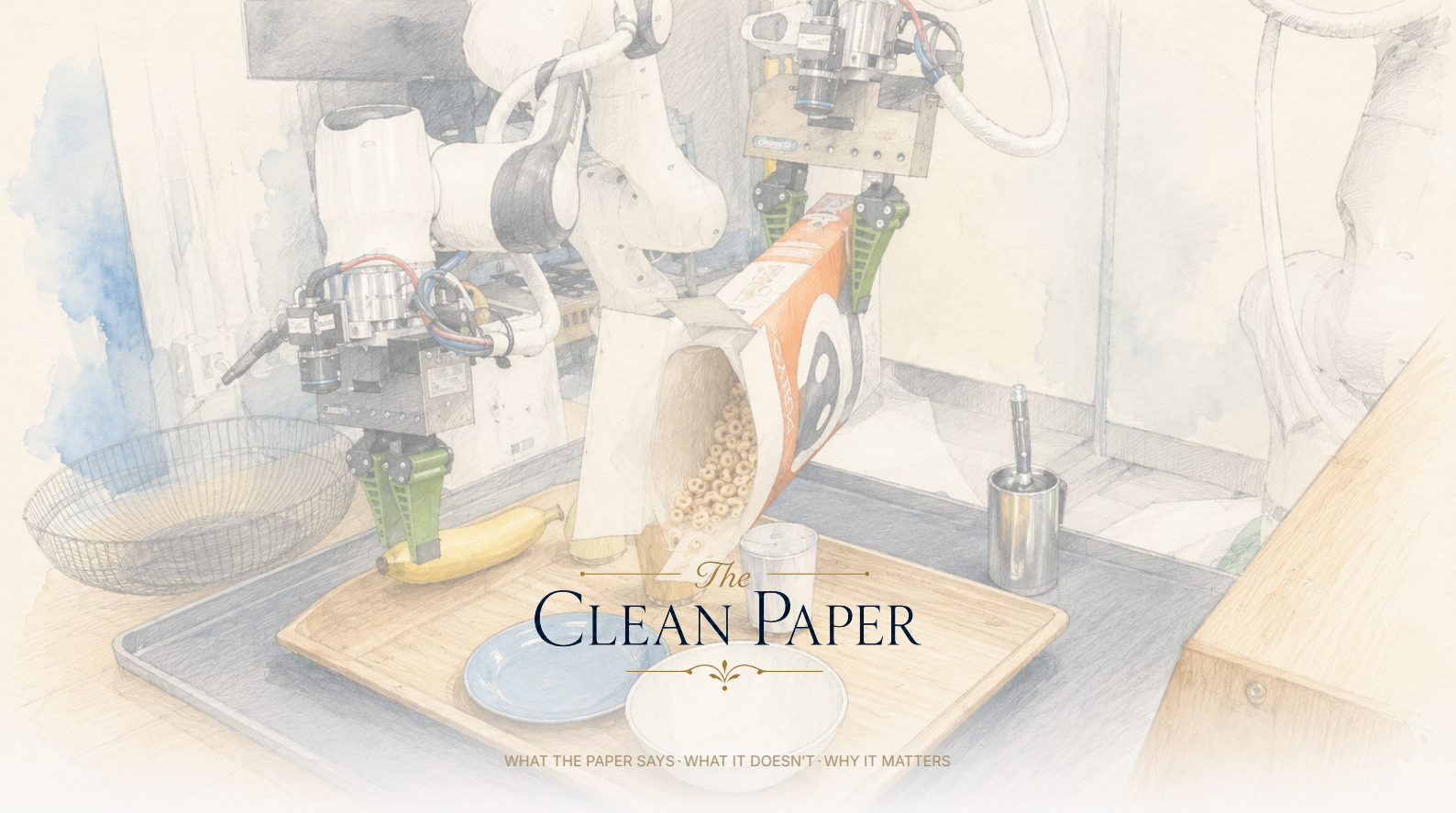


Чи справді великі «фундаментальні моделі» роботів працюють краще? Обережна відповідь: помірно так — і більшість досліджень не здатні цього надійно побачити

Toyota Research Institute навчив «великі моделі поведінки» — роботизовані політики з попереднім навчанням на ~1 700 годинах різноманітних маніпуляцій — і порівняв їх із моделями для однієї задачі, навченими з нуля, з незвичною суворістю: засліплені рандомізовані випробування великого масштабу (~1 800 реальних і 47 000+ симуляційних) зі справжньою статистикою. Після донавчання на конкретній задачі великі моделі в середньому працювали краще, потребували приблизно у 3–5 разів менше специфічних даних і були стійкішими до зміни умов; продуктивність плавно зростала з обсягом pretraining-даних. Але без донавчання вони не мали стабільної переваги над моделями для однієї задачі, деякі ефекти були настільки малі, що їх було видно лише у великих вибірках, а звичайний вибір нормалізації даних важив більше за архітектуру. Це стримана підтримка напряму фундаментальних моделей для роботів — не універсальний робот, не zero-shot універсал і не «емерджентний стрибок» — плюс гостре попередження, що значна частина робототехніки може вимірювати шум.

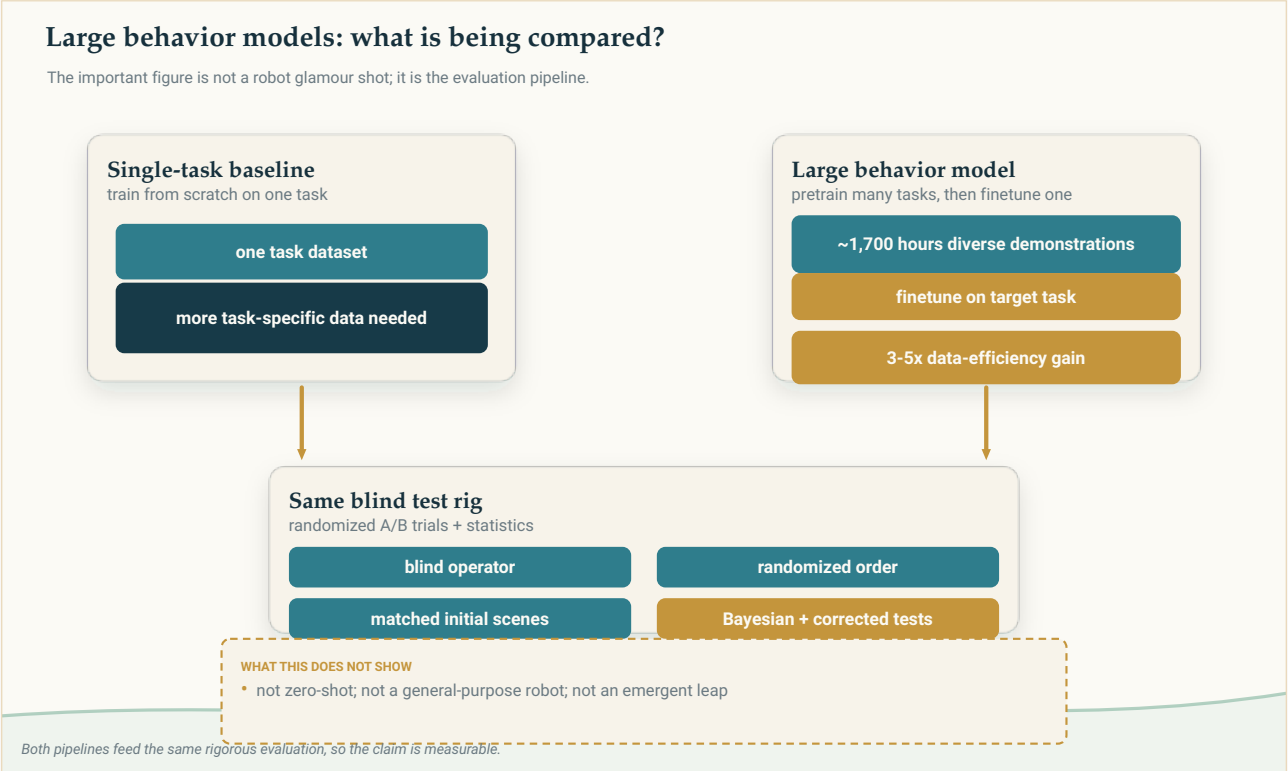
Written by Lucio Vaglio · Cross-checked by Laura Nesso · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

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

Стаття про роботів, справжня тема якої — чесність

Робототехніка переживає бум «фундаментальних моделей». Ідея, запозичена з мовного та візуального AI, дуже приваблива: замість того щоб навчати робота окремо для кожної задачі, навчити одну велику **«велику модель поведінки» (large behavior model, LBM)** на величезній і різноманітній колекції демонстрацій та отримати систему, яка багато що вміє і швидко адаптується. Ентузіазм — і інвестиції — величезні. Заголовок пишеться сам: *універсальні мізки для роботів уже тут.*

Ця стаття Toyota Research Institute цікава саме тим, що відмовляється писати такий заголовок. Її справжній внесок — не ефектніший робот. Це жорсткий погляд на оманливо просте запитання — *чи справді підхід із великою моделлю працює краще і як ми взагалі можемо це знати?* — із рівнем статистичної акуратності, який, за словами самих авторів, незвичний для галузі. Результат — справді корисне «так, але» й попередження, що значна частина робототехніки, можливо, вимірює шум.



Обидва підходи проходять однакове засліплене рандомізоване оцінювання. Твердження статті не в тому, що фундаментальні моделі роботів уже є zero-shot універсалами, а в тому, що попереднє навчання може підвищувати ефективність даних і стійкість, якщо вимірювати це належним чином.

Original The Clean Paper diagram CC BY 4.0

Що зробили автори

Вони побудували LBM у конкретному сенсі: **дифузійні візуально-моторні політики** — Diffusion Transformer, який читає зображення з камер, коротку мовну інструкцію та положення суглобів самого робота, а потім із частотою 10 Гц видає короткі послідовності моторних команд. Ці моделі **попередньо навчали приблизно на 1 700 годинах** демонстрацій роботів — понад 500 різних задач, зібраних усередині лабораторії, плюс публічні набори даних, — а потім **донавчали** на окремих задачах. У всіх порівняннях базовою системою є **політика для однієї задачі, навчена з нуля** лише на даних цієї задачі.

Однак серце статті — саме *оцінювання*, яке автори подають як головний результат. Щоб не обдурити самих себе, вони використали:

- **Засліплене рандомізоване A/B-тестування** в реальному світі: людина, яка запускала робота, не знала, яку політику тестують, а порядок був випадковим.
- **Контрольовані й відтворювані початкові умови**: перед кожною спробою оператори вирівнювали сцену за накладеним еталонним зображенням.
- **Велику кількість спроб**: 50 реальних запусків на кожну задачу, політику й умову; 200 на задачу в симуляції. Загалом — близько **1 800 засліплених реальних запусків і понад 47 000 симуляційних**.
- **Належну статистику**: байєсівські оцінки ймовірності успіху та попарні перевірки гіпотез із поправками на множинні порівняння, замість візуального порівняння перекривних похибок. Вони навіть провели контроль якості на чверті спроб, оцінених людьми, щоб виміряти помилки самого оцінювання.

[video pending]

Порівняння моделей, що сервірують стіл для сніданку: ліворуч — базова політика для однієї задачі, праворуч — LBM. Обидва відео відтворюються зі швидкістю 1×. Це одна оцінювана задача, а не доказ універсальної автономності.

Уся ця машина оцінювання і є суттю. Стаття загалом стверджує: без неї неможливо відрізнити справжнє поліпшення від удачі.

Що вони виявили

Великі моделі після донавчання в середньому перевершують моделі для однієї задачі, навчені з нуля. У сукупності задач LBM, попередньо навчена й потім донавчена на конкретній задачі, надійно випереджала політику, навчену з нуля на тій самій задачі, і в симуляції, і в реальному світі; різниця була статистично значущою. На окремих задачах донавчена LBM була статистично не гіршою або кращою майже завжди: 3/3 реальних задач і 15/16 симуляційних.

Найбільший і найчистіший виграш — ефективність даних. Донавчена LBM досягала продуктивності моделі, навченої з нуля, використовуючи приблизно **у 3–5 разів менше специфічних для задачі даних**. В одній реальній задачі — сервіруванні столу для сніданку — LBM, донавчена лише на **15%** демонстрацій, перевершила політику, навчену з нуля на **100%**.

Попереднє навчання найбільше допомагає, коли умови змінюються. Коли тестове середовище навмисно відхиляли від тренувальних умов («distribution shift»), перевага донавченої LBM *зростала*. В одному наборі симуляцій вона статистично перевершувала модель із нуля на 3 із 16 задач за звичайних умов, але на 10 із 16 під час зсуву розподілу. Оскільки реальні середовища завжди відрізняються від тренувальних, ця стійкість, мабуть, є практично найважливішим результатом.

Більше даних попереднього навчання допомагало плавно. Продуктивність стабільно зростала зі збільшенням pretraining-даних — **без раптового стрибка чи «емерджентного» переходу** в дослідженому діапазоні. Корисно, передбачувано, не драматично.

А от історія про універсала без донавчання не підтвердилася. Попередньо навчена LBM у режимі *zero-shot* — без специфічного донавчання на задачі — *не* перевершувала політики для однієї задачі послідовно. Одна мережа могла виконувати багато задач, але мрія «просто скажи їй, що робити» тут не справдилася; частково автори пов'язують це з крихкістю невеликого мовного енкодера.

І виграш був достатньо малим, щоб його легко не помітити — або випадково намалювати з шуму. Багато ефектів стали видимими лише завдяки більшому, ніж зазвичай, числу спроб і ретельним тестам. Автори прямо пишуть, що з огляду на розмір ефектів і шум **існує значний ризик, що багато статей із робототехніки вимірюють статистичний шум**. Вони також виявили, що буденне рішення — як саме *нормалізувати* дані — впливало на результат більше, ніж архітектурні зміни, а **помилку** нормалізації під час pretraining виявили лише *після* завершення оцінювань.

Що це, ймовірно, означає

Захищене читання: масштабне попереднє навчання на різноманітних роботизованих даних — справді корисний компонент. Воно зменшує кількість даних, потрібних для нової задачі, і робить політики стійкішими, коли світ не збігається з тренувальними умовами. Це реально підтримує напрямок, на який ставить галузь. Але переваги *помірні й умовні*: переважно проявляються після донавчання й найвиразніші в агрегованих результатах та під стресом, а не означають появу готового універсального робота.

Тихіший і, можливо, важливіший сенс — методологічний. Стаття фактично пропонує мірну лінійку: показує, скільки доказів насправді потрібно, щоб надійно заявити про поліпшення роботизованої політики, і натякає, що значна частина опублікованого ентузіазму спирається на надто малі вибірки. Такий коректив галузі потрібніший за

ще одну модель.

Чого це *не* доводить

- Це **не універсальний робот**. Переваги показані для конкретної архітектури (дифузійних політик), донавченої окремо для кожної задачі, у контрольованих умовах і на демонстраціях телеманіпуляції — а не для автономного робота, що виконує довільні нові завдання за командою.
- Це **не підтверджує zero-shot використання**. Без донавчання велика модель не мала стабільної переваги над базовими моделями для однієї задачі.
- Це **не свідчення «емерджентного стрибка»**. Масштабування давало плавні поліпшення; немає жодного розриву, який підтримував би наратив «і раптом вона стала здатною».
- Числа **відносні й прив'язані до лабораторії**. Абсолютні показники успіху навмисно налаштовували приблизно до 50%, щоб зробити порівняння чутливішими; це не оцінка реальної надійності. І це одна архітектура з однієї лабораторії.
- Робота **не пояснює, чому** конкретна політика успішна або невдала; кілька окремих задач, де велика модель працювала *гірше*, повідомлені, але не пояснені.
- Вона **нічого не говорить про безпеку, автономність чи розгортання** поза тестовою установкою.

Наскільки сильні докази?

Для центральних порівняльних тверджень — що донавчені LBM в агрегаті перевершують моделі з нуля, потребують у кілька разів менше даних і краще витримують зсув розподілу — докази сильні *і* незвично добре контрольовані: засліплення, рандомізація, великі вибірки, статистичні тести й перевірка якості людського оцінювання. Це рідкісний випадок, коли методологія достатньо надійна, щоб головні висновки можна було сприймати майже буквально.

Чесні застереження автори висувають самі. Їхні інтервали невизначеності враховують випадковість *оцінювання*, але **не** випадковість *навчання*: навчіть ту саму модель двічі — й можете отримати суттєво іншу політику, а ця варіативність у статистику не входить. Реальні задачі мали по 50 спроб — достатньо для ефектів середнього розміру, але дрібні можуть лишитися непоміченими. Мовне кондиціювання використовувало скромний енкодер, тому висновки про «просто скажіть роботу, що робити» можуть бути іншими для більших систем. І є відверте повідомлення про постфактум знайдену помилку нормалізації. Ніщо з цього не топить головних результатів, але це саме ті речі, які, за аргументом статті, галузь часто змітає під килим.

І одна примітка про джерело в тому самому дусі: цей матеріал спирається на препринт

авторів. Нам не вдалося отримати опубліковану журнальну версію, тому ми не перевіряли, чи є між препринтом і фінальним текстом зміни.

Чому це важливо

«Фундаментальні моделі для роботів» — словосполучення, створене для перебільшень, і таку роботу легко неправильно прочитати в обидва боки: як тріумфальне *«працює!»* або як зневажливе *«усе роздуте»*. Точна відповідь корисніша за обидві: попереднє навчання на різноманітних даних дає реальні, вимірювані, але помірні переваги — насамперед менше даних на задачу й більшу стійкість — і масштабування поліпшує їх передбачувано.

Глибша причина важливості в тому, що стаття спрямовує свою суворість на власну галузь. Показуючи, що справжні ефекти достатньо малі, аби зникати при недбалому оцінюванні, а нудний вибір нормалізації даних може важити більше за хитру нову архітектуру, вона доводить: значна частина прогресу в навчанні роботів потребує міцнішого вимірювання, перш ніж йому можна повірити. Стаття, яка витрачає свою довіру на охорону межі між результатом і бажанням, робить дещо рідкісніше й цінніше, ніж очолити таблицю лідерів.

Короткий підсумок

Дослідники Toyota Research Institute навчили «великі моделі поведінки» — дифузійні політики для роботів, попередньо навчені приблизно на 1 700 годинах різноманітних даних маніпуляцій, — і порівняли їх із політиками для однієї задачі, навченими з нуля, використовуючи незвично суворий протокол: засліплені, рандомізовані випробування великого масштабу (≈1 800 реальних і понад 47 000 симуляційних) зі справжньою статистикою. Після донавчання на конкретній задачі великі моделі надійно працювали краще в сукупності, потребували приблизно **у 3–5 разів менше специфічних даних** і були **стійкішими до зміни умов**, а продуктивність плавно зростала зі збільшенням pretraining-даних. Але **без донавчання** вони не мали стабільної переваги над моделями для однієї задачі; кілька ефектів були настільки малими, що стали видимими лише завдяки великим вибіркам, а звичайний вибір нормалізації даних впливав більше за архітектуру. Це сильна й стримана підтримка напряму роботизованих фундаментальних моделей — **не** універсальний робот, не zero-shot універсал і не «емерджентний стрибок» — плюс гостре попередження, що значна частина робототехніки може вимірювати шум.

Твереза оцінка

Що показує стаття: За суворого, засліпленого й статистично потужного оцінювання (≈1 800 реальних + 47 000+ симуляційних запусків) дифузійні політики, попередньо навчені на багатьох задачах і потім донавчені (LBM), в агрегаті перевершують політики для однієї задачі, навчені з нуля; досягають еквівалентної продуктивності з приблизно у 3–5 разів меншою кількістю специфічних даних; краще витримують зсув розподілу; продуктивність плавно масштабується з обсягом pretraining-даних.

Що правдоподібне, але не доведене: Що ці переваги перенесуться на значно більші vision-language-action моделі (їхній мовний енкодер був невеликим); що плавне масштабування продовжиться поза дослідженим діапазоном даних.

Чого робота не показує: Універсального або zero-shot робота (без донавчання немає стабільної переваги); будь-якого «емерджентного» стрибка можливостей; пояснення конкретних невдач на окремих задачах; абсолютної реальної надійності (успішність навмисно тримали близько 50% для чутливості); нічого про безпеку чи автономне розгортання.

Основні обмеження: Статистика враховує випадковість оцінювання, але не варіативність між окремими запусками навчання; 50 реальних спроб на задачу можуть не виявляти дрібні ефекти; одна архітектура й одна лабораторія; скромний мовний енкодер; помилку нормалізації даних знайшли після оцінювань; аналіз базується на препринті (опубліковану версію не звірено).

Скільки довіри варто мати звичайному читачеві? Високу, що багатозадачне попереднє навчання плюс донавчання дає реальні, помірні переваги — особливо в ефективності даних і стійкості — і що їх виміряли незвично ретельно. Високу, що це **не** універсальний або zero-shot робот і **не** емерджентний стрибок. Середню щодо того, наскільки ці переваги масштабуватимуться на більші моделі. І варто серйозно поставитися до попередження самих авторів: ефекти в галузі достатньо малі, щоб недостатньо потужні дослідження могли повідомляти шум. Доречна позиція: стриманий оптимізм щодо підходу й здорова недовіра до результатів robot-AI без такого статистичного фундаменту.

Джерела

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/