा को किस तरह normalise किया जाता है — कुछ architectural बदलावों से अधिक असर डाल सकता है। और pretraining में normalisation की एक bug मूल्यंकन पूरा होने के बाद सामने आई।

इसका शब्द क्या मतलब है

सबसे बचपन योग्य व्याख्या यह है: विविध रैबेटे डेटा पर बड़े पैमाने का pretraining सचमुच उपयोगी है — नए काम के लिए कम डेटा चाहिए और जब वस्तुतः दुनिया प्रशिक्षण से मेल नहीं खाती तो policy अधिक मजबूत रहती है। यह उस दिशा को वस्तुतः समर्थन देता है जिस पर क्षेत्र दौड़ लगा रहा है। लेकिन फ़ग़दामध्यम और शर्तों के साथ है: वह मुख्यतः fine-tuning के बाद दिखता है और सभी कामों को जोड़कर या कठिन परिस्थितियों में सबसे सफ़ा होता है। यह तैयार-उपयोग समान्य रैबेटे का आगमन नहीं है।

शंत लेकिन शब्द अधिक महत्वपूर्ण अर्थ पद्धतगित है। यह शोधपत्र व्यवहार में एक मज़दंड है: भरोसेमंद robot-policy दक्षता करने के लिए वस्तुतः में कितने सक्षम चाहिए, यह दिखता है — और संकेत देता है कि प्रकृति उत्सह का अच्छा हिस्सा बहुत कम सक्षम पर टिका हो सकता है। इस क्षेत्र को एक और मॉडल से अधिक ऐसे सुधारात्मक अध्ययन की जरूरत है।

यह क्या सच नहीं करता

- यह समान्य-उद्देश्य रैबेटे नहीं है। बढ़त एक खास architecture — diffusion policies — के लिए, हर काम पर fine-tuning, नियंत्रित परिस्थितियों और teleoperated demonstrations के साथ दिखाई गई है; ऐसा स्वयत्त रैबेटे नहीं जो आदेश मिलते ही कोई भी नया काम कर दे।
- यह zero-shot उपयोग को प्रमत्ति नहीं करता। Fine-tuning के बिना बड़ा मॉडल single-task baselines को लागू नहीं कर पाया।
- यह किसी “emergent छल्ला” का सक्षम नहीं है। Scaling से प्रदर्शन धीरे-धीरे सुधरा यहाँ ऐसा discontinuity नहीं है जो “और फ़रि अचानक क्षमता आ गई” वाली कहानी का समर्थन करे।
- संख्याएँ समेक और प्रयोगशाला सीमित हैं। तुलना को संवेदनशील बनाने के लिए absolute success rates जमाबूझकर

लगभग 50% के आसपास रखे गए; वे वास्तविक-world reliability काम में नहीं हैं। और यह एक लैब की एक architecture है।

- यह यह नहीं बताता कि कोई policy खराब कम पर क्यों सफल या असफल होती है। कई ऐसे कम जहाँ बड़ा मॉडल खराब रहा रिपोर्ट किए गए लेकिन समझाए नहीं गए।
- मूल्यंकन व्यवस्था के बहर सुरक्षा स्वयं प्रतीतितायें नतीके बारे में यह कुछ नहीं बताता।

समृद्ध कतिनामजबूत है?

मुख्य तुलनात्मक दलों के लिए — कि fine-tuned LLMs सभी कमों को जोड़ने पर शुरू से प्रशिक्षित baselines से बेहतर हैं, कई गुना कम डेटा चाहते हैं और distribution shift में अधिक मजबूत हैं — समृद्ध मजबूत और असमर्थ रूप से अच्छी तरह नियंत्रित है: blind, randomized, बड़े sample वला संख्यिकीय 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 जिनमें लेखक खुद उठते हैं। Error bars मूल्यंकन की randomness को पकड़ते हैं, लेकिन **training-run randomness** को नहीं — एक ही मॉडल को दो बार प्रशिक्षित करने पर अर्थपूर्ण रूप से अलग policy मल सकती है और वह variation इन आँकड़ों में नहीं है। वास्तविक दुनिया के हर कम में 50 trials थे, जो मध्यम प्रभाव पकड़ने के लिए पर्याप्त हैं लेकिन छोटे प्रभाव छूट सकते हैं। भाषा conditioning के लिए अपेक्षित छोटा encoder था इसलिए “रेग्रेट को बस बता दो क्या करना है” वला दबाव बड़े systems में अलग निकल सकता है। और evaluation के बग मल्टी normalisation bug की सामा disclosure भी है। इनमें से कई मुख्य निष्कर्षों को गिराना नहीं लेकिन यही वैसे हैं जिनमें यह शोध पत्र कहता है कि क्षेत्र अक्सर नज़रअंदाज़ कर देता है।

उसी भ्रमना में स्रोत संबंधी एक टिप्पणी यह व्याख्या लेखकों के preprint पर आधारित है। हमें journal में प्रकाशित संस्करण नहीं मिला इसलिए preprint और प्रकाशित पत्र के बीच संभावित बदलाव जाँचे नहीं गए।

यह क्यों मयने रखता है

“Robot foundation models” ऐसा वाक्यंश है जिसमें बढ़ाचढ़कर दवा करना आसानी है, और इस अध्ययन को दो मोदशियों में गलत पढ़ा जा सकता है — वजिरी “यह कम करता है!” या dismissive “सब hype है।” सही व्याख्या दोनों से अधिक उपयोगी है: विविध डेटा पर pretraining वास्तविक, मयने योग्य लेकिन मध्यम लाभ देता है — मुख्यतः हर कम 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 शोध पत्र अपनी कठोरता अपने ही क्षेत्र पर लागू करता है। यह दिखाकर कि वास्तविक प्रभाव इतने छोटे हैं कि कमजोर मूल्यंकन में गायब हो सकते हैं, और data normalisation जैसा उबका चुनना किसी चतुर नई architecture से अधिक असर डाल सकता है, वह यह मसलाबना है कि robot-learning की बहुत-सी कथित प्रगति पर विश्वास करने से पहले बेहतर मयन चाहिए। परिणाम और इच्छा के बीच फर्क की पहरेदारी पर अपनी विश्वासनीयता खर्च करने वाला शोध पत्र लीडरबोर्ड में पहला स्थान पाने से अधिक दुर्लभ और अधिक मूल्यवान कम 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 असमर्थ रूप से कठोर protocol में शुरू से प्रशिक्षित single-task policies से तुलना की blind, randomized, बड़े sample वलामूल्यंकन, जिसमें लगभग 1,800 वस्तु-विश्व और 47,000 से अधिक simulation trials तथा उचित संख्या की शामिल थी। हर कम के लिए fine-tuning के बड़े बड़े मॉडल सभी कम को जोड़कर विश्वसनीय रूप से बेहतर रहे, लगभग 3-5 गुना कम task-specific डेटा से समान प्रदर्शन तक पहुँचे और परस्थितियाँ बदलने पर अधिक मजबूत रहे। Pretraining डेटा बढ़ने के साथ प्रदर्शन धीरे-धीरे सुधरा। लेकिन fine-tuning के बिना उन्हें single-task मॉडलों को लगभग नहीं हरया कई प्रभाव इतने छोटे थे कि केवल बड़े sample ने उन्हें उजागर किया और सघन data-normalisation का चुनौतीपूर्ण architecture से अधिक महत्वपूर्ण नकिला। यह robot-foundation-model दिशा के लिए ठोस लेकिन मसाहूआ समर्थन है — समर्थ-उद्देश्य रेबोट नहीं zero-shot generalist नहीं और कोई “emergent छल्ला” नहीं — साथ ही तीखी चेतनी की रेबोटिक्स का अच्छा हिस्सा शायद शुरुआत रहा हो।

बनाबढ़ाकर जँच

शोधपत्र क्या दिखाता है: कठोर, blind और पर्याप्त statistical power वाले मूल्यंकन — लगभग 1,800 वस्तु-विश्व और 47,000+ simulation rollouts — में multitask pretraining के बड़े fine-tune की गई diffusion policies (LBMs) सभी कम 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 सन्नति नहीं कीये लम बहुत बड़े vision-language-action मॉडलों तक भी उसी तरह पहुँचेंगे — इस अध्ययन का language encoder छोटा था — और कि जँच की गई सीमा से बहुत आगे भी यही smooth scaling जारी रहेगी।

यह क्या नहीं दिखाता समर्थ-उद्देश्य या zero-shot रेबोट — fine-tuning नहीं तो लगभग बढ़त नहीं कोई “emergent” क्षमता छल्ला; खस task-level failures की वजह से absolute वस्तु-विश्व reliability — सफलता दरें तुलना को संवेदनशील बनाने के लिए लगभग 50% के पास रखी गई; या सुरक्षा और स्वयत्त तैयारी के बारे में कुछ भी।

मुख्य सीमाएँ: संख्या की evaluation randomness को पकड़ती है, training-run randomness को नहीं हर वस्तु-विश्व कम पर 50 trials छोटे प्रभाव चूक सकते हैं; एक architecture और एक लैब; अपेक्षित छोटा language encoder; evaluation के बड़े मल्टी data-normalisation bug; और विश्लेषण preprint पर आधारित है, प्रकृति संस्करण की जँच नहीं हुई।

समर्थ पठक को कतिना भरोसा रखना चाहिए? इस बात पर उच्च कि multitask pretraining + fine-tuning वस्तु-विश्व लेकिन मध्यम लम देता है — खसकर डेटा दृढ़ता और robustness में — और इन्हें असमर्थ सन्नधि से मसाहूआ। इस बात पर भी उच्च कि यह समर्थ-उद्देश्य या zero-shot रेबोट नहीं और emergent छल्ला नहीं है। बहुत बड़े मॉडलों तक लम कतिना बढ़ेगा इस पर मध्यम भरोसा रखें। और लेखकों की चेतनी गंभीरता से लें: इस क्षेत्र में प्रभाव इतने छोटे हो सकते हैं कि कम शक्ति वाले अध्ययन शुरुआत को परिणाम समझ लें। उचित रुख है — इस दिशा पर मसाहूआ आशावाद, और ऐसे robot-AI दलों पर स्वस्थ संदेह जिनके पीछे इस स्तर की संख्या की जँच नहीं है।

सूत्रे

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/>