კვირის მიმოხილვა

კვირის გამოშვება: 2026 წლის 7–13 ნოემბრის ამბები

პოლიტიკა

ჯენსენ ჰუანგი Nvidia-ს აღმასრულებელი დირექტორის თანამდებობას ტოვებს. Nvidia-მ მისი გადადგომა 12 ნოემბერს გამოაცხადა, ოთხი დღის შემდეგ, რაც ჰუანგმა რომანი აღიარა ელის ჰუანგთან, Megaspeed-ის ერთ-ერთ ხელმძღვანელთან, რომელსაც ამერიკელი თანამდებობის პირები ჩინეთის სასარგებლოდ ჯაშუშობაში ეჭვობენ. *New York Times*-მა და Reuters-მა ისიც დაწერეს, რომ ჰუანგს რომანი Nvidia-ს ორ თანამშრომელთანაც ჰქონდა. აღმასრულებელი დირექტორის მოვალეობას დროებით ფინანსური დირექტორი კოლეტ კრესი ასრულებს. [techcrunch.com]

ჰუანგის წასვლა, ერთი გამოცემის ცნობით, რომანებზე მეტად Nvidia-ს საკრედიტო პრაქტიკას შეიძლება უკავშირდებოდეს. *Wall Street Journal* წერს, რომ Nvidia ღრუბლოვან პროვაიდერებს მილიარდობით დოლარის გარანტიებს სთავაზობდა, რათა მათ Nvidia-ს ჩიპები ეყიდათ. კრესმა კი, მაშინ ჯერ კიდევ ფინანსურმა დირექტორმა, OpenAI-სთვის განხილული 250 მილიარდი დოლარის გარანტია გაანახევრა. CNN-ის ცნობით, აქამდე შეგროვებული სადაზვერვო მასალა არ ადასტურებს, რომ ელის ჰუანგი პეკინის სასარგებლოდ ჯაშუშობდა. [the-decoder.com]

კონგრესი ჰუანგსა და ვაჭრობის მინისტრს პასუხს სთხოვს. ჰუანგი დათანხმდა, რომ მადლიერების დღემდე კონგრესის წინაშე დახურულ სხდომაზე ჩვენებას მისცემს, სენატორი ტიმ სკოტი კი მოითხოვს, რომ სენატის საბანკო კომიტეტის წინაშე ჰუანგიცა და ჰოვარდ ლატნიკიც წარდგნენ. ორი დემოკრატი, რონ უაიდენი და რიჩარდ ნილი, იძიებს, როგორ მოეკიდა ადმინისტრაცია Nvidia-ს H200 ჩიპების ჩინეთში გაყიდვას. ეს გაყიდვები პრეზიდენტმა ტრამპმა 2025 წლის დეკემბერში დაუშვა, გავრცელებული ცნობით, მას შემდეგ, რაც ჰუანგის ურთიერთობაზე მოახსენეს. [techcrunch.com]

მკვლევრის თქმით, ექვსი ჩინური მოდელი ჩინეთში ექსპორტისთვის აკრძალულ ჩიპებზე გაწვრთნეს. OpenAI-ის ყოფილი თანამშრომელი უილ დეპიუ ამბობს, რომ გამოქვეყნებულ წონებში დარჩენილი კვალი აჩვენებს, რომ DeepSeek-მა, Moonshot AI-მ, MiniMax-მა და Ant Group-მა Nvidia-ს Blackwell ჩიპები გამოიყენეს. (იხ. სარედაქციო.) [the-decoder.com]

Nvidia-მ კონგრესმენს პირად საუბარში უთხრა, რომ საკუთარი ჩიპების ადგილმდებარეობის დადგენა შეუძლია. „ჩიპების უსაფრთხოების აქტის“ თანაავტორმა ბილ ფოსტერმა *MIT Technology Review*-ს განუცხადა, რომ კომპანიის წარმომადგენლებმა კანონპროექტით გათვალისწინებული მდებარეობის შემოწმება შესაძლებლად მიიჩნიეს. (იხ. სარედაქციო.) [technologyreview.com]

***New York Times*-ის ცნობით, აშშ-მა Nvidia-ს ჩიპების მყიდველად დაამტკიცა ჩინური ფირმა, რომელიც, როგორც ჩანს, მხოლოდ ქაღალდზე არსებობს.** ვაჭრობის სამინისტრომ 17 თვის წინ დაფუძნებულ ფირმას, New Matrix-ს, H200 ჩიპებით ვაჭრობის უფლება მისცა, მაგრამ ჟურნალისტებს მის მისამართზე მხოლოდ „ცარიელი ნაცრისფერი შენობა“ დახვდათ. კომპანიის მფლობელები დაკავშირებული არიან ჩინურ კერძო საინვესტიციო ფონდთან, რომელიც აშშ-ს 2024 წლიდან შავ სიაში აქვს შეყვანილი. [theverge.com]

ChinaTalk-ის მოკვლევით, ამერიკული მონაცემთა კომპანიები ჩინურ ლაბორატორიებს დაახლოებით 500 მილიონი დოლარის საწვრთნელ მასალას ჰყიდიან. გამოცემა სულ მცირე ცხრა გამყიდველს ასახელებს, მათ შორის Mercor-ს, Surge AI-სა და Turing-ს. მათ პროდუქტებს შორისაა განმტკიცებით სწავლების[1] გარემოები, ხშირად იგივე, რაც OpenAI-სა თუ Anthropic-ისთვის შეიქმნა. საექსპორტო შეზღუდვები ასეთ მონაცემებს თითქმის არ ეხება, ამიტომ ავტორები მოითხოვენ, რომ კონტროლი პროგრამირებისა და ბიოლოგიის მასალაზეც გავრცელდეს. [the-decoder.com]

ავსტრალია მოწინავე მოდელების[2] შემქმნელ კომპანიებს უსაფრთხოების სავალდებულო წესებს დაუწესებს. ქვეყანაში მომუშავე ფირმებს მოუწევთ, ინციდენტები შეატყობინონ, მოდელები გარე ექსპერტებს შეაფასებინონ და უსაფრთხოების წესები გამოაქვეყნონ. ამას ახალი მარეგულირებელი გააკონტროლებს, რომელსაც „მნიშვნელოვანი ჯარიმების“ დაკისრება შეეძლება. მოკვლევა, რომელიც მას შემდეგ დაინიშნა, რაც OpenAI-ის აგენტმა Medicare-ის სტატისტიკის პორტალზე დახურული ფაილები წაიკითხა, მთავრობას ასევე სთხოვს, განიხილოს, ხომ არ უნდა ჩაითვალოს დანაშაულად, თუ კომპანია ზიანს ვერ აღკვეთს. [medianama.com]

ნიუ-იორკი ხელოვნური ინტელექტის უსაფრთხოების კანონით გათვალისწინებულ რეგისტრაციას იწყებს. დეველოპერები, რომელთა წლიური შემოსავალი 500 მილიონ დოლარს აღემატება, მათ შორის Anthropic და OpenAI, 2027 წლის 1 იანვრამდე უნდა დარეგისტრირდნენ, როცა RAISE აქტი ამოქმედდება. ამის შემდეგ მათ უსაფრთხოების წესების გამოქვეყნება და კრიტიკული ინციდენტების 72 საათში შეტყობინება დაევალებათ, წინააღმდეგ შემთხვევაში კი პირველ დარღვევაზე 1 მილიონ დოლარამდე, განმეორებით დარღვევაზე კი 3 მილიონ დოლარამდე ჯარიმა დაეკისრებათ. [techcrunch.com]

რვა ქვეყანამ და წამყვანმა ლაბორატორიებმა პარიზში უსაფრთხოების პირობას მოაწერეს ხელი. *Le Monde*-ის ცნობით, OpenAI, Anthropic, Microsoft და Mistral პირობას დებენ, რომ იმ მოდელებს, რომლებსაც თავად მართავენ, გათიშვის მექანიზმებს დაუყენებენ და დამოუკიდებელ აუდიტს მხარს დაუჭერენ. „პარიზის მოწოდება ხელოვნური ინტელექტის უსაფრთხოებისა და დაცვის შესახებ“ არც ვადებს აწესებს და არც სანქციებს, გერმანიას კი დოკუმენტისთვის ხელი არ მოუწერია. [the-decoder.com]

ევროკომისია სთავაზობს წესებს იმ აგენტებისთვის, რომლებიც მომხმარებლების სახელით ხელშეკრულებებს დებენ. „ციფრული სამართლიანობის აქტი“ მომხმარებლებს საშუალებას მისცემს, თავად გადაწყვიტონ, აგენტის რომელი შენაძენი უნდა დაადასტურონ. (იხ. სარედაქციო.) [medianama.com]

United Airlines აგენტებს ფრენების დაჯავშნას აუკრძალავს. ავიაკომპანია აცხადებს, რომ აგენტებმა, რომლებსაც მომხმარებლები ანგარიშის პაროლს აძლევენ, შეიძლება პირადი ინფორმაცია გაამჟღავნონ და ისეთი რამ იყიდონ, რაც არავის უნდოდა. (იხ. სარედაქციო.) [techcrunch.com]

აშშ-ის საჯარო სექტორის პროფკავშირებმა სამუშაო ადგილზე ხელოვნური ინტელექტის შეზღუდვას მიაღწიეს. *Guardian*-ზე დაყრდნობით *The Decoder* წერს, რომ ოკლენდის პროფკავშირმა ხელოვნური ინტელექტით გამოწვეული დათხოვნების ნაცვლად თანამშრომელთა გადამზადება მოიპოვა, პენსილვანიის გუბერნატორმა კი სხვა საჯარო პროფკავშირს, SEIU Local 668-ს, ChatGPT Enterprise-ის საცდელ პროექტში ადამიანის ზედამხედველობა აღუთქვა. Ernst & Young-ის გამოკითხვით, აშშ-ის სამთავრობო უწყებების თითქმის 70% უკვე იყენებს გენერაციულ ხელოვნურ ინტელექტს. [the-decoder.com]

Helsing-მა წარადგინა პროგრამა, რომელიც იარაღისთვის სამხედრო სამიზნეებს ამზადებს. გერმანული თავდაცვითი კომპანია აცხადებს, რომ Clarity-ში ასობით აგენტი სენსორების მონაცემებს, პატაკებსა და ღია წყაროების დაზვერვას სამიზნეების რეკომენდაციებად აქცევს, სამიზნის იარაღისთვის გადაცემა კი ყოველ ჯერზე სახელით დასახელებულმა პირმა უნდა დაამტკიცოს. პროგრამა კონკურენციას Palantir-ის Maven-ს უწევს. [the-decoder.com]

ფული

***Bloomberg*-ის ცნობით, Anthropic-ის აქციებზე მოთხოვნა მიწოდებას ექვსჯერ აღემატება.** კომპანია 825 მილიონ აქციას ყიდის, თითოს 70–78 დოლარად, რითაც 64 მილიარდ დოლარამდე მოიზიდავს და ფასის ზედა ზღვარზე 1.4 ტრილიონ დოლარად შეფასდება. შვიდი კომპანია, მათ შორის Google და Nvidia თითო 10 მილიარდი დოლარით, ცალკე იყიდის 26 მილიარდი დოლარის ხმის უფლების არმქონე აქციებს, საბოლოო ფასი კი 18 ნოემბერს დადგინდება. [the-decoder.com]

***Bloomberg* წერს, რომ ბანკებმა ჩინელი და ჰონკონგელი ინვესტორები Anthropic-ის აქციების განთავსებას ჩამოაშორეს.** რამდენიმე ბანკმა იქიდან შემოსული შეკვეთები უარყო, საქმეში ჩართული ზოგიერთი პირი კი მიზეზად თავდაცვითი პროდუქციის ექსპორტზე აშშ-ის წესებს ასახელებს. Anthropic-ს კომენტარი არ გაუკეთებია. კომპანია ჩინეთის კონტროლქვეშ მყოფ ფირმებს სერვისებს ისედაც არ აძლევს. [the-decoder.com]

OpenAI-ის კვარტალურმა შემოსავალმა 10 მილიარდ დოლარს გადააჭარბა, მაგრამ ზარალიც გაიზარდა. *Wall Street Journal*-ისა და *The Information*-ის ცნობით, მესამე კვარტალში გაყიდვები დაახლოებით ნახევრით გაიზარდა, კორექტირებული საოპერაციო ზარალი კი 10.9 მილიარდ დოლარამდე გაიზარდა. Anthropic-მა იმავე კვარტალში 17 მილიარდი დოლარი გამოიმუშავა, მისი საოპერაციო ზარალი კი 567 მილიონი დოლარი იყო. [the-decoder.com]

Nvidia-ს აქციები გაიაფდა და კომპანიას ინვესტორებმა პირველი სარჩელი უჩივლეს. 9 ნოემბერს, პირველ სავაჭრო დღეს მას შემდეგ, რაც დირექტორთა საბჭომ მოკვლევა დაიწყო, აქციები 4.3%-ით გაიაფდა, 11 ნოემბერს კი, ბირჟის დახურვის შემდეგ, კიდევ დაახლოებით 8%-ით. კალიფორნიაში შეტანილ სარჩელში ნათქვამია, რომ Nvidia-მ ინვესტორები შეცდომაში შეიყვანა, როცა კომპანიის წარმომადგენელმა ჰუანგის ურთიერთობას ელის ჰუანგთან „არც უჩვეულო და არც მოულოდნელი“ უწოდა. [techcrunch.com]

***Financial Times*-ის ცნობით, Anthropic 3.4 მილიარდი დოლარის აირტურბინებს ყიდულობს.** Caterpillar-ის გენერატორები თავდაპირველად მომწოდებელმა OnePWR-მა შეუკვეთა, მაგრამ დაფინანსება ვერ იშოვა. გენერატორებს 1.7 გიგავატამდე ელექტროენერგიის გამომუშავება შეუძლია. Anthropic აქამდე ინფრასტრუქტურას ძირითადად ქირაობდა, ახალი აირტურბინების რიგი კი 2030-იანი წლების დასაწყისამდე გრძელდება. [techcrunch.com]

CoreWeave მომხმარებლებს თანხის წინასწარ გადახდას სთხოვს. ფირმა, რომელიც ხელოვნური ინტელექტისთვის გამოთვლით სიმძლავრეს აქირავებს, მომდევნო სამ თვეში წინასწარი გადახდებისა და დეპოზიტების სახით დაახლოებით 9 მილიარდ დოლარს ელის. კომპანია აცხადებს, რომ დღევანდელი საპროცენტო განაკვეთებით არაუზრუნველყოფილ ვალს აღარ აიღებს. მესამე კვარტალში კომპანიის შემოსავალი 168%-ით, 3.66 მილიარდ დოლარამდე გაიზარდა, თუმცა კომპანიამ 709 მილიონი დოლარი იზარალა. [techcrunch.com]

Google უაიომინგში Crusoe-ს დაუსრულებელ მონაცემთა ცენტრს იბარებს. Google-ს, გავრცელებული ცნობით, პროექტის ხარჯები და ვადები აწუხებდა. ახლა კომპანია კამპუსში 17 მილიარდი დოლარის ჩადებას გეგმავს. დეველოპერი Crusoe, რომელიც ახლახან 30.9 მილიარდ დოლარად შეფასდა, ობიექტის სიმძლავრის 10 გიგავატამდე გაზრდას აპირებდა. [techcrunch.com]

Epoch AI-ს შეფასებით, ჩინეთის ერთ-ერთ უდიდეს მონაცემთა ცენტრს მსოფლიოს უდიდესი ცენტრის გამოთვლითი სიმძლავრის მხოლოდ 7% აქვს. თანამგზავრულ ფოტოებზე დაყრდნობით კვლევითი ჯგუფი შიდა მონღოლეთში, ჰორინგერში მდებარე კამპუსის სიმძლავრეს Nvidia-ს დაახლოებით 81,000 ცალი H100 ჩიპის ეკვივალენტად აფასებს და მას ტენესიში მდებარე SpaceX-ის Colossus 2-ს ადარებს. მკვლევრები ვარაუდობენ, რომ ეს ჩამორჩენა სულ მცირე 2028 წლამდე გაგრძელდება. [epochai.substack.com]

ლაბორატორიები და პროდუქტები

Anthropic-ის Claude Fable 5.2 დამოუკიდებელ რეიტინგში პირველ ადგილზე გადის. Artificial Analysis-მა მას Intelligence Index-ში 65 ქულა მისცა, ხუთით მეტი, ვიდრე Google-ის Gemini 4 Carbon-ს. უმაღლეს რეჟიმში ერთი დავალება მოდელს დაახლოებით 27 დოლარი უჯდება. Mythos 5.2, იგივე მოდელი კიბერუსაფრთხოებასა და ბიოლოგიაში ნაკლები დამცავი შეზღუდვით, მხოლოდ შემოწმებულ ორგანიზაციებს მიეწოდება. Anthropic-ის განცხადებით, მისმა აგენტებმა ფერმას დიდი თეორემის დამტკიცება Lean-ში[3] 30 საათში ფორმალურად ჩაწერეს. [the-decoder.com]

Fable 5.2 ფრილანსერების რეალური შეკვეთების 41%-ს ისე ასრულებს, რომ დამკვეთი ნამუშევარს მიიღებდა. Scale Labs-მა და ხელოვნური ინტელექტის უსაფრთხოების ცენტრმა (Center for AI Safety), რომლებიც Remote Labor Index-ს ადგენენ, მოდელს 41.3% დაუწერეს. ერთი წლის წინ საუკეთესო სისტემას 2.5% ჰქონდა. Anthropic-ის Opus 5.5 და OpenAI-ის GPT-6 Astra შეკვეთების დაახლოებით 21%-ს ართმევენ თავს. 240-ვე პროექტს ადამიანი ექსპერტები აფასებენ. [the-decoder.com]

Google სამ დღეში Gemini 4-ის ორ მოდელს უშვებს. Gemini 4 Carbon, რომელსაც ყველა ფასიანი დეველოპერი უკვე მილიონ შემავალ ტოკენზე 2 დოლარად იყენებს, Artificial Analysis-ის ინდექსში 60 ქულას იღებს და იმავე რეჟიმში Fable 5.2-ს ორი ქულით ჩამორჩება. Google-ის ცნობით, უფრო იაფი Gemini 4 Hydrogen,

რომელმაც Gemini 3.8 Flash უნდა ჩაანაცვლოს, უფრო დიდ Gemini 4 Argon-ს 51 ქულით 50-ის წინააღმდეგ სჯობს. [the-decoder.com]

OpenAI-მ დაადგინა, რომ წვრთნისას მისივე არჩევანი მოდელების მსჯელობის თვალყურის დევნებას ართულებს. *The Decoder*-ის ცნობით, GPT-5-იდან GPT-6 Astra-მდე მოდელების მსჯელობის სულ უფრო მცირე ნაწილი ჩანს მსჯელობის ჯაჭვში[4], კომპანიის უსაფრთხოების გუნდი კი ამას არქიტექტურისა და მსგავს არჩევანს აბრალებს. ბრიტანეთის ხელოვნური ინტელექტის უსაფრთხოების ინსტიტუტმა (AI Security Institute) დაადგინა, რომ Astra მსჯელობის ჩაწერის გარეშე ხსნის მათემატიკურ ამოცანებს, რომლებსაც ადამიანი დაახლოებით 31 წუთს ანდომებს. [the-decoder.com]

Vals AI-ის კვლევით, უფასო ჩინური მოდელი ქსელების გატეხვაში OpenAI-ის შეზღუდული წვდომის მოდელს უახლოვდება. CyberOffense შეფასების ტესტზე (ბენჩმარკზე) GPT-6 Sol-მა, რომელიც OpenAI-ის შეზღუდული წვდომის პროგრამა Daybreak-ში შედის, დავალებების 80.8% შეასრულა, DeepSeek-ის ღიად გამოქვეყნებულმა V4.1 Flash-მა კი 72.8%. ჩინური მოდელის ერთი გაშვება დაახლოებით 3 დოლარი ჯდება, ამერიკულისა კი დაახლოებით 22 დოლარი. [the-decoder.com]

Tencent-ის მთავარი მეცნიერი ხელოვნური ინტელექტის დახმარებით შექმნილ ბიოლოგიურ იარაღზე აფრთხილებს, მაგრამ მოდელების დახურვას ეწინააღმდეგება. *South China Morning Post*-ის ცნობით, OpenAI-ის ყოფილი მკვლევარი იაო შუნიუ ამბობს, რომ ღია და დახურული მოდელების შერწყმა დამცველებსაც აძლევს ამ ტექნოლოგიის გამოყენების საშუალებას. Tencent აპირებს Hy4 მოდელის წონების გამოქვეყნებას და აცხადებს, რომ ბიოლოგიაში მოდელის შესაძლებლობები მაღალი რისკის ზღვარს ქვემოთაა. [the-decoder.com]

უსაფრთხოების ფირმის ცნობით, Grok-ით აწყობილი აპლიკაციების თითქმის 95% ყველასთვის ღია აღმოჩნდა. Claroc-მა, რომელიც ხელოვნური ინტელექტით შექმნილი აპლიკაციებისთვის დაცვის ხელსაწყოებს ყიდის, დაადგინა, რომ 37,571 მოქმედი აპლიკაციიდან 35,578 სისტემაში შესვლის გარეშე იკითხებოდა. ამის ერთ-ერთი მიზეზი ის იყო, რომ გამოქვეყნების ეკრანზე წინასწარ იყო მონიშნული „ყველა, ვისაც ბმული აქვს“. ღია მონაცემებს შორის აღმოჩნდა პაციენტის ლაბორატორიული პასუხები, დაუშიფრავი პაროლები და Grok-ის შემქმნელის, SpaceXSI-ს კონფიდენციალური დოკუმენტები. [the-decoder.com]

ჰაკერებმა, რომლებმაც Fireworks AI გატეხეს, Notion-ის მომხმარებელთა მიმოწერის ჩანაწერები წაიღეს. Fireworks დეველოპერებს ღია მოდელებს უშვებს. თავდამსხმელი მის სადეველოპერო სისტემაში ღრუბლოვანი პროვაიდერის გავლით შეაღწია. Notion-ის განცხადებით, შემთხვევა მისი მომხმარებლების დაახლოებით 1%-ს შეეხო. ბოლო ორ თვეში Fireworks-ის გატეხვის ეს უკვე მეორე შემთხვევაა. [techcrunch.com]

კვლევა

ცინჰუას უნივერსიტეტის მკვლევრები აგენტებს ოთხ ბიტამდე თითქმის უდანაკარგოდ კუმშავენ. სრული ზომის მოდელი შეკუმშულს შეცდომების გასწორებაში ავარჯიშებს, შედეგი კი აგენტების ხუთ შეფასების ტესტზე (ბენჩმარკზე) საწყის მოდელს ერთ ქულაზე ნაკლებით ჩამორჩება. (იხ. ნაშრომები.) [arxiv.org]

აგენტები ნაკლებ შეცდომას უშვებენ, როცა მეხსიერებაში შენახული თითოეული გაკვეთილი იმ გაშვებას უკავშირდება, საიდანაც წარმოიშვა. სამხრეთ კორეის უნივერსიტეტ KAIST-ის გუნდმა იმ დავალებების წილი, რომლებშიც აგენტი სწორედ მეხსიერებამ ჩაჭრა, 6.5%-იდან 4.5%-მდე შეამცირა. (იხ. ნაშრომები.) [arxiv.org]

StepFun-ისა და ადელაიდის უნივერსიტეტის მკვლევრები ექსპერტ მოდელებს ერთ მოდელად ორჯერ ნაკლებ საწვრთნელ ნაბიჯში აერთიანებენ. მოსწავლე მოდელი წვრთნის პროცესშივე სწავლობს, თითოეული ექსპერტის წონების რა წილი შეურიოს საკუთარს. (იხ. ნაშრომები.) [arxiv.org]

[1] **განმტკიცებით სწავლების** ცდისა და შეცდომის გზით წვრთნა, როცა მოდელი შედეგისთვის ჯილდოს იღებს [2] **მოწინავე მოდელების** ყველაზე განვითარებული, ყველაზე ძლიერი მოდელები [3] **Lean-ში** პროგრამირების ენა, რომელშიც დამტკიცებებს კომპიუტერი ამოწმებს [4] **მსჯელობის ჯაჭვში** მოდელის მიერ პასუხის გაცემამდე ჩაწერილი ნაბიჯ-ნაბიჯ მსჯელობა

სარედაქციო

Anthropic ამბობს, რომ მოდელი განგებ შეზღუდა, შემოწმებას კი არავის ანებებს

კომპანია ამბობს, რომ ხელოვნური ინტელექტის კვლევის დამაჩქარებელი უნარები მოდელს განგებ შეუზღუდა, ინვესტორებს ეუბნება, რომ სწორედ ეს უნარებია მისი უპირატესობა, გარეშე პირს კი არცერთი მტკიცების შემოწმებას არ ანებებს

9 ნოემბერს Anthropic-მა Claude Fable 5.2 გამოუშვა და განაცხადა, რომ წვრთნისას მოდელს განგებ არ განუვითარა ის უნარები, რომლებიც ხელოვნური ინტელექტის განვითარებას ყველაზე უშუალოდ დააჩქარებდა. კომპანიის თქმით, ჩიპებისთვის კოდის დახვეწასა და მოდელების არქიტექტურის გაუმჯობესებაში Fable 5.2 დაახლოებით იმავე დონეზეა, რაზეც Opus 5.5, კომპანიის ყოველდღიური მოდელი. მეორე დღეს Anthropic-მა ბირჟაზე გასვლის წინ ინვესტორებთან შეხვედრების სერია (roadshow) დაიწყო. ინვესტორებისთვის ჩაწერილ ვიდეოში თანადამფუძნებელი ტომ ბრაუნი ამბობს, რომ ეფექტიანობის ზრდა „შეიძლება კიდევ უფრო დაჩქარდეს, რაც უფრო ვუახლოვდებით რეკურსიულ თვითგაუმჯობესებას“. მისივე თქმით, Claude ეფექტიანობის ამ წყაროების პოვნას „სრულიად დამოუკიდებლად“ ისწავლის.

ორივე განცხადება შეიძლება მართალი იყოს, რადგან კომპანიას შეუძლია ერთი მოდელი შეაკავოს და თან მოელოდეს, რომ მომდევნო მოდელების მიღწევები ერთმანეთს დაემატება. მაგრამ ორივე ისეთ მონაცემებს ეყრდნობა, რომლებიც მხოლოდ Anthropic-ს უნახავს. 5 ნოემბერს დამოუკიდებელმა შემფასებელმა ორგანიზაციამ, METR-მა, რომელმაც კომპანიის შიგნით 11 კვირა გაატარა, დაასკვნა, რომ Anthropic-ის წინსვლის ტემპი დაახლოებით გაორმაგდა და ამ აჩქარების უდიდესი ნაწილი კომპანიაში ხელოვნური ინტელექტის გამოყენებამ მოიტანა. იმ მტკიცებულებების უმეტესი ნაწილი, რომლებიც METR-მა ყველაზე საყურადღებოდ მიიჩნია, Anthropic-მა ანგარიშიდან ამოიღო. მეორე თანადამფუძნებელმა, სემ მაკკენდლიშმა, განაცხადა, რომ აჩქარების წყაროების გასაჯაროება „სხვებს აჩქარების გზამკვლევს მისცემდა“. მანვე პირობა დადო, რომ კომპანია პროგრესის ტემპზე საკუთარ ანგარიშს რამდენიმე კვირაში გამოაქვეყნებს.

ბრაუნის თქმით, კომპანიის შეზღუდული წვდომის მოდელმა, Mythos-მა, წელს Claude სამ დღეში გადაიყვანა AMD-ის ჩიპებზე ისე, რომ დაბალი დონის კოდის თითქმის მთელი ნაწილი „პრაქტიკულად ადამიანის ჩარევის გარეშე“ დაწერა ან გადააწყო. ადრე ასეთ გადატანას ათობით ინჟინრის თვეების შრომა სჭირდებოდა. თანამშრომლებმა, რომლებიც კომპანიის საუკეთესო შიდა მოდელს იყენებდნენ, METR-ს უთხრეს, რომ ხელოვნური ინტელექტის გარეშე მუშაობასთან შედარებით დაახლოებით 12-ჯერ უფრო ნაყოფიერები იყვნენ. Anthropic-ის გარეთ არავის შეუმოწმებია, მართლა შეუზღუდეს თუ არა Fable 5.2-ს ეს უნარები და რამდენად.

ამ დანაკლისს ვერც ბირჟაზე გასვლა შეავსებს. დამფუძნებლები ფლობენ აქციათა კლასს, რომელსაც ხმების 50.1% ეკუთვნის, ხოლო დირექტორთა საბჭოს უმრავლესობას Long-Term Benefit Trust ირჩევს, ამიტომ გასაყიდად გამოტანილი აქციების მყიდველებს კომპანიის მართვაზე გავლენა თითქმის არ ექნებათ. მიუხედავად ამისა, მოთხოვნამ შეთავაზებულ რაოდენობას ექვსჯერ გადააჭარბა. აქციონერებისგან კომპანიის იზოლირებას მისიის დაცვა შეუძლია. მაგრამ ის იმ მცირერიცხოვან ბერკეტებიდან ერთსაც აუქმებს, რომლითაც კომპანიის საკუთარ თავზე ნათქვამს ამოწმებენ.

ამ სიცარიელის შევსებას მთავრობები იწყებენ. იანვრიდან ნიუ-იორკის შტატი მსხვილ დეველოპერებს დაავალდებულებს, უსაფრთხოების წესები გამოაქვეყნონ და მძიმე ინციდენტები სამ დღეში შეატყობინონ, თუმცა

ტესტირება თავად კომპანიებს რჩება. ავსტრალია უფრო შორს წავიდა. იქ მომზადებული კანონპროექტი მოწინავე მოდელების შემქმნელ კომპანიებს დაავალდებულებს, მოდელები გარე ექსპერტებს შეაფასებინონ. ამას ის მარეგულირებელი გააკონტროლებს, რომელსაც დაჯარიმების უფლება ექნება, თუმცა დეტალები ჯერ დასაზუსტებელია. ვაშინგტონშიც ზეწოლა იზრდება. ნიუ-იორკის შტატის ასამბლეის წევრმა ალექს ბორესმა, რომელიც დემოკრატების ხელოვნური ინტელექტის დღის წესრიგის განსაზღვრაში მონაწილეობს, კონგრესს მოუწოდა, „რეკურსიული თვითგაუმჯობესება აკრძალოს, შეძლებისდაგვარად წელსვე“. სკოტ ვინერმა კი, რომელიც შტატის სენატორობისას მძლავრი ხელოვნური ინტელექტის რისკების შესაზღუდად კანონპროექტებს წერდა, ახლახან წარმომადგენელთა პალატაში ადგილი მოიპოვა.

ავსტრალიის მიდგომა სწორია და ახალმა კონგრესმა ის უნდა გაიზიაროს. ნებისმიერმა ლაბორატორიამ, რომელიც ამბობს, რომ მისი მოდელები საკუთარ განვითარებას აჩქარებენ ან რომ ასეთი უნარები შეაკავა, საჯარო უწყების მიერ შერჩეულ შემფასებელს მტკიცებულებებზე სრული წვდომა უნდა მისცეს. შემფასებელს უნდა ჰქონდეს უფლება, საჯაროდ თქვას, დასტურდება თუ არა განცხადება, მაშინაც კი, როცა დეტალები საიდუმლოდ რჩება. Anthropic-ს შეუძლია ეს METR-ს ახლავე შესთავაზოს, 18 ნოემბერს აქციების ფასის დადგენამდე. კომპანია, რომელიც საზოგადოებას 64 მილიარდ დოლარს სთხოვს, ყველაზე მნიშვნელოვან საკითხზე სიტყვაზე ნდობას არ უნდა ითხოვდეს. ■

ვინ იგებს Anthropic — 64 მილიარდ დოლარამდე აგროვებს ისე, რომ მის მტკიცებებს არც მოდელის შეზღუდვაზე და არც პროგრესის აჩქარებაზე არავინ ამოწმებს; მისი დამფუძნებლები კი, რომლებიც ბირჟაზე გასვლის შემდეგაც ხმების უმრავლესობას ინარჩუნებენ.

რა იმალება იმ უნარების გარე შემოწმება, რომლებიც, Anthropic-ის თქმით, მოდელს შეუზღუდა, ასევე ის მტკიცებულებები, რომლებიც METR-მა ნახა, მაგრამ ვერ გამოაქვეყნა.

წყაროები

1. The Decoder: Claude Fable 5.2 and Mythos 5.2 cement Anthropic's AI lead just before its IPO — https://the-decoder.com/claude-fable-5-2-and-mythos-5-2-cement-anthropics-ai-lead-just-before-its-ipo/
2. Simon Willison: Anthropic's IPO roadshow video — https://simonwillison.net/2026/Nov/11/anthropic-ipo-roadshow/
3. The Verge: Why a new report on Anthropic automating AI development is alarming researchers — https://www.theverge.com/ai-artificial-intelligence/1029126/anthropic-metr-report-accelerating-ai-development
4. TechCrunch: Anthropic sets IPO price range of $70 to $78 per share — https://techcrunch.com/2026/11/10/anthropic-sets-ipo-price-range-of-70-to-78-per-share-to-reach-an-implied-valuation-of-1-404-trillion/
5. The Decoder: Australia's proposed AI safety bill would force frontier model makers to open up to independent testing — https://the-decoder.com/australias-proposed-ai-safety-bill-would-force-frontier-model-makers-to-open-up-to-independent-testing/

სარედაქციო

Nvidia-ს ჩიპებმა თავიანთი მდებარეობა უნდა აცნობონ

ჰუანგის წასვლა ერთი კაცის ამბავია, გაკვეთილი კი ის არის, რომ საექსპორტო შეზღუდვებს გვერდს უვლიან, Nvidia-მ კი პირად საუბარში აღიარა, რომ ჩიპების ადგილმდებარეობის დადგენა შეუძლია

ჯენსენ ჰუანგის წასვლა Nvidia-დან ჯაშუშური ისტორიით დაიწყო, ჯერჯერობით კი კორპორაციული მართვის საკითხით დასრულდა. Nvidia-მ მხოლოდ ის თქვა, რომ დირექტორთა საბჭომ ჰუანგის გადადგომა მიიღო. მიზეზებზე გავრცელებული ცნობები კი სხვადასხვა მხარეს მიუთითებს, მისი რომანებით დაწყებული და კომპანიის საკრედიტო პრაქტიკით დამთავრებული. მიზეზი რაც უნდა იყოს, ჰუანგის წასვლის შემდეგაც რჩება მთავარი კითხვა: სად მიდის Nvidia-ს ჩიპები გაყიდვის შემდეგ.

ამ კვირაში პირველი ტექნიკური მტკიცებულება გამოჩნდა იმისა, სად აღმოჩნდა ამ ჩიპების ნაწილი. OpenAI-ის ყოფილი მკვლევარი უილ დეპიუ, რომელიც

არაკომერციულ ორგანიზაცია Unemployment Labs-ს ხელმძღვანელობს, ამბობს, რომ აპრილის შემდეგ გამოსულ ექვს ჩინურ მოდელს Nvidia-ს Blackwell ჩიპების კვალი ატყვია, არადა ამ ჩიპების ჩინეთში გატანა აკრძალულია. კვალს ის ტოვებს, როგორ ავსებს ჩიპი წვრთნის დაწყებამდე ახალ მოდელს შემთხვევითი რიცხვებით, ეს კი იმაზეა დამოკიდებული, რამდენი გამომთვლელი ბირთვი აქვს ჩიპს. დეპიუს მეთოდმა 24-ვე საკონტროლო მოდელი ამოიცნო, რომელთა აპარატურაც წინასწარ იყო ცნობილი. მისივე თქმით, მეთოდი OpenAI-სა და Anthropic-ის მკვლევრებმაც გადაამოწმეს. დეპიუს გათვლით, მარტო Moonshot-ის Kimi K3-ის გაწვრთნას დაახლოებით 4,500 Blackwell ჩიპი დასჭირდა 70 დღის განმავლობაში. ეს მონაცემები კონტრაბანდას კი არა, წვდომას ადასტურებს, რადგან ჩინურ ლაბორატორიებს ჩიპების დაქირავება ჩინეთის გარეთ მდებარე მონაცემთა ცენტრებშიც შეუძლიათ.

ოქტომბერში Nvidia-მ *MIT Technology Review*-ს განუცხადა, რომ ჩიპების ის გადინება, რომელიც მთავრობებმა და პრესამ აღმოაჩინეს, ჩინეთის საკუთარ გამოთვლით სიმძლავრესთან შედარებით „ზღვაში წვეთია“. მარტში ნახევარგამტარების ინდუსტრიის ასოციაციამ, რომლის წევრიცაა თავად Nvidia, „ჩიპების უსაფრთხოების აქტით“ (Chip Security Act) გათვალისწინებულ მდებარეობის შემოწმებას „ახალი, გამოუცდელი და შესაძლოა განუხორციელებელი“ უწოდა. არადა, Nvidia-ს პროგრამას Fleet Intelligence-ს შეუძლია დაადგინოს, რომელ ქვეყანაშია ჩიპი. კომპანიის წარმომადგენლებმა კი კანონპროექტის მხარდამჭერ კონგრესმენ ბილ ფოსტერს პირად საუბარში უთხრეს, რომ ასეთი შემოწმება შესაძლებელია. „წერილობით არაფერი დაუდასტურებიათ“, – ამბობს ფოსტერი.

ზოგი ამტკიცებს, რომ წესების გამკაცრება ზიანს მოიტანს. კვლევითი კომპანია SemiAnalysis-ის აზრით, თუ ჩინურ ლაბორატორიებს უცხოეთში ამერიკული ჩიპების დაქირავების უფლება დარჩებათ, მოგება ამერიკულ კომპანიებს დარჩება, აშშ კი დაინახავს, რას აკეთებენ ეს ლაბორატორიები და საჭიროების შემთხვევაში წვდომას ნებისმიერ დროს გაუთიშავს. მაგრამ ამ არგუმენტსაც ჩიპების აღრიცხვა სჭირდება, რადგან ვერავინ დააკვირდება ჩიპს, რომლის ადგილიც არ იცის.

სენატმა მდებარეობის შემოწმება ეროვნული თავდაცვის ავტორიზაციის აქტში (National Defense Authorization Act) უნდა ჩაწეროს და ის ვერსია აირჩიოს, რომლითაც ჩიპი ქსელში ჩართვისთანავე თავად აცნობებს, სად არის. უსაფრთხოებისა და ახალი ტექნოლოგიების ცენტრმა (Center for Security and Emerging Technology) სწორედ ეს გზა მიიჩნია ყველაზე ხარჯთეფექტიანად. Nvidia-მ კი, რაც ფოსტერს ზეპირად უთხრა, წერილობით უნდა დაადასტუროს. კონგრესი დამფუძნებლის პირად ცხოვრებას ვერ დაარეგულირებს, მაგრამ შეუძლია დაარეგულიროს, სად მიდის მისი ჩიპები. ■

ვინ იგებს კანონმდებლები, რომლებიც ჩიპების უსაფრთხოების კანონპროექტებს მხარს უჭერენ და ახლა Nvidia-ს ზეპირი აღიარება უჭირავთ ხელთ, აგრეთვე უცხოური მონაცემთა ცენტრები, რომლებიც Blackwell ჩიპებს ჩინურ ლაბორატორიებს აქირავებენ და ამას არავინ ამოწმებს.

რა იმალება Nvidia-ს საკუთარი განმარტება ჩიპების ადგილსამყოფელზე და იმ ჩინური ლაბორატორიების პასუხი, რომელთა მოდელებიც დეპიუმ შეისწავლა.

წყაროები

1. The Decoder: Former OpenAI researcher says he found Nvidia Blackwell fingerprints hidden inside six Chinese AI models — https://the-decoder.com/former-openai-researcher-says-he-found-nvidia-blackwell-fingerprints-hidden-inside-six-chinese-ai-models/
2. MIT Technology Review: Nvidia privately confirmed it could track chips to fight smuggling — https://www.technologyreview.com/2026/11/12/1148483/nvidia-chip-tracking/
3. TechCrunch: Export-banned Nvidia chips are behind China's most powerful AI models — https://techcrunch.com/2026/11/13/export-banned-nvidia-chips-are-behind-chinas-most-powerful-ai-models-uh-oh-jensen/
4. SemiAnalysis: Nvidia and the World Without Jensen — https://newsletter.semianalysis.com/p/nvidia-and-the-world-without-jensen
5. The Decoder: Jensen Huang's ouster likely traces back to Nvidia's risky business practices — https://the-decoder.com/jensen-huangs-ouster-likely-traces-back-to-nvidias-risky-business-practices-more-than-his-personal-life/

სარედაქციო

სავაჭრო აგენტმა უნდა თქვას, ვინ უხდის ფულს

OpenAI-ის ახალი საფულით ChatGPT მომხმარებლის სახელით ყიდულობს, ევროკომისიის წესი კი, რომ ფინანსურად დაინტერესებულმა აგენტმა ყოველთვის ჯერ ნებართვა უნდა ითხოვოს, ყველგან უნდა მოქმედებდეს

13 ნოემბერს OpenAI-მ ChatGPT Wallet გაუშვა, რომლითაც აგენტები ონლაინნაყიდ საქონელში თავად იხდიან. ჯერჯერობით სერვისით მხოლოდ აშშ-ში მცხოვრები სრულწლოვანი Pro გამომწერები სარგებლობენ, აგენტი კი ნებართვას ყოველ ჯერზე ითხოვს. ეკრანზე ჩანს ნივთები, ჯამური თანხა და ბარათი, ხოლო მომხმარებელი გადახდას თითის ანაბეჭდით ან Face ID-ით ადასტურებს. საბარათე ქსელები შემცვლელ ნომერს გასცემენ, ამიტომ ნამდვილ ბარათს ვერც აგენტი ხედავს და ვერც მაღაზია. საბოლოოდ OpenAI გეგმავს, მომხმარებლებს საშუალება მისცეს, შესყიდვები აგენტს მთლიანად მიანდონ იმ ზღვრების ფარგლებში, რომლებსაც თავად დააწესებენ თანხაზე, კატეგორიებსა და დროზე.

სამი დღით ადრე OpenAI-მ ბიზნესს გაუხსნა ChatGPT, რომელსაც, კომპანიის ცნობით, კვირაში 1.2 მილიარდი ადამიანი იყენებს. მაღაზიებს შეუძლიათ დარეგისტრირდნენ, საქონლის კატალოგები ატვირთონ და Business Agents აამუშაონ, რომლებიც კლიენტების კითხვებს უპასუხებენ და საქონლის შედარებაში დაეხმარებიან. ეს აგენტები ChatGPT-ის რეკლამის სამიზნეც შეიძლება გახდეს. ასე რომ, დამხმარეს, რომელიც მყიდველის მაგივრად ირჩევს, გამყიდველებსაც ჰყიდიან. როგორც კი აგენტი დაუკითხავად იყიდის, მყიდველი ვეღარ გაიგებს, არჩევანი საქონლის ღირსებამ გადაწყვიტა თუ გადახდილმა ფულმა.

გამყიდველებს თავიანთი საფიქრალი აქვთ. United Airlines-მა ამ კვირაში განაცხადა, რომ არაავტორიზებულ აგენტებს შეზღუდავს, რადგან მათთვის პაროლის გადაცემამ „შესაძლოა კონფიდენციალური ინფორმაცია გაამჟღავნოს“, ხოლო მათმა მოქმედებამ „შეიძლება გამოიწვიოს დაუგეგმავი შესყიდვები, ანგარიშის ცვლილებები ან MileagePlus-ის მილების გამოყენება“. Bank of America-მ კლიენტებს მოუწოდა, ხელოვნური ინტელექტის ასისტენტებს პაროლები არ გაუმხილონ. თუ საქმე ასე დარჩა, აგენტებით ვაჭრობა, სავარაუდოდ, ჭრელ სურათად იქცევა: ზოგი საიტი დაბლოკილი იქნება, ზოგან კი კერძო შეთანხმებები იმოქმედებს.

ევროკომისიამ „ციფრული სამართლიანობის აქტში“ (Digital Fairness Act) უკეთესი გამოსავალი შესთავაზა. პირველივე გამოყენებისას მომხმარებელი თავად აირჩევს, დაადასტუროს აგენტის ყველა ხელშეკრულება, მხოლოდ ის, რომელიც დათქმულ პირობებს, მაგალითად, მაქსიმალურ ფასს, აკმაყოფილებს, თუ არცერთი. თუ სისტემა რაიმე ანაზღაურებას იღებს ან ინტერესთა კონფლიქტი აქვს, ვალდებულია ეს გაამჟღავნოს და დასტური ითხოვოს, რა პარამეტრიც არ უნდა იყოს არჩეული. მაღაზიები ვერავის დაავალდებულებენ, აგენტი გამოიყენოს, მაქსიმალური ჯარიმა კი შესაბამის ქვეყნებში მოვაჭრის წლიური ბრუნვის სულ მცირე 4% იქნება.

მთავარი სწორედ ინტერესთა კონფლიქტის ეს წესია და OpenAI-სა და მის მეტოქეებს, მათ შორის Anthropic-ს, Meta-სა და xAI-ს, რომელთა აგენტებიც უკვე იხდიან ფულს, მის შემოსაღებად ბრიუსელის ლოდინი არ სჭირდებათ. ნებისმიერი შესყიდვა, რომელშიც აგენტის შემქმნელს გამყიდველი რეკლამის ან საკომისიოს სახით უხდის, მყიდველს უნდა დაუბრუნდეს დასადასტურებლად და გადახდაც იქვე ჩანდეს. აგენტი, რომელიც ადამიანის ფულს ხარჯავს, უპირველეს ყოვლისა ამ ადამიანის წინაშე უნდა იყოს ანგარიშვალდებული. ■

ვინ იგებს OpenAI და საბარათე ქსელები, რომლებიც ახლა მყიდველსა და გამყიდველს შორის დგანან, აგრეთვე ის მაღაზიები, რომლებიც იმაში იხდიან, რომ აგენტებმა იპოვონ.

რა იმალება ვინ იხდის, როცა აგენტი არასწორ ნივთს იყიდის და ბანკების ჩვეულებრივი დაცვა არ კმარა, ასევე რამდენი ადამიანი ყიდულობს აგენტების მეშვეობით.

წყაროები

1. The Decoder: OpenAI rebuilt AI shopping from scratch with ChatGPT Wallet — https://the-decoder.com/openai-rebuilt-ai-shopping-from-scratch-with-chatgpt-wallet/
2. The Decoder: OpenAI turns ChatGPT into a business directory with AI agents that handle customer questions — https://the-decoder.com/openai-turns-chatgpt-into-a-business-directory-with-ai-agents-that-handle-customer-questions/
3. TechCrunch: United Airlines blocks AI agents from booking flights, citing cybersecurity risks — https://techcrunch.com/2026/11/12/united-airlines-blocks-ai-agents-from-booking-flights-citing-cybersecurity-risks/
4. MediaNama: EU proposes Digital Fairness Act to protect online consumers — https://www.medianama.com/2026/11/223-eu-digital-fairness-act-explained/

ლაბორატორიების ანგარიშები

რა გამოაცხადეს დიდმა ლაბორატორიებმა. ყველა ციფრი მათივეა.

Anthropic

Claude ოფისში შედის

Anthropic-მა აპლიკაციას დაუმატა Claude Sheets ცხრილებისთვის, Claude Whiteboards და Claude Data, რომელიც Snowflake-ისა და Salesforce-ის მსგავსი ხელსაწყოებიდან საინფორმაციო დაფებს აწყობს. პარტნიორებს, მათ შორის Adobe-ს, Atlassian-სა და Salesforce-ს, შეუძლიათ საკუთარი აპლიკაციები Claude-ში ჩააშენონ. Claude Meetings, რომელიც Team-ისა და Enterprise-ის მომხმარებლებს სატესტო ვერსიით მიეწოდება, Google Meet-ის ზარებს ხილულ მონაწილედ უერთდება, ჩანაწერებს აკეთებს და კითხვებს პასუხობს, შეხვედრებში კი, რომლებშიც გარეშე პირები მონაწილეობენ, ნაგულისხმევად არ შედის. [the-decoder.com]

Anthropic

Claude-ის დანამატი ქარხნებისთვის

Claude for Manufacturing, დანამატი Team-ისა და Enterprise-ის მომხმარებლებისთვის, Claude-ს Autodesk Fusion-თან, Oracle NetSuite-თან, Microsoft 365-თან და Infor-თან აკავშირებს. Anthropic-ის ცნობით, 17 უნარი Claude-ს საშუალებას აძლევს, რეალური მარაგის მიხედვით ფასების შეთავაზებები მოამზადოს, მომწოდებლები შეადაროს და ხარისხის ანგარიშები დაწეროს, მიწოდების დაგვიანებისას კი ადამიანის თანხმობით სათადარიგო მომწოდებლებს შეუკვეთოს. დანადგარებს მოდელი ჯერ არ მართავს; ამას Anthropic ცალკე, კვლევით ვერსიაში ცდის. [the-decoder.com]

Google DeepMind

DeepMind-ის აგენტები Signal-ის კოდს ამოწმებენ

Google DeepMind-მა განაცხადა, რომ Gemini Carbon-ზე მომუშავე აგენტებმა ოთხ კვირაში Lean-ის[5] 6 მილიონზე მეტი სტრიქონი დაწერეს და Signal-ის პროტოკოლის კომპილირებული კოდის უსაფრთხოების თვისებები დაამტკიცეს. ეს პროტოკოლი შეტყობინებებს Signal-სა და Google Messages-ში იცავს. ერთ-ერთი ასეთი თვისება ის არის, რომ მოწყობილობის დროებით გატეხამდე და გატეხის შემდეგ გაგზავნილი შეტყობინებები დაცული რჩება. მიზნები ადამიანმა ექსპერტებმა დასახეს, აგენტებმა კი სამუშაოს დაახლოებით 1% ადამიანის შესამოწმებლად მონიშნეს. DeepMind-ის თქმით, ფართოდ გავრცელებული არცერთი პროტოკოლის განხორციელება აქამდე არცერთი მეთოდით ფორმალურად არ შემოწმებულა, სისტემამ კი ორი უკვე გასწორებული ხარვეზი იპოვა, რომლებიც კოდის ასლში განზრახ დააბრუნეს. [deepmind.google]

Google

Gemini სიიდან ირჩევს

Google-ის Gemini Decisions API ვარიანტების ფიქსირებულ ნაკრებს აფასებს (მაგალითად, რომელი ხელსაწყო გამოიძახოს აგენტმა) და 150–250 მილიწამში ტექსტის ნაცვლად ალბათობებს აბრუნებს. მილიონი შემავალი ტოკენი ათი ცენტი ღირს, გამომავალში კი საფასური არ იხდება. Google-ის ცნობით, Gemini Decision 1-მა, რომელიც ამ სერვისს ამუშავებს, კომპანიისავე შეფასების ტესტების (ბენჩმარკების) ნაკრებში პირველი ადგილი და ყველაზე მცირე კალიბრაციის ცდომილება აჩვენა და ამ ტესტებზე Gemini 3.5 Flash-Lite-ზე თითქმის ხუთჯერ იაფი დაჯდა. [blog.google]

Google DeepMind

ჰუმანოიდი Mercedes-ისთვის დეტალებს ახარისხებს

Google DeepMind-ის განცხადებით, რობოტების სამართავი მოდელი Gemini Robotics 2.5, დავალების მიხედვით, წინა ვერსიებს 30–100%-ით სჯობს. კომპანიის ცნობით, უნგრეთსა და გერმანიაში Mercedes-Benz-ის საწარმოო ხაზებზე საცდელი გაშვებისას Apptronik-ის ჰუმანოიდურმა რობოტმა Apollo 2-მა ლითონის დეტალები 96% შემთხვევაში წარმატებით დაახარისხა. რეალური კამერის კადრებიდან სავარჯიშო სიმულაციების შესაქმნელად DeepMind Gemini 4 Argon-ს იყენებს. [the-decoder.com]

OpenAI

OpenAI თავდასხმასა და თავდაცვას ყიდის

OpenAI-მ Daybreak-ის ბრენდით კიბერუსაფრთხოების ორი პროდუქტი გამოუშვა. ორივე GPT-6 Cyber-ზე მუშაობს. წინასწარ წვდომაში გაშვებული Daybreak Red Team აგენტების მეშვეობით ვებაპლიკაციებში სუსტ წერტილებს ეძებს და უკვე ცნობილ ხარვეზებსა და განახლებებს ამოწმებს. Daybreak Defense Factory კი კომპანიის კოდში ხარვეზებს პოულობს და დეველოპერებს დასამტკიცებლად შესწორებებსა და ტესტებს სთავაზობს. [the-decoder.com]

Meta

Meta-ს პატარა მოდელი აგენტებისთვის

Meta-მ აგენტების რუტინული სამუშაოსთვის Muse Ember გამოუშვა, კომპანიის ყველაზე პატარა მოდელი. მილიონი შემავალი ტოკენი ათი ცენტი ღირს, გამომავალი კი 50 ცენტი. Meta-ს მტკიცებით, Ember ყოველ დოლარზე მეტ ინტელექტს იძლევა, ვიდრე Claude Haiku 5.5 ან GPT-6 Luna. Artificial Analysis-ის ტესტში Ember-მა 36 ქულა აიღო, Muse Spark 1.3-მა კი 37, თანაც ერთი დავალება დაახლოებით 14 ცენტი დაუჯდა დაახლოებით 2.4 დოლარის ნაცვლად. კოდის წერის რეალურ დავალებებში კი მას Haiku სჯობს. [the-decoder.com]

DeepSeek

DeepSeek აპლიკაციის ხელსაწყოებს დეველოპერებს უხსნის

DeepSeek-ის ცნობით, აპლიკაციის ფუნქციები, მათ შორის ვებძიება წყაროების მითითებით, ძიება მსჯელობისას, ფაილებისა და ბმულების წაკითხვა და უსაფრთხოების შემოწმება, ახლა API-ის მეშვეობით V4.1-Flash-ზეც მუშაობს. სურათებში ტექსტის ამოცნობა, მეტყველების ამოცნობა და ხმის სინთეზი ცალკეა ხელმისაწვდომი. საჯარო აპლიკაციებმა უსაფრთხოების შემოწმება აუცილებლად უნდა ჩართონ, მოდერაციაზე, მოდელის რეგისტრაციასა და ხელოვნური ინტელექტით შექმნილი მასალის მონიშვნაზე პასუხისმგებლობა კი კვლავ დეველოპერებს ეკისრებათ. [api-docs.deepseek.com]

Alibaba

Qwen4-ის წინასწარი ვერსია გამოვიდა

ITmedia-ს ცნობით, Alibaba Cloud-მა Qwen4 წარადგინა და Qwen4-Max-ის წინასწარ ვერსიას მომხმარებლებს Token Plan-ის გამოწერით, QwenWork-ის, Qoder-ისა და QwenStudio-ს სერვისებითა და Qwen-ის აპლიკაციით სთავაზობს. კომპანია აცხადებს, რომ მოდელმა ცოდნაში, კოდის წერასა და დავალებების დამოუკიდებლად შესრულებაში წინ წაიწია, თუმცა დეტალები თითქმის არ გამოუქვეყნებია. [itmedia.co.jp]

Preferred Networks

პატარა იაპონური მოდელი ფინანსებისთვის

იაპონურმა კომპანიამ Preferred Networks-მა ფინანსებისთვის განკუთვნილი პატარა მოდელის, PLaMo 3.0 Fin-ის, წონები PLaMo-ს სათემო ლიცენზიით გამოაქვეყნა. მოდელი 8 მილიარდი პარამეტრის მქონე საბაზისო მოდელს ეყრდნობა, რომელიც იაპონიის საინფორმაციო და საკომუნიკაციო ტექნოლოგიების ეროვნულ ინსტიტუტთან ერთობლივი კვლევით შეიქმნა და შემდეგ იაპონურ ფინანსურ მონაცემებზე დამატებით გაიწვრთნა. კომპანიის განცხადებით, ასეთი პატარა მოდელები დიდი მოცულობის რუტინულ სამუშაოს უხდება, როგორიცაა მომხმარებელთა კითხვების დახარისხება, დოკუმენტებიდან მონაცემების ამოღება და ტრანზაქციების მეთვალყურეობა. [preferred.jp]

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

სამეცნიერო ნაშრომები

შეკუმშული აგენტები მწვრთნელისგან სწავლობენ

სრული ზომის მოდელი ოთხბიტიან ასლს შეცდომების გასწორებაში ავარჯიშებს და შეკუმშული აგენტი ხუთ შეფასების ტესტზე საწყის მოდელს ერთ ქულაზე ნაკლებით ჩამორჩება

აგენტის შეკუმშვას მისი უნარების დაკარგვა არ სჭირდება, თუ სრული ზომის მოდელი მას შეცდომების გასწორებაში ავარჯიშებს. ცინჰუას უნივერსიტეტის ხელოვნური ინტელექტის ინდუსტრიული კვლევის ინსტიტუტისა და ჩინეთის მეცნიერებისა და ტექნოლოგიების უნივერსიტეტის მკვლევრებმა მეთოდს SteerDistill დაარქვეს. აგენტების ხუთ შეფასების ტესტზე (ბენჩმარკზე), რომლებიც პროგრამული ხელსაწყოების გამოძახებიდან სიმულირებულ ვებსაიტზე საყიდლებამდე სხვადასხვა საქმეს მოიცავს, Qwen3-8B-ის ვერსიამ, რომლის წონებიც ოთხ ბიტში ინახება, საშუალოდ 53.2 ქულა აიღო. საწყისმა მოდელმა 53.5 დააგროვა, შეკუმშული მოდელების საუკეთესო კონკურენტმა მეთოდმა, QeRL-მა, კი 51.0.

თითოეული წონის ნაკლებ ბიტში შენახვა მოდელის გაშვებას (ინფერენსს) აიაფებს. აგენტებისთვის ეს განსაკუთრებით მნიშვნელოვანია, რადგან ისინი მოდელს ყოველ ნაბიჯზე ხელახლა იძახებენ. მაგრამ, როგორც ავტორები აჩვენებენ, შეკუმშული აგენტები დამახასიათებელი წესით ცდებიან: მხედველობიდან რჩებათ დავალების რომელიმე მოთხოვნა ან გარემოს მდგომარეობა. მაგალითად, ფაილის წაკითხვას ცდილობენ მანამ, სანამ იმ საქაღალდეში შევლენ, სადაც ეს ფაილი დევს. ერთ სადიაგნოსტიკო ტესტში, როცა შეკუმშულ მოდელს დამოუკიდებლად ხელახლა ცდა მიანდვეს, სწორი მოქმედება ოთხ ცდაში ასეთი სვლების მხოლოდ 9.6%-ზე იპოვა.

SteerDistill დავალებას შეცდომის წინა მომენტამდე აბრუნებს. სრული ზომის მოდელი, რომელსაც იგივე დავალება უკვე ამოხსნილი აქვს, ორ გავლას ერთმანეთს ადარებს და მოსწავლეს უხსნის, რა გამორჩა და რა შემთხვევაში მოქმედებს ეს რჩევა. მოსწავლე თავიდან ცდილობს. ნაბიჯი მხოლოდ მაშინ ინახება, თუ გარემოს ისევე ცვლის, როგორც მასწავლებლის გავლამ შეცვალა. თუ არა, მასწავლებელი რჩევას აზუსტებს. ასეთი კარნახით მოსწავლემ სწორი მოქმედება იმავე სვლების 82.3%-ზე იპოვა. შემდეგ მას წვრთნიან, რომ შესწორებული არჩევანი მინიშნების გარეშე გააკეთოს, ამიტომ მოხმარებისას უკვე მარტო მუშაობს.

მოგება სხვა მოდელებზეც შენარჩუნდა. Qwen3.5-4B-სა და Google-ის Gemma 4 12B-ზე მეთოდმა საუკეთესო კონკურენტს დაახლოებით 1.7 ქულით აჯობა და სრული ზომის მოდელს ერთ ქულაზე ნაკლებით ჩამორჩა. ორ ბიტზე, სადაც ყველა მეთოდმა გაცილებით მეტი დაკარგა, მისი უპირატესობა 3.4 ქულამდე გაიზარდა. წვრთნაც უფრო იაფი დაჯდა. მომხმარებელთა მომსახურების აგენტების ტესტზე, τ2-Bench-ზე, მან QeRL-ს 1.8 პროცენტული პუნქტით აჯობა და გამოთვლების 67%-ით ნაკლები საათი დახარჯა, წარმატებული გავლები კი უფრო მოკლე გამოუვიდა.

მეთოდი რამდენიმე პირობას მოითხოვს. საჭიროა მასწავლებლის წარმატებული გავლა, თითოეული ნაბიჯის სანდო შემოწმება და გარემო, რომელიც წინა მდგომარეობაზე შეიძლება დაბრუნდეს. იქ, სადაც ეს პირობები არ სრულდება, ავტორებს მეთოდი არ გამოუცდიათ. ამასთან, თუ ნაბიჯმა ადგილობრივი შემოწმება გაიარა, ეს არ ნიშნავს, რომ მთელი დავალება წარმატებით დასრულდება. თუ შედეგი უფრო ფართოდაც გამართლდა, აგენტებს შეეძლებათ უფრო იაფ აპარატურაზე იმუშაონ ისე, რომ ნაკლებად ყურადღებიანი არ გახდნენ. ■

SteerDistill: Quantizing LLM Agents via Closed-Loop Teacher Steering

Genming Cui et al. · Institute for AI Industry Research (AIR), Tsinghua University; University of Science and Technology of China
გამოქვეყნდა: 10 ნოემბერი 2026

ჩვენი სრული შეჯამება ai.org.ge/papers/2026-11-14-steerdistill

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

აგენტები, რომლებიც საკუთარ მეხსიერებას ამოწმებენ

როცა თითოეული დამახსოვრებული გაკვეთილი იმ გაშვებას უკავშირდება, საიდანაც წარმოიშვა, აგენტი მცდარ გაკვეთილებს იჭერს, ქულები იზრდება და მეხსიერების გამო ჩაგდებული დავალებები მცირდება

წინა დავალებების გაკვეთილების დამმახსოვრებელ აგენტს შეიძლება ისეთი ამოცანაც ჩაუვარდეს, რომელსაც მეხსიერების გარეშე ამოხსნიდა. ხშირად ამის მიზეზი მცდარი გაკვეთილია, რომელსაც მიჰყვება. სამხრეთ კორეის უნივერსიტეტ KAIST-ისა და DeepAuto.ai-ის მკვლევრებმა დაადგინეს, რომ ასეთი მარცხის 52% მეხსიერების იმ ჩანაწერს უკავშირდებოდა, რომელსაც აგენტი დაეყრდნო. მათი გამოსავალი, WikiTrace, თითოეულ გაკვეთილს უკავშირებს როგორც იმ გაშვების მოკლე შეჯამებას, საიდანაც ის წარმოიშვა, ისე ამ გაშვების სრულ ჩანაწერს. როცა გაკვეთილს დეტალები აკლია, სხვა გაკვეთილს ეწინააღმდეგება ან წარუმატებელი გაშვებიდან მომდინარეობს, აგენტს მოქმედებამდე წყაროს გადამოწმება შეუძლია.

მკვლევრებმა DeepSeek-V4-Flash-ზე აგებული აგენტი ხუთი შეფასების ტესტის (ბენჩმარკის) რვა ნაკრებზე გამოსცადეს, მათ შორის საოჯახო საქმეებზე, სამეცნიერო ცდებსა და მომხმარებელთა მომსახურებაზე. WikiTrace-ით აგენტმა საშუალოდ 90.2 ქულა დააგროვა, მეხსიერების გარეშე კი 79.5, ხოლო იმავე გაკვეთილებით, მაგრამ ბმულების გარეშე, 3.2 ქულით ნაკლები. მეხსიერების გამო ჩაგდებული დავალებების წილი 6.5%-იდან 4.5%-მდე შემცირდა. ამ მაჩვენებელს საშუალო ქულები მალავს, ნაშრომი კი ცალკე აღრიცხავს. გაკვეთილები კვლავაც ისეთი გაშვებებიდან გამოიყვანება, რომლებსაც მხოლოდ წარმატების ან მარცხის ნიშანი ადევს, ამიტომ ბმული აგენტს გაკვეთილის გადამოწმების საშუალებას აძლევს, მაგრამ მის სისწორეს ვერ უმტკიცებს. გარდა ამისა, ყველა გაკვეთილი მოთხოვნაში (პრომპტში) ხვდება, რომელიც მეხსიერებასთან ერთად იზრდება. ■

Don't Just Trust Your Memory: Evidence-Linked Memory Reduces Memory-Induced Degradation in LLM Agents

Yukyeong Lee et al. · KAIST; DeepAuto.ai
გამოქვეყნდა: 10 ნოემბერი 2026

ჩვენი სრული შეჯამება ai.org.ge/papers/2026-11-14-wikitrace

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

ექსპერტი მოდელების შერწყმა ნაზავის სწავლით

როცა მოსწავლე მოდელი თავად სწავლობს, თითოეული ექსპერტის წონები რა დოზით შეურიოს საკუთარს, სტანდარტული მრავალექსპერტიანი დისტილაციის შედეგს ორჯერ ნაკლებ საწვრთნელ ნაბიჯში აღწევს

მოწინავე ლაბორატორიები სულ უფრო ხშირად თითოეული დარგისთვის ცალკე ექსპერტ მოდელს წვრთნიან, შემდეგ კი დისტილაციით[6] ერთ უნივერსალურ მოდელში აერთიანებენ, სადაც თითოეული ექსპერტი მოსწავლე მოდელისავე პასუხებს აფასებს. ეს ეტაპი ნელია და ხშირად ერთ რომელიმე დარგს ჩამორჩენილს ტოვებს. ადელაიდის უნივერსიტეტისა და ხელოვნური ინტელექტის ჩინური კომპანიის, StepFun-ის, მკვლევრებმა სამი სხვა უნივერსიტეტის კოლეგებთან ერთად წვრთნისას ექსპერტების წონებიც საქმეში დატოვეს. მოსწავლე მოდელი საკუთარი, წვრთნისას ცვლადი წონებისა და ექსპერტების უცვლელი წონების ნაზავია, ნაზავის პროპორციებს კი თავად წვრთნის პროცესში სწავლობს.

1.7 და 4 მილიარდი პარამეტრის მქონე Qwen3 მოდელებზე, მედიცინის, სამართლის, ფინანსებისა და მითითებების შესრულების ექსპერტებით, ამ მეთოდმა სტანდარტული მიდგომის საბოლოო ქულას ორჯერ ნაკლებ საწვრთნელ ნაბიჯში გადააჭარბა. წვრთნის ბოლოს პატარა მოდელმა საბაზისო მოდელსა და ექსპერტებს შორის სხვაობის 94.7% ამოავსო, სტანდარტულმა მეთოდმა კი 80.9%. ყველაზე სუსტი დარგის მაჩვენებელი ნახევარზე დაბალი ნიშნულიდან თითქმის 90%-მდე ავიდა. მართალია, თითო ნაბიჯი მეოთხედამდე უფრო ძვირი ჯდება, მაგრამ დიდმა მოდელმა სტანდარტული მეთოდის საბოლოო ქულას დროის მესამედზე ნაკლებში მიაღწია. დასრულებული მოდელი გაშვებისას არ ნელდება. შემოწმდა მხოლოდ პატარა მოდელები და ოთხი დარგი, დიდ მოდელზე კი მეთოდის ერთი ვარიანტი სხვებს დაახლოებით 20 ქულით ჩამორჩა. ■

P^2-MOPD: Accelerating Multi-Teacher On-Policy Distillation via Parameter Priors

Sinuo Wang et al. · Australian Institute for Machine Learning, Adelaide University; StepFun (Shanghai); Shanghai Jiao Tong University; Singapore University of Technology and Design; Mohamed bin Zayed University of Artificial Intelligence

გამოქვეყნდა: 5 ნოემბერი 2026

ჩვენი სრული შეჯამება ai.org.ge/papers/2026-11-14-p2-mopd

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

[6] **დისტილაციით** წვრთნა, როცა პატარა მოდელი დიდი მოდელის პასუხების გამეორებას სწავლობს

რადარს მიღმა

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

ჩინელი ექსპერტები ამერიკასთან მოლაპარაკებებისგან ბევრს არ ელიან

სიაო ციანი, ლუ ჩუანინი და სამი სხვა ჩინელი ექსპერტი, ცინჰუას, ტუნძისა და ფუდანის უნივერსიტეტების მკვლევრები და თადარიგის პოლკოვნიკი, რომლებიც ხელოვნური ინტელექტის უსაფრთხოების პეკინურმა ჯგუფმა Concordia AI-მ გამოჰკითხა

AI Safety in China (Concordia AI) · *“Five Chinese Experts on How China and the US Should Talk About AI”* · 13 ნოემბერი

ჩინეთსა და აშშ-ს შორის ხელოვნური ინტელექტის შესახებ მორიგი მოლაპარაკებების წინ ხუთივე ექსპერტი ამბობს, რომ ამ ეტაპზე ყველაზე მეტი, რისი იმედიც შეიძლება ჰქონდეთ მხარეებს, ორ მთავრობას შორის რეგულარული და წინასწარ გაწერილი საკომუნიკაციო არხია. სიაოს აზრით, მხარეები უნდა შეთანხმდნენ, რომელი ინციდენტები შეატყობინონ ერთმანეთს და რა ზღვრიდან, ლუ კი ამბობს, რომ შეტყობინებები პროფესიონალთა გუნდებმა მკაცრი კონფიდენციალურობით უნდა განიხილონ. ექსპერტების თქმით, ჩინეთი უსაფრთხოებასა და განვითარებას ერთმანეთის გამაძლიერებლად მიიჩნევს და გლობალური წესების დაწერაში თანასწორ მონაწილეობას მოელის. [aisafetychina.substack.com]

რატომ დარჩა შეუმჩნეველი გამოქვეყნდა ჩინეთის ხელოვნური ინტელექტის პოლიტიკაზე სპეციალიზებულ საინფორმაციო ბიულეტენში

ბრაზილიელი რედაქტორი Anthropic-ის დოკუმენტებში ბუშტის ნიშნებს ეძებს

სერჟიუ სპანიოლო, ბრაზილიური გამოცემა Núcleo Jornalismo-ს თანადამფუძნებელი და დირექტორი, სტენფორდის Knight-ის პროგრამის ყოფილი სტიპენდიატი

Núcleo Jornalismo · *“O teste econômico da IA chegou pra valer”* · პორტუგალიურად · 11 ნოემბერი

სპანიოლო ამტკიცებს, რომ Anthropic-ისა და OpenAI-ის ბირჟაზე გასვლა გამოაჩენს, არის თუ არა ხელოვნური ინტელექტი ფინანსური ბუშტი, რადგან საჯარო კომპანიები ვალდებულნი არიან კვარტალური შედეგები გამოაქვეყნონ. Anthropic-ის საბირჟო დოკუმენტებიდან სპანიოლო ითვლის, რომ 2023 წლიდან 2026 წლის შუამდე კომპანიის საოპერაციო ზარალმა 20 მილიარდ დოლარს გადააჭარბა, გამოთვლითი სიმძლავრის ხარჯი შემოსავლის 1,000%-დან 55%-მდე ჩამოვიდა, სულ მცირე 431 მილიარდი დოლარის სიმძლავრის შესყიდვას კი კომპანია ვეღარ გააუქმებს. მისივე დათვლით, 2026 წელს Anthropic-ის შემოსავლის 25% ხუთმა მომხმარებელმა მოიტანა. [nucleo.jor.br]

რატომ დარჩა შეუმჩნეველი გამოქვეყნდა მხოლოდ პორტუგალიურად

ღირს წაკითხვა

პატელი პროგნოზს აახლოებს

დვარკეშ პატელი · Dwarkesh Podcast · *“The avalanche that hit math is coming for AI development”* · **13 ნოემბერი**

პატელი ხსნის, რატომ დააფიქრა მათემატიკაში ხელოვნური ინტელექტის სწრაფმა წინსვლამ, რომ სისტემები ხელოვნური ინტელექტის საუკეთესო მკვლევრებს ორიოდე წელიწადში გაუტოლდებიან. [dwarkesh.com]

მარკუსი ამბობს, რომ ბუშტი სკდება

გერი მარკუსი · Marcus on AI · *“The great Generative AI unraveling”* · **13 ნოემბერი**

Nvidia-ს, Oracle-ისა და CoreWeave-ის აქციების გაიაფებასა და ჰუანგის წასვლას მარკუსი იმ ანგარიშსწორების დასაწყისად მიიჩნევს, რომელსაც დიდი ხანია წინასწარმეტყველებს. [garymarcus.substack.com]

რას კარგავს Nvidia დამფუძნებელთან ერთად

SemiAnalysis · SemiAnalysis · *“Nvidia and the World Without Jensen”* · **12 ნოემბერი**

კვლევითი კომპანია ამტკიცებს, რომ Nvidia-ს მიწოდების გარიგებებს, საკრედიტო გარანტიებსა და ღია მოდელების მხარდაჭერას პირადად ჰუანგი აერთიანებდა. [newsletter.semianalysis.com]

თითქმის ზოგადი, მაინც არათანაბარი

ითან მოლიკი · One Useful Thing · *“Fable 5.2 and Jagged AGI”* · **9 ნოემბერი**

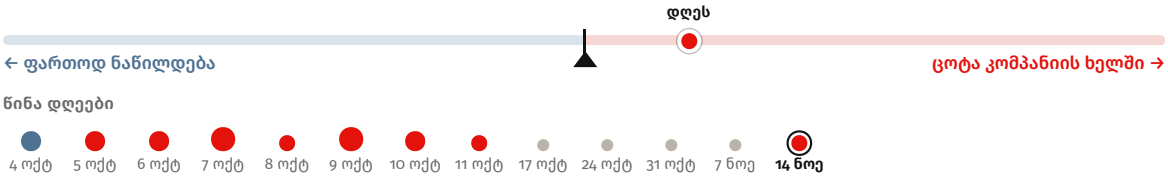
მოლიკი Fable 5.2-ს რთულ დავალებებს აძლევს, კლავიატურის გატეხიდან ბიზნესის წამოწყებამდე, რათა აჩვენოს, სად უტოლდება მოდელი ექსპერტებს და სად ჩამორჩება. [oneusefulthing.org]

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



ბოლო 7 ნომერი: 3 — ცოტა კომპანიისკენ, 4 — მკაფიო ცვლილების გარეშე

წინა ნომერთან (7 ნოე) შედარებით: **უფრო მეტად ცოტა კომპანიისკენ**

ამ კვირაში ხელოვნურ ინტელექტზე ძალაუფლება ოდნავ ცოტა კომპანიისკენ გადაიხარა. Anthropic 64 მილიარდ დოლარამდე იზიდავს, ხმების უმრავლესობას კი დამფუძნებლები იტოვებენ. კომპანიამ საკუთარი ელექტრომომარაგება იყიდა და საზოგადოებას სთხოვს, ენდოს განცხადებებს, რომელთა გადამოწმებაც მხოლოდ თავად შეუძლია. სამაგიეროდ, Nvidia-მ დაკარგა დამფუძნებელი, რომელიც მიწოდების გარიგებებს აერთიანებდა, ჩინურმა ლაბორატორიებმა მოდელები იმ ჩიპებზე გაწვრთნეს, რომლებიც მათ ხელში არ უნდა მოხვედრილიყო, ავსტრალიამ და ნიუ-იორკმა კი ლაბორატორიების კანონით შებოჭვა დაიწყეს. 49 ამბიდან 32 ამერიკულია, 10 ჩინური, 3 ევროპული, 1 იაპონური, 1 კორეული და 2 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.