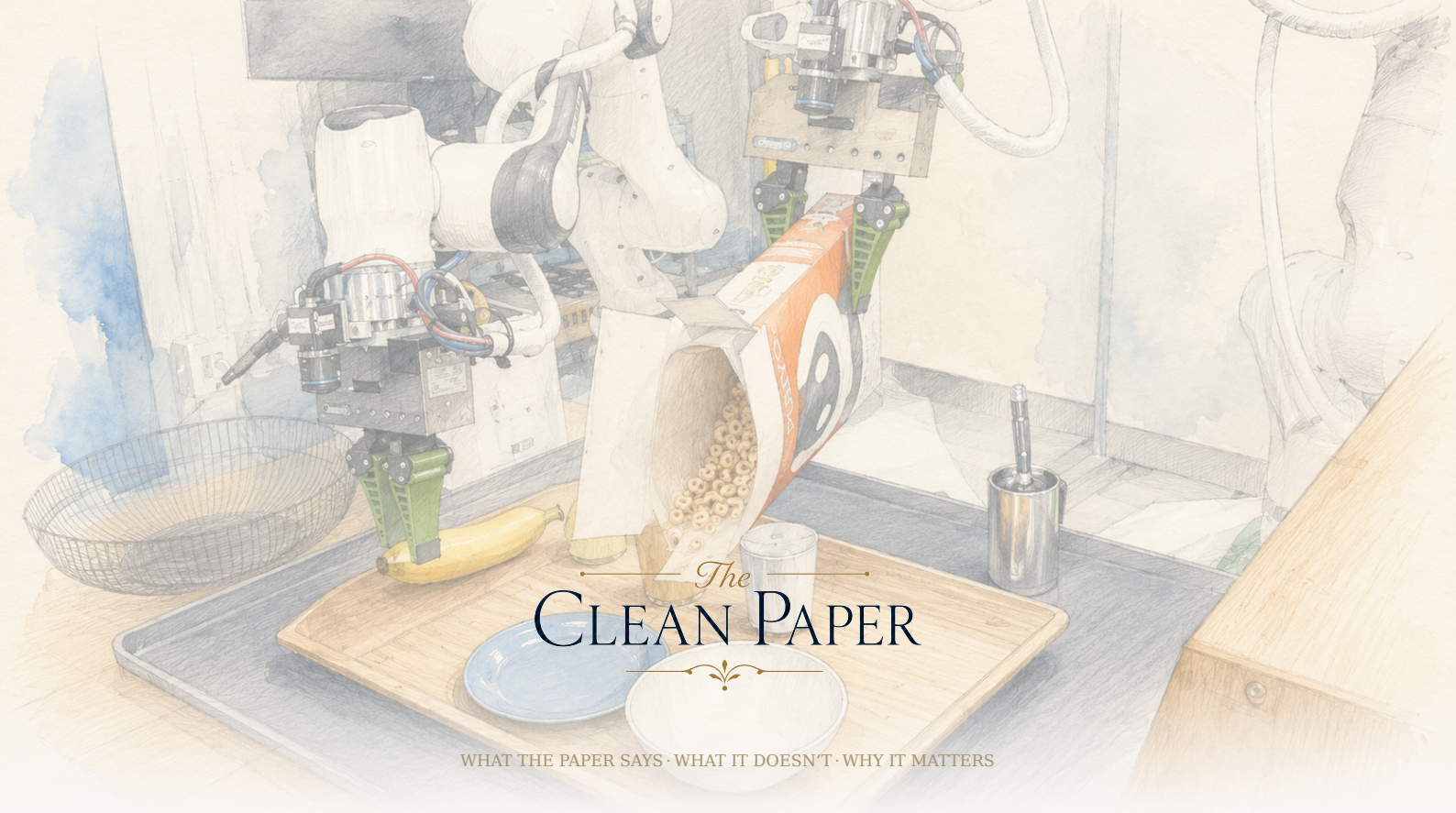


Մեծ robot «foundation model»-ներն իսկապե՞ս ավելի լավ են աշխատում: Զգույշ պատասխանը՝ չափավոր այո, իսկ շատ ուսումնասիրություններ դա չեն կարող տարբերել noise-ից

Toyota Research Institute-ը train է արել large behavior model-ներ՝ մոտ 1,700 ժամ տարբեր manipulation data-ի վրա pretrain արված robot policy-ներ, և դրանք համեմատել from-scratch single-task policy-ների հետ անսովոր խիստ blind, randomized, large-sample փորձերով՝ մոտ 1,800 real-world և 47,000+ simulation trial: Per-task finetuning-ից հետո մեծ մոդելները միջինում ավելի լավ էին, մոտ 3–5× պակաս task-specific data էին պահանջում և ավելի robust էին պայմանների փոփոխության դեպքում, իսկ performance-ը pretraining data-ի հետ սահուն աճեց: Բայց առանց finetuning-ի դրանք single-task model-ներին հետևողականորեն չէին գերազանցում, մի շարք effect-ներ այնքան փոքր էին, որ մեծ sample էր պետք, իսկ data-normalisation-ի սովորական ընտրությունը architecture-ից ավելի կարևոր էր: Սա չափված աջակցություն է robot-foundation-model ուղղությանը՝ ոչ general-purpose robot, ոչ zero-shot generalist, ոչ emergent leap, և զգուշացնում, որ robotics-ի մեծ մասը կարող է noise չափել:

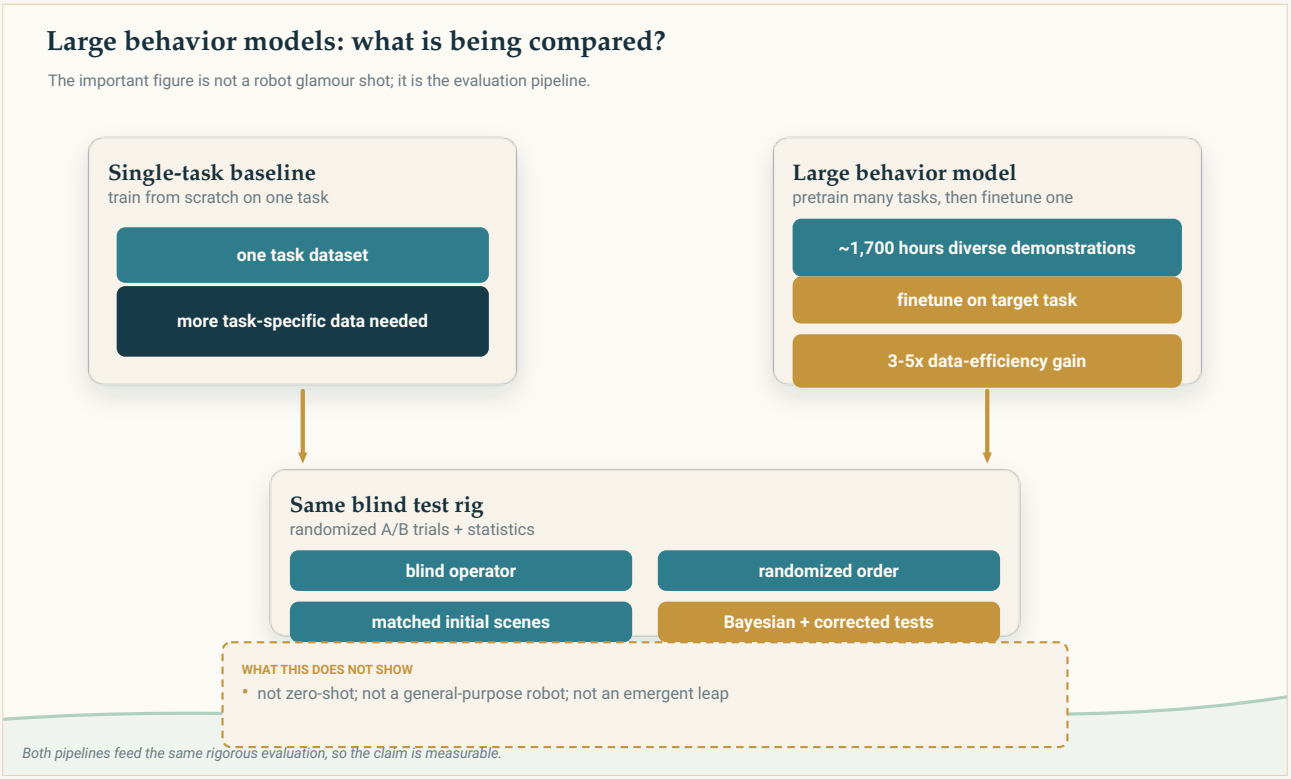
Written by Lucio Vaglio · Cross-checked by Laura Nesso · Edited by Michele Renda · Figures by Laura Nesso and Nova Vela · AI-assisted, human-reviewed

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

Ռոբոտների մասին հոդված, որի իրական թեման ազնվությունն է

Ռոբոտիկան հիմա «հիմք մոդել»-ների մրցավազքի մեջ է: Լեզվային և պատկեր ԱԲից փոխառված գաղափարը գայթակղիչ է. ամեն առաջադրանքի համար առանձին ռոբոտ ուսուցանել անելու փոխարեն մեկ մեծ **մեծ վարք մոդել (LBM)** ուսուցանել անել հսկայական ու բազմազան ցուցադրումների վրա և ստանալ համակարգ, որը լայն կարողություններ ունի ու արագ է հարմարվում: Էնտուզիազմը, ինչպես նաև ներդրումները, հսկայական են: վերնագիրը ինքն իրեն է գրում. *ընդհանուրնպատակ ռոբոտ ուղեղերը արդեն այստեղ են*:

Toyota Research Institute-ի այս հոդվածը հետաքրքիր է հենց այն պատճառով, որ հրաժարվում է այդ վերնագիրը գրել: Նրա հիմնական ներդրումը ավելի փայլուն ռոբոտը չէ: Այն շատ խիստ է նայում խաբուսիկորեն պարզ հարցին՝ *մեծ մոդելի մոտեցումն իսկապես ավելի լավ է աշխատո՞ւմ, և ընդհանրապես ինչպե՞ս պետք է դա իմանանք*: Պատասխանը տրված է այնպիսի վիճակագրական բուժօգնությունով, որն, ըստ հեղինակների, այս ոլորտում սովորական չէ: Արդյունքը իսկապես օգտակար «այո, բայց» է և միաժամանակ զգուշացում, որ robotics-ի բազմաթիվ հոդվածներ կարող են աղմուկ չափել:



Երկու մոտեցումներն էլ անցնում են նույն կույր, պատահականացված գնահատումով: Հոդվածի պնդումը այն չէ, որ ռոբոտ հիմք մոդելները զրոյական օրինակներով generalist են, այլ այն, որ նախնական ուսուցումը կարող է բարձրացնել տվյալներ efficiency-ն և կայունությունը, երբ դա զգուշորեն է չափվում:

Original The Clean Paper diagram CC BY 4.0

Ինչ արեցին հեղինակները

Նրանք LBM-ներ են կառուցել կոնկրետ իմաստով՝ **դիֆուզիայի վրա հիմնված տեսաշարժողական քաղաքականություններ**՝ Diffusion Transformer, որը կարդում է տեսախցիկի պատկերները, կարճ լեզվային հրահանգը և ռոբոտի սեփական հոդերի դիրքերը, ապա 10 Hz հաճախականությամբ տալիս է շարժիչ հրամանների կարճ հաջորդականություններ: Այդ մոդելները **նախապես ուսուցանել են մոտ 1,700 ժամ** ռոբոտ ցուցադրումի վրա՝ ավելի քան 500 տարբեր առաջադրանք, որոնք հավաքվել են ներսում, և հանրային տվյալների հավաքածուներ, ապա յուրաքանչյուր առաջադրանքի համար **լրացուցիչ ուսուցանել** են արել: Համեմատության ամբողջ ընթացքում ելակետայինը մեկ առաջադրանքի տվյալների վրա զրոյից ուսուցանել արված **մեկ առաջադրանքի քաղաքականություն** է:

Բայց հոդվածի սիրտը հենց *գնահատում*-ն է, որը հեղինակները վերաբերվում են որպես գլխավոր արդյունք: Իրենց չխաբելու համար նրանք օգտագործել են.

- Իրական աշխարհում **կույր, պատահականացված A/B փորձարկում**. ռոբոտը վարող մարդը չգիտեր, թե որ քաղաքականությունն է փորձարկվում, իսկ հերթականությունը պատահականացված էր:
- **Վերահսկվող, կրկնելի սկզբնական պայմաններ**. յուրաքանչյուր փորձարկումից առաջ օպերատորը տեսարանը համընկեցրել է պատկերի վերադրման հետ:
- **Մեծ փորձարկում count-եր**. իրական աշխարհում յուրաքանչյուր առաջադրանքի, քաղաքականությունի և պայմանի համար 50 փորձարկում, սիմուլյացիաում՝ առաջադրանքի համար 200: Ընդհանուր՝ մոտ **1,800 կույր իրական աշխարհի փորձարկում և ավելի քան 47,000 սիմուլյացիա փորձարկում**:
- **Իրական վիճակագրություն**. հաջողության հավանականության բայեսյան գնահատումներ և զույգային հիպոթեզների թեստեր՝ բազմաթիվհամեմատություն ուղղումներով, ոչ համընկնող սխալի գծերին աչքով նայելը: Նրանք նույնիսկ մարդկանց կողմից գնահատված փորձարկումների մեկ քառորդի վրա որակի ապահովման ստուգում են արել՝ գնահատում սխալը չափելու համար:

[video pending]

Breakfast table պատրաստելու մոդելների կողքկողքի համեմատություն՝ ձախում մեկառաջադրանք ելակետային, աջում LBM: Երկուսն էլ 1× արագությամբ են: Սա մեկ գնահատված առաջադրանք է, ոչ ընդհանուրնպատակ autonomy-ի ապացույց:

Այս ամբողջ machinery-ն հենց բուն միտքն է: Հոդվածի փաստարկն այն է, որ առանց նման գնահատման չեք կարող իրական բարելավումը բախտից տարբերել:

Ինչ գտան

լրացուցիչ ուսուցանել արված մեծ մոդելները միջինում հաղթում են զրոյից ուսուցանել արված մեկառաջադրանք մոդելներին: Task-երի վրա ընդհանրացված անելիս pretrain + լրացուցիչ ուսուցանել LBM-ը վստահորեն գերազանցել է նույն առաջադրանքի տվյալների վրա զրոյից ուսուցանել արված քաղաքականությունին՝ և սիմուլյացիաում, և իրական աշխարհում, իսկ տարբերությունը վիճակագրորեն նշանակալի էր: Առանձին առաջադրանքերում լրացուցիչ ուսուցմամբ հարմարեցված LBM-ը գրեթե միշտ վիճակագրորեն as-good-or-better էր՝ իրական աշխարհում 3/3 առաջադրանք և սիմուլյացիաում 15/16 առաջադրանք:

Ամենամեծ ու ամենամաքուր շահումը տվյալներ efficiency-ն է: Finetuned LBM-ը զրոյից-equivalent արդյունավետության հասել է մոտ **3–5× պակաս առաջադրանքկոնկրետ տվյալներ** օգտագործելով: Իրական աշխարհի մեկ առաջադրանքում՝ breakfast table պատրաստելիս, ցուցադրումների միայն **15%**-ով լրացուցիչ ուսուցանել արված LBM-ը գերազանցել է զրոյից քաղաքականությունին, որը ուսուցանել էր արված **100%** ցուցադրումներով:

նախնական ուսուցումը ամենաշատը օգնում է, երբ պայմանները փոխվում են: Երբ թեստ միջավայրը դիտավորյալ հեռացվել է ուսուցում պայմաններից՝ «բաշխում շեղում», լրացուցիչ ուսուցմամբ հարմարեցված LBM-ի առավելությունը *աճել է*: Սիմուլյացիայի մի set-ում նորմալ պայմաններում այն վիճակագրորեն հաղթել է զրոյից քաղաքականությունին 16 առաջադրանքից 3-ում, իսկ բաշխում շեղումի դեպքում՝ 16-ից 10-ում: Քանի որ իրական կիրառումները միշտ ինչոր չափով շեղվում են ուսուցում պայմաններից, կայունությունը հավանաբար գործնականորեն ամենակարևոր արդյունքն է:

Ավելի շատ preուսուցման տվյալներն օգնել է սահուն կերպով: արդյունավետությունը շարունակաբար աճել է preուսուցման տվյալներ ավելացնելիս՝ փորձարկված սանդղակներում **առանց հանկարծակի jump-ի կամ «ինքնաբերաբար առաջացող» թռիչքի**: Օգտակար, կանխատեսելի և ոչ դրամատիկ:

Բայց առանց լրացուցիչ ուսուցում ընդհանուր նշանակության մոդելի պատմությունը չաշխատեց: Pretrained LBM-ը *զրոյական օրինակներով* օգտագործելիս՝ առանց առաջադրանքկոնկրետ լրացուցիչ ուսուցումի, մեկ առաջադրանքի քաղաքականությունը հետևողականորեն չի գերազանցել: Մեկ ցանցը կարող էր միաժամանակ շատ առաջադրանք անել, բայց «պարզապես ասա՝ ինչ անի» երազանքը այստեղ չի հաստատվել: Հեղինակները դրա մի մասը կապում են իրենց փոքր լեզու encoder-ի փխրունության հետ:

Եվ շահումները այնքան փոքր էին, որ հեշտ էր դրանք բաց թողնել կամ կեղծել: Բազմաթիվ ազդեցություններ տեսանելի դարձան միայն սովորականից ավելի մեծ նմուշի չափներով և ճիշտ թեստերով: Հեղինակները բացահայտ գրում են, որ ազդեցության չափների և աղմուկի պայմաններում **Էական ռիսկ կա, որ robotics-ի շատ հոդվածներ պարզապես վիճակագրական աղմուկ են չափում**: Նրանք նաև գտել են, որ սովորական մի ընտրություն՝ տվյալների

նորմավորում-ը, արդյունքի վրա ավելի մեծ ազդեցություն է ունեցել, քան ճարտարապետությունը՝ փոփոխությունները, և նախնական ուսուցումի նորմավորում **bug**-ը հայտնաբերվել է միայն գնահատումները ավարտելուց *հետո*:

Ինչ է սա, ամենայն հավանականությամբ, նշանակում

Պաշտպանելի ընթերցումը սա է. տարբեր ռոբոտ տվյալների լայնամասշտաբ նախնական ուսուցումը իրական և օգտակար բաղադրիչ է: Այն նվազեցնում է նոր առաջադրանքի համար անհրաժեշտ տվյալները և քաղաքականություններին ավելի դիմացկուն է դարձնում, երբ իրական աշխարհը ուսուցումին չի համապատասխանում: Սա իսկապես աջակցում է այն ուղղությանը, որի վրա ոլորտը խաղադրույք է անում: Բայց շահումները *չափավոր և պայմանական* են՝ հիմնականում լրացուցիչ ուսուցումից հետո, առավել հստակ ընդհանրացվածում ու stress պայմաններում: Սա drop-in ընդհանուր ռոբոտի գալուստը չէ:

Ավելի լուռ, գուցե ավելի կարևոր հետևությունը մեթոդաբանական է: Հոդվածը գործնականում չափման քանոն է. ցույց է տալիս, թե ռոբոտ քաղաքականության մասին վստահելի պնդում անելու համար իրականում որքան ապացույցներ է պետք, և ակնարկում, որ հրապարակված ոգևորության մի մասը չափազանց քիչ ապացույցների վրա է կանգնած: Ոլորտին այս ուղղումն ավելի շատ է պետք, քան մեկ այլ մոդել:

Ինչ սա *չի* ապացուցում

- Սա **ընդհանուրնպատակ ռոբոտ չէ**: Շահումները ցուցադրված են կոնկրետ ճարտարապետությունի՝ diffusion քաղաքականությունների համար, որոնք յուրաքանչյուր առաջադրանքի համար լրացուցիչ ուսուցանել են արվել վերահսկվող միջավայրներում և teleoperated ցուցադրումներից: Սա կամայական նոր աշխատանքներ հրամանով կատարող ինքնավար ռոբոտ չէ:
- Սա **զրոյական օրինակներով օգտագործումը չի վավերացնել անում**: Առանց լրացուցիչ ուսուցումի մեծ մոդելը մեկառաջադրանք ելակետայիններին հետևողականորեն չի հաղթել:
- Սա **«ինքնաբերաբար առաջացող leap»-ի ապացույցներ չէ**: Scaling-ը արդյունավետությունը սահուն է բարձրացրել. այստեղ discontinuity չկա, որը կարող է «հանկարծ դարձավ ունակ» պատմություն կրել:
- Թվերը **հարաբերական և լաբորատորիասահման են**: Absolute success հաճախականությունները դիտավորյալ tune են արվել մոտ ~50%-ի, որպեսզի համեմատությունները զգայուն լինեն: Դրանք իրական աշխարհի հուսալիությունի չափ չեն, իսկ աշխատանքը մեկ լաբորատորիաի մեկ ճարտարապետությունն

է:
- Սա չի բացատրում **ինչու** որևէ քաղաքականություն հաջողվում կամ ձախողվում է, և մի քանի կոնկրետ առաջադրանքեր, որտեղ մեծ մոդելն ավելի վատ է եղել, հաղորդվում են առանց պատճառի բացատրության:
- Գնահատման համակարգից դուրս սա **ոչինչ չի ասում անվտանգության, ինքնավարության կամ իրական գործարկման մասին**:

Որքանո՞վ է ուժեղ ապացույցը

Կենտրոնական համեմատական պնդումների համար՝ լրացուցիչ ուսուցմամբ հարմարեցված LBM-ները ընդհանրացվածում հաղթում են զրոյից ելակետայիններին, մի քանի անգամ պակաս տվյալներ են պահանջում և բաշխում շեղումի դեպքում ավելի կայուն են, ապացույցները ուժեղ է և անսովոր լավ վերահսկված. կույր, պատահականացված, մեծնմուշ, վիճակագրորեն փորձարկված, գնահատումի QA pass-ով: Սա այն հազվադեպ դեպքն է, երբ մեթոդաբանությունը բավական ուժեղ է, որպեսզի վերնագիր եզրակացությունները մոտավորապես ուղիղ ընդունվեն:

Ազնիվ վերապահումները հեղինակներն իրենք են բարձրացնում: Նրանց սխալ bar-երը ներառում են *գնահատում*-ի պատահականությունը, բայց **ոչ ուսուցումի պատահականությունը**. նույն մոդելը երկու անգամ ուսուցանել անելու դեպքում կարող են բավական տարբեր քաղաքականություններ ստացվել, և այդ variation-ը վիճակագրության մեջ չկա: Իրական աշխարհի առաջադրանքերը յուրաքանչյուրն ունեին 50 փորձարկում՝ միջին ազդեցություններ գտնելու համար բավարար, բայց փոքր ազդեցությունները բաց թողնելու հավանականությամբ: Language conditioning-ը համեստ encoder էր օգտագործում, ուստի «պարզապես ասեք ռոբոտին՝ ինչ անել» պնդումները ավելի մեծ համակարգերի համար կարող են այլ լինել: Եվ կա նորմավորում bug-ի post-hoc բացահայտումը: Սրանցից ոչ մեկը հիմնական արդյունքը չի տապալում, բայց հենց այն մանրուքներն են, որոնք, ըստ հոդվածի, ոլորտը հաճախ անտեսում է:

Sourcing-ի մեկ նշում՝ նույն ոգով. այս explainer-ը հիմնված է հեղինակների նախատպագիրի վրա: Journal-հրապարակված տարբերակը հասանելի չդարձավ, այնպես որ նախատպագիրից հրապարակում անցման ժամանակ հնարավոր փոփոխությունները չենք ստուգել:

Ինչու է սա կարևոր

«ռոբոտ հիմք մոդել» արտահայտությունը չափազանց պնդումների համար գրեթե պատրաստված է, և այսպիսի աշխատանքը հեշտ է սխալ կարդալ երկու ուղղությամբ՝ հաղթական «աշխատեց» կամ dismissive «overhyped է»: Ճիշտ ընթերցումը երկուսից էլ օգտակար է. տարբեր տվյալների վրա նախնական ուսուցումը տալիս է իրական, չափելի, բայց moderate օգուտներ՝ հիմնականում քիչ տվյալներ per առաջադրանք և

ավելի կայունություն, և սանդղակի հետ ուղին կանխատեսելիորեն լավանում է։

Ավելի խոր կարևորությունն այն է, որ հոդվածը իր խստությունը շրջում է սեփական ոլորտի վրա։ Ցույց տալով, որ իրական ազդեցությունները այնքան փոքր են, որ sloppy գնահատումի մեջ կարող են անհետանալ, և որ տվյալների նորմավորումի նման ձանձրալի ընտրությունը կարող է clever ճարտարապետությունից ավելի կարևոր լինել, այն փաստարկում է, որ ռոբոտ learning-ի մեծ մասը ավելի ամուր չափում է պահանջում, մինչև առաջընթացին հավատալը։ Հոդվածը, որը իր credibility-ն ծախսում է արդյունքի և ցանկության տարբերությունը վերահսկելու վրա, ավելի հազվադեպ ու ավելի արժեքավոր բան է անում, քան leaderboard գլխավորելը։

Կարճ ամփոփում

Toyota Research Institute-ի հետազոտողները ուսուցանել են արել «մեծ վարք մոդել»-ներ՝ diffusion-based ռոբոտ քաղաքականություններ, որոնք pretrain են արվել մոտ 1,700 ժամ տարբեր manipulation տվյալների վրա, և դրանք համեմատել զրոյից մեկ առաջադրանքի քաղաքականությունների հետ անսովոր խիստ արձանագրությունով՝ կույր, պատահականացված, մեծնմուշ մոտ 1,800 իրական աշխարհի և 47,000+ սիմուլյացիա փորձարկումներով և իրական վիճակագրությամբ։ Task-կոնկրետ լրացուցիչ ուսուցումից հետո մեծ մոդելները ընդհանրացվածում վստահորեն ավելի լավ էին, մոտ **3–5× պակաս առաջադրանքկոնկրետ տվյալներ** էին պահանջում և **ավելի կայուն էին բաշխում շեղումի դեպքում**, իսկ preուսուցման տվյալներ ավելացնելիս արդյունավետությունը սահուն աճում էր։ Բայց **առանց լրացուցիչ ուսուցումի** դրանք մեկառաջադրանք մոդելներին հետևողականորեն չէին հաղթում, մի քանի ազդեցություն այնքան փոքր էին, որ միայն մեծ նմուշներն էին ցույց տալիս, իսկ տվյալներնորմավորումի սովորական ընտրությունը ճարտարապետությունից ավելի մեծ ազդեցություն ուներ։ Սա ռոբոտհիմքմոդել ուղղությանը ուժեղ, չափավոր աջակցություն է՝ **ոչ** ընդհանուրնպատակ ռոբոտ, ոչ զրոյական օրինակներով generalist, ոչ «ինքնաբերաբար առաջացող leap», և միաժամանակ սուր զգուշացում, որ robotics-ի շատ արդյունքներ կարող են աղմուկ չափել։

Առանց չափազանցության

Ինչ է ցույց տալիս հոդվածը. խիստ կույր և վիճակագրորեն-powered գնահատումով մոտ 1,800 իրական աշխարհի + 47,000+ սիմուլյացիա փորձարկում-ներում multitask-նախապես ուսուցված-then-լրացուցիչ ուսուցմամբ հարմարեցված diffusion քաղաքականությունը ընդհանրացվածում գերազանցում են զրոյից մեկ առաջադրանքի քաղաքականություններին, մոտ 3–5× պակաս առաջադրանքկոնկրետ տվյալներով հասնում համարժեք արդյունավետության և բաշխում շեղումի դեպքում ավելի կայուն են։ արդյունավետությունը preուսուցման տվյալների հետ սահուն սանդղակ է անում։

Ինչն է հավանական, բայց ապացուցված չէ. որ նույն օգուտները կտեղափոխվեն շատ ավելի մեծ vision-լեզու-action մոդելների վրա՝ այստեղ լեզու encoder-ը փոքր էր, և որ smooth մասշտաբավորումը կշարունակվի փորձարկված տվյալներ միջակայքից դուրս:

Ինչ սա չի ցույց տալիս. ընդհանուրնպատակ կամ զրոյական օրինակներով ռոբոտ, առանց լրացուցիչ ուսուցումի հետևողական առավելություն, որևէ «ինքնաբերաբար առաջացող» կարողություն jump, առաջադրանքմակարդակ ձախողումների ամբողջական պատճառաբանություն, բացարձակ իրական աշխարհի հուսալիություն կամ անվտանգությունն կիրառումի մասին ապացույցներ:

Հիմնական սահմանափակումները. վիճակագրությունը ներառում է գնահատում պատահականությունը, ոչ ուսուցում-run պատահականությունը, 50 իրական աշխարհի փորձարկում per առաջադրանքը կարող է փոքր ազդեցություն բաց թողնել, մեկ ճարտարապետություն և մեկ լաբորատորիա, համեստ լեզու encoder, գնահատումից հետո հայտնաբերված տվյալներնորմավորում bug և վերլուծություն նախատպագիրի հիման վրա՝ հրապարակված տարբերակը չստուգված:

Որքա՞ն վստահություն պետք է ունենա ընդհանուր ընթերցողը: Բարձր՝ որ multitask նախնական ուսուցում + լրացուցիչ ուսուցումը իրական, moderate օգուտներ է տալիս, հատկապես տվյալներ efficiency և կայունություն, և որ դրանք անսովոր զգուշորեն են չափվել: Բարձր՝ որ սա **ընդհանուրնպատակ կամ զրոյական օրինակներով ռոբոտ չէ** և **ինքնաբերաբար առաջացող leap չէ**: Միջին՝ թե օգուտները որքան հեռու կսանդղակվեն ավելի մեծ մոդելների վրա: Եվ լուրջ արժե ընդունել հեղինակների սեփական զգուշացումը. ազդեցությունները այնքան փոքր են, որ անբավարար վիճակագրական հզորությամբ ուսումնասիրությունները կարող են աղմուկ հրապարակել: Ճիշտ դիրքորոշումը՝ չափավոր լավատեսություն մոտեցման նկատմամբ և առողջ թերահավատություն այն ռոբոտԱԲ արդյունքների հանդեպ, որոնք այս կարգի վիճակագրական հիմնավորում չունեն:

Աղբյուրներ

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/