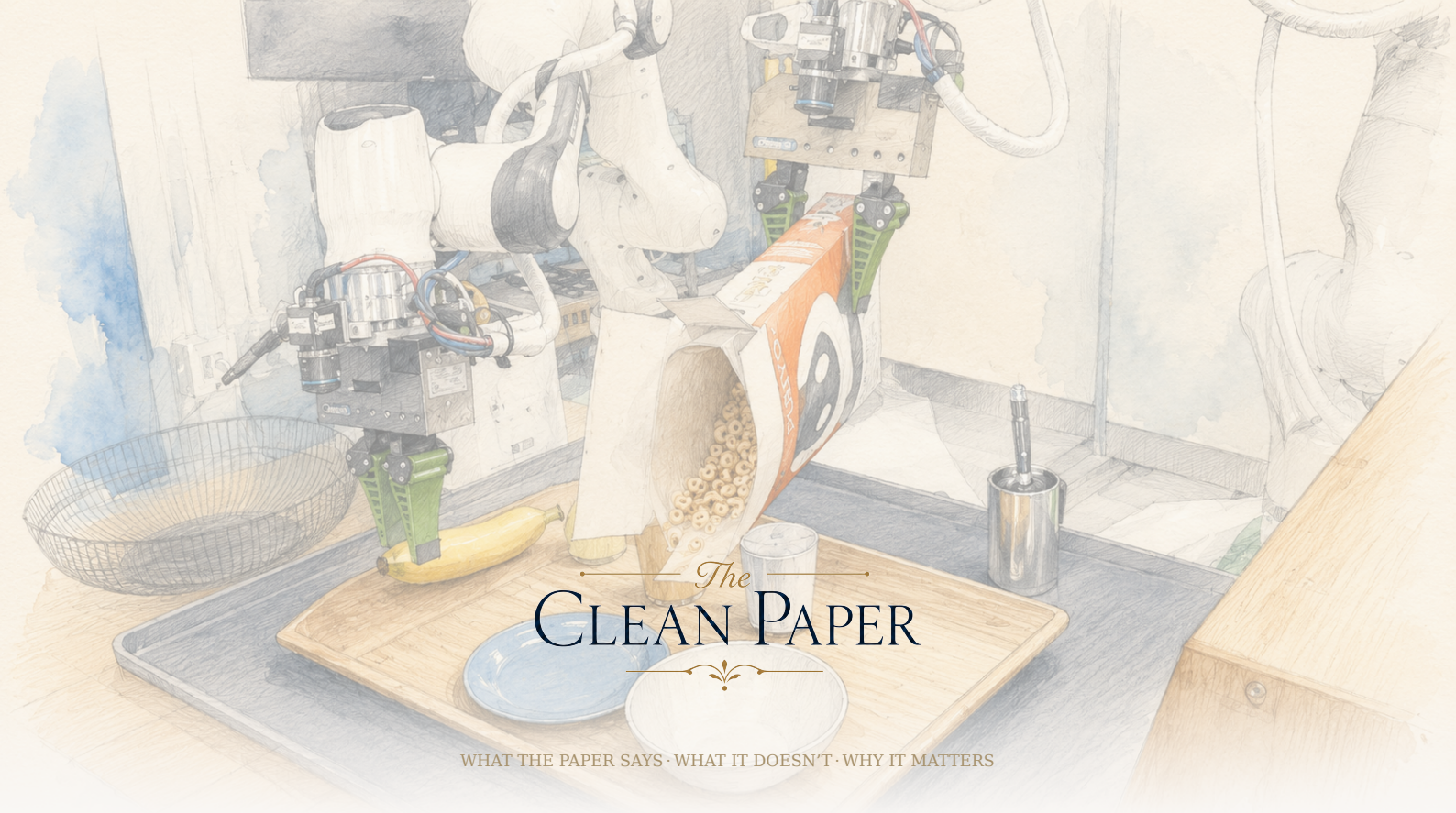


The Clean Paper

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

# დიდი რობოტული „foundation model“-ები მართლა უკეთ მუშაობს? ფრთხილი პასუხი: ზომიერად კი — და ბევრი კვლევა ამას სანდოდ ვერ ზომავს

*Toyota Research Institute-მა „large behavior models“ — დაახლოებით 1,700 საათის მრავალფეროვან manipulation მონაცემზე წინასწარ გაწვრთნილი რობოტული policy-ები — ნულიდან გაწვრთნილ ერთამოცანიან baseline-ებს უჩვეულო სიმკაცრით შეადარა: ბრმა, რანდომიზებული, დიდი ნიმუშის ცდები (~1,800 რეალური, 47,000+ სიმულაციური) ნამდვილი სტატისტიკით. თითო ამოცანაზე finetuning-ის შემდეგ დიდი მოდელები საშუალოდ უკეთ მუშაობდა, დაახლოებით 3–5-ჯერ ნაკლები ამოცანა-სპეციფიკური მონაცემი სჭირდებოდა და პირობების ცვლილებისას უფრო robust იყო; შესრულება pretraining მონაცემთან ერთად გლუვად იზრდებოდა. მაგრამ finetuning-ის გარეშე ისინი ერთამოცანიან მოდელებს თანმიმდევრულად არ აჯობებდა, რამდენიმე ეფექტი მხოლოდ დიდი ნიმუშით გამოჩნდა და ჩვეულებრივმა data-normalisation არჩევანმა არქიტექტურაზე მეტი გავლენა მოახდინა. ეს robot-foundation-model მიმართულების გაზომილი მხარდაჭერაა — არა ზოგადი რობოტი, zero-shot გენერალისტი ან „emergent ნახტომი“ — და გაფრთხილება, რომ რობოტიკის ნაწილი შესაძლოა ხმაურს ზომავდეს.*

---

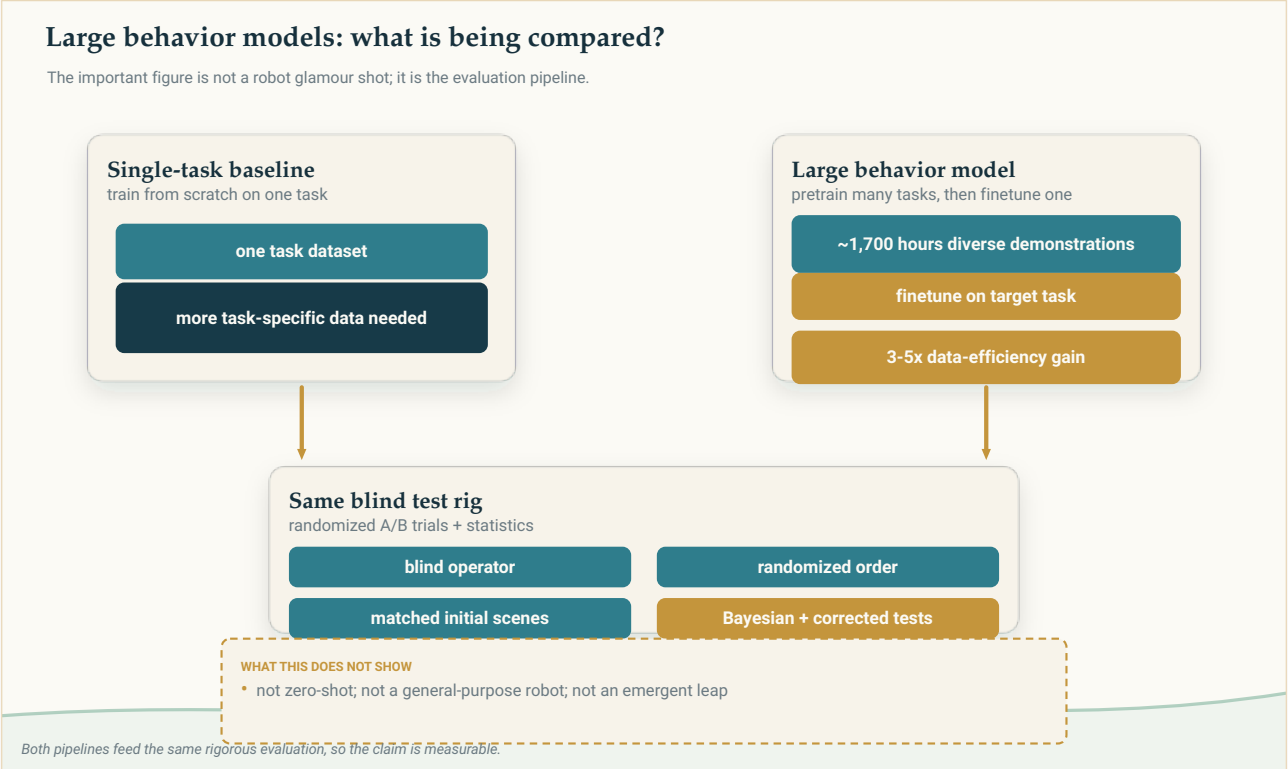
Written by Lucio Vaglio · Cross-checked by Laura Nesso · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

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

# რობოტიკის ნაშრომი, რომლის ნამდვილი თემა გულწრფელობაა

რობოტიკა „foundation მოდელი“-ების აჩქარებულ ტალღაშია. ენობრივი და ვიზუალური AI-დან ნასესხები იდეა მიმზიდველია: იმის ნაცვლად, რომ რობოტი თითო ამოცანაზე ცალ-ცალკე გაწვრთნა, ერთი დიდი **დიდი behavior მოდელი (LBM)** ასწავლო ძალიან მრავალფეროვან დემონსტრაციებზე და მიიღო სისტემა, რომელსაც ბევრი რამ შეუძლია და ახალ ამოცანებს სწრაფად ერგება. ენთუზიაზმიც და ინვესტიციაც უზარმაზარია. სათაური თითქოს თავისით იწერება: *ზოგადი დანიშნულების რობოტის ტვინი უკვე აქაა*.

Toyota Research Institute-ის ეს ნაშრომი საინტერესო სწორედ იმიტომაა, რომ ამ სათაურს **არ** წერს. მისი მთავარი წვლილი უფრო შთამბეჭდავი რობოტი კი არა, deceptively მარტივი კითხვის მკაცრი შემოწმებაა: *დიდი მოდელის მიდგომა ნამდვილად უკეთ მუშაობს? და ამას საერთოდ როგორ გავიგებთ?* ავტორები ამ კითხვას ისეთი სტატისტიკური სიფრთხილით პასუხობენ, რომელიც, მათი თქმით, სფეროში უჩვეულოა. შედეგი სასარგებლო „კი, მაგრამ“-ია და გაფრთხილება, რომ რობოტიკის დიდი ნაწილი შეიძლება უბრალოდ ხმაურს ზომავდეს.




ორივე მიდგომა ერთსა და იმავე ბრმა, რანდომიზებულ შეფასებაში შედის. ნაშრომის მტკიცება არაა, რომ რობოტის foundation მოდელები zero-shot გენერალისტებია; მტკიცება ისაა, რომ pretraining-ს შეუძლია მონაცემთა ეფექტიანობა და robustness გააუმჯობესოს, როცა ამას ფრთხილად ვზომავთ.

# რა გააკეთეს ავტორებმა

მათ LBM კონკრეტული მნიშვნელობით ააგეს: **diffusion-based visuomotor პოლიტიკა-ები** — Diffusion Transformer, რომელიც კადრებს კამერიდან, მოკლე ენობრივ ინსტრუქციას და რობოტის საკუთარი სახსრების პოზიციებს კითხულობს და 10 Hz სიხშირით მოკლე მოტორული ბრძანებების სერიებს აბრუნებს. მოდელები **დაახლოებით 1,700 საათის** რობოტული დემონსტრაციებით წინასწარ გაწვრთნეს — 500-ზე მეტი განსხვავებული შიდა ამოცანით და საჯარო მონაცემთა ნაკრებებით — შემდეგ კი ცალკეულ ამოცანებზე **finetune** გაუკეთეს. მთელი ნაშრომის განმავლობაში შედარების ობიექტია **ერთამოცანიანი პოლიტიკა, რომელიც იმავე ამოცანის მონაცემებზე ნულიდანაა გაწვრთნილი**.

მაგრამ ნაშრომის გული *შეფასებაა*, რომელსაც ავტორები თითქმის მთავარ შედეგად განიხილავენ. საკუთარი თავის მოტყუების თავიდან ასაცილებლად მათ გამოიყენეს:

- **ბრმა, რანდომიზებული A/B ტესტირება რეალურ სამყაროში** — რობოტის ოპერატორმა არ იცოდა, რომელ პოლიტიკა-ს ამოწმებდა, ხოლო რიგი შემთხვევით განისაზღვრა.
- **კონტროლირებული, გამეორებადი საწყისი პირობები** — თითო ცდის წინ ოპერატორები სცენას გამოსახულების overlay-ს უთანაბრებდნენ.
- **დიდი ცდების რაოდენობა** — რეალურ სამყაროში 50 rollout თითო ამოცანაზე, პოლიტიკა-ზე და პირობაზე; სიმულაციაში 200 თითო ამოცანაზე. ჯამში დაახლოებით **1,800 ბრმა რეალური rollout და 47,000-ზე მეტი სიმულაციური rollout**.
- **სწორი სტატისტიკა** — წარმატების ალბათობის ბაიესური შეფასებები და წყვილური ჰიპოთეზების ტესტები მრავალი შედარების კორექციით, ნაცვლად შეცდომა bar-ების თვალით შეფასებისა. ადამიანების მიერ შეფასებული ცდების მეოთხედზე მათ QA შემოწმებაც ჩაატარეს, რათა სკორინგის შეცდომა გაეზომათ.

[video pending]

საუზმის მაგიდის გაწყობის მოდელების გვერდიგვერდ შედარება: მარცხნივ ერთამოცანიანი საწყისი დონე, მარჯვნივ LBM. ორივე ვიდეო 1× სიჩქარით მიდის. ეს ერთი შეფასებული ამოცანაა და არა ზოგადი დანიშნულების ავტონომიის მტკიცებულება.

სწორედ ეს მექანიზმია მნიშვნელოვანი. მთელი ნაშრომი ფაქტობრივად ამტკიცებს, რომ ასეთი დისციპლინის გარეშე რეალურ გაუმჯობესებას იღბლისგან ვერ გაარჩევთ.

# რა აღმოაჩინეს

**Finetune-გაკეთებულმა დიდმა მოდელებმა საშუალოდ აჯობა ნულიდან გაწვრთნილ ერთამოცანიან მოდელებს.** ამოცანების ერთობლიობაში წინასწარ გაწვრთნილმა და მერე კონკრეტულ ამოცანაზე finetune-გაკეთებულმა LBM-მა იმავე ამოცანაზე ნულიდან გაწვრთნილ პოლიტიკა-ს როგორც სიმულაციაში, ისე რეალურ სამყაროში

სანდოდ აჯობა; განსხვავება სტატისტიკურად მნიშვნელოვანი იყო. ცალკეულ ამოცანებშიც finetuned LBM თითქმის ყოველთვის სტატისტიკურად ისეთივე კარგი ან უკეთესი აღმოჩნდა (3/3 რეალურ ამოცანაზე, 15/16 სიმულაციაში).

**ყველაზე დიდი და მკაფიო მოგება მონაცემთა ეფექტიანობაშია.** Finetuned LBM-მა ნულიდან გაწვრთნილი მოდელის ეკვივალენტურ შესრულებას დაახლოებით **3–5-ჯერ ნაკლები ამოცანა-სპეციფიკური მონაცემით** მიაღწია. ერთ რეალურ ამოცანაში — საუზმის მაგიდის გაწყობაში — მხოლოდ დემონსტრაციების **15%**-ზე finetune-გაკეთებულმა LBM-მა ნულიდან, დემონსტრაციების **100%**-ზე გაწვრთნილ პოლიტიკა-ს აჯობა.

**წინასწარი წვრთნა ყველაზე მეტად მაშინ ეხმარება, როცა პირობები იცვლება.** როცა ტესტის გარემო განზრახ გადაიყვანეს გაწვრთნის პირობებიდან განსხვავებულ მდგომარეობაში („distribution shift“), finetuned LBM-ის უპირატესობა *გაიზარდა*. ერთ სიმულაციურ ჯგუფში ნორმალურ პირობებში ნულიდან გაწვრთნილ მოდელს 16 ამოცანიდან 3-ზე სტატისტიკურად აჯობა, distribution shift-ისას კი 10/16-ზე. რადგან რეალურ დანერგვაში გარემო თითქმის ყოველთვის აცდენილია ტრენინგის ზუსტ პირობებს, ეს robustness შესაძლოა ყველაზე პრაქტიკულად მნიშვნელოვანი შედეგი იყოს.

**მეტი წინასწარი წვრთნა მონაცემი თანმიმდევრულად ეხმარებოდა.** შესრულება მონაცემების დამატებასთან ერთად თანდათან იზრდებოდა — ტესტირებულ მასშტაბებზე **არ ყოფილა უეცარი ნახტომი ან „წარმოქმნილი“ გარდატეხა**. სასარგებლო, პროგნოზირებადი და არადრამატული შედეგი.

**მაგრამ finetuning-ის გარეშე „გენერალისტის“ ამბავი არ გამოვიდა.** Pre-trained LBM-ის *zero-shot* გამოყენება — ამოცანა-სპეციფიკური finetuning-ის გარეშე — ერთამოცანიან პოლიტიკა-ებს თანმიმდევრულად **არ** აჯობებდა. ერთ ქსელს ერთდროულად მრავალი ამოცანა შეეძლო, მაგრამ „უბრალოდ უთხარი, რა გააკეთოს“ ოცნება აქ არ დადასტურდა; ავტორები ამის ნაწილს საკუთარ შედარებით პატარა ენობრივ encoder-ს უკავშირებენ.

**და მოგება იმდენად მცირე იყო, რომ ადვილად შეიძლებოდა დაგეკარგათ — ან „გაგეკეთებინათ“.** ბევრი ეფექტი მხოლოდ უჩვეულოდ დიდი ნიმუშებითა და ფრთხილი ტესტებით გახდა ხილული. ავტორები პირდაპირ წერენ, რომ ეფექტების ზომისა და ხმაურის გათვალისწინებით, **არსებობს მნიშვნელოვანი რისკი, რომ რობოტიკის ბევრი ნაშრომი სტატისტიკურ ხმაურს ზომავს**. მათ ისიც ნახეს, რომ ჩვეულებრივი გადაწყვეტილება — მონაცემების *ნორმალიზაციის* გზა — შედეგებზე არქიტექტურულ ცვლილებებზე მეტად მოქმედებდა; pretraining-ში ნორმალიზაციის **bug** კი მხოლოდ შეფასებების დასრულების *შემდეგ* აღმოაჩინეს.

# რას ნიშნავს ეს, სავარაუდოდ

დასაცავი წაკითხვა: მრავალფეროვან რობოტულ მონაცემებზე ფართომასშტაბიანი წინასწარი წვრთნა რეალური და ღირებული ინგრედიენტია — ახალ ამოცანაზე

ნაკლები მონაცემი გჭირდებათ და პოლიტიკა უფრო გამძლეა, როცა გარემო ტრენინგს ზუსტად არ ემთხვევა. ეს ნამდვილად მხარს უჭერს მიმართულებას, რომელზეც სფერო ფსონს დებს. მაგრამ მოგება *ზომიერი და პირობითია* — ძირითადად finetuning-ის შემდეგ ჩანს და ყველაზე ნათელია აგრეგირებულად და stress პირობებში — არა „ჩასასმელი“ ზოგადი რობოტის გამოჩენა.

უფრო ჩუმი და შესაძლოა უფრო მნიშვნელოვანი მნიშვნელობა მეთოდოლოგიურია. ნაშრომი ფაქტობრივად საზომი ჯოხია: აჩვენებს, რამდენი მტკიცებულება სჭირდება რობოტის პოლიტიკა-ზე სანდო მტკიცებას და მიანიშნებს, რომ გამოქვეყნებული ენთუზიაზმის დიდი ნაწილი ზედმეტად მცირე მონაცემს ეყრდნობა. სფეროს ასეთი კორექცია კიდევ ერთ მოდელზე მეტად სჭირდება.

# რას არ ამტკიცებს ეს

- ეს **არ არის ზოგადი დანიშნულების რობოტი**. მოგება ნაჩვენებია კონკრეტულ არქიტექტურაზე (diffusion პოლიტიკა), თითო ამოცანაზე finetuning-ით, კონტროლირებულ პირობებში და teleoperation-ის დემონსტრაციებზე — არა ავტონომიურ რობოტზე, რომელიც ბრძანებით ნებისმიერ ახალ საქმეს აკეთებს.
- ის **არ ადასტურებს zero-shot გამოყენებას**. Finetuning-ის გარეშე დიდი მოდელი ერთამოცანიან საწყისი დონე-ებს თანმიმდევრულად არ აჯობებდა.
- ეს **არ არის „წარმოქმნილი ნახტომის“ მტკიცებულება**. მასშტაბირება შესრულებას გლუვად აუმჯობესებდა; აქ უეცარი დისკონტინუიტეტი არ ჩანს.
- რიცხვები **ფარდობითი და ლაბორატორიულია**. აბსოლუტური წარმატების მაჩვენებლები შეგნებულად ~50%-ის მიდამოში მოარგეს, რათა შედარება მგრძნობიარე ყოფილიყო; ისინი რეალურ სამყაროში საიმედოობას არ ზომავს და სამუშაო ერთი ლაბორატორიის ერთი არქიტექტურაა.
- ის **არ ხსნის, რატომ** გამოსდის ან არ გამოსდის პოლიტიკა-ს კონკრეტული ამოცანა; რამდენიმე შემთხვევა, სადაც დიდი მოდელი *უარესი* იყო, აღწერილია, მაგრამ ახსნილი არაა.
- **უსაფრთხოებაზე, ავტონომიაზე ან შეფასების სისტემის გარეთ დანერგვაზე** არაფერი ამბობს.

# რამდენად ძლიერია მტკიცებულება?

ძირითადი შედარებითი მტკიცებებისთვის — finetuned LBM აგრეგირებულად აჯობებს ნულიდან გაწვრთნილ საწყისი დონე-ს, რამდენიმე ჯერ ნაკლებ მონაცემს საჭიროებს და distribution shift-ისას უფრო მდგრადი-ია — მტკიცებულება ძლიერია **და** უჩვეულოდ კარგად კონტროლირებული: ბრმა, რანდომიზებული, დიდი ნიმუშით, სტატისტიკურად შემოწმებული, სკორინგის QA-ით. ეს იშვიათი შემთხვევაა, როცა მეთოდოლოგია იმდენად კარგია, რომ მთავარ დასკვნებს თითქმის პირდაპირ შეგვიძლია ვენდოთ.

სანდო caveat-ებს ავტორები თავადაც აღნიშნავენ. მათი შეცდომა bar-ები შეფასების შემთხვევითობას ასახავს, მაგრამ **არა ტრენინგის შემთხვევითობას** — იგივე მოდელი ორჯერ რომ გაწვრთნათ, შეიძლება მნიშვნელოვნად განსხვავებული პოლიტიკა მიიღოთ და ეს ვარიანტულობა სტატისტიკაში არაა. რეალურ ამოცანებზე 50 ცდა საშუალო ეფექტების დასაჭერად საკმარისია, მაგრამ მცირეს შეიძლება ვერ ხედავდეს. ენობრივი conditioning შედარებით მოკრძალებულ encoder-ს იყენებდა, ამიტომ „უბრალოდ უთხარი რობოტს რა გააკეთოს“ ტიპის მტკიცება უფრო დიდ სისტემებში სხვაგვარად გამოიყურებოდეს. და არის გულწრფელად გამჟღავნებული post-hoc ნორმალიზაციის bug-იც. ეს ყველაფერი მთავარ შედეგს არ აქრობს — და ზუსტად ასეთი დეტალებია, რომლის იგნორირებასაც ნაშრომი სფეროს საყვედურობს.

წყაროს ერთი შენიშვნაც იმავე სულით: ეს განმარტება ავტორების პრეპრინტი-ზეა დაფუძნებული. ჟურნალში გამოქვეყნებული ვერსიის მიღება ვერ მოვახერხეთ, ამიტომ პრეპრინტი-სა და საბოლოო ტექსტს შორის ცვლილებები არ შეგვიმოწმებია.

# რატომ არის ეს მნიშვნელოვანი

„Robot foundation მოდელი“ გადაჭარბებული მტკიცებისათვის თითქმის იდეალური ფრაზაა და ეს კვლევა ორივე მხარეს შეიძლება ცუდად წავიკითხოთ — ტრიუმფალურ „იმუშავა!“ ან უარყოფით „ჰაიპია“. ზუსტი დასკვნა ორივეზე სასარგებლოა: მრავალფეროვან მონაცემებზე წინასწარი წვრთნა რეალურ, გაზომვად, მაგრამ ზომიერ სარგებელს იძლევა — განსაკუთრებით ნაკლებ მონაცემს თითო ამოცანაზე და მეტ robustness-ს — და მასშტაბთან ერთად გზა პროგნოზირებადად უმჯობესდება.

უფრო ღრმა მიზეზი, რატომაც ნაშრომი მნიშვნელოვანია, არის ის, რომ სიმკაცრეს **საკუთარ სფეროს** მიმართ იყენებს. ის აჩვენებს, რომ ნამდვილი ეფექტები იმდენად მცირეა, რომ ცუდ შეფასებაში ქრება, და რომ მოსაწყენმა რამემ — მაგალითად მონაცემების ნორმალიზაციამ — „ჭკვიან“ ახალ არქიტექტურაზე მეტი შეიძლება შეცვალოს. ამით იგი ამტკიცებს, რომ robot learning-ის პროგრესის დიდი ნაწილი უფრო მყარ გაზომვას მოითხოვს, სანამ დავიჯერებთ. ნაშრომი, რომელიც საკუთარ სანდოობას შედეგსა და სურვილს შორის საზღვრის დასაცავად ხარჯავს, leaderboard-ის პირველ ადგილზე გასვლაზე იშვიათ და უფრო ღირებულ საქმეს აკეთებს.

# სუფთა შეჯამება

Toyota Research Institute-ის მკვლევრებმა გაწვრთნეს „დიდი behavior მოდელები“ — diffusion-based რობოტული პოლიტიკა-ები, წინასწარ დაახლოებით **1,700 საათის** მრავალფეროვან manipulation მონაცემზე — და ნულიდან გაწვრთნილ ერთამოცანიან პოლიტიკა-ებს უჩვეულოდ მკაცრი პროტოკოლით შეადარეს: ბრმა, რანდომიზებული, დიდი ნიმუში (≈1,800 რეალური და 47,000+ სიმულაციური ცდა) ნამდვილი სტატისტიკით. თითო ამოცანაზე finetuning-ის შემდეგ დიდი მოდელები აგრეგირებულად სანდოდ

უკეთ მუშაობდა, დაახლოებით **3–5-ჯერ ნაკლები ამოცანა-სპეციფიკური მონაცემი** სჭირდებოდა და **პირობების ცვლილებისას უფრო მდგრადი იყო**; წინასწარი წვრთნა მონაცემის გაზრდასთან ერთად შესრულება გლუვად უმჯობესდებოდა. მაგრამ **finetuning-ის გარეშე** ისინი ერთამოცანიან მოდელებს თანმიმდევრულად არ აჯობებდა, რამდენიმე ეფექტი იმდენად მცირე იყო, რომ მხოლოდ დიდი ნიმუშებით გამოჩნდა, და ჩვეულებრივმა მონაცემთა ნორმალიზაციის არჩევანმა არქიტექტურაზე მეტი გავლენა მოახდინა. ეს robot-foundation-მოდელი მიმართულების მყარი, გაზომილი მხარდაჭერაა — **არა** ზოგადი დანიშნულების რობოტი, არა zero-shot გენერალისტი და არა „წარმოქმნილი ნახტომი“ — პლუს მკვეთრი გაფრთხილება, რომ რობოტიკის ნაწილი შეიძლება ხმაურს ზომავდეს.

# პირდაპირი შემოწმება

**რას აჩვენებს ნაშრომი:** მკაცრი, ბრმა და სტატისტიკურად საკმარისი შეფასებით (≈1,800 რეალური + 47,000+ სიმულაციური rollout), მრავალამოცანიან pretraining-სა და შემდგომ finetuning-ზე აგებული diffusion პოლიტიკა-ები (LBM) აგრეგირებულად აჯობებს ნულიდან გაწვრთნილ ერთამოცანიან პოლიტიკა-ებს, ეკვივალენტურ შესრულებას ~3–5-ჯერ ნაკლები ამოცანა-სპეციფიკური მონაცემით აღწევს და distribution shift-ისას უფრო მდგრადი-ია; შესრულება წინასწარი წვრთნა მონაცემის მატებასთან ერთად გლუვად იზრდება.

**რა არის სარწმუნო, მაგრამ დაუმტკიცებელი:** რომ ეს სარგებელი ბევრად უფრო დიდ vision-ენა-ქმედება მოდელებზეც გადავა (მათი ენა encoder პატარა იყო); რომ გლუვი scaling ტესტირებულ დიაპაზონს გაცდება.

**რას არ აჩვენებს:** ზოგადი დანიშნულების ან zero-shot რობოტს (finetuning-ის გარეშე თანმიმდევრული უპირატესობა არაა); რაიმე „წარმოქმნილი“ უნარის ნახტომს; კონკრეტული ამოცანის მარცხის ახსნას; რეალურ სამყაროში აბსოლუტურ საიმედოობას (წარმატების მაჩვენებლები მგრძნობელობისთვის ~50%-ზე იყო მორგებული); უსაფრთხოებას ან ავტონომიურ დანერგვას.

**მთავარი შეზღუდვები:** სტატისტიკა შეფასების შემთხვევითობას მოიცავს, მაგრამ არა წვრთნა-run-ის შემთხვევითობას; 50 რეალური ცდა თითო ამოცანაზე პატარა ეფექტებს შეიძლება ვერ დაიჭერდეს; ერთი არქიტექტურა და ერთი ლაბორატორია; მოკრძალებული ენა encoder; შეფასებების შემდეგ აღმოჩენილი data-normalisation bug; ანალიზი პრეპრინტი-ზეა დაფუძნებული და გამოქვეყნებული ვერსია არ შეგვიმოწმებია.

**რამდენად უნდა ენდოს ზოგადი მკითხველი?** მაღალი ნდობა, რომ მრავალამოცანიანი წინასწარი წვრთნა + finetuning რეალურ, ზომიერ სარგებელს იძლევა — განსაკუთრებით მონაცემთა ეფექტიანობასა და robustness-ში — და რომ ეს უჩვეულოდ ფრთხილად გაზომეს. ასევე მაღალი ნდობა, რომ ეს **არ არის** ზოგადი დანიშნულების ან zero-shot რობოტი და **არ არის** წარმოქმნილი ნახტომი. საშუალო ნდობა იმაზე, რამდენად შორს მასშტაბირდება სარგებელი. და სერიოზულად მისაღებია ავტორების საკუთარი გაფრთხილება: ეფექტები იმდენად მცირეა, რომ under-powered კვლევები შესაძლოა

ხმაურს აქვეყნებდეს. სწორი პოზიციაა გაზომილი ოპტიმიზმი მიდგომის მიმართ და ჯანსაღი სკეპტიციზმი robot-AI შედეგებისადმი, რომლებსაც მსგავსი სტატისტიკური საფუძველი აკლია.

---

# წყაროები

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/