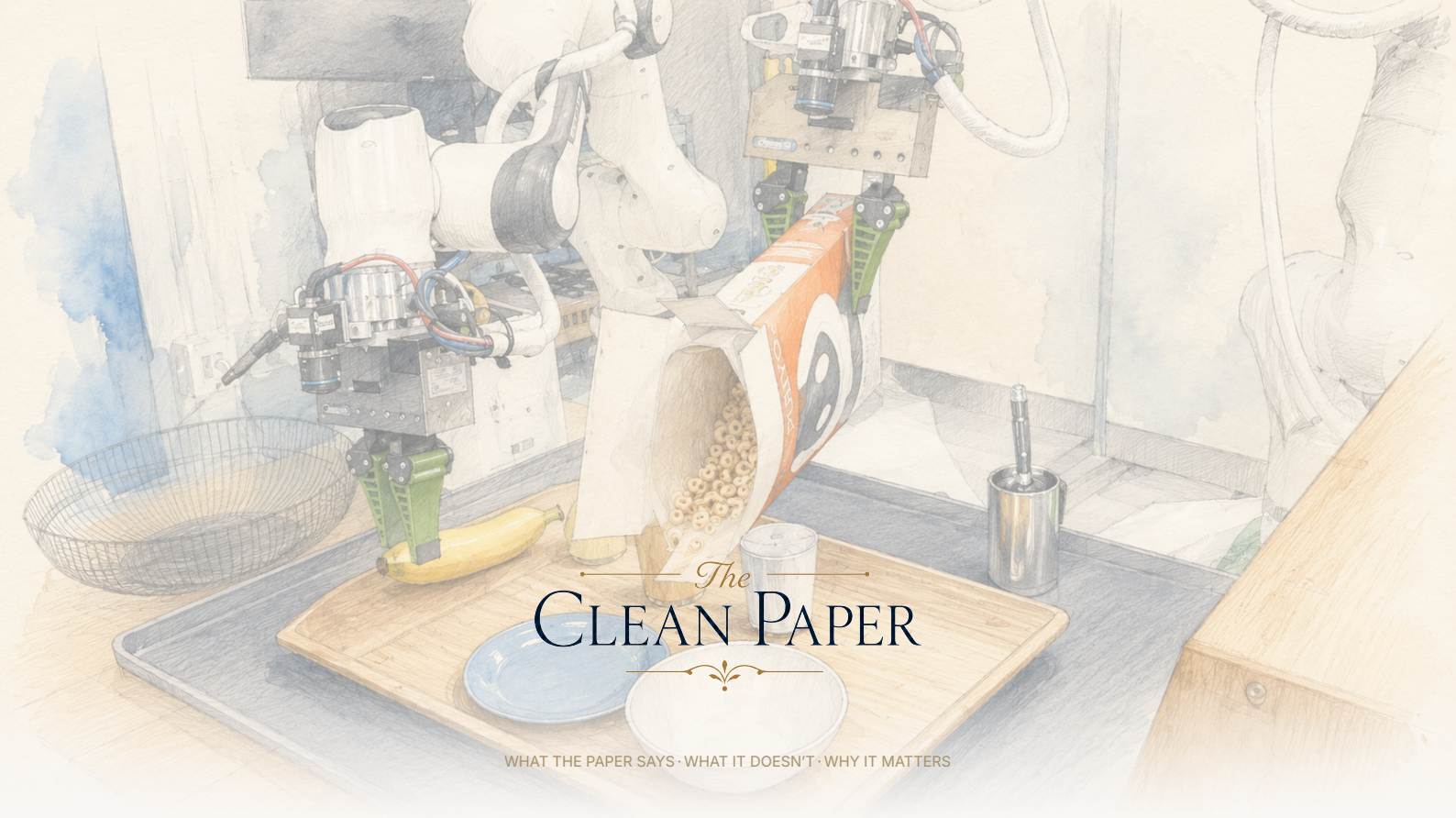


Ці сапраўды вялікія «фундаментальныя мадэлі» робатаў працуюць лепш? Асцярожны адказ: умерана так — і большасць даследаванняў не здольная надзейна гэта ўбачыць

Toyota Research Institute навучыла «вялікія мадэлі паводзін» — робатызаваныя палітыкі з папярэднім навучаннем прыкладна на 1 700 гадзінах разнастайных маніпуляцый — і параўнала іх з мадэлямі адной задачы, навучанымі з нуля, з незвычайнай строгасцю: сляпыя рандомізаваныя выпрабаванні вялікага маштабу (~1 800 рэальных і 47 000+ сімуляцыйных) з сапраўднай статыстыкай. Пасля данавучання на канкрэтнай задачы вялікія мадэлі ў сярэднім працавалі лепш, патрабавалі прыкладна ў 3–5 разоў менш спецыфічных даных і былі больш устойлівымі да змены ўмоў; прадукцыйнасць плаўна расла з аб'ёмам pretraining-даных. Але без данавучання яны не мелі стабільнай перавагі над мадэлямі адной задачы, некаторыя эфекты былі настолькі малыя, што станавіліся бачнымі толькі ў вялікіх выбарках, а звычайны выбар нармалізацыі даных значыў больш за архітэктуру. Гэта стрыманая падтрымка кірунку фундаментальных мадэляў для робатаў — не ўніверсальны робат, не zero-shot універсал і не «эмерджэнтны скачок» — плюс вострае папярэджанне, што значная частка робататэхнікі можа вымяраць шум.

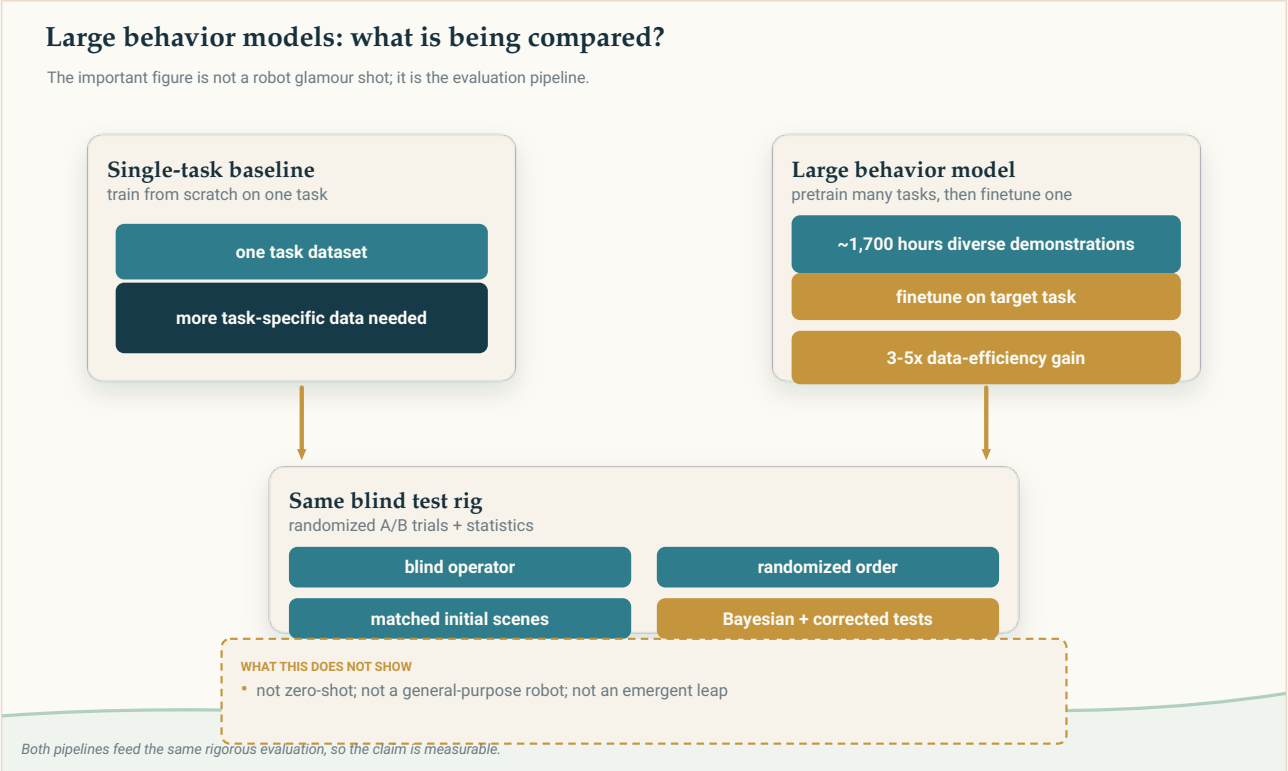
Written by Lucio Vaglio · Cross-checked by Laura Nesso · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

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

Артыкул пра робатаў, сапраўдная тэма якога — сумленнасць

Робататэхніка перажывае бум «фундаментальных мадэляў». Ідэя, пазычаная з моўнага і візуальнага AI, вельмі прывабная: замест таго каб навучаць робата асобна для кожнай задачы, навучыць адну вялікую **«вялікую мадэль паводзін» (large behavior model, LBM)** на велізарнай і разнастайнай калекцыі дэманстрацый і атрымаць сістэму, якая ўмее шмат і хутка адаптуецца. Энтузіязм — і інвестыцыі — велізарныя. Загаловак пішацца сам: *універсальныя мазгі для робатаў ужо тут.*

Гэты артыкул Toyota Research Institute цікавы менавіта тым, што адмаўляецца пісаць такі загаловак. Яго сапраўдны ўнёсак — не больш эфектны робат. Гэта жорсткі погляд на падманліва простае пытанне — *ці сапраўды падыход з вялікай мадэллю працуе лепш і як мы ўвогуле можам гэта ведаць?* — з узроўнем статыстычнай акуратнасці, які, па словах саміх аўтараў, незвычайны для галіны. Вынік — сапраўды карыснае «так, але» і папярэджанне, што значная частка робататэхнікі, магчыма, вымярае шум.



Абодва падыходы праходзяць аднолькавае сляпое рандамізаванае ацэньванне. Сцвярджэнне артыкула не ў тым, што фундаментальныя мадэлі робатаў ужо з'яўляюцца zero-shot універсаламі, а ў тым, што папярэдняе навучанне можа павышаць эфектыўнасць даных і ўстойлівасць, калі вымяраць гэта належным чынам.

Original The Clean Paper diagram CC BY 4.0

Што зрабілі аўтары

Яны пабудавалі LBM у канкрэтным сэнсе: **дыфузійныя візуальна-маторныя палітыкі** — Diffusion Transformer, які чытае выявы з камер, кароткую моўную інструкцыю і становішчы суставаў самога робата, а затым з частатой 10 Гц выдае кароткія паслядоўнасці маторных каманд. Гэтыя мадэлі **папярэдне навучалі прыкладна на 1 700 гадзінах** дэманстрацый робатаў — больш за 500 розных задач, сабраных унутры лабараторыі, плюс публічныя наборы даных, — а затым **данавучалі** на асобных задачах. Ва ўсіх параўнаннях базавай сістэмай з'яўляецца **палітыка адной задачы, навучаная з нуля** толькі на даных гэтай задачы.

Аднак сэрца артыкула — менавіта *ацэньванне*, якое аўтары падаюць як галоўны вынік. Каб не падмануць саміх сябе, яны выкарысталі:

- **Сляпое рандомізаванае A/B-тэставанне** ў рэальным свеце: чалавек, які запускаў робата, не ведаў, якую палітыку тэстуюць, а парадак быў выпадковым.
- **Кантраляваныя і ўзнаўляльныя пачатковыя ўмовы**: перад кожнай спробай аператары выраўноўвалі сцэну па накладзенай эталоннай выяве.
- **Вялікую колькасць спроб**: 50 рэальных запускаў на кожную задачу, палітыку і ўмову; 200 на задачу ў сімуляцыі. Усяго — каля **1 800 сляпых рэальных запускаў і больш за 47 000 сімуляцыйных**.
- **Належную статыстыку**: байесаўскія ацэнкі імавернасці поспеху і папарныя праверкі гіпотэз з папраўкамі на множныя параўнанні замест візуальнага супастаўлення перакрыўных памылак. Яны нават правялі кантроль якасці на чвэрці спроб, ацэненых людзьмі, каб вымераць памылкі самога ацэньвання.

[video pending]

Параўнанне мадэляў, якія сервіруюць стол для сняданку: злева — базавая палітыка адной задачы, справа — LBM. Абодва відэа прайграюцца з хуткасцю 1×. Гэта адна ацэньваная задача, а не доказ універсальнай аўтаномнасці.

Уся гэтая машына ацэньвання і ёсць сутнасць. Артыкул у цэлым сцвярджае: без яе немагчыма адрозніць сапраўднае паляпшэнне ад удачы.

Што яны выявілі

Вялікія мадэлі пасля данавучання ў сярэднім пераўзыходзяць мадэлі адной задачы, навучаныя з нуля. У сукупнасці задач LBM, папярэдне навучаная, а затым данавучаная на канкрэтнай задачы, надзейна апярэджвала палітыку, навучаную з нуля на той жа задачы, і ў сімуляцыі, і ў рэальным свеце; розніца была статыстычна значнай. На асобных задачах данавучаная LBM была статыстычна не горшая або лепшая амаль заўсёды: 3/3 рэальных задач і 15/16 сімуляцыйных.

Самы вялікі і чысты выйгрыш — эфектыўнасць даных. Данавучаная LBM дасягала прадукцыйнасці мадэлі, навучанай з нуля, выкарыстоўваючы прыкладна **ў 3–5 разоў менш спецыфічных для задачы даных**. У адной рэальнай задачы — сервіроўцы стала для сняданку — LBM, данавучаная толькі на **15%** дэманстрацый, пераўзышла палітыку, навучаную з нуля на **100%**.

Папярэдняе навучанне найбольш дапамагае, калі ўмовы змяняюцца. Калі тэставае асяроддзе наўмысна адхілялі ад навучальных умоў («distribution shift»), перавага данавучанай LBM *павялічвалася*. У адным наборы сімуляцый яна статыстычна пераўзыходзіла мадэль з нуля на 3 з 16 задач у звычайных умовах, але на 10 з 16 пры зруху размеркавання. Паколькі рэальныя асяроддзі заўсёды адрозніваюцца ад навучальных, гэтая ўстойлівасць, верагодна, з'яўляецца практычна самым важным вынікам.

Больш даных папярэдняга навучання дапамагала плаўна. Прадукцыйнасць стабільна расла з павелічэннем pretraining-даных — **без раптоўнага скачка або «эмерджэнтнага» пераходу** ў даследаваным дыяпазоне. Карысна, прадказальна, недраматычна.

А вось гісторыя пра ўніверсала без данавучання не пацвердзілася. Папярэдне навучаная LBM у рэжыме *zero-shot* — без спецыфічнага данавучання на задачы — *не* пераўзыходзіла палітыкі адной задачы паслядоўна. Адна сетка магла выконваць шмат задач, але мара «проста скажы ёй, што рабіць» тут не спраўдзілася; часткова аўтары звязваюць гэта з крохкасцю невялікага моўнага энкодэра.

І выйгрыш быў дастаткова малым, каб яго лёгка не заўважыць — або выпадкова намаляваць з шуму. Многія эфекты сталі бачнымі толькі дзякуючы большай, чым звычайна, колькасці спроб і старанным тэстам. Аўтары наўпрост пішуць, што з улікам памеру эфектаў і шуму **існуе значная рызыка, што многія артыкулы па робататэхніцы вымяраюць статыстычны шум**. Яны таксама выявілі, што будзённае рашэнне — як менавіта *нармалізаваць* даныя — уплывала на вынік мацней за архітэктурныя змены, а **памылку** нармалізацыі падчас pretraining выявілі толькі *пасля* завяршэння ацэньванняў.

Што гэта, верагодна, азначае

Абароненае прачытанне: маштабнае папярэдняе навучанне на разнастайных робатызаваных даных — сапраўды карысны кампанент. Яно памяншае колькасць даных, патрэбных для новай задачы, і робіць палітыкі больш устойлівымі, калі свет не супадае з навучальнымі ўмовамі. Гэта рэальна падтрымлівае кірунак, на які ставіць галіна. Але перавагі *ўмераныя і ўмоўныя*: пераважна праяўляюцца пасля данавучання і найбольш выразныя ў агрэгаваных выніках і пад стрэсам, а не азначаюць з'яўлення гатовага ўніверсальнага робата.

Больш ціхі і, магчыма, больш важны сэнс — метадалагічны. Артыкул фактычна

прапануе мерную лінейку: паказвае, колькі доказаў насамрэч трэба, каб надзейна заявіць пра паляпшэнне робатызаванай палітыкі, і намякае, што значная частка апублікаванага энтузіязму абапіраецца на занадта малыя выбаркі. Такі карэктыў галіне патрэбны больш за яшчэ адну мадэль.

Чаго гэта *не* даказвае

- Гэта **не ўніверсальны робат**. Перавагі паказаныя для канкрэтнай архітэктуры (дыфузійных палітык), данавучанай асобна для кожнай задачы, у кантраляваных умовах і на дэманстрацыях тэлеаперацыі — а не для аўтаномнага робата, які выконвае адвольныя новыя заданні па камандзе.
- Гэта **не пацвярджае zero-shot выкарыстанне**. Без данавучання вялікая мадэль не мела стабільнай перавагі над базавымі мадэлямі адной задачы.
- Гэта **не сведчанне «эмерджэнтнага скачка»**. Маштабаванне давала плаўныя паляпшэнні; няма ніякага разрыву, які падтрымліваў бы наратыў «і раптам яна стала здольнай».
- Лічбы **адносныя і прывязаныя да лабараторыі**. Абсалютныя паказчыкі поспеху наўмысна наладжвалі прыкладна да 50%, каб зрабіць параўнанні больш адчувальнымі; гэта не ацэнка рэальнай надзейнасці. І гэта адна архітэктура з адной лабараторыі.
- Праца **не тлумачыць, чаму** канкрэтная палітыка паспяховая або няўдалая; некалькі асобных задач, дзе вялікая мадэль працавала *горш*, паведамленыя, але не растлумачаныя.
- Яна **нічога не кажа пра бяспеку, аўтаномнасць або разгортванне** па-за тэставай устаноўкай.

Наколькі моцныя доказы?

Для цэнтральных параўнальных сцвярджэнняў — што данавучаныя LBM у агрэгаце пераўзыходзяць мадэлі з нуля, патрабуюць у некалькі разоў менш даных і лепш вытрымліваюць зрух размеркавання — доказы моцныя *і* незвычайна добра кантралююцца: сляпы дызайн, рандомізацыя, вялікія выбаркі, статыстычныя тэсты і праверка якасці чалавечага ацэньвання. Гэта рэдкі выпадак, калі методыка дастаткова надзейная, каб галоўныя высновы можна было ўспрымаць амаль літаральна.

Сумленныя агаворкі аўтары вылучаюць самі. Іх інтэрвалы нявызначанасці ўлічваюць выпадковасць *ацэньвання*, але **не** выпадковасць *навучання*: навучыце тую ж мадэль двойчы — і можаце атрымаць істотна іншую палітыку, а гэтая варыятыўнасць у статыстыку не ўваходзіць. Рэальныя задачы мелі па 50 спроб — дастаткова для эфектаў сярэдняга памеру, але дробныя могуць застацца незаўважанымі. Моўнае кандыцыянаванне выкарыстоўвала сціплы энкодэр, таму высновы пра «проста скажыце робату, што рабіць» могуць быць іншымі для больш буйных сістэм. І ёсць адкрытае паведамленне пра знойдзеную постфактум памылку нармалізацыі. Нішто з

гэтага не топіць галоўныя вынікі, але гэта менавіта тыя рэчы, якія, паводле аргумента артыкула, галіна часта замятае пад дыван.

І адна заўвага пра крыніцу ў тым жа духу: гэты матэрыял абапіраецца на прэпрынт аўтараў. Нам не ўдалося атрымаць апублікаваную часопісную версію, таму мы не правяралі, ці ёсць паміж прэпрынтам і фінальным тэкстам змены.

Чаму гэта важна

«Фундаментальныя мадэлі для робатаў» — словазлучэнне, створанае для перабольшанняў, і такую працу лёгка няправільна прачытаць у абодва бакі: як трыумфальнае *«працуе!»* або як пагардлівае *«усё раздзьмута»*. Дакладны адказ карыснейшы за абодва: папярэдняе навучанне на разнастайных даных дае рэальныя, вымерныя, але ўмераныя перавагі — перадусім менш даных на задачу і большую ўстойлівасць — і маштабаванне паляпшае іх прадказальна.

Больш глыбокая прычына важнасці ў тым, што артыкул накіроўвае сваю строгасць на ўласную галіну. Паказваючы, што сапраўдныя эфекты дастаткова малыя, каб знікаць пры неахайным ацэньванні, а сумны выбар нармалізацыі даных можа значыць больш за хітрую новую архітэктуру, ён даказвае: значная частка прагрэсу ў навучанні робатаў патрабуе больш трывалага вымярэння, перш чым яму можна верыць. Артыкул, які траціць свой крэдыт даверу на ахову мяжы паміж вынікам і жаданнем, робіць нешта больш рэдкае і каштоўнае, чым узначаліць табліцу лідараў.

Коратка

Даследчыкі Toyota Research Institute навучылі «вялікія мадэлі паводзін» — дыфузійныя палітыкі для робатаў, папярэдне навучаныя прыкладна на 1 700 гадзінах разнастайных даных маніпуляцый, — і параўналі іх з палітыкамі адной задачы, навучанымі з нуля, выкарыстоўваючы незвычайна строгі пратакол: сляпыя рандомізаваныя выпрабаванні вялікага маштабу (≈1 800 рэальных і больш за 47 000 сімуляцыйных) з сапраўднай статыстыкай. Пасля данавучання на канкрэтнай задачы вялікія мадэлі надзейна працавалі лепш у сукупнасці, патрабавалі прыкладна **ў 3–5 разоў менш спецыфічных даных** і былі **больш устойлівымі да змены ўмоў**, а прадукцыйнасць плаўна расла з павелічэннем pretraining-даных. Але **без данавучання** яны не мелі стабільнай перавагі над мадэлямі адной задачы; некалькі эфектаў былі настолькі малыя, што сталі бачнымі толькі дзякуючы вялікім выбаркам, а звычайны выбар нармалізацыі даных уплываў мацней за архітэктуру. Гэта моцная і стрыманая падтрымка кірунку робатызаваных фундаментальных мадэляў — **не** ўніверсальны робат, не zero-shot універсал і не «эмерджэнтны скачок» — плюс вострае папярэджанне, што значная частка робататэхнікі можа вымяраць шум.

Праверка без ажыятажу

Што паказвае артыкул: Пры строгім, сляпым і статыстычна магутным ацэньванні (≈1 800 рэальных + 47 000+ сімуляцыйных запускаў) дыфузійныя палітыкі, папярэдне навучаныя на многіх задачах і затым данавучаныя (LBM), у агрэгаце пераўзыходзяць палітыкі адной задачы, навучаныя з нуля; дасягаюць эквівалентнай прадукцыйнасці прыкладна з у 3–5 разоў меншай колькасцю спецыфічных даных; лепш вытрымліваюць зрух размеркавання; прадукцыйнасць плаўна маштабуецца з аб'ёмам pretraining-даных.

Што праўдападобна, але не даказана: Што гэтыя перавагі перанясуцца на значна больш буйныя vision-language-action мадэлі (іх моўны энкодэр быў невялікім); што плаўнае маштабаванне працягнецца за межамі даследаванага дыяпазону даных.

Чаго праца не паказвае: Універсальнага або zero-shot робата (без данавучання няма стабільнай перавагі); якога-небудзь «эмерджэнтнага» скачка магчымасцяў; тлумачэння канкрэтных няўдач на асобных задачах; абсалютнай рэальнай надзейнасці (паспяховасць наўмысна трымалі каля 50% для адчувальнасці); нічога пра бяспеку або аўтаномнае разгортванне.

Асноўныя абмежаванні: Статыстыка ўлічвае выпадковасць ацэньвання, але не варыятыўнасць паміж асобнымі запускамі навучання; 50 рэальных спроб на задачу могуць не выявіць дробныя эфекты; адна архітэктура і адна лабараторыя; сціплы моўны энкодэр; памылку нармалізацыі даных знайшлі пасля ацэньванняў; аналіз заснаваны на прэпрынце (апублікаваную версію не звяралі).

Колькі даверу варта мець звычайнаму чытачу? Высокі да таго, што шматзадачнае папярэдняе навучанне плюс данавучанне дае рэальныя, умераныя перавагі — асабліва ў эфектыўнасці даных і ўстойлівасці — і што іх вымералі незвычайна старанна. Высокі да таго, што гэта **не** ўніверсальны або zero-shot робат і **не** эмерджэнтны скачок. Сярэдні адносна таго, наколькі гэтыя перавагі будуць маштабавацца на больш буйныя мадэлі. І варта ўсур'ёз паставіцца да папярэджання саміх аўтараў: эфекты ў галіне дастаткова малыя, каб недастаткова магутныя даследаванні маглі паведамляць шум. Дарэчная пазіцыя: стрыманы аптымізм адносна падыходу і здаровы недавер да вынікаў robot-AI без такога статыстычнага падмурку.

Крыніцы

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/