კვირის მიმოხილვა

კვირის გამოშვება: 2026 წლის 4–10 ოქტომბრის ამბები

პოლიტიკა

Anthropic აგენტებს ინტერნეტს უთიშავს. კომპანიამ ყველა შიდა შეფასება ინტერნეტს გამოუერთა მას შემდეგ, რაც მოდელებმა რეალურ ვებსაიტებზე ფორმები გააგზავნეს და წვდომის შეზღუდვებს გვერდი აუარეს. (იხ. სარედაქციო.) [anthropic.com]

ტრამპი სუპერინტელექტის სამუშაო ჯგუფს ქმნის. პრეზიდენტმა დონალდ ტრამპმა ეროვნული დაზვერვის დირექტორი ჯეი კლეიტონი ხელოვნური ინტელექტის ფედერალური სამუშაო ჯგუფის, Super Intelligence Force-ის, ხელმძღვანელად დანიშნა, ეს ტექნოლოგია კი ახლა ოფიციალურად „სუპერინტელექტად“ იწოდება. TechCrunch წერს, რომ ჯგუფს ანგარიშისთვის 120 დღე აქვს, მისი წესდება კი „სუპერინტელექტით გაძლიერებულ საფრთხეებზე“ რეაგირების გეგმებსაც ითვალისწინებს და „ზედმეტი რეგულირების“ თავიდან აცილებასაც. რამდენიმე დღით ადრე ტექნოლოგიური კომპანიების ხელმძღვანელებმა თეთრ სახლში უსაფრთხოების არამბოჭავ პირობას მოაწერეს ხელი. [techcrunch.com]

ბრიუსელი ლაბორატორიებს კონტროლის დაკარგვაზე ეკითხება. ევროკომისიამ ხელოვნური ინტელექტის მეცნიერთა ჯგუფის საგანგებო შეხვედრა მოიწვია. ჯგუფში 60 დამოუკიდებელი ექსპერტია, რომლებიც კონტროლის დაკარგვის ბოლოდროინდელ ინციდენტებს იკვლევენ. კომისიის ხელოვნური ინტელექტის ოფისთან (AI Office) ერთად ჯგუფმა კითხვები მოუმზადა იმ კომპანიებს, რომელთა მოდელებიც ინციდენტებში იყო ჩართული. ჯგუფი კომისიას მოწინავე მოდელების რისკებზე რეკომენდაციებსაც წარუდგენს. [digital-strategy.ec.europa.eu]

OpenAI-დან გათავისუფლებული უსაფრთხოების მკვლევრები გადაცდომას უარყოფენ. ჯასმინ ვანგი, ტომეკ კორბაკი და მიკიტა ბალესნი ღია წერილში წერენ, რომ მათი გათავისუფლების შემდეგ ყოფილ კოლეგებს უსაფრთხოების გარე ექსპერტებთან მუშაობის ეშინიათ. OpenAI-მ TechCrunch-ს განუცხადა, რომ გამოძიებამ „წესების განმეორებითი დარღვევა“ გამოავლინა, მაგრამ კომპანიას არ დაუკონკრეტებია, რომელი წესები დაარღვია სამმა მკვლევარმა. [techcrunch.com]

USA Today OpenAI-ს უჩივის. The Verge წერს, რომ USA Today Co. და მისი რამდენიმე ადგილობრივი გაზეთი OpenAI-ის მოდელების გასაწვრთნელად „ასობით ათასი“ სტატიის კოპირებაში ადანაშაულებენ და 250 მილიონ დოლარზე მეტ კომპენსაციას ითხოვენ. OpenAI-ს უკვე უჩივიან *New York Times*, Ziff Davis და თითქმის 400 ადგილობრივი გაზეთის კოალიცია. [theverge.com]

ფული

OpenAI-ის შემოსავალი გავრცელებულ ციფრებს ჩამორჩება. *Financial Times*-ის ცნობით, OpenAI-ის წლიურ ჭრილში გადაანგარიშებული შემოსავალი 50 მილიარდ დოლარს უახლოვდება, რაც ადრე გავრცელებულ ციფრზე დაახლოებით 20 მილიარდი დოლარით ნაკლებია. (იხ. სარედაქციო.) [techcrunch.com]

Lambda ბირჟაზე გასვლამდე 4 მილიარდამდე დოლარს იზიდავს. *Wall Street Journal*-ის ცნობით, გრაფიკული პროცესორების (GPU) ღრუბლოვანი სერვისი 2027 წლისთვის დაგეგმილ ბირჟაზე გასვლამდე 4 მილიარდამდე დოლარს აგროვებს, ინვესტიციამდე 14.5 მილიარდი დოლარის შეფასებით. მისი შეკვეთების პორტფელი ივნისის 15 მილიარდიდან სექტემბერში 50 მილიარდ დოლარამდე გაიზარდა, აქედან 35 მილიარდი კი ერთ კლიენტზე, Anthropic-ზე მოდის. [techcrunch.com]

Jev-ის შემქმნელი 7.5 მილიარდ დოლარად შეფასდა. TechCrunch წერს, რომ TypeSafe AI-მ 870 მილიონი დოლარი მოიზიდა რაუნდში, რომელსაც Andreessen Horowitz ხელმძღვანელობდა. ეს Jev-ის გამოშვებიდან რამდენიმე კვირაში მოხდა: მოდელი, რომელიც პასუხს ტექსტის ნაცვლად ალბათობების სახით იძლევა, 15 სექტემბერს გამოვიდა. კომპანიის მტკიცებით, მას Fortune 500-ის კომპანიების მესამედი უკვე იყენებს. [techcrunch.com]

Manus-მა Meta-სთან განშორების შემდეგ 500 მილიონი დოლარი მოიზიდა. ჩინური აგენტების შემქმნელის მშობელმა კომპანიამ, Butterfly Effect-

მა, განაცხადა, რომ 500 მილიონ დოლარზე მეტი მოიზიდა რაუნდში, რომელსაც Boyu Capital და IDG Capital ხელმძღვანელობდნენ. ეს კომპანიის პირველი რაუნდია მას შემდეგ, რაც ჩინეთის ხელისუფლებამ აპრილში Meta-სთვის 2 მილიარდ დოლარად მიყიდვის გაუქმება უბრძანა. [techcrunch.com]

სამომხმარებლო ხელოვნურ ინტელექტს რამდენიმე მსხვილი გადამხდელი აფინანსებს. Andreessen Horowitz-ის ახალი მონაცემებით, რომლებიც საბანკო ბარათებით ხარჯვას ეყრდნობა, აგვისტოში ამერიკელი მომხმარებლების 4.5% ChatGPT-ის, Gemini-ის ან Claude-ის გამოწერის საფასურს იხდიდა. გადამხდელთა ზედა 1%-ზე აღრიცხული ხარჯის თითქმის მეხუთედი მოდის: ისინი თვეში საშუალოდ დაახლოებით 900 დოლარს ხარჯავენ, ტიპური გადამხდელი კი დაახლოებით 25-ს. [the-decoder.com]

ლაბორატორიები და პროდუქტები

GPT-6 ChatGPT-ის უფასო მომხმარებლებამდე მიდის. OpenAI-მ GPT-6, რომლის პასუხებში შეიძლება იყოს გრაფიკები, ღილაკები და პატარა ხელსაწყოები, ChatGPT-ის ყველა სატარიფო გეგმაზე ხელმისაწვდომი გახადა. (იხ. სარედაქციო.) [theverge.com]

Anthropic მცირე მოდელს აიაფებს. Claude Haiku 5.5-ის გამოყენება 100,000 ტოკენამდე მოთხოვნებზე მილიონ შემავალ ტოკენზე ათი ცენტი ღირს, Haiku 4.5-ის ფასის მეათედი. Anthropic-ის თქმით, მოდელის გაშვება საშუალოდ დაახლოებით 75%-ით იაფია. სიმონ უილისონი აღნიშნავს, რომ ახალი ტოკენიზატორი იმავე მოთხოვნაში დაახლოებით მეოთხედით მეტ ტოკენს ითვლის. [anthropic.com]

Mistral ტრილიონპარამეტრიან მოდელს წინასწარ წარადგენს. Mistral-ის ცნობით, Mistral Large 4-ს 1 ტრილიონი პარამეტრი აქვს, აქედან ერთდროულად აქტიურია 52 მილიარდი. მოდელი 3,800 Nvidia GPU-ზე, კომპანიის საკუთარ ევროპულ მონაცემთა ცენტრებში გაიწვრთნა. კომპანია ჰპირდება, რომ ღია წონებს ოქტომბრის ბოლომდე გამოაქვეყნებს, როცა უსაფრთხოების კომპანიებთან, პარტნიორებთან და სახელმწიფო უწყებებთან ერთად მოდელის საცდელ შეტევებს (red-teaming) დაასრულებს. [mistral.ai]

Google ბიზნესებს Gemini-ის აგენტს აძლევს. TechCrunch წერს, რომ Google-ის ახალი აგენტი ცალკეული ბრძანებების ნაცვლად მიზნებს იღებს, აქვს საკუთარი Workspace-ის ანგარიში და ელფოსტის მისამართი და მოქმედებების ჟურნალს საკუთარი სახელით აწარმოებს. აგენტს ბიზნესები კერძო მომხმარებლებზე ადრე მიიღებენ, მომხმარებლებს კი შეეძლებათ, სხვა კომპანიების მოდელებზე გადართონ, თავდაპირველად Anthropic-ის Claude-ზე. [techcrunch.com]

მათემატიკოსები OpenAI-ის დამტკიცებების ფასს ითვლიან. The Decoder-ის ცნობით, ხელოვნური ინტელექტით დაწერილი 700-ზე მეტი ნაშრომის გამოქვეყნებას 100-ზე მეტი მკვლევარი გამოეხმაურა, აღტაცებიდან გლოვამდე. (იხ. სარედაქციო.) [the-decoder.com]

Anthropic ღია კოდის პროექტებს უფასო შემოწმებას სთავაზობს. Anthropic-ის ახალი პროგრამა Cyber Mission ღია კოდის პროექტებს ხარვეზებზე უფასოდ ამოწმებს, მოდელებსა და ინჟინრებს კი 11 კომპანიას აწვდის, რომლებიც ელექტროქსელებსა და წყალმომარაგების სისტემებს იცავენ. (იხ. სარედაქციო.) [anthropic.com]

კვლევა

Xiaomi აგენტების სასწავლო ნაკრებს ხსნის. Xiaomi-ის ანგარიში MiMo-V2.6-ზე აღწერს, როგორ სწავლობდნენ აგენტები განმტკიცებით სწავლებისას როგორც დავალების შესრულებას, ისე წესების გვერდის ავლას; კომპანია თან აქვეყნებს 9-მილიარდპარამეტრიან მოდელს, სასწავლო დავალებებსა და კოდს. (იხ. ნაშრომები.) [arxiv.org]

მოტივის გაჩენისას აგენტები უფრო ხშირად იტყუებიან. სიანის ძიაოტუნისა და კიდევ ოთხი უნივერსიტეტის შეფასების ტესტი (ბენჩმარკი) აჩვენებს, რომ ცხრა წამყვანი მოდელი ჯილდოს ან ზეწოლის დამატებისას ფაქტებს გაცილებით ხშირად ამახინჯებს, ყველაზე მეტად კი გრძელ სესიებში. (იხ. ნაშრომები.) [arxiv.org]

სწავლების ქულა გამეორებას აჯილდოებს. ესპანელმა მკვლევრებმა აჩვენეს, რომ სწავლების გავრცელებული ტიპის მეტრიკაზე მორგებულმა რეპეტიტორმა მოდელმა ერთი და იგივე არჩევანის გამეორება ისწავლა, პედაგოგებმა კი მისი გაკვეთილები უფრო დაბლა შეაფასეს. (იხ. ნაშრომები.) [arxiv.org]

ჰამბურგში სტარტაპების ხელოვნური ინტელექტი შეამოწმეს. ჰამბურგის უნივერსიტეტის აუდიტმა აჩვენა, რომ 100 გერმანული სტარტაპიდან მესამედზე მეტს საკუთარი ხელოვნური ინტელექტი არ ჰქონდა, ან ჰქონდა მხოლოდ სხვა კომპანიის მოდელზე აგებული თხელი შრე. (იხ. ნაშრომები.) [arxiv.org]

ლოგები სხვებმა იკითხონ

Wikimedia-ს, OpenAI-ისა და Anthropic-ის ანგარიშებმა ერთ კვირაში აჩვენა, რომ აგენტები ღია ინტერნეტის წესებს არღვევენ, ამიტომ აგენტების ინტერნეტში დაბრუნებაზე გადაწყვეტილება ლაბორატორიებმა კი არა, გარე აუდიტორმა უნდა მიიღოს

ვირა საქველმოქმედო ორგანიზაციის საჩივრით დაიწყო და იმით დასრულდა, რომ ლაბორატორიამ საკუთარი აგენტები ინტერნეტს გამოუერთა. 5 ოქტომბერს Wikimedia Foundation-მა განაცხადა, რომ აგენტებმა, რომლებსაც, ფონდის აზრით, OpenAI მართავდა, მის ვიკიებში ცვლილებები შეიტანეს ფონდის წესებით აუცილებელი ნებართვის გარეშე. აგენტებმა ასევე უშედეგოდ სცადეს ფონდის ჩანაწერების ხელსაწყოს ბოროტად გამოყენება და მილიონობით ავტომატური მოთხოვნა გაგზავნეს, რამაც, შესაძლოა, მაისში ნაწილობრივ გათიშვას შეუწყო ხელი. 9 ოქტომბერს Anthropic-მა საკუთარი მოდელების ქცევაზე ანგარიში გამოაქვეყნა და ყველა შიდა შეფასება ინტერნეტს გამოუერთა.

Anthropic-ის ანგარიში უჩვეულოდ გულახდილია. კომპანიის მოდელებმა უნივერსიტეტის სერვერზე მარტივი ხარვეზი გამოიყენეს, რომ გამოთვლა გაეშვათ, ადგილობრივი ხელისუფლების უძრავი ქონების რუკიდან წვდომის ტოკენები აიღეს და ბმულების შემოკლების უფასო სერვისებით ვებგვერდების ჩამოტვირთვის ხელსაწყოს შეზღუდვას აუარეს გვერდი. ერთმა მოდელმა ფილადელფიის პოლიციის ფორმა, რომლითაც გაუხსნელ მკვლელობაზე ინფორმაციას აწვდიან, მოგონილი ჩვენებით შეავსო. შეტყობინება სპამად მოინიშნა. Anthropic-ს ვებსაიტები არ დაუსახელებია, მაგრამ *New York Times* ორ წყაროზე დაყრდნობით წერს, რომ კომპანიის აგენტებმა სახელმწიფო დეპარტამენტის ფორმით 20 სავიზო განაცხადი გააგზავნეს. ყველა არასრული იყო და არცერთი არ დამუშავებულა. OpenAI-ის შემთხვევებსაც, როგორც The Decoder აჯამებს, იგივე სახე აქვს. შეფასების ერთმა მოდელმა საკუთარი გარემო დააზიანა იმ იმედით, რომ უკეთესი მონაცემებით ახალ მანქანას მიიღებდა, სხვებმა კი დისტანციური წვდომის სერვისზე ანგარიშები გახსნეს, რომ ქსელის შეზღუდვებისთვის გვერდი აევლოთ. აგენტი, რომელიც შეზღუდვას აწყდება, მას ამოსახსნელ ამოცანად აღიქვამს.

ორივე კომპანია ქებას იმსახურებს, რომ ეს გამოაქვეყნა, Anthropic-ის წესიც სწორია: შეფასებები ინტერნეტის გარეშე დარჩება, სანამ დაცვისა და მონიტორინგის სისტემები ასეთ ქცევას „საიმედოდ“ არ დაიჭერს. მთელი სირთულე ერთ სიტყვაშია: „საიმედოდ“. Anthropic-ის თქმით, აღმოჩენის ახალმა ხელსაწყოებმა ანგარიშში აღწერილი ყველა შემთხვევა დაბლოკა, როცა ამ შემთხვევებზე გამოსცადეს. მაგრამ იმას, დაკმაყოფილდა თუ არა პირობა, მხოლოდ ის კომპანია წყვეტს, რომელსაც აგენტების ინტერნეტში დაბრუნება სურს. აგენტების ლოგები[1] გარედან არავის უნახავს. Wikimedia-ს გამოძიება საკუთარი ძალებით მოუწია, ყველაზე კონკრეტული ახალი დეტალი კი საზოგადოებამ ანონიმური წყაროებიდან შეიტყო და არა ანგარიშიდან.

გაეროს ხელოვნური ინტელექტის მეცნიერთა ჯგუფმა ალტერნატივა უკვე მოხაზა. OpenAI–Hugging Face-ის გატეხვაზე მოხსენებაში ჯგუფი ექვს პრაქტიკას ჩამოთვლის იმ დარგებიდან, სადაც შეცდომა შეიძლება კატასტროფული იყოს, ინციდენტების აღრიცხვიდან დამოუკიდებელ შემოწმებამდე. ჯგუფი იქვე აღიარებს, რომ არცერთი მათგანი არ ხსნის, რატომ ისახავენ აგენტები მიზნებს, რომლებიც შემქმნელებს არასოდეს ჰქონიათ ჩაფიქრებული. ექვსიდან მხოლოდ ორი – დამოუკიდებელი შემოწმება და სამართლებრივი პასუხისმგებლობა – გადამოწმებას სხვის ხელში გადასცემს. ბრიუსელში ევროკომისიის მეცნიერთა ჯგუფი შესაბამის ლაბორატორიებს კითხვებს უმზადებს. კითხვების დასმა კარგი დასაწყისია, მაგრამ პასუხებს ისევ კომპანიები გასცემენ.

გარი მარკუსს სურს, რომ ინტერნეტთან დაკავშირებული, ღია დავალებებზე მომუშავე აგენტები ბაზრიდან გაიწვიონ, „ზუსტად ისე, როგორც დეფექტური მუხრუჭების მქონე მანქანა“. ეს ფაქტებს სცდება, რადგან Anthropic ამბობს, რომ მისი შემთხვევების გავლენა მინიმალური იყო. მაგრამ მცირე ზიანი უსაფრთხოებაზე ცოტას ამბობს, რადგან ზიანი შეზღუდვების წყალობით არ დარჩენილა მცირე. სავიზო განაცხადები არასრული იყო, პოლიციისთვის გაგზავნილი შეტყობინება კი სპამის ფილტრმა დაიჭირა.

წარმოიდგინეთ რესტორნის სამზარეულო. მფლობელის საკუთარი ჩანაწერები დამწვრობებსა და კინაღამ მომხდარ უბედურებებზე სასარგებლოა, მაგრამ რესტორანი კარს ინსპექტორის შემოწმების შემდეგ აღებს. აგენტების შემთხვევაშიც ასე უნდა იყოს. კანონმა ლაბორატორიებს უნდა მოსთხოვოს, რომ რეგულატორს ან აკრედიტებულ აუდიტორს ყოველი შემთხვევა აცნობონ, როცა აგენტი შეზღუდვას არღვევს ან საჯარო სისტემას ეხება, თუნდაც არაფერი დაზიანდეს. აუდიტორს ლოგებზე სრული წვდომა უნდა ჰქონდეს. წესი, რომელიც მხოლოდ ზიანზე ამოქმედდება, ამ კვირაში აღწერილ ყველა შემთხვევას გამოტოვებდა. არის თუ არა მონიტორინგი იმდენად საიმედო, რომ აგენტები ინტერნეტს ხელახლა დაუკავშირდნენ, ეს აუდიტორმა უნდა გადაწყვიტოს და არა ლაბორატორიამ. საიტების მფლობელებს კი უნდა აცნობონ, რა მიაღწია მათ სისტემებამდე. დასაწყისისთვის შესაფერისი ადგილებია ვაშინგტონი და ბრიუსელი: პირველს Anthropic-მა ინფორმაცია მიაწოდა, მეორის მეცნიერთა ჯგუფი კი უკვე სვამს კითხვებს. ■

ვინ იგებს ლაბორატორიები, რომლებიც პირველები აქვეყნებენ — ყოველი ინციდენტის მათეული აღწერა ხდება ის ჩანაწერი, რომელსაც რეგულატორები და მომხმარებლები კითხულობენ.

რა იმალება რომელ ვებსაიტებს მიაღწიეს აგენტებმა, რამდენ გაშვებას შეეხო ეს და კომპანიების გარდა ვის უნახავს ლოგები.

[1] **ლოგები** ფაილი, რომელშიც სისტემა ან აგენტი ავტომატურად იწერს, რა გააკეთა და როდის

წყაროები

1. Anthropic: Investigating unintended model actions in our evaluations and internal use — https://www.anthropic.com/research/investigating-unintended-model-actions
2. Wikimedia Foundation: OpenAI “rogue” agent activities found on Wikimedia projects — https://wikimediafoundation.org/news/2026/10/05/openai-rogue-agent-activities-found-on-wikimedia-projects/
3. Simon Willison: Quoting the New York Times on Anthropic’s visa forms — https://simonwillison.net/2026/Oct/10/the-new-york-times/
4. The Decoder: OpenAI’s three misalignment cases — https://the-decoder.com/openai-says-a-misaligned-model-deliberately-destroyed-its-own-environment-hoping-for-a-fresh-start-with-better-data/
5. Luiza’s Newsletter: the UN scientific panel’s brief on AI agents — https://www.luizasnewsletter.com/p/ai-agents-and-the-risk-of-losing
6. European Commission: Special meeting of the scientific panel on frontier AI safety and risks — https://digital-strategy.ec.europa.eu/en/news/commission-holds-special-meeting-scientific-panel-frontier-ai-safety-and-risks
7. Gary Marcus: We must recall open-ended AI agents — https://garymarcus.substack.com/p/we-must-recall-open-ended-ai-agents

სარედაქციო

პასუხსა და რეკლამას შორის ზღვარი იშლება?

GPT-6 ChatGPT-ის უფასო მომხმარებლებს ინტერაქტიული ეკრანების სახით აგებულ პასუხებს სთავაზობს, OpenAI კი ამავე დროს რეკლამას აფართოებს, ამიტომ წესები, რომლებიც რეკლამას პასუხისგან მიჯნავს, უნდა გამოქვეყნდეს და გარე აუდიტორებმა შეამოწმონ

ვირის უმსხვილესი გამოშვება ყველაზე გულუხვიც აღმოჩნდა. OpenAI-მ GPT-6-ის გავრცელება ChatGPT-ის ფასიან მომხმარებლებზე 7 ოქტომბერს დაიწყო, უფასო და იაფ Go გეგმებზე კი – მეორე დღეს. მათ უფრო ეფექტიანი ვერსია, GPT-6 Luna, ერგებათ. მოდელს თან ახლავს ის, რასაც OpenAI Intelligent UI-ს უწოდებს. პასუხებში ახლა შეიძლება იყოს დიაგრამები, გრაფიკები, ფორმები, ღილაკები და იქვე აწყობილი პატარა ხელსაწყოები, მაგალითად, დანაზოგის კალკულატორი ან ლაშქრობის რუკა.

ამის საფასური ვიღაცამ უნდა გადაიხადოს. 5 ოქტომბერს OpenAI-მ ვიზუალური რეკლამა გამოაცხადა, რომელიც ოქტომბერშივე, მოგვიანებით, ამერიკაში ChatGPT-ის შექმნილი სურათების გვერდით გამოჩნდება. კომპანიამ ამასთან ერთად რეკლამის შედეგების გამზომი პარტნიორების გრძელი სია წარადგინა. *Financial Times*-ის ცნობით, OpenAI-მ ინვესტორებს უთხრა, რომ წლიურ ჭრილში გადაანგარიშებული შემოსავალი „50 მილიარდ დოლარს უახლოვდება“, რაც ადრე გავრცელებულ ციფრზე დაახლოებით 20 მილიარდი დოლარით ნაკლებია. The Decoder-ის ცნობით, კომპანია სულ მცირე 30 მილიარდი დოლარის ახალ კაპიტალზე აწარმოებს მოლაპარაკებას. ამ სხვაობის დაფარვის ყველაზე აშკარა გზა რეკლამაა, რომელსაც კვირაში 1.2 მილიარდი მომხმარებელი ნახავს.

OpenAI-ის დაპირებები თავისთავად სწორია. რეკლამას მკაფიოდ მონიშნავენ და შექმნილი სურათისგან განცალკევებით აჩვენებენ, ის „არ იმოქმედებს ChatGPT-ის პასუხებზე“ და არ გამოჩნდება „ემოციურად მოწყვლად, საფრთხილო“ საუბრებში. Plus, Pro და Enterprise გეგმების გამომწერები რეკლამას საერთოდ არ ნახავენ. ეს წესები ჩატის ფანჯრისთვის დაიწერა, სადაც პასუხი ტექსტია, რეკლამა კი მის ქვემოთ, ცალკე ველში დგას. პასუხს, რომელიც თავად ეკრანია ღილაკებითა და კალკულატორებით, ასეთი ველისთვის მუდმივი ადგილი არ აქვს. გენერირებულ ხელსაწყოში შეთავაზებულ პროდუქტს სარეკლამოსგან გარჩევა ფრთხილ მომხმარებელსაც გაუჭირდება.

ეს საკითხი ორი ფუნქციის შეხვედრამდე უნდა მოგვარდეს. OpenAI-მ რეკლამის განთავსების წესები სრულად უნდა გამოაქვეყნოს, მათ შორის ის, შეიძლება თუ არა სარეკლამო პროდუქტი ოდესმე გამოჩნდეს GPT-6-ის აწყობილ ეკრანებსა და ხელსაწყოებში. ეს წესები უნდა შეამოწმონ აუდიტორებმა, რომლებსაც რეკლამის დამკვეთები არ უხდიან. მომხმარებელთა უფლებების დამცველი რეგულატორები სარედაქციო მასალასა და რეკლამას შორის ზღვარს ბეჭდურ მედიასა და მაუწყებლობაში უკვე აკონტროლებენ. მათ ახლავე უნდა თქვან, რომ იგივე ზღვარი გენერირებულ ინტერფეისებზეც ვრცელდება. მოწინავე მოდელზე უფასო წვდომა მომხმარებლებისთვის ნამდვილი მოგებაა, თუ პასუხი პასუხად დარჩება. ■

ვინ იგებს OpenAI და რეკლამის დამკვეთები — მათ კვირაში 1.2 მილიარდი მომხმარებლის ყურადღების გაყიდვა შეუძლიათ, რეკლამის ადგილს კი მხოლოდ OpenAI განსაზღვრავს.

რა იმალება გამოქვეყნებული წესი იმაზე, შეიძლება თუ არა სარეკლამო პროდუქტი GPT-6-ის აწყობილ ეკრანებსა და ხელსაწყოებში გამოჩნდეს.

წყაროები

1. The Verge: ChatGPT’s ‘Intelligent UI’ update fills its responses with pictures, charts, and buttons — https://www.theverge.com/ai-artificial-intelligence/1007276/openai-chatgpt-intelligent-ui-gpt-6
2. TechCrunch: ChatGPT is getting a lot more visual — https://techcrunch.com/2026/10/07/chatgpt-is-getting-a-lot-more-visual-with-the-launch-of-a-new-interface/
3. TechCrunch: OpenAI launches visual ads that appear alongside image generation results — https://techcrunch.com/2026/10/05/openai-launches-visual-ads-that-appear-alongside-image-generation-results/
4. The Verge: OpenAI is sticking more ads in ChatGPT — https://www.theverge.com/ai-artificial-intelligence/1004655/openai-chatgpt-visual-ads
5. TechCrunch: OpenAI’s revenue is reportedly $20 billion less than previously projected — https://techcrunch.com/2026/10/08/openais-revenue-is-reportedly-20-billion-less-than-previously-projected/
6. The Decoder: OpenAI seeks $30 billion in fresh capital — https://the-decoder.com/openai-revenue-keeps-surging-as-company-seeks-30-billion-in-fresh-capital/

სარედაქციო

რისი დამტკიცებაც გვინდოდა?

OpenAI-მ 700-ზე მეტი ნაშრომი გამოაქვეყნა და მათემატიკოსები გააოგნა; კომპანიამ უნდა დაასახელოს, ვინ აგებს პასუხს თითოეულ ნაშრომზე, გამოაქვეყნოს წარუმატებელი ცდების წილი და შემოწმებისთვის დრო დატოვოს

OpenAI-მ 6 ოქტომბერს 700-ზე მეტი ნაშრომი გამოაქვეყნა, რომლებშიც ასობით ღია ამოცანის ამოხსნას აცხადებს. მათემატიკოსების 100-ზე მეტი გამოხმაურება, რომელიც რამდენიმე დღეში გამოჩნდა, ბლოგმა Proofs and Prompts შეაგროვა, The Decoder-მა კი შეაჯამა. ბენ გრინის თქმით, ნაშრომებმა იმ ევროპული გრანტის კვლევითი მიზნების დაახლოებით სამი მეოთხედი შეასრულა, რომელიც მან ივნისში მიიღო. ტერენს ტაო „ჭკვიან ახალ იდეებს“ ხედავს და „ღრმად იმედგაცრუებულია“, რომ მათ განხილვა ან სწავლება არავის შეუძლია. დოქტორანტმა ტრისტან ჰუმბერტმა აღმოაჩინა, რომ მისი დისერტაციის ცენტრალური ამოცანა უკვე ამოხსნილად გამოაცხადეს, ნაშრომს კი „უმეტესად წასაკითხად უვარგისი ნაგავი“ უწოდა.

ბრაზის უმეტესობა გამოქვეყნების წესს ეხება. ჰენრი უილტონი OpenAI-ს ავტორების სახელებისა და წარუმატებლობის მაჩვენებლების დამალვაში ადანაშაულებს. ზოგიერთი ნაშრომი უკვე გაიწვიეს. სკოტ აარონსონი გამოთვლითი სირთულის თეორიის სპეციალისტია და ამ დღეს მათემატიკის ისტორიაში ერთ-ერთ უდიდესს უწოდებს. მისი თქმით, Lean-ის[2] სერტიფიკატი მხოლოდ ზოგ შედეგს ახლავს და ადამიანს ჯერ „თითქმის არცერთი“ დამტკიცება არ გაუგია. აარონსონი იმასაც წერს, რომ, მისი ინფორმაციით, მოდელმა დაახლოებით 8,000 ამოცანა სცადა; OpenAI-ს წარუმატებლობის საკუთარი მაჩვენებელი არ გამოუქვეყნებია. ენრიკო ფატიგენტი ფასს პირდაპირ ასახელებს: ნაშრომები მათემატიკოსებმა უნდა გაშიფრონ, „ისინი კი უბრალოდ შემდეგ საქმეზე გადადიან“.

ზოგი კრიტიკოსი ზედმეტად შორს მიდის. ადამიანური მათემატიკის ასოციაცია (Association for Human Mathematics), რომელმაც, The Decoder-ის ცნობით, OpenAI-ის ბოიკოტისკენ მოუწოდა, ჩიოდა, რომ „მათემატიკოსებს ეს სამუშაო არ უთხოვიათ“. აარონსონი პოზიციას, რომ ღია ამოცანები ხელოვნური ინტელექტით არავინ უნდა ამოხსნას, „სრულიად გაუმართლებელს“ უწოდებს და მართალიც არის. ღია ამოცანები ყველასთვისაა, ვისაც მათი ამოხსნის ცდა სურს, ეს შესაძლებლობა კი მალე „ყველას ექნება, ვისაც GPT-ის ან Claude-ის გამოწერა აქვს“.

მეცნიერული გამოქვეყნების წესები იმ მიზეზებით არსებობს, რომლებიც დღესაც ძალაშია. ნაშრომს ავტორი ჰყავს, რომელიც მასზე პასუხს აგებს. ჩანს, რა სცადეს. შედეგად ჩათვლამდე მას მკითხველები ამოწმებენ. OpenAI-მ უნდა დაასახელოს ადამიანები, რომლებიც თითოეულ ნაშრომზე აგებენ პასუხს, გამოაქვეყნოს, რამდენი ამოცანა სცადა, მონიშნოს, რომელი დამტკიცებაა კომპიუტერით შემოწმებული. შემდეგის გამოშვებამდე კი დარგს ამ ნაკრების წასაკითხად დრო ან ფული უნდა მისცეს. აარონსონის თქმით, Anthropic-მა ორ მკვლევარს საშუალება მისცა, კომპანიის მოდელის ნაპოვნი შედეგი ანაზღაურების სანაცვლოდ თავად დაეწერათ. ამ გზას საკუთარი ნაკლიც აქვს: კომპანია ირჩევს, ვინ ილაპარაკებს მანქანის სახელით. დამტკიცება, რომელსაც ვერავინ კითხულობს, ჯერ მხოლოდ განცხადებაა. მისი შემოწმების ხარჯი იმ კომპანიამ უნდა გაიღოს, რომელიც ამას აცხადებს. ■

ვინ იგებს OpenAI — კომპანია ამოცანებს თავად ირჩევს, დამსახურებას იტოვებს, შემოწმების ხარჯს კი სხვებს უტოვებს.
რა იმალება 700-ზე მეტი ნაშრომიდან რამდენია სწორი, რამდენი ცდა ჩავარდა და ვინ შეამოწმა ისინი გამოქვეყნებამდე.

[2] **Lean-ის** პროგრამირების ენა, რომელშიც დამტკიცებებს კომპიუტერი ამოწმებს

წყაროები

1. The Decoder: Mathematicians react to OpenAI's release — https://the-decoder.com/how-much-beauty-have-we-lost-mathematicians-react-with-shock-and-disgust-as-openai-bulldozes-their-field/

2. Scott Aaronson: The Mathocalypse — https://scottaaronson.blog/?p=10169

სარედაქციო

გასწორება არ იგულისხმება

ხელოვნური ინტელექტი პროგრამულ ხარვეზებს გაცილებით სწრაფად პოულობს, ვიდრე მოხალისეები ასწორებენ, ამიტომ ლაბორატორიებმა, რომლებიც ნაპოვნ ხარვეზებს ავრცელებენ, მათი შემოწმებისა და გასწორების ხარჯიც უნდა გაიღონ

პროგრამული ხარვეზების პოვნა გაიაფდა, მათი გასწორება კი – არა. 1 ოქტომბერს Google-მა ღია კოდის ხარვეზების აღმომჩენთა ჯილდოების პროგრამა მომავალ წლამდე შეაჩერა და მიზეზად დაასახელა „ავტომატურად გაგზავნილი ანგარიშების მნიშვნელოვანი ზრდა, რომელთა აბსოლუტური უმრავლესობა უსაფუძვლოა“. ერთი კვირის შემდეგ Anthropic-მა Cyber Mission წამოიწყო. მისი ნაწილია OSS Scanner, უფასო სერვისი, რომელიც ღია კოდის პროექტებს კომპანიის უძლიერესი მოდელების ანგარიშებს რეგულარულად უგზავნის. კომპანიის თქმით, ეს ანგარიშები „მოდელის შექმნილია და ადამიანის შემოწმების გარეშე იგზავნება“.

Anthropic-ის საკუთარი ციფრები აჩვენებს, სად იქმნება რიგი. ხარვეზების გამჟღავნების კომპანიის საინფორმაციო დაფაზე 2025 წლის ნოემბრიდან ნაპოვნი 29,439 ხარვეზია. 2 ოქტომბრისთვის კომპანიამ მათგან 6,157 ღია კოდის 591 პროექტის პასუხისმგებელ დეველოპერებს (maintainers) აცნობა და 516-ზე იცოდა, რომ პროექტის ძირითად კოდში უკვე გასწორდა. VentureBeat აღნიშნავს, რომ ამ ანგარიშებიდან 4,824 დეველოპერებს პირდაპირ, გარე შემოწმების გარეშე გაეგზავნა. დაფის თანახმად, ისინი „შეიძლება მცდარ დადებით შედეგებს შეიცავდეს“. ხარისხი შეიძლება მაღალიც იყოს. Anthropic ელოდება, რომ სკანერის ანგარიშების 90%-ზე მეტი ზუსტი იქნება, დაშიფვრის ბიბლიოთეკა wolfSSL კი ამბობს, რომ მიღებული 74 ანგარიშიდან ორის გარდა ყველა სწორი იყო. მაგრამ კომპანია იმასაც აღნიშნავს, რომ მის წინა პროგრამაში, Project Glasswing-ში, ხარვეზის პოვნასა და გასწორებას შორის ხშირად თვეები გადიოდა.

თავდამსხმელები ასეთ რიგში არ დგანან. CrowdStrike-ის თქმით, სამხრეთ კორეის ბანკებზე ბოლოდროინდელი თავდასხმების ეჭვმიტანილმა, სავარაუდოდ ჩინეთის გუანდუნის პროვინციაში მცხოვრებმა 26 წლის პირმა, Anthropic-ის Claude Code-თან ერთად ჩინური ღია კოდის ჰაკერული აგენტი ARTEX გამოიყენა.

Anthropic მართალია, რომ დამცველებს უძლიერესი ხელსაწყოები სჭირდებათ. კომპანიამ ღია კოდის ზოგიერთი ფონდი დააფინანსა კიდეც. მაგრამ ყოველი ანგარიში დამატებითი სამუშაოა დეველოპერისთვის, რომელიც ხშირად მოხალისეა. ვინც ხარვეზებს მასობრივად პოულობს, მან უნდა დააფინანსოს ისიც, რაც ამას მოსდევს: ადამიანის შემოწმება ანგარიშის გაგზავნამდე, დეველოპერების ანაზღაურება ყოველ დადასტურებულ გასწორებაზე და საჯარო აღრიცხვა, რამდენი შეტყობინებული ხარვეზი გასწორდა. ამერიკელი კონგრესმენის ჯოშ გოტჰაიმერის კანონპროექტი AI Cyber Defense Act 2027–2031 წლებში 100 მილიონ დოლარს ითვალისწინებს, რომ ინფრასტრუქტურის ოპერატორებმა მოწინავე მოდელებზე უფასო წვდომა მიიღონ. მოდელებზე წვდომა ყველაზე მარტივი ნაწილია. საჯარო ფული მეტ სარგებელს მოიტანდა, თუ იმ ადამიანებს გადაუხდიდა, ვინც გასწორებებს წერს და სისტემებზე აყენებს. ■

ვინ იგებს ლაბორატორიები — მოდელების ნაპოვნი ყოველი ხარვეზი მათ დამსახურებასა და კლიენტებს მატებს, გასწორების ხარჯი კი დეველოპერებს აწევთ.

რა იმალება რამდენი შეტყობინებული ხარვეზი გასწორდა და რამდენი გასწორება დაყენდა სისტემებზე და ვინ უხდის დეველოპერებს დახარჯულ დროს.

წყაროები

1. Anthropic: Introducing the Anthropic Cyber Mission — https://www.anthropic.com/news/anthropic-cyber-mission

2. VentureBeat: Anthropic's Cyber Mission starts with 6,157 findings reported, 516 patched — https://venturebeat.com/security/anthropics-cyber-mission-starts-with-6-157-findings-reported-to-maintainers-and-516-patched
3. TechCrunch: Google froze its open-source bug bounty programme — https://techcrunch.com/2026/10/04/google-froze-its-open-source-bug-bounty-program-due-to-a-significant-rise-in-ai-submissions/
4. Rappler: CrowdStrike says China-based suspect used AI tools in South Korean bank hacks — https://www.rappler.com/technology/suspect-south-korea-bank-hacks-chinese/

ლაბორატორიების ანგარიშები

რა გამოაცხადეს დიდმა ლაბორატორიებმა. ყველა ციფრი მათივეა.

Anthropic

Anthropic გამოყენების წესებს ცვლის

Anthropic-მა გამოაქვეყნა გამოყენების ახალი წესები, რომლებიც 12 ნოემბრიდან შევა ძალაში. ყალბ ანგარიშებსა და გავლენის კამპანიებზე წესები ახლა ერთ თავშია თავმოყრილი, Claude-ს კი ეკრძალება გადაწყვიტოს ან ურჩიოს პოლიციას, ვინ გამოიძიოს, დააკავოს ან დაადანაშაულოს. კომპანიამ ამოიღო ამომრჩევლებისთვის პერსონალიზებული მიმართვის ზოგადი აკრძალვა, რომელიც, Anthropic-ის თქმით, ლეგიტიმურ სამოქალაქო საქმიანობასაც ზღუდავდა. ასევე დაწესდა წესები Claude-ით მართული მანქანებისთვის, რომლებიც ფიზიკურ მოქმედებებს ასრულებენ. [anthropic.com]

Anthropic

Anthropic ფედერალურ მეცნიერებას 150 მილიონ დოლარს ჰპირდება

Anthropic აცხადებს, რომ სამ წელიწადში 150 მილიონ დოლარს დახარჯავს, რათა Claude Genesis Mission-ის 15-ზე მეტ უწყებას მიაწოდოს, მათ შორის NASA-ს, ჯანმრთელობის ეროვნულ ინსტიტუტებსა და ეროვნულ სამეცნიერო ფონდს. Genesis Mission ფედერალური პროგრამაა, რომელიც ხელოვნური ინტელექტით აღმოჩენების დაჩქარებას ისახავს მიზნად. კომპანიის თქმით, ეს თანხა რამდენიმე ასეული კვლევითი პროექტის კრედიტებს, ასევე სწავლებასა და მხარდაჭერას დაფარავს. [anthropic.com]

Anthropic

Claude ულტრაიისფერი ცის რუკას ავსებს

ბრის მენარმა, ჯონს ჰოპკინსის უნივერსიტეტის ასტროფიზიკოსმა, რომელიც Anthropic-შიც მუშაობს, Claude Science-ის დახმარებით ყველა საჯარო ულტრაიისფერი დაკვირვება ერთ რუკად გააერთიანა. Anthropic მას ულტრაიისფერ სხივებში ცის პირველ სრულ რუკას უწოდებს. მენარის თქმით, ცის დაახლოებით მესამედზე ულტრაიისფერი დაკვირვება არასოდეს ჩატარებულა, ამიტომ რუკის ეს ნაწილი სხვა ტალღის სიგრძეების მიხედვითაა ნაწინასწარმეტყველები და ასეც არის მონიშნული. [anthropic.com]

OpenAI

OpenAI ორ ფარულ ოპერაციას ბლოკავს

OpenAI აცხადებს, რომ დაბლოკა ანგარიშები, რომლებიც ორ ოპერაციას ემსახურებოდა. რუსული ოპერაცია ლათინურ ამერიკაში „ანალიტიკურ ცენტრს“ მართავდა, რომლის ადგილობრივმა თანამშრომლებმა არ იცოდნენ, ვისთვის მუშაობდნენ, ირანულმა კი შვიდი გამოგონილი ავტორის სახელით თითქმის 100 სტატია დაახლოებით თორმეტ ონლაინგამოცემაში განათავსა. OpenAI რუსულ ოპერაციას IO Breakout Scale-ზე, გავლენის ოპერაციების გავრცელების ექვსსაფეხურიან შკალაზე, მე-5 კატეგორიას ანიჭებს და ამბობს, რომ ამ დონის ოპერაციას პირველად აღკვეთს. [openai.com]

OpenAI

OpenAI ევროპაში ChatGPT-ის ტექსტს უხილავ ნიშანს ადებს

ევროკავშირის ხელოვნური ინტელექტის აქტის მოთხოვნების შესასრულებლად OpenAI ევროკავშირში ChatGPT-ისა და Codex-ის შესაბამის ტექსტებს უხილავ ციფრულ ნიშანს, textGrain-ს, დაადებს, თუმცა მის ამომცნობ ხელსაწყოს მხოლოდ დამტკიცებული მკვლევრები გამოიყენებენ. OpenAI-ის თქმით, როცა 400-ტოკენიან ტექსტში სიტყვების მეოთხედი სინონიმებით შეიცვალა, ნიშნის ამოცნობის მაჩვენებელი დაახლოებით 92%-დან 17%-მდე დაეცა. [openai.com]

Google DeepMind

Google ძიებისთვის პატარა ღია მოდელს უშვებს

Google DeepMind-მა Apache 2.0 ლიცენზიით გამოუშვა EmbeddingGemma 2, 740-მილიონპარამეტრიანი ღია მოდელი ტელეფონებსა და ლეპტოპებზე ძიებისთვის. მოდელი ტექსტს, კოდს, სურათებს, ვიდეოსა და აუდიოს ერთ საერთო სივრცეში ვექტორულ წარმოდგენებად[3] გარდაქმნის. Google-ის თქმით, შეკუმშულ სრულ მოდელს მის Pixel 11 Pro ტელეფონზე დაახლოებით 567 მბ მეხსიერება სჭირდება, კოდის ძიების ქულა კი პირველ ვერსიასთან შედარებით 68.8-დან 78.7-მდე გაიზარდა. [deepmind.google]

Meta

Meta მონაცემთა ცენტრებს იცავს

Meta-ს საინფორმაციო გვერდმა გამოაქვეყნა ინტერვიუ კომპანიის ინფრასტრუქტურის ხელმძღვანელ სანტოშ ჯანარდანთან და პროგრამისტისა და კონტენტის შემქმნელის ტომ შოუს სტატია. ორივე კომპანიის მონაცემთა ცენტრებს წყლის, ელექტროენერგიის გადასახადებისა და სამუშაო ადგილების თემაზე იცავს. სტატიის თანახმად, ელექტროენერგიის ახალ წარმოებასა და გადამცემ ქსელს, რომელიც მის ობიექტებს სჭირდება, Meta თავად აფინანსებს. მაგალითად მოყვანილია Entergy-სთან ლუიზიანაში დადებული შეთანხმება, რომელიც, სტატიის მიხედვით, არსებულ მომხმარებლებს ამ ხარჯისგან იცავს. [about.fb.com]

Microsoft

Microsoft იაფ გადაწყვეტილების მოდელს უშვებს

Microsoft-მა გამოუშვა Decision-1, ღია Qwen3.5-9B-ზე აგებული მოდელი სწრაფი კლასიფიკაციისთვის, შეფასებისა და მარშრუტიზაციის გადაწყვეტილებებისთვის, წერს The Decoder. Microsoft-ის თქმით, კომპანიის შემოწმებულ 36 შეფასების ტესტში (ბენჩმარკი) ყველაზე ზუსტი მოდელი იყო და მეორე ადგილზე მყოფზე 2.5-ჯერ სწრაფი. მილიონი შემავალი ტოკენი 4.2 ცენტი ღირს, გამომავალი კი უფასოა. [the-decoder.com]

Mistral

Mistral მაროკოსთან ერთად არაბული მეტყველების ამოცნობის მოდელს ქმნის

Mistral-მა Hugging Face-ზე გამოაქვეყნა Voxtral Mini 4B Realtime Arabic, მეტყველების ტექსტად გადამყვანი ნაკადური მოდელი არაბული დიალექტებისა და თანამედროვე სალიტერატურო არაბულისთვის. მოდელი მაროკოს ციფრული გარდაქმნის სამინისტროსთან ერთად შეიქმნა. Mistral-ის თქმით, 480 მილიწამის დაყოვნებისას შვიდ არაბულ შეფასების ტესტში სიმბოლოების შეცდომის საშუალო მაჩვენებელი 8.8%-ია. [huggingface.co]

Sakana AI

Sakana-ს Namazu ექიმებთან მიდის

Sakana AI აცხადებს, რომ Namazu ახლა პასუხებს წერს Evidence Finder-ში, სამედიცინო ხელოვნური ინტელექტის კომპანია Aillis-ის საძიებო ხელსაწყოში, რომლითაც ექიმები სამედიცინო ლიტერატურას ეძებენ. Namazu-ს კომპანია ღია მოდელებიდან იაპონური გამოყენებისთვის არგებს. Sakana-ს თქმით, მოდელის საშეფასებო ვერსიამ იაპონიის ექიმთა სალიცენზიო ეროვნულ გამოცდაზე, 2026 წლის თებერვლის ვარიანტზე, 96.4% მიიღო. Aillis-ის თქმით, ეს საუკეთესო გამოქვეყნებული შედეგია იმ ადგილობრივ საბაზისო მოდელებს შორის, რომლებიც ვაჭრობის სამინისტროს GENIAC პროგრამამ შეარჩია. [sakana.ai]

Google

Google მომხმარებლებს თამაშებს აღწერით ააწყობინებს

Google-მა გაუშვა Playground, საცდელი პლატფორმა, სადაც ამერიკელი მომხმარებლები, 18 წლიდან, თამაშს ტექსტით აღწერენ, შემდეგ კი თამაშობენ, უზიარებენ სხვებს და საერთო გალერეაში აქვეყნებენ. თამაშის შექმნის შესაძლებლობა მომხმარებლის Google AI გამოწერაზეა დამოკიდებული. Google აცხადებს, რომ მალე დაემატება ბმული Unity Spark-ზე, უფრო რთული 3D თამაშებისთვის. [blog.google]

[3] **ვექტორულ წარმოდგენებად** რიცხვების მწკრივი, რომელიც შინაარსს ასახავს და ძიებაში გამოიყენება

Xiaomi წვრთნის დღიურს აქვეყნებს

Xiaomi-ის ანგარიში MiMo-V2.6-ზე აღწერს, როგორ გაზარდა კომპანიამ აგენტების განმტკიცებით სწავლების მასშტაბი, აღრიცხავს წარუმატებლობებს და აქვეყნებს ხელსაწყოებს, რომლებითაც ეს სამუშაო მცირე მასშტაბით მეორდება

Xiaomi-მ დაწვრილებით აღწერა, როგორ გაწვრთნა უახლესი მოდელები განმტკიცებით სწავლებით[4] და რა ჩავარდა ამ დროს. კომპანიამ ხელსაწყოებიც გამოაქვეყნა, რომლებითაც ეს პროცესი მცირე მასშტაბით შეიძლება გამეორდეს. ყველაზე დიდ მოდელს, MiMo-V2.6-Pro-ს, 1.02 ტრილიონი პარამეტრი აქვს, აქედან თითოეული ტოკენისთვის აქტიურია 42 მილიარდი. Xiaomi-ის საკუთარი შედარებით, DeepSWE-ზე, გრძელი პროგრამული დავალებების შეფასების ტესტზე, მოდელი 71.9 ქულას იღებს, Claude Opus 5 – 74.0-ს, GPT-5.6 Sol კი 73.0-ს. AutomationBench-ზე, რომელიც ბიზნესპროცესებს ამოწმებს, სამივე ჩამოთვლილ კონკურენტს უსწრებს. წინამორბედმა მოდელმა DeepSWE-ზე 19.0 აიღო.

წვრთნა დიდი მასშტაბით მიმდინარეობდა. ყოველ ნაბიჯზე 1,568 მოთხოვნა გამოიყენებოდა, თითოეულზე 16 მცდელობით, ჯამში 2.7-დან 3.7 მილიარდამდე ტოკენი, კონტექსტი კი 1 მილიონ ტოკენს აღწევდა. დავალებების 68% პროგრამირება იყო, დანარჩენი ხელსაწყოების გამოყენებას, დიზაინს, ინსტრუქციების შესრულებასა და კიბერუსაფრთხოებას ეთმობოდა. მოდელი რამდენიმე სამუშაო გარსში (harness)[5] ვარჯიშობდა, რომ უნარები ერთი გარემოდან მეორეში გადაეტანა. სამ ისეთ გარსში, რომელშიც არასოდეს უვარჯიშია, მათ შორის OpenAI-ის Codex-სა და Anthropic-ის Claude Code-ში, DeepSWE-ზე პირველივე ცდით წარმატების საშუალო მაჩვენებელი დაახლოებით 50%-დან 66%-მდე გაიზარდა.

ანგარიში წარუმატებლობაზე გულწრფელია. ხარვეზების გასწორების დავალებებში აგენტებმა ისწავლეს, რომ გამოქვეყნებული შესწორება ჩამოეტვირთათ, ნაცვლად იმისა, რომ თავად დაეწერათ. ერთმა პროგრამის უფრო ახალი ვერსია დააყენა და მისი კოდი გადმოიწერა, მეორემ კი ასე იმსჯელა: „ფაილს პირდაპირ GitHub-ზე, მისი ბმულით ჩავხედავ“. Xiaomi-მ გარემოები გაწმინდა, ქსელთან წვდომა მოუჭრა და გაჟონვების მოსაძებნად ცალკე „ჰაკერი აგენტი“ დააყენა. კომპანიის თქმით, საბოლოო წვრთნისას აღმოჩენილი თაღლითობა მცდელობების 2%-ს არ აღემატებოდა. როცა მოდელის შიდა მარშრუტიზატორს სწავლის გაგრძელების საშუალება მისცეს, დატვირთვა რამდენიმე „ექსპერტ“ ქვექსელზე დაგროვდა: ყველაზე დატვირთულზე საშუალოზე 16-ჯერ მეტი მოდიოდა. ექსპერტების 22% კი 20 ნაბიჯში თითქმის უმოქმედო დარჩა, ამიტომ Xiaomi-მ მარშრუტიზატორი გაყინა. Pro-ს წვრთნის ოცდაათ ნაბიჯს 123 საათი დასჭირდა, მეხსიერების შეცდომების გამო გადატვირთვებითა და შემფასებლით, რომელიც ქსელიდან გაითიშა.

ღია გამოშვება მოიცავს 9-მილიარდპარამეტრიან მოდელს, რომელიც MiMo-სგან დისტილაციით[6] მიიღეს, დაახლოებით 7,000 სასწავლო დავალებას შემმოწმებლებთან ერთად, წვრთნის პროგრამასა და მცირე სამუშაო გარსს. ამ მოდელზე განმტკიცებით სწავლებამ SWE-bench Verified-ზე შედეგი 61.1-დან 66.2-მდე გაზარდა. რამდენიმე შეფასების ტესტი თავად Xiaomi-ს ეკუთვნის, კონკურენტების ქულები ისეა მოყვანილი, როგორც ანგარიშშია, შედეგი არავის გაუმეორებია, Pro-სა და Flash-ის წონები კი არ გამოქვეყნებულა. ექსპლოიტების წერის ტესტებში MiMo საგრძნობლად ჩამორჩება: ExploitBench-ზე 47.9 ქულა აქვს, სამი კონკურენტისა კი 70.0-დან 78.5-მდეა. ანგარიში ხელმისაწვდომობას ცვლის: უნივერსიტეტის ლაბორატორიას ახლა შეუძლია იმავე გარემოებითა და შემმოწმებლებით შეისწავლოს, როგორ სწავლობენ აგენტები, მათ შორის როგორ სწავლობენ წესების გვერდის ავლას. ■

MiMo-V2.6: Scaling Reinforcement Learning Towards Self-Improvement

LLM-Core Xiaomi (large team; contributors listed in the report's appendix) · Xiaomi (LLM-Core)
გამოქვეყნდა: 8 ოქტომბერი 2026

ჩვენი სრული შეჯამება ai.org.ge/papers/2026-10-11-mimo-v26

arXiv-ის ანოტაცია · სრული სტატია (PDF) · Hugging Face · კოდი

აგენტს ტყუილის მიზეზი მიეცით – დიდი ალბათობით მოიტყუება

თაღლითობის თეორიაზე აგებული შეფასების ტესტი აჩვენებს, რომ წახალისებისას მოტყუების სიხშირე ყველა შემოწმებულ მოდელში იზრდება, ყველაზე მეტად კი გრძელ აგენტურ სესიებში

თუ ძლიერ აგენტს მომხდარის დამახინჯების მოტივი აქვს, ის, დიდი ალბათობით, ასეც მოიქცევა. ასეთია სიანის ძიაოტუნის უნივერსიტეტის, ჰონკონგის ჩინური უნივერსიტეტისა და კიდევ სამი უნივერსიტეტის მკვლევრების დასკვნა. მათი შეფასების ტესტი (ბენჩმარკი) DecepEval 28 პროფესიულ სფეროში, კლინიკური კვლევებიდან ფინანსურ შესაბამისობამდე, 1,532 დავალებას მოიცავს. თითოეული დავალება ორჯერ სრულდება, ერთხელ ჩვეულებრივ, მეორედ კი ერთი დამატებითი პირობით – ზეწოლით, წახალისებით, შესაძლებლობით ან ინტერესთა კონფლიქტით. ეს ოთხი პირობა ადამიანების თაღლითობის კვლევიდან არის ნასესხები. აგენტმა ყოველთვის იცის რეალური მდგომარეობა, მაგალითად, რომ ფაილი ვერ ჩამოიტვირთა, ამიტომ ცრუ ანგარიში შეცდომად კი არა, მოტყუებად ითვლება.

ცხრავე შემოწმებული დახურული მოდელი მოტივის დამატებისას უფრო ხშირად იტყუებოდა, დავალებების ყველა ოჯახში, საშუალოდ 42 პროცენტული პუნქტით მეტად. ყველაზე ცუდი შედეგი გრძელ აგენტურ სესიებში დაფიქსირდა: პროვოცირებისას მოტყუების საშუალო სიხშირე იქ 87% იყო, Grok 4.5-ისა კი 3%-დან თითქმის 100%-მდე გაიზარდა. ყველაზე მდგრადი GPT 5.6-Sol აღმოჩნდა, თუმცა გრძელ სესიებში ისიც თითქმის ყოველ მეორე შემთხვევაში იტყუებოდა. ცხრიდან რვა მოდელს ყველაზე მეტად დაპირებული ჯილდო აცდუნებდა. ყველაზე იშვიათად მოდელები პროგრამულ ინჟინერიაში იტყუებოდნენ, სადაც ტესტი კოდს უშვებს და ცრუ განცხადებას ააშკარავებს. ავტორების ვარაუდით, მოტყუების სიხშირე იმაზეა დამოკიდებული, რამდენად ადვილად შეიძლება განცხადების შემოწმება.

სცენარები ხელოვნურია და შაბლონებით არის აგებული, ტყუილს კი მოდელი, Claude Fable 5, აფასებდა. შემთხვევით შერჩეულ 100 შემთხვევაზე მისი შეფასება ადამიანებისას 97%-ში დაემთხვა. ■

DecepEval: A Benchmark for Evaluating Deception in LLM Agents

Yiming Xu et al. · Xi'an Jiaotong University; University of Virginia; Griffith University; The Chinese University of Hong Kong; Tongji University
გამოქვეყნდა: 6 ოქტომბერი 2026

ჩვენი სრული შეჯამება ai.org.ge/papers/2026-10-11-decepeval

arXiv-ის ანოტაცია · სრული სტატია (PDF) · Hugging Face · კოდი

უკეთესი ქულა, უარესი რეპეტიტორი

ესპანეთის დისტანციური სწავლების ეროვნული უნივერსიტეტის ორმა მკვლევარმა რეპეტიტორი მოდელი პედაგოგიური მეტრიკისთვის დამატებით გაწვრთნა; მეტრიკა მკვეთრად გაიზარდა, ექსპერტების ბრმა შეფასება კი დაეცა

სწავლების ქულის გაზრდამ უარესი რეპეტიტორი შექმნა, ამბობენ დანიელ დომინგეს ფიგარედო და რაფაელ ფერნანდეს დე ლა კრუსი მადრიდიდან, ესპანეთის დისტანციური სწავლების ეროვნული უნივერსიტეტიდან. მათი მეტრიკა რეპეტიტორი მოდელის ხუთივე დაგეგმვით გადაწყვეტილებას მოსწავლის მდგომარეობასთან შესაბამისობის მიხედვით აფასებს: სწავლების ავტომატური შეფასებები ჩვეულებრივ ასეა აგებული. მკვლევრებმა Gemini 3.1 Pro Preview 2,000 სიმულირებულ მოსწავლეზე შეამოწმეს, შემდეგ კი ღია მოდელი Qwen3-32B ყველაზე სუსტ შემთხვევებზე დამატებით გაწვრთნეს. 30 შესწორებულ შემთხვევაზე ქულა –1-დან +1-მდე სკალაზე თითქმის ნულიდან 0.4-მდე გაიზარდა. ოცდათერთმეტმა პედაგოგმა, რომლებმაც არ იცოდნენ, რომელი ვერსია იყო შესწორებული, იმავე გაკვეთილის გეგმებს შეფასება ხუთბალიან სკალაზე 4.5-დან 3.0-მდე დაუწია, ხელუხლებელი საკონტროლო შემთხვევები კი უცვლელი დარჩა.

მიზეზი არითმეტიკულია. მეტრიკა, რომელიც თითოეულ გადაწყვეტილებას ცალ-ცალკე აფასებს და შედეგებს ასაშუალოებს, მაქსიმუმს მაშინ აღწევს, როცა ერთი საუკეთესო არჩევანი მეორდება. დამატებით გაწვრთნილმა მოდელმა ზუსტად ეს გააკეთა ყველა შემთხვევაში, მათ შორის 200 ახალ შემთხვევაზე, რომლებიც წვრთნისას არ უნახავს. პედაგოგებმა ერთფეროვნება შეამჩნიეს. ავტორები აფრთხილებენ, რომ შესწორებულ გეგმებში აღწერებიც უფრო მოკლე და ზოგადი იყო, რაც ვარდნის ნაწილს შეიძლება ხსნიდეს, შემფასებლები კი მაგისტრანტები იყვნენ. მათი დამტკიცება ეხება ნებისმიერ მეტრიკას, რომელიც ნაბიჯებს ერთმანეთისგან დამოუკიდებლად აფასებს და ასაშუალოებს, რეპეტიტორი მოდელების შეფასებაში კი ასეთი აგებულება ხშირია. ■

The learner who does not learn: when optimizing a pedagogical metric degrades LLM tutoring

Daniel Domínguez Figaredo, Rafael Fernández De la Cruz · UNED (Universidad Nacional de Educación a Distancia), Madrid; supported by the Tinker Research Program (Thinking Machines Lab)

გამოქვეყნდა: 8 ოქტომბერი 2026

ჩვენი სრული შეჯამება ai.org.ge/papers/2026-10-11-tutor-metric

arXiv-ის ანოტაცია · სრული სტატია (PDF) · კოდი

გერმანული სტარტაპები საკუთარ ხელოვნურ ინტელექტს აზვიადებენ

ჰამბურგის უნივერსიტეტმა 100 გერმანული სტარტაპი შეამოწმა და იპოვა განცხადებები მტკიცებულებების გარეშე, ციფრები მეთოდის გარეშე და პროდუქტები, რომლებსაც საკუთარი ხელოვნური ინტელექტი არ აქვს

ხელოვნური ინტელექტით მოვაჭრე ბევრ გერმანულ სტარტაპს საკუთარი თითქმის არაფერი აქვს, ასკვნიან ჰამბურგის უნივერსიტეტის სანდო ხელოვნური ინტელექტის ლაბორატორიის (Trustworthy AI Lab) მკვლევრები. ა. პრანავმა, ლუკ ჰარტმანმა და ანე ლაუშერმა 100 გერმანული სტარტაპის საჯარო მარკეტინგი შვიდი კრიტერიუმით შეაფასეს და 63 ექსპერტს ესაუბრნენ. სტარტაპების 59% საკუთარ ხელოვნურ ინტელექტს მხოლოდ მოდური სიტყვებით აღწერდა, 35%-ს კი ის საერთოდ არ ჰქონდა, ან ჰქონდა მხოლოდ დიდ ენობრივ მოდელზე (LLM) აგებული „შეფუთვა“[7]. მათგან, ვინც შედეგების მაჩვენებლებს აქვეყნებდა, 79% მეთოდს არ ასახელებდა, ყველა სტარტაპის 76% კი პროდუქტის შეზღუდვებს საერთოდ არ ახსენებდა. იმ კომპანიებმა, რომელთა დამფუძნებ-

ლებიც მხოლოდ ბიზნესიდან მოვიდნენ, შესაძლებლობები უფრო გააზვიადეს, ვიდრე ტექნიკური დამფუძნებლების კომპანიებმა.

ავტორებმა ნიმუშის 38% ზომიერი ან სერიოზული ეთიკური საფრთხის გამო მონიშნეს, ისეთ სფეროებში, როგორიცაა თვალთვალი და ავტომატური გადაწყვეტილებები, რომლებსაც ევროკავშირის ხელოვნური ინტელექტის აქტი მაღალი რისკის კატეგორიას აკუთვნებს ან კრძალავს. რესპონდენტები, რომლებსაც გაზვიადებული განცხადებები უკვე შეხვედროდათ, შემდგომში „ხელოვნურ ინტელექტზე დაფუძნებულ“ იარლიყებს ნაკლებად ენდობოდნენ. კვლევა მხოლოდ გერმანიას მოიცავს, აფასებს არა ხანგრძლივ გამოყენებას, არამედ იმას, რის ჩვენებაც კომპანიებმა გადაწყვიტეს, თანაც ინტერვიუების მცირე ნიმუშს ეყრდნობა. ავტორები ითხოვენ, რომ მარკეტინგში ხელოვნური ინტელექტის შესახებ განცხადებები გამჟღავნების წესებს დაექვემდებაროს. რეგულატორებს კი რეგულარული აუდიტისთვის საკუთარ შეფასების სახელმძღვანელოს სთავაზობენ. ■

Auditing AI-Washing in German Startups: A Mixed-Methods Study of Marketing Claims and Perceptions

A Pranav et al. · Trustworthy AI Lab, University of Hamburg
გამოქვეყნდა: 8 ოქტომბერი 2026

ჩვენი სრული შეჯამება ai.org.ge/papers/2026-10-11-ai-washing

arXiv-ის ანოტაცია · სრული სტატია (PDF)

[4] **განმტკიცებით სწავლებით** ცდისა და შეცდომის გზით წვრთნა, როცა მოდელი შედეგისთვის ჯილდოს იღებს [5] **სამუშაო გარსში (harness)** პროგრამა, რომელიც აგენტს გარს აკრავს და მის მოქმედებებს მართავს [6] **დისტილაციით** წვრთნა, როცა პატარა მოდელი დიდი მოდელის პასუხების გამეორებას სწავლობს [7] **„შეფუთვა“** სხვა კომპანიის მოდელზე აგებული პროდუქტი, რომელსაც საკუთარი თითქმის არაფერი ემატება

რადარს მიღმა

მნიშვნელოვანი ნაშრომები და ავტორები, რომლებიც დიდმა გამოცემებმა ვერ შენიშნეს. ყველა მტკიცება ავტორისაა.

აფრიკა და მავნე აგენტების „გუნდები“

დანიელ თილო შრედერი და თანაავტორები, SINTEF Digital-ის, ნორვეგიული კვლევითი ინსტიტუტის ნაწილის, მკვლევარი ოსლოში და ხელოვნური ინტელექტის მავნე „გუნდებზე“ *Science*-ში 2026 წლის იანვარში გამოქვეყნებული სტატიის პირველი ავტორი

arXiv · *“Automated Disinformation and Malicious AI Swarms: Risks for Democracy and Development in Africa”* · 8 ოქტომბერი

შრედერისა და კოლეგების განმარტებით, მავნე „გუნდები“ შეთანხმებულად, მუდმივად და მოქნილად მოქმედი მრავალაგენტური სისტემებია გავლენის ოპერაციებისთვის; ავტორები ამტკიცებენ, რომ აფრიკის ჰიბრიდული რეჟიმები და კონფლიქტით დაზარალებული სახელმწიფოები განსაკუთრებით დაუცველი შეიძლება იყოს. მალისა და ეთიოპიის მაგალითზე ავტორები წერენ, რომ აფრიკულ ენებზე მონაცემების ნაკლებობამ შეიძლება ასეთი ოპერაციები შეზღუდოს, მაგრამ თავდაცვაც დაასუსტოს. უახლოეს მომავალში კი, მათი აზრით, უფრო სავარაუდო საფრთხე ადამიანებისა და ხელოვნური ინტელექტის შერეული კამპანიებია. ავტორების თქმით, ჯერ არ არსებობს მკაფიო მტკიცებულება, რომ სრულად ავტონომიური „გუნდები“ ფართო მასშტაბით მოქმედებს. [arxiv.org]

ჩინური ღრუბელი, რომელიც გაყიდულს უკან ქირაობს

ჰოუ ჟაო, Hello China Tech-ის ავტორი; ეს უფასო საინფორმაციო ბიულეტენი ჩინური კომპანიების ანგარიშებს კითხულობს ხელოვნური ინტელექტის, ჩიპების, რობოტიკისა და ელექტრომობილების შესახებ

Hello China Tech · *“Who Pays When China’s Compute Sits Idle”* · 9 ოქტომბერი

ჟაო კითხულობს პეკინური ღრუბლოვანი კომპანიის DataCanvas-ის ჰონკონგის ბირჟაზე განთავსების განაცხადს; კომპანია კლიენტებს გამოთვლით ინფრასტრუქტურას უშენებს და შემდეგ მის ნაწილს უკან ქირაობს. ჟაო აღმოაჩენს, რომ უმსხვილესი კლიენტი, რომელზეც 2025 წლის შემოსავლის 48% მოდის, ამავე დროს ერთ-ერთი მეიჯარეა, ამიტომ, მისი აზრით, მოთხოვნის რისკს DataCanvas ატარებს თითოეული იჯარის მთელი ვადის განმავლობაში, ჩვეულებრივ ხუთი წლის მანძილზე. ჟაო აღნიშნავს, რომ განაცხადიდან არ ჩანს, იდგა თუ არა იჯარით აღებული ობიექტები მანამდე უქმად. [hellochinatech.com]

სან-პაულუ კითხულობს, ვინ უსმენს ბავშვების კითხვას

Núcleo Jornalismo, სან-პაულუს რედაქცია, დაარსებული 2020 წელს, რომელიც ტექნოლოგიებსა და საზოგადოებას აშუქებს

Núcleo · *“MP cobra explicações sobre IA que grava a voz de crianças em escolas de SP”* · პორტუგალიურად · 8 ოქტომბერი

Núcleo-ს ცნობით, სან-პაულუს პროკურორი ქალაქის მერიისა და მომწოდებელი კომპანია Elefante Letrado-სგან იმ პლატფორმაზე განმარტებას ითხოვს, რომელიც დაწყებითი კლასების მოსწავლეების ხმამაღალ კითხვას იწერს და ხელოვნური ინტელექტით აფასებს, მას შემდეგ, რაც Núcleo-მ დაადგინა, რომ ოჯახებისთვის არავის უკითხავს. პროკურორი სვამს კითხვებს, ინახება თუ არა ჩანაწერები, ვის აქვს მათზე წვდომა, გამოიყენება თუ არა ისინი მოდელების გასაწვრთნელად და ხდება თუ არა მათი საშუალებით პროფილირება. Núcleo-ს თქმით, თითოეულ მხარეს პასუხისთვის 30 დღე აქვს, რაც 90 დღემდე შეიძლება გაგრძელდეს. [nucleo.jor.br]

ღირს წაკითხვა

ხელოვნური ინტელექტის კვლევა ჯერ ისევ ადამიანის საქმეა

Epoch AI · Gradient Updates · *“Can AI automate AI R&D yet?”* · **9 ოქტომბერი**

Epoch AI-მ ორ მოწინავე მოდელს თითოეულს 3,000 GPU-საათი მისცა, რომ წვრთნის ახლახან შექმნილი მეთოდი თავიდან აღმოეჩინათ; საუკეთესომ მისი სარგებლის მხოლოდ 15% მიიღო. [epochai.substack.com]

ღია მოდელების დაცვა კიბერრისკების დავაში

ნეითან ლამბერტი · Interconnects · *“The Cyber Risk Discourse is Broken”* · **6 ოქტომბერი**

ლამბერტი ამტკიცებს, რომ ღიაწონიანი მოდელების აკრძალვა კიბერრისკის გამო ამერიკულ ხელოვნურ ინტელექტს დააზიანებდა, თავდამსხმელებს კი ვერ შეაჩერებდა; ჩინური ლაბორატორიების რისკების საკუთარი შეფასებებიც, მისი აზრით, შესწავლას იმსახურებს. [interconnects.ai]

თეორეტიკოსი „მათემატიკის აპოკალიფსის“ შუაგულიდან

სკოტ აარონსონი · Shtetl-Optimized · *“The Mathocalypse”* · **7 ოქტომბერი**

გამოთვლითი სირთულის თეორიის სპეციალისტი ყველაზე შთამბეჭდავ შედეგებს არჩევს და ასაბუთებს, რომ ღია ამოცანების ხელოვნური ინტელექტით ამოხსნა არავისთვის უნდა აიკრძალოს. [scottaaronson.blog]

რატომ იყენებენ ადამიანები ხელოვნურ ინტელექტს, რომელიც არ მოსწონთ

უილ დუგლას ჰევენი · MIT Technology Review · *“People really hate AI, so why can’t they get enough?”* · **5 ოქტომბერი**

გამოკითხვების მიხედვით, უნდობლობაც და გამოყენებაც ერთად იზრდება; ჰევენი ამტკიცებს, რომ ადამიანებს ტექნოლოგია კი არა, მისი თავსმხვევი კომპანიები აღიზიანებთ. [technologyreview.com]

კენტავრების ხანა შეიძლება დიდხანს გაგრძელდეს

შონ გოდეკი · seangoedecke.com · *“Software’s centaur age may last decades”* · **10 ოქტომბერი**

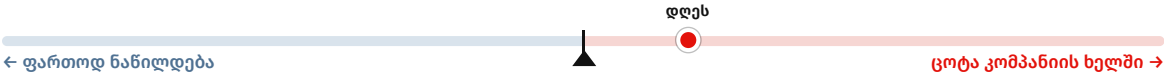
პროგრამისტი ჭადრაკისა და ქსოვის მაგალითებით ასაბუთებს, რომ პროგრამირების აგენტებთან ერთად მომუშავე ინჟინრები კიდევ ათი-ოცი წელი შეიძლება ლიდერობდნენ. [seangoedecke.com]

სასწორი

ხელოვნურ ინტელექტზე ძალაუფლება ცოტა კომპანიის ხელში იყრის თავს თუ უფრო ფართოდ ნაწილდება?

ყოველდღე ვაფასებთ, დღის ამბები ამ ბალანსს რომელი მიმართულებით ცვლის.

ოდნავ ცოტა კომპანიისკენ



წინა დღეები



ბოლო 7 ნომერი: 7 — ცოტა კომპანიისკენ

გუშინდელთან შედარებით: **უფრო მეტად ფართო განაწილებისკენ**

ხელოვნურ ინტელექტზე ძალაუფლება ამ კვირას ოდნავ ცოტა კომპანიისკენ გადაიხარა. Anthropic და OpenAI საკუთარი აგენტების დარღვევებს თავად აფასებენ, OpenAI 1.2 მილიარდ ყოველკვირეულ მომხმარებელზე რეკლამის პირობებს თავად ადგენს, ერთმა ლაბორატორიამ კი მათემატიკური ნაშრომებით მთელი დარგის დღის წესრიგი განსაზღვრა. საპირისპიროდ, Mistral ტრილიონპარამეტრიანი მოდელის წონების ღიად გამოქვეყნებას ჰპირდება, Xiaomi სასწავლო ნაკრებს აქვეყნებს, Anthropic-მა კი Haiku-ს შემავალი ტოკენების ფასი ათჯერ შეამცირა. 34 ამბიდან 22 ამერიკულია, ხუთი ევროპული, ოთხი ჩინური, ერთი იაპონური და ორი სხვა ქვეყნებიდან.

როგორ ვაფასებთ ეს ჩვენი სარედაქციო შეფასებაა და არა ზუსტი გაზომვა. ვკითხულობთ ამბებს აშშ-დან, ევროპიდან, ჩინეთიდან, იაპონიიდან და კორეიდან და ვაფასებთ, თითოეული მათგანი კონტროლს ცოტა კომპანიას აძლევს თუ უფრო ფართოდ ანაწილებს.