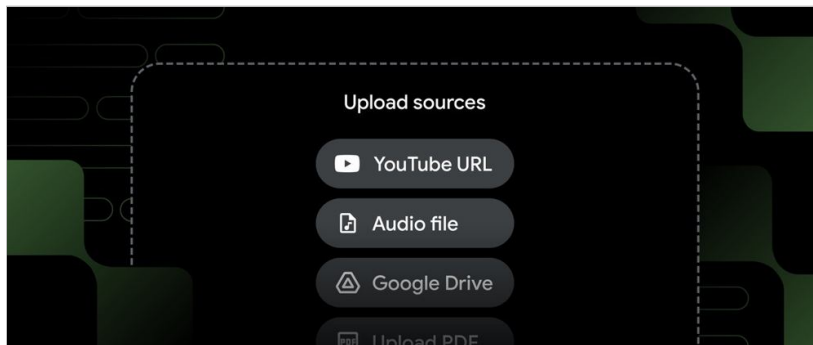


א טיפער איבערבליק אויף אידיש

TLDR AI

טייל 2 פון 4: יאנואר – דעצעמבער 2024



260 אויסגאבעס · 3,825 ידיעות

2 יאנואר 2024 – 31 דעצעמבער 2024

tldr.tech/ai/archives

3	וועגן דעם בוך
4	2024: איבערבליק פונעם יאר
7	יאנואר 2024
19	פערואר 2024
33	מערץ 2024
45	אפריל 2024
58	מיי 2024
73	יוני 2024
84	יולי 2024
99	אויגוסט 2024
113	סעפטעמבער 2024
126	אקטאבער 2024
140	נאוועמבער 2024
156	דעצעמבער 2024

TLDR AI איז א טעגלעכער newsletter (נייעס-בריוו) פון tldr.tech, וואס ווערט געשיקט אומזיסט דורך אימעיל. לויט זיין אייגענעם וועבזייטל ברענגט ער די אינטערעסאנטסטע ידיעות און דורכברוכן אין דער וועלט פון AI (קינסטלעכע אינטעליגענץ). יעדע אויסגאבע ברענגט א רייע קורצע ידיעות: א קעפל, א סיכום פון עטליכע שורות, און א לינק צום מקור. די ידיעות זענען צעטיילט אין אפטיילונגען, ווי Headlines & Launches (הויפט-ידיעות און לאנסירונגען), Research & Innovation (פארשונג און חידושים), Engineering & Research (טעכניק און פארשונג), Miscellaneous (פארשידענעס) און Quick Links (קורצע לינקס).

די ערשטע אויסגאבע אינעם ארכיוו איז פונעם 13טן פעברואר 2023, און דער ארכיוו אויפן וועבזייטל (tldr.tech/ai/archives) גייט ביזן 15טן אקטאבער 2026. ער האלט 941 אויסגאבעס מיט 13,988 ידיעות. דער טייל וואס איר האלט דא, טייל 2, נעמט ארום 2 יאנואר 2024 ביז 31 דעצעמבער 2024: 260 אויסגאבעס מיט 3,825 ידיעות.

דאס בוך איז צעטיילט אין 4 טיילן, איינס פאר יעדעס יאר. יעדער חודש איז א קאפיטל: עס הייבט זיך אן מיט א קורצן סיכום פונעם חודש, און דערנאך קומען די ידיעות, צעטיילט לויט טעמעס – מאדעלן און לאנסירונגען, פארשונג, געצייג פאר פראגראמירער, פירמעס און געלט, רעגירונגען און זיכערהייט, אנאליזן און מיינונגען, און ווייטערע ידיעות. ענליכע ידיעות זענען צוזאמענגענומען אין איין פונקט, און ביי יעדן פונקט שטייט די דאטע ווען דער newsletter האט עס געברענגט. די בילדער קומען פון די ארטיקלען וואס דער newsletter לינקט.

אלעס דא איז אזוי ווי דער newsletter האט עס געמאלדן אין יענער צייט; שפעטערדיגע ענטוויקלונגען זענען נישט צוגעלייגט, און מיינונגען און פאראויסזאגונגען ווערן געברענגט אין נאמען פון דעם וואס האט זיי געזאגט.

2024: איבערבליק פונעם יאר

אינעם יאר 2024 איז די וועלט פון קינסטלעכע אינטעליגענץ געלאפן אין א טעמפא וואס מען האט פריער נישט געזען, און דער טעגליכער בולעטין "TLDR AI" האט כמעט יעדן טאג געהאט וואס צו באריכטן. דאס יאר האט זיך אנגעהויבן מיט OpenAI אין צענטער. אין יאנואר האט זי געעפנט דעם GPT Store, לאנסירט ChatGPT Team, ארויסגעלאזט נייע embedding מאדעלן און א פארבעסערטע GPT-4 Turbo, און איר יערליכע איינקונפט איז געשטיגן איבער 1.6 ביליאן דאלאר. אבער גלייך פון אנהייב האט זי זיך געמוזט שלאגן מיט קאפירעכט-קלאגעס פון דער New York Times און פון שרייבער, און מיט אויפזיכט פון רעגולירער אין אייראפע און אמעריקע. Microsoft האט לאנסירט Copilot Pro און א נייעם Copilot קנעפל אויף קלאוויאטורן. Google DeepMind האט ארויסגעלאזט AlphaGeometry, וואס ליזט אלימפיאדע-פראבלעמען אין געאמעטריע. Meta האט ארויסגעלאזט CodeLlama 70B, און Anthropic האט געוויזן אז מען קען טריינירן "שלאפנדיגע אנגענטן" וואס מען קען שפעטער נישט אויסרייניגן.

דער שטראם פון גרויסע מאדעלן איז נאר געוואקסן. אין פעברואר האט Google לאנסירט Gemini Advanced מיט Gemini Ultra 1.0, דערנאך Gemini 1.5 Pro מיט א קאנטעקסט פון ביז 1 מיליאן טאקענס, און די אפענע Gemma מאדעלן. OpenAI האט געוויזן Sora, וואס שאפט ווידיאס אין א שטוינענדיגער קוואליטעט, און Mistral האט ארויסגעלאזט Mistral Large מיט א שותפות מיט Azure. אין מערץ איז געקומען Anthropic's Claude 3 פאמיליע, וואס מען האט פארגליכן מיט GPT-4, און אין אפריל איז Opus 3 צום ערשטן מאל ארויף איבער GPT-4 אויפן Chatbot Arena. אין מאי האט OpenAI לאנסירט GPT-4o, וואס פארשטייט טעקסט, בילדער און קול אינאיינעם, און Google האט ביי I/O געענטפערט מיט Claude 3.5, Gemini Flash, Veo, Imagen 3 און Project Astra. אין יוני האט Anthropic ארויסגעלאזט Claude 3.5 Sonnet, שטארקער ווי Opus און פאר א פינפטל פונעם פרייז, און Apple האט אויף WWDC אנטפלעקט "Apple Intelligence" צוזאמען מיט א שותפות מיט OpenAI. אין יולי איז ארויס GPT-4o Mini און OpenAI האט אנגעהויבן פרובירן SearchGPT. אין אקטאבער האט Anthropic ארויסגעלאזט א פארבעסערטן Claude 3.5 Sonnet וואס קען אליין באדינען א קאמפיוטער, צוזאמען מיט Haiku 3.5.

פאראלעל צו די ריזן איז די אפן-סאורס וועלט געוואקסן פון חודש צו חודש. אין מערץ האט xAI געעפנט Grok-1 מיט 314 ביליאן פאראמעטערס, און Databricks און AI21 האבן ארויסגעגעבן DBRX און Jamba. אין אפריל איז געקומען Meta's Llama 3 מיט 8B און 70B ווערסיעס, בשעת א 400B מאדעל איז נאך געווען אין טריינירן, צוזאמען מיט Mixtral 8x22B פון Mistral און Phi 3. אין יוני זענען ארויס Qwen 2, Nemotron-4, DeepSeek Coder V2 און Gemma 2. אין יולי האט Meta ארויסגעלאזט Llama 3.1 מיטן 405B מאדעל, Mistral האט געענטפערט מיט Mistral Large 2 און Mistral Nemo, און Apple האט ארויסגעגעבן אלעס פון איר DCLM מאדעל. אין אויגוסט זענען געקומען Flux פון Black Forest Labs, Phi-3.5, Qwen2-Math און Hermes 3. אין סעפטעמבער זענען ארויס Llama 3.2 מיט ווייזע, Qwen 2.5 פון Alibaba מיט איבער 100 מאדעלן, Pixtral 12B און Molmo פון Allen AI. אין נאוועמבער האט DeepMind אפן געמאכט AlphaFold 3, און עס זענען ארויס Pixtral Large און Qwen 2.5 Coder, OLMo און Qwen 2.5. צום סוף יאר זענען געקומען Phi-4, ModernBERT און Llama 3.3, DeepSeek-V3 מיט 671 ביליאן פאראמעטערס.

הינטער דעם גאנצן טומל זענען געשטאנען טשיפס און ריזיגע סומעס געלט. אין פעברואר איז OpenAI געשאצט געווארן איבער 80 ביליאן דאלאר, איר יערליכע איינקונפט איז אריבער 2 ביליאן, און Sam Altman

האט גערעדט פון אויפנעמען ביז 7 טריליאן דאלאר פאר טשיפ-פאבריקן. Nvidia איז דעמאלט געקומען צו א ווערט פון 1.7 טריליאן, און אין יוני איז זי געווארן די ווערטפולסטע פובליק-פירמע אויף דער וועלט מיט 3.34 טריליאן דאלאר. אין מערץ האט Microsoft אריינגענומען די גרינדער פון Inflection, און Amazon האט פארענדיגט א 4 ביליאן דאלאר אינוועסטירונג אין Anthropic. אין נאוועמבער איז צוגעקומען נאך 4 ביליאן פון Amazon. אין אפריל האבן Google, Intel, Meta און באוויזן נייע טשיפס, Nvidia האט אפגעקויפט Run:ai פאר בערך 700 מיליאן דאלאר, און לויט באריכטן האבן Microsoft און OpenAI פלאנירט א סופער-קאמפיוטער פון 100 ביליאן דאלאר. די געלט-רונדעס זענען נאר געוואקסן: xAI מיט 6 ביליאן אין מאי און נאכאמאל 6 ביליאן אין דעצעמבער, Scale AI מיט 1 ביליאן, Coreweave מיט 1.1 ביליאן, Groq מיט 640 מיליאן און Magic מיט 320 מיליאן. אין אקטאבער האט OpenAI אויפגעהויבן 6.6 ביליאן דאלאר ביי א ווערט פון 157 ביליאן, באקומען א קרעדיט-ליניע פון 4 ביליאן, און זיך גענומען איבערבויען אויף א פירמע פאר פראפיט. דאך האבן באריכטן אין יולי געזאגט אז די פירמע קען פארלירן 5 ביליאן דאלאר, און מען האט זיך ברייט געטענהט צי AI איז א בלאז.

דער גרעסטער שינוי אין פארשונג איז געווען דער אריבערגאנג צו מאדעלן וואס "טראכטן". שוין אין פעברואר האבן פארשער געארבעט אויף קלענערע און שפארעוודיגערע מאדעלן, ביז צו מאדעלן מיט בלויז 1.58 ביטס פאר יעדן פאראמעטער, און אין מאי האט Anthropic געמאפירט מיליאנען באגריפן אינעווייניג אין Claude Sonnet. אין יולי האט DeepMind's AlphaProof דערגרייכט א זילבערנעם מעדאל-שטאפל ביי דער מאטעמאטיק אלימפיאדע, און אין אויגוסט האט מען שטארק געארבעט אויף נוצן מער רעכן-קראפט בשעת'ן ענטפערן. אין סעפטעמבער איז דער געדאנק געווארן א פראדוקט מיט OpenAI's o1 מאדעלן, וואס זענען טריינירט צו "טראכטן" איידער זיי ענטפערן. אין נאוועמבער האבן באריכטן געזאגט אז OpenAI's קומענדיגער מאדעל "Orion" ווייזט ווייניגער פארשריט, און די אינדוסטריע האט אנגעהויבן רעדן פון "טעסט-צייט סקעילינג". DeepSeek מיט r1 און Alibaba מיט QwQ-32B האבן פרובירט נאכטון o1. אין דעצעמבער האט OpenAI אין אירע "12 טעג" ארויסגעלאזט דעם פולן o1 מיט, ChatGPT Pro, Sora, Projects, א טעלעפאן-נומער פאר ChatGPT, און צום סוף דעם שטארקסטן טראכט-מאדעל o3. Google האט געענטפערט מיט Gemini 2.0 Flash און א "טראכט"-ווערסיע, Veo 2, דעם קוואנטום-טשיפ Willow און א ריי AI אגענטן. אין אקטאבער זענען די Nobel פרייזן אין פיזיק און כעמיע געגאנגען צו פארשער וואס זענען פארבונדן מיט AI.

פאר פראגראמירער איז דאס געווען א יאר פון נייע געצייג און פון AI אגענטן. אין מערץ האט Cognition באוויזן Devin, אן AI ווייכווארג-אינזשעניר, און אין דעצעמבער איז Devin געעפנט געווארן פאר אלעמען. אין אפריל איז ארויס GitHub Copilot Workspace, אין מאי האט Mistral לאנסירט Codestral, און אין אויגוסט האט Cursor אויפגענומען 60 מיליאן דאלאר. OpenAI האט אין אויגוסט לאנסירט Structured Outputs און פיין-טיונינג פאר GPT-4o, און אין אקטאבער ביי DevDay די Realtime API און באלד דערנאך Canvas. אין סעפטעמבער איז געקומען א כוואליע פון AI אגענטן ביי Microsoft, Salesforce, און T-Mobile. אין נאוועמבער האט Anthropic ארויסגעלאזט דעם Model Context Protocol, וואס פארבינדט מאדעלן מיט געצייג און דאטא.

ביי רעגירונגען און אויף דער זיכערהייט-פראנט איז דאס יאר געווען נישט ווייניגער שטורמיש. אין מערץ האט דאס אייראפעאישע פארלאמענט באשטעטיגט דעם ערשטן גרויסן AI געזעץ, און אין אויגוסט איז דער EU AI Act אריין אין קראפט. אין דער זעלבער צייט איז קאליפארניע'ס געזעץ SB 1047 דורך די לעגיסלאטור, אבער אין אקטאבער האט Newsom עס אפגעווארפן. Elon Musk האט אין מערץ געקלאגט OpenAI,

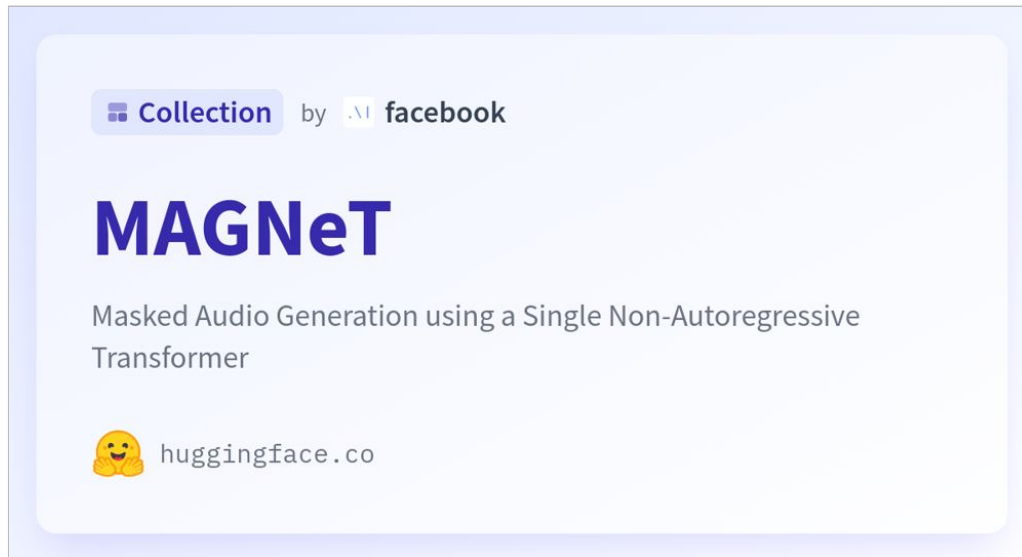
און אין אויגוסט האט ער ווידער אנגעהויבן זיין קלאגע. אין מאי זענען Ilya Sutskever און Jan Leike אוועק פון OpenAI, דער superalignment-טיים איז צעלייגט געווארן, און אכט צייטונגען האבן די פירמע פארקלאגט. אין יוני האט Sutskever געגרינדעט א נייע פירמע פאר זיכערע סופער-אינטעליגענץ, און א גרופע אנגעשטעלטע האבן געפאדערט שוץ פאר די וואס ווארענען פון סכנות. אין אויגוסט האבן דריי גרינדער פארלאזט OpenAI, און אין סעפטעמבער זענען די CTO און צוויי שפיץ-פארשער אוועק. צום סוף יאר האבן פארשער ארויסגעברענגט נאך ערנסטערע פראגן, ווי "שווינדלען" ביים אויסגלייכן (alignment faking) און מאדעלן וואס פלאנירן צו פארפירן. אזוי האט זיך דאס יאר 2024 פארענדיגט: מיט מאדעלן וואס טראכטן טיפער ווי אמאל, און מיט נאך גרעסערע פראגן וועגן וואו דאס אלעס פירט.

דעם יאנואר 2024 איז OpenAI געשטאנען אין צענטער: זי האט געעפנט דעם GPT Store, לאנסירט ChatGPT Team, ארויסגעלאזט נייע embedding מאדעלן און א פארבעסערטע GPT-4 Turbo, און איר יערליכע איינקונפט איז געשטיגן איבער 1.6 ביליאן דאלאר. דערביי האט זי זיך געמוזט שלאגן מיט קלאגס איבער קאפירעכט פון דער New York Times און פון שרייבער, און מיט אויפזיכט פון רעגולירער אין אייראפע און אמעריקע. Microsoft האט לאנסירט Copilot Pro און א נייעם Copilot קנעפל אויף קלאוויאטורן. Google DeepMind האט ארויסגעלאזט AlphaGeometry, וואס לייזט אלימפיאדע-פראבלעמען אין געאמעטריע, און Google האט געוויזן Lumiere פאר ווידעא. Meta האט ארויסגעלאזט CodeLlama 70B, Anthropic האט געוויזן אז מען קען טריינירן "שלאפנדיגע אגענטן" וועלכע מען קען שפעטער נישט אויסרייניגן, און אין פארשונג האט מען אסאך גערעדט פון self-play און DPO.

מאדעלן און נייע לאנסירונגען

- **2 יאנואר: Google's VideoPoet** – Google האט פארעפנטליכט א פארשונג-ארבעט וועגן VideoPoet. דאס איז א שפראך-מאדעל וואס שאפט ווידעאס פון הויכע קוואליטעט, צוזאמען מיט פאסיגע קלאנג, פון א ברייטע ריי סארטן אריינגאבעס.
- **3 יאנואר: Copilot אפליקאציע אויף iOS – Microsoft** – האט ארויסגעלאזט די Copilot טשעטבאט-אפליקאציע פאר iOS און iPadOS. מען קען דארט באנוצן GPT-4 אן צאלן א מנוי: פרעגן פראגעס, שרייבן אימעילס, צוזאמענפאסן טעקסט און שאפן בילדער. מיט דעם צוריק-טויפן פון Bing Chat אויף Copilot ציעלט Microsoft צו א זעלבשטענדיגע איבערלעבונג ענליך צו ChatGPT.
- **5 יאנואר: Bard Advanced** – לויט באריכטן גרייט Google צו "Bard Advanced", א בעסערע ווערסיע פון Bard. זי וועט ארבעטן אויף Gemini Ultra און וועט זיין צוגענגליך דורך א באצאלטן Google One מנוי. Google וועט זי פארקויפן מיט'ן צוואג פון "מער קאמפליצירטע, בעסערע ענטפערס" ווי די היינטיגע ווערסיע, וואס לויפט אויף Gemini Pro און קאנקורירט מיט GPT-3.5.
- **5 יאנואר: דער GPT Store – OpenAI** – האט אנגעזאגט און דערנאך טאקע געעפנט דעם GPT Store. דארט קענען אנטוויקלער באקומען באצאלט פאר GPTs וואס זיי בויען, אויב זיי באשטעטיגן זייער פראפיל און מאכן די GPTs עפנטליך. OpenAI פלאנט צו ארויסהייבן GPTs יעדע וואך. שפעטער אין חודש האבן באצאלטע באניצער באקומען די מעגליכקייט אריינצורופן GPTs גלייך אין א שמועס, און יעדער GPT פארשטייט דעם גאנצן קאנטעקסט פונעם שמועס.
- **8 יאנואר: OpenInterpreter OS מאדעל – Open Interpreter** – א פראיעקט וואס איז נאכגעמאכט נאך OpenAI's interpreter, האט באקומען אן OS מאדעל און א וויזועלע מאדעל. אזוי קען א שפראך-מאדעל קאנטראלירן דעם קאמפיוטער: קוקן אויפ'ן עקראן און קליקן אויף קנעפלעך.
- **9 יאנואר: Generative AI by iStock – Getty Images** – און Nvidia האבן לאנסירט Generative AI by iStock, א פלאטפארמע וואס שאפט סטאק-בילדער פאר קליינע און מיטל-גרויסע ביזנעסער. זי ניצט Nvidia's Picasso מאדעל, טריינירט אויף די Getty און iStock ביבליאטעקן.

- **9 יאנואר: Copilot קנעפל און Windows** – Microsoft לייגט צו א Copilot קנעפל צו קלאוויאטורן, די ערשטע גרויסע ענדערונג אין דער Windows קלאוויאטור אין כמעט דריי צענדליג יאר. מיט'ן קנעפל עפנט מען גלייך Microsoft Copilot. Windows Copilot פרובירט אויך אויס צו עפענען Copilot אויטאמאטיש אויף ברייטע עקראנען פון לכל הפחות 27 אינטש.
- **10 יאנואר: Rabbit R1** – Rabbit האט געשאפן א "large action model" (LAM) פאר איר R1, א קליינער קעשענע-קאמפיוטער. די מאשין איז א נייער געדאנק פון קאמפיוטער און סמארטפאון, און זי ארבעט כמעט אינגאנצן אויף דעם LAM.
- **10 יאנואר: Intel טשיפ פאר קארס** – Intel האט אנגעזאגט אן AI טשיפ פאר אויטאס, "דער ערשטער דור AI system-on-chip פאר א ווייכווארג-באשטימטן אויטא".
- **11 יאנואר: OpenAI – ChatGPT Team** האט לאנסירט ChatGPT Team פאר קליינע גרופעס ביז 149 מיטגלידער. דער פלאן גיט א באזונדערן ארבעטס-ארט, צוטריט צו GPT-4 און DALL-E 3, געצייג צו אנאליזירן טעקעס, און די מעגליכקייט צו בויען אייגענע AI אפליקאציעס אן קאוד. דער פרייז איז 25 ביז 30 דאלאר פער באניצער א חודש.
- **12 יאנואר: נייע פראדוקטן ביי CES** – Swarovski האט געמאכט AI שפאקטיוון וואס דערקענען 9000 סארטן פייגל און חיות און קענען כאפן בילדער און ווידעאס. Google Cloud האט ארויסגעלאזט AI געצייג פאר קרעמעס: א טשעטבאט מיט פערזענליכע פארשלאגן פאר פראדוקטן, און בעסערע זוכונג פון פראדוקטן. Amazon האט ביי CES געוויזן נייע אנטוויקלונגען פאר Alexa פון Character.AI, Splash און Volley.
- **16 יאנואר: Copilot Pro** – Microsoft האט לאנסירט Copilot Pro פאר 20 דאלאר א חודש פער באניצער. עס ארבעט אויף אלע מכשירים, גיט צוטריט צו 365 Microsoft אפליקאציעס און צו GPT-4 Turbo, און האט א בעסערע מעגליכקייט צו שאפן בילדער.
- **16 יאנואר: Nous Hermes 2 Mixtral** – Nous Research האט ארויסגעלאזט SFT און DPO ווערסיעס פון Mixtral. זיי זענען שטארק אין א ריי אויפגאבן און זענען פריי צוגענגליך דורך Together API.
- **17 יאנואר: Stability AI – Stablecode 3B** האט ארויסגעלאזט א 3B קאוד-מאדעל וואס שלאגט אריבער CodeLlama 7B. ער איז קליין גענוג צו לויפן לאקאל אויף א MacBook.
- **17 יאנואר: Magnet מאדעלן** – Meta האט ארויסגעלאזט אירע מאדעלן פאר שאפן מוזיק און קלאנגען, צוזאמען מיט קאוד צו לויפן זיי.



Magnet מאדעלן · 17 יאנואר 2024

- **18 יאנואר: AlphaGeometry** - AlphaGeometry ס'DeepMind האט געלייזט 25 אלימפיאדע-פראבלעמען אין דער באשטימטער צייט, אויפ'ן ניווא פון א גאלד-מעדאל. די פריערדיגע בעסטע מעטאד האט געלייזט 10. דער מאדעל איז טריינירט אויף 100 מיליאן סינטעטישע פראבלעמען, און קיין מענטשליכע ביישפילן, און אזוי ווערט אוועקגענומען א מאנגל אין דאטא. DeepMind האט ארויסגעגעבן דעם קאוד און די וואגן.
- **18 יאנואר: AI טעלעפאנען** - Samsung ס'Galaxy S24 און S24 Plus, וואס קומען ארויס דעם 31טן יאנואר, האבן פארבעסערטע AI מעגליכקייטן. מען זעט דערין א ברייטערע ריכטונג פון AI אין טעלעפאנען, מיט ביישפילן ווי Google Pixel 8 Pro און Rabbit R1.
- **19 יאנואר: Mega Dolphin 120B** - א 120B מאדעל, געשאפן דורך צוזאמענשמעלצן צוויי קלענערע Dolphin מאדעלן, וועלכע זענען געבויט בעיקר אויף Mixtral. Dolphin ווערט גערעכנט ווי איינער פון די שטארקסטע אפן-מאדעלן אן צענזור. ביי Goliath 120B האט מען געזען אז אזא צוזאמענשמעלצונג קען פארבעסערן די לייסטונג.
- **23 יאנואר: Stable LM 2 1.6B** - Stability AI האט ארויסגעלאזט א קאמפאקטן 1.6B מאדעל וואס שטיצט זיבן שפראכן. ער שלאגט אריבער אנדערע מאדעלן אונטער 2 ביליאן פאראמעטערס, אריינגערעכנט Stability ס'אייגענעם פריערדיגן 3B מאדעל. עס קומט אויך א "האלב-געקאכטע" ווערסיע, כדי מען זאל קענען מער צופאסן.
- **23 יאנואר: Tesla FSD v12** - Tesla האט אנגעהויבן ארויסצולאזן Full Self-Driving Beta v12. די ווערסיע ביטן איבער די קאנטראל פונעם אויטא פון קלארן C++ קאוד אויף איין נוירישע נעץ. עס איז נאך אלץ אין beta, און מען וועט בלייבן ספקותדיג ביז מען וועט זען ממשות'דיגע פארבעסערונגען.
- **24 יאנואר: AI אין Chrome** - Google לייגט צו צו Chrome א שרייב-געהילף, אן אויטאמאטישן ארגאניזער פאר טאבס, און א מעגליכקייט צו שאפן אייגענע טעמעס מיט א טעקסט-צו-בילד מאדעל.
- **25 יאנואר: Lumiere** - Google, Weizmann Institute און Tel Aviv University, האט געוויזן Lumiere. דאס איז א space-time diffusion מאדעל פאר רעאליסטישע ווידעאס, וואס צוזאגט

מער צוזאמענגעהאלטענע באוועגונג און בעסערע קוואליטעט ווי די היינטיגע מאדעלן. ער איז נאך נישט עפנטליך צוגעגליך.

• **25 יאנואר: Adept – Fuyu-Heavy** האט פארגרעסערט איר Fuyu ארכיטעקטור. דער נייער מאדעל איז דער דריט-בעסטער אויף MMMU, שלאגט אריבער Gemini Pro, און באקומט איבער 40 אויף HumanEval. צוליב דער פשוטקייט פון דער ארכיטעקטור זענען געווען ווייניגער שוועריגקייטן ביים פארגרעסערן, אבער פראבלעמען מיט דאטא i/o, זכרון-ברייט און אייגענע kernels זענען געבליבן.

• **26 יאנואר: נייע GPT-4 Turbo און OpenAI – embeddings** האט ארויסגעלאזט א נייעם embeddings מאדעל, א בעסערע און ביליגערע GPT-3.5 Turbo, און א GPT-4 Turbo preview וואס איז ווייניגער "פויל" ביים שרייבן קאוד. די פרייזן זענען אויך בעסער. די נייע embeddings קענען טוישן זייער גרייס לויט'ן באדערפעניש, און מען האלט אז דאס קומט פון דער "Matryoshka" מעטאד, וואס לערנט אייגנשאפטן אויף פארשידענע מדרגות.

• **29 יאנואר: AMIE – Google DeepMind** האט אנטוויקלט AMIE, א מאדעל וואס דיאגנאזירט פאציענטן. ער איז טריינירט אויף מעדיצינישע טעקסטן און שמועסן. אין א קאנטראלירטן עקספערימענט האט ער אריבערגעשלאגן מענטשליכע דאקטוירים, אבער ווי ער ארבעט אין דער ווירקליכקייט מוז מען נאך אויספרובירן.

• **29 יאנואר: Imp v1 3B – א מולטי-מאדאלער מאדעל** געבויט אויף Phi-2 און SigLIP. ער ווייזט זייער גוטע רעזולטאטן און איז קליין גענוג צו לויפן אויף א מכשיר.

• **30 יאנואר: CodeLlama 70B – Meta** האט ארויסגעלאזט CodeLlama 70B מאדעלן, טריינירט און צוגעפאסט צו אינסטרוקציעס מיט פיל מער קאוד. זיי באקומען שטארקע רעזולטאטן אויף HumanEval און קומען אין געוויסע הינזיכטן נאענט צו GPT-4. זיי זענען צוגעגליך אונטער דער Meta לייסענס אויף Hugging Face.

• **30 יאנואר: RWKV Eagle 7B – RWKV** איז אן אלטערנאטיוו צו Transformer, מיט א לינעארע attention וואס איז זייער עפעקטיוו. דער נייער Eagle 7B איז טריינירט אויף 1T טאקעס און שלאגט אריבער אסאך 7B Transformer מאדעלן.

• **31 יאנואר: Opera און Poem/1 – Opera** וועט ארויסלאזן איר AI בראווער Opera One פאר iOS אין אייראפע, נאכדעם וואס Apple האט, צוליב דעם Digital Markets Act, ערלויבט בראווערס וואס זענען נישט געבויט אויף WebKit. אויך איז לאנסירט געווארן Poem/1, א זיגער וואס ווייזט די צייט דורך א ליד און ניצט קלוגע prompt caching.

פארשונג און טעכניק

• **2 יאנואר: בילדער און סעגמענטאציע – Paint3D** שאפט 2K טעקסטורן פאר 3D נעצן אן אייגענעבויטע באלייכטונג, אזוי אז מען קען זיי שפעטער לייכט איבערבאלייכטן. HD-Painter פארריכט בילדער לויט א טעקסט ביז 2K, אן קיין טריינירן. AnyText שאפט און רעדאגירט טעקסט אין בילדער אין פיל שפראכן, מיט שפייץ-גענויקייט אויף כינעזיש און ענגליש. ProLab פארגרעסערט סעגמענטאציע מיט לעיבלס וואס באשרייבן אייגנשאפטן, געבויט אויף "פשוטן שכל" פון שפראך-מאדעלן.

- **3 יאנואר:** בילדער וואס פארפירן אויך מענטשן – DeepMind האט געפונען אז קליינע ענדערונגען אין בילדער, געמאכט כדי צו פארפירן AI זעה-סיסטעמען, קענען אונטער קאנטראלירטע באדינגונגען אויך סיסטעמאטיש באאיינפלוסן מענטשליכע אויפנעמונג.
- **3 יאנואר:** embeddings – Microsoft פארשער האבן טריינירט א decoder-only מאדעל, באזירט אויף Mistral, פאר embeddings מיט סינטעטישע דאטא, און ער איז דער בעסטער אין זיין קלאס. די סינטעטישע דאטא האבן זיי געשאפן מיט א צוויי-שטאפלדיגן prompt צו GPT-4. Mosaic האט אריינגעלייגט FlashAttention, GLU און אנדערע חידושים אין MosaicBERT, וואס מאכט דאס פאר-טריינירן פיל שנעלער.
- **3 יאנואר:** פארבעסערן פאטאגראפיעס – א נייע מעטאד פארגרעסערט סטאבילער א בילד פון א נידעריגער רעזאלוציע, מיט פאקוס אויפ'ן בעסטן צענטער פון מעגליכע רעזולטאטן און אויף ראש. TMRNet פארבעסערט בילדער געכאפט אין טונקל, דורך "bracketing" פון עטליכע בילדער.
- **4 יאנואר:** Boundary Attention – א נייע געלערנטע מעטאד צו געפינען ראנדן אין בילדער, מיט בלויז 207 טויזנט פאראמעטערס. זי ארבעט זייער גוט אפילו ווען עס איז דא ראש, און פארבעסערט שטארק די בעסטע רעזולטאטן ביז איצט.
- **4 יאנואר:** self-play פאר אליינמענט – SPIN לאזט א שפראך-מאדעל זיך פארבעסערן דורך "שפילן מיט זיך אליין", מיט בלויז SFT דאטא, אן טייערע דאטא פון מענטשליכע פרעפערענצן. Google פארשער האבן אריינגעברענגט SPO, א פשוטערע אלטערנאטיוו צו RLHF באזירט אויף שפיל-טעאריע, וואס איז שטארק אקעגן פרעפערענצן מיט ראש.
- **4 יאנואר:** שאפן קלאנג – Auffusion שאפט הויך-קוואליטעט קלאנג פון טעקסט, אינספירירט פון טעקסט-צו-בילד מאדעלן. א צווייטע ארבעט שאפט קלאנג מיט איין masked Transformer אויף קלאנג-טאקענס, אן diffusion און אן קיין autoregressive מאדעל.
- **5 יאנואר:** זעה-מאדעלן – 3D-aware VQA ענטפערט פראגעס וועגן דריי-דימענסיאנאלע סצענעס. CAINet פארבעסערט RGB-T סעגמענטאציע פאר אויטאמאטישע פארן. Hugging Face און Stability AI האבן אפן נאכגעמאכט MUSE און ארויסגעגעבן קאוד, checkpoints און א פארשונג-ארבעט.
- **8 יאנואר:** קאנטראלירן מאדעלן – א ארבעט שטעלט פאר צו פירן Llama 2 דורך צולייגן נעגאטיווע embeddings צו די bias טערמינען. אן אנדערע ווייזט אז מען קען צולייגן נייע בלאקן צו א שפראך-מאדעל און זיי טריינירן אויף א נייער אויפגאבע אן פארלירן אלטע וויסן, סיי ביי קאוד און סיי ביי שפראך.
- **8 יאנואר:** עטליכע זעה-ארבעטן – GUESS שאפט מענטשליכע באוועגונג פון טעקסט. DiffusionEdge געפינט ראנדן מיט diffusion, גענויער און עפעקטיוו. א מעטאד פאסט צו CLIP מאדעלן מיט prompts געלערנט פון LLM דאטא, אן געצייכנטע בילדער. LLaVA-Phi איז א מולטי-מאדאלער געהילף געבויט אויף Phi-2 מיט בלויז 2.7B פאראמעטערס.
- **9 יאנואר:** DeepSeek און Mixtral פארשונג-ארבעטן – DeepSeek האט ארויסגעגעבן א טעכנישן באריכט וועגן טריינירן, צאל טאקענס און ארכיטעקטור פון איר LLM, וואס איז נאענט געקומען צו GPT-3.5 אויף אסאך בענטשמארקס. די Mixtral פארשונג-ארבעט רעדט נישט וועגן די פאר-טריינירונג דאטא, אבער ברענגט א נייעם איינבליק וועגן ראטינג צווישן די עקספערטן.

- **9 יאנואר: ViT און פנימער** – א פארבעסערונג אין די positional embeddings פון Vision Transformers נעמט אוועק אומגעוואונטשענע "גראטע"-פלעקן און גיט איבער 25% פארבעסערונג אין ווייטערדיגע אויפגאבן. GAN-NeRF רעדאגירט ווידעאס פון פנימער קאנסיסטענט. Open-Vocabulary SAM פארבינדט CLIP און SAM. Sharp-NeRF מאכט שארפע בילדער פון פארשמירטע.
- **10 יאנואר: ווייטערע ארבעטן** – א נייע מאס פאר טעקסט-צו-3D מאדעלן ניצט GPT-4V און שטימט אויך מיט מענטשליכע פרעפערענצן. SpeechAgents סימולירט שמועסן צווישן ביז 25 אגענטן. STL פארשטארקט FAN ViT מאדעלן. א פראיעקט לאזט AI אגענטן שפילן ראלעס פון דאקטוירים פון פארשידענע געביטן, כדי צו פארבעסערן מעדיצינישע לאגיק.
- **11 יאנואר: TechGPT-2.0 און DPO דאטא** – TechGPT-2.0 פארבעסערט שפראך-מאדעלן אין בויען knowledge graphs, און איז א שריט פאר דער כינעזישער אפן-סאורס קהילה. א נייע DPO דאטא-סעט, געשאפן מיט Distilabel, פילטערט אויס פרעפערענצן-פארן פון הויכע קוואליטעט און פארבעסערט שטארק דעם באזיס-מאדעל.
- **12 יאנואר: prompts, אויסזוכן און דיסטילאציע** – א פארשונג ווייזט אז לויט ווי מען שטעלט צו דעם prompt קען די גענויקייט שוואנקען פון 4% ביז Monarch Mixer 88%. Together איז א 32k retrieval מאדעל וואס שלאגט אריבער אסאך פארמאכטע מאדעלן. פשוטע ערקלערונגען אין שפראך פארבעסערן איבערזעצן קאוד צווישן 19 פראגראמירונג-שפראכן. MiniMA איז א 3B מאדעל, דיסטילירט פון א 7B לערער.
- **15 יאנואר: Sleeper Agents און מער** – Anthropic האט אומיסטן "פארסמט" א שפראך-מאדעל מיט א פראצעדור וואס ברענגט ארויס א געוויס פירונג. דערנאך האט זי נישט געקענט אויסרייניגן דאס פירונג אפילו מיט די זיכערהייט-טריינירונגען וואס זי ניצט פאר אירע פראדוקציע-מאדעלן. די פארשער האבן געפונען אז מען קען טריינירן AI צו פארפירן, און אז זי טוט עס זייער גוט. אנדערע ארבעטן: א דאטא-סעט פון פראצעס וואס רעדן וועגן זיך אליין, וואס זענען שווער פאר מאדעלן, און δ -PIXART, וואס שאפט בילדער שנעלער און מיט מער קאנטראל.
- **16 יאנואר: פארשידענע ארבעטן** – ConvFormer פארגרעסערט בילדער מיט גרויסע convolutions אנשטאט attention. INTERS איז א דאטא-סעט וואס פארבעסערט ווי LLaMA און Mistral זוכן אינפארמאציע. א graph distillation מעטאד פארבעסערט ביאמעדישע סעגמענטאציע. MCQ קוואנטיזאציע פרעסט צוזאם שפראך-מאדעלן מיט העכערע גענויקייט ביי ווייניגער ביטס. LEGO איז א grounding מאדעל פאר בילדער, קלאנג און ווידעא.
- **17 יאנואר: טריינירן אויף גרינגע דאטא** – א פארשונג ווייזט אז טריינירן אויף פארבינדענע אבער גרינגע דאטא פארבעסערט שטארק ווי מאדעלן לייען שווערע פראבלעמען. דאס שטיצט די מיינונג אז fine-tuning ברענגט ארויס וויסן וואס איז שוין דא, אנשטאט צולייגן נייע. ווייטער: א בלאג וועגן צולייגן הערן צו שפראך-מאדעלן דורך צופאסן FILCO, embeddings, פילטערט קאנטעקסט, און Tencent's PhotoMaker שאפט פערזענליכע סטיליזירטע בילדער פון איין בילד מיט ID embeddings, אן LoRA.
- **18 יאנואר: SGLang און AlphaCodium** – LMSYS האט ארויסגעלאזט Radix Attention, וואס שנעלערט טאקענס פער סעקונדע, און SGLang, א שפראך פאר קאמפליצירטע prompts מיט פאראלעליזם. AlphaCodium פארבעסערט מיט א מערשטאפלדיגן פראצעס, באזירט אויף טעסטן, ווי GPT-4 לייזט פראגראמירונג-פראבלעמען אויף CPO. CodeContests פארבעסערט איבערזעצונג ביי

ALMA 13B. DoraemonGPT פארשטייט ווידעאס. א איבערבליק פון איבער 250 ארבעטן באהאנדלט זעה-מאדעלן פאר אויטאס.

- **19 יאנואר: Apple בילד-מאדעלן און TACO** – Apple האט געוויזן אז גרויסע בילד-מאדעלן, פאר-טריינירט מיט autoregressive loss, וואקסן אויף א זייער פאראויסזאגבארן אופן, פונקט ווי שפראך-מאדעלן. TACO איז א נייער גרעסערער בענטשמארק פאר קאוד: GPT-4 באקומט 30% אויף גרינגע פראבלעמען און בלויז 2% אויף די שווערסטע. ווייטער: Vlogger שאפט ווידעא-בלאגס פון א מינוט לאנג, און Compose and Conquer שטעלט אוועק אביעקטן אין 3D אין געשאפענע בילדער.
- **22 יאנואר: טעקסט-צו-ווידעא און באטעריע** – Hugging Face האט ארויסגעגעבן אן איבערבליק וועגן טעקסט-צו-ווידעא. OMG-Seg טוט אלע סארטן סעגמענטאציע מיט איין מאדעל. Microsoft און Pacific Northwest National Laboratory האבן מיט AI געפונען א נייעם הארטן עלעקטראליט פאר באטעריעס, וואס קען זיין ווייניגער געפערליך ווי ליטיום-איאן, אבער מען מוז נאך אויספרובירן אויב ער טויג.
- **23 יאנואר: Self-Rewarding מאדעלן** – א מאדעל וואס באורטיילט זיך אליין ביים טריינירן האט אריבערגעשלאגן סיסטעמען ווי DiverseEvol. GPT-4. LAZARUS א מאדעל אליין אויסקלייבן זיין טריינירונג-דאטא. Depth Anything שאצט טיפקייט פון איין בילד, מיט א דאטא-סעט פון בערך 62 מיליאן בילדער. GroupAnything גרופירט 3D אביעקטן אויף עטליכע שטאפלען.
- **24 יאנואר: צוזאמענשמעלצן reward מאדעלן** – א מאדעל וואס איז צוגעפאסט מיט עטליכע צוזאמענגעשמאלצענע reward מאדעלן ווערט בעסער באטראכט אין 79% פון די פעלער. CPO איז מער עפעקטיוו ווי DPO ביים איבערזעצן. MMCBench פרובירט אויס איבער 100 מולטי-מאדאלע מאדעלן. RPG פארבעסערט בילדער מיט אסאך אביעקטן.
- **25 יאנואר: Mechanistic Interpretability** – Neel Nanda, איינער פון די פיאנירן פון דעם פעלד, האט ארויסגעגעבן א פירער מיט 200 אפענע פראגעס. דער פעלד פאדערט ווייניג רעכן-קראפט, אבער ביז איצט איז נאך ווייניג געלונגען. ווייטער: HEDNet פאר 3D דעטעקציע אין אויטאס, און א פארשונג וועגן asynchronous Local-SGD פאר טריינירן שפראך-מאדעלן.
- **26 יאנואר: פארשידענע** – SpeechGPT-Gen שאפט רייד. DittoGym איז א סימולאטאר פאר ווייכע ראבאטן וואס טוישן זייער פארעם. SegMamba איז אן אלטערנאטיוו צו Transformer פאר 3D מעדיצינישע בילדער. א נייע מאס פאר ענליכקייט צווישן בילדער ניצט סינטעטישע דאטא און מענטשליכע פרעפערענצן.
- **29 יאנואר: Multimodal Pathway** – א מעטאד פארבעסערט א Transformer מיט נישט-פארבינדענע דאטא פון אן אנדער מאדאליטעט, למשל א בילד-מאדעל מיט קלאנג-דאטא. Vivim סעגמענטירט מעדיצינישע ווידעאס מיט א state space מאדעל. Pix2Gestalt שטעלט צוריק אביעקטן וואס זענען טיילווייז פארדעקט.
- **30 יאנואר: SliceGPT** – SliceGPT איז א מעטאד וואס מאכט גרויסע שפראך-מאדעלן שיטערער נאכ'ן טריינירן. זי טוישט אויס די וואגן-מאטריצן אויף קלענערע און קען אראפנעמען ביז 25% פאראמעטערס פון LLAMA2-70B און OPT 66B, בשעת מען האלט ביז 99% פון דער לייסטונג. זי ארבעט אפילו אויף מאדעלן אזוי קליין ווי Phi-2. אויך: Google פארבעסערט meta learning מיט סינטעטישע דאטא פון א Universal Turing Machine, און א 3D diffusion מאדעל מיט א פיל גרעסערן אוצר פון אביעקטן.

• **31 יאנואר: Rephrasing the web** – Apple האט איבערגעשריבן C4 אלס פראגעס און ענטפערס, אינסטרוקציעס און אזוי ווייטער פאר'ן פאר-טריינירן. דאס האט געשנעלערט די קאנווערגענץ צען מאל, אויף קאסט פון דעם איבערשרייב-שריט. אויך איז ארויסגעקומען אן איבערבליק פון continual learning.

געצייג פאר פראגראמירער און אפן-סאורס

• **2 יאנואר: Diffusers JS און Mixtral Offloading** – Hugging Face Diffusers לויפט Mixtral Offloading מיט JavaScript. Mixtral-8x7B לויפט אויף Colab אדער א היים-קאמפיוטער, מיט HQQ קוואנטיזאציע און offloading צווישן GPU און CPU.

• **3 יאנואר: Llamafile און LLM קורס** – Llamafile פאקט איין דעם מאדעל און קאוד אין איין טעקע וואס מען קען לויפן, און מען קען אריינשיקן אויסגאבע פון דער קאמאנד-ליין אין מאדעל. א LLM קורס לערנט merging, GGUF, קוואנטיזאציע און Pykoi. DPO. Pykoi איז א Python ביבליאטעק פאר RLHF.

• **4 יאנואר: GPT-fast און מער** – GPT-fast פון PyTorch שטיצט Mixtral און דערגרייכט 218 טאקענס פער סעקונדע אויף TinyGPT-V 8 GPUs. איז א קליינער מולטי-מאדאלער מאדעל. Marigold מאכט 3D VR ווידעאס פון געוויינטליכע ווידעאס, אונטער MIT לייסענס. עס איז אויך דא א notebook פאר ColBERT, און Intel האט ארויסגעלאזט א געצייג צו לויפן Transformers אויף CPU.

• **5 יאנואר: RAGatouille און Copilot Chat** – RAGatouille איז א ביבליאטעק פאר טריינירן RAG סיסטעמען. GitHub האט געמאכט Copilot Chat צוגענגליך פאר אלעמען, צו פרעגן פראגעס וועגן קאוד. ODTrack גייט נאך אביעקטן אין ווידעאס. א ארטיקל ערקלערט ניין באגריפן ווען מען גייט איבער פון PyTorch צו Jax.

• **9 יאנואר: Portkey, ALOHA, CrewAI** – Portkey AI האט געעפנט איר gateway, וואס פירט פארקער צווישן אסאך מאדעלן. עס איז ארויס קאוד פאר דעם ALOHA ראבאט אין סימולאציע. CrewAI ארגאניזירט AI אגענטן וואס ארבעטן צוזאמען. Mixtral אין Tinygrad.

• **10 יאנואר: Switch Transformer** – Google האט ארויסגעגעבן אויף Hugging Face איר 1.6T MoE מאדעל, איינער פון די ערשטע געלונגענע Mixture-of-Experts מאדעלן. AI Toolkit איז א ++C ביבליאטעק פאר NPCs אין שפילן, מיט finite state machines, behavior trees און פלאנירונג.

• **11 יאנואר: Unsloth און Ollama** – Unsloth שנעלערט fine-tuning און ארבעט איצט מיט TRL. Ollama שטיצט זעה-מאדעלן, אזוי אז מען קען לויפן Llava אויף א MacBook Pro. Chatbot UI איז אן אפענער קאוד פאר א טשעט-אויבערפלאך.

• **12 יאנואר: 33B סייבער-זיכערהייט מאדעל** – א מאדעל טריינירט סיי אויף אטאקירן סיי אויף פארטיידיגן, וואס העלפט אין סייבער-זיכערהייט אויפגאבן.

• **15 יאנואר: Surya און ווייטער** – Surya געפינט שורות טעקסט אין בילדער און PDFs אין פיל שפראכן. א פראיעקט מיט GPT-4V און text-to-speech לאזט "David Attenborough" דערציילן דיין לעבן. עס איז דא אן אפענע נאכמאכונג פון Vanna. AnimateAnyone לאזט ארבעטן מיט א דאטאבאזע דורך AI. Meta ווייזט אן איבערזעצונג מיט קלאנירטע שטימע.

• **16 יאנואר: GGUF קוואנטיזאציע** – א notebook וואס קוואנטיזירט אויטאמאטיש מאדעלן צו GGUF פאר llama cpp.

- **17 יאנואר: דאטא-סעטס** – א DPO דאטא-סעט פאר שרייבן Haiku, און Tiny Narrations, אן אויפגעלייענטע ווערסיע פון Tiny Stories מיט XTTS2.
- **18 יאנואר: Interconnect Tools** – Nathan Lambert האט ארויסגעלאזט געצייג וואס פארוואנדלען אויטאמאטיש א בלאג אין א פאדקאסט און YouTube. Privy איז א קאוד-געהילף וואס שטעלט פרייוואטקייט ערשט.
- **19 יאנואר: Hippo Attention** – א 8bit attention, לויט די שאפער ביז 3 מאל שנעלער ווי Flash Attention v2, מיט 700 TFLOPs אויף MI-300X. די פרטים זענען אבער קנאפ. ווייטער: א טריק צו פאסן Mixtral מיט GPT-4 כדי צו פארקלענערן קאסטן.
- **22 יאנואר: Nanotron און Datatrove** – Hugging Face האט ארויסגעלאזט Nanotron פאר טריינירן מיט 3D פאראלעליזם און Datatrove פאר פילטערן גרויסע טעקסט דאטא-סעטס. אויך דא: א געצייג צו צולייגן באשרייבונגען צו בילדער.
- **24 יאנואר: PairRM און Nano ColBERT** – א שטארקער Mistral מאדעל טריינירט מיט סינטעטישע DPO דאטא. Nano ColBERT איז א פשוטע נאכמאכונג וואס גיט כמעט די זעלבע לייסטונג ווי דער אריגינעל. א טאבעלע פארגלייכט ווי אפט מאדעלן "כאלוצינירן" ביים צוזאמענפאסן דאקומענטן.
- **25 יאנואר: embeddings און מעסטונגען** – מען קען מאכן embeddings פון דער גאנצער ענגלישער Wikipedia אין 15 מינוט מיט Modal. TensorDict איז א ביבליאטעק פאר PyTorch. א ארטיקל באשרייבט מעסטונגען פאר LLM אפליקאציעס אן "גרונט-אמת" דאטא.
- **26 יאנואר: LLM Steer** – LLM Steer פירט א מאדעל מיט וועקטארן פון באגריפן, למשל "באטראכטזאם", און ווערט אפט בעסער ווי prompts. RoMa איז פאר דריינגען אין PyTorch. AgentBoard איז א בענטשמארק פאר אגענטן. עס איז אויך דא א פירער ווי צו בויען א sparse MoE מאדעל פון אנפאנג אן.
- **29 יאנואר: WebDataset און Mamba אין Rust** – Hugging Face שטיצט WebDataset פאר שנעלע לאדונג פון דאטא. LLMs from Scratch לערנט בויען מאדעלן. Simon Willison פארשט ColBERT און RAGatouille. Mamba לויפט אויף א CPU אין ריינעם Rust.
- **31 יאנואר: Lilac Garden און float8** – Lilac Garden צעטיילט דאטא אין גרופעס 100 מאל שנעלער. PyTorch האט קאוד פאר טריינירן מיט float8. Enchanted איז אן iOS אפליקאציע פאר Ollama. Plock לאזט ניצן א מאדעל אומעטום וואו מען שרייבט.

פירמעס, געלט און ביוזעס

- **2 יאנואר: OpenAI איינקונפט** – OpenAI'ס יערליכע איינקונפט איז אריבער 1.6 ביליאן דאלאר, בלויז צוויי חדשים נאכדעם וואס זי האט דערגרייכט 1.3 ביליאן.
- **2 יאנואר: Apple און AI** – Apple, וואס האט אנטוויקלט Ferret מיט Cornell, וועט לויט באריכטן ווייזן ביי WWDC אין יוני נייע AI געצייג, אריינגערעכנט א בעסערע Siri אין iOS 18. זיי זענען געבויט אויף איר Ajax מאדעל. Apple קויפט AI פירמעס, גיט אויס מער AI שטעלעס, און פרובירט אינערליך מיט ChatGPT API. אין iOS 17.4 האט מען אנטדעקט א SiriSummarization framework פאר Messages.

- **2 יאנואר:** Google Ads – Google ארגאניזירט איבער איר Ads ביזנעס, ווייל אירע AI געצייג נעמען אוועק אסאך ארבעט פון ארבעטער וואס פארקויפן מודעות.
- **3 יאנואר:** וויטרוועלע אינפלוענסער – פירמעס דינגען "אינפלוענסער" געשאפן דורך AI און באריכטן גוטע רעזולטאטן מיט ביליגערע קאסטן.
- **4 יאנואר:** Intel – Articul8 AI שפאלט אפ א זעלבשטענדיגע AI וויכווארג פירמע, Articul8 AI, מיט אינוועסטירונג פון DigitalBridge.
- **5 יאנואר:** Perplexity און Nabla – Perplexity האט אויפגעהויבן 73.6 מיליאן דאלאר. Nabla, וואס מאכט א געהילף פאר דאקטוירים, האט אויפגעהויבן 24 מיליאן.
- **5 יאנואר:** Intel Gaudi 2 – לויט Databricks שלאגט Intel'ס Gaudi 2 טשיפ אריבער Nvidia'ס A100.
- **8 יאנואר:** Microsoft און OpenAI – Microsoft VP איז אריין אין OpenAI'ס ברעט אלס באאבאכטער אן א שטים. EU רעגולירער פארשן Microsoft'ס אינוועסטירונג פון 13 מיליאן דאלאר, און Satya Nadella זאגט אז ער וויל סטאביליטעט ביי OpenAI, נישט קאנטראל.
- **10 יאנואר:** OpenAI און צייטונגען – OpenAI פאַרשלאגט צייטונגען צווישן 1 און 5 מיליאן דאלאר א יאר פאר זייער אינהאלט.
- **12 יאנואר:** Nous Research – Nous Research האט אויפגעהויבן 5 מיליאן דאלאר seed.
- **17 יאנואר:** Sakana AI – די פירמע, געגרינדעט דורך געוועזענע Google פארשער David Ha און Llion Jones, האט אויפגעהויבן 30 מיליאן דאלאר seed פון Lux Capital און Khosla Ventures, צו בויען קלענערע מאדעלן אין טאקיא.
- **19 יאנואר:** OpenAI, Meta, Agrawal – OpenAI'ס ערשטע שותפות מיט אן אוניווערסיטעט: Arizona State University באקומט Mark Zuckerberg ChatGPT Enterprise. ציעלט צו AGI און באטראכט אן אפן-סאורס וועג. Parag Agrawal האט אנגעהויבן AI סטארטאפס. Microsoft גיט אומזיסט Reading Coach.
- **22 יאנואר:** Altman און טשיפס – לויט באריכטן הייבט Sam Altman אויף געלט פאר א פארשפרייטע נעץ פון פאבריקן פאר AI טשיפס. צוויי DeepMind פארשער פירן שמועסן צו עפענען א סטארטאפ אין פאריז.
- **24 יאנואר:** ElevenLabs – ElevenLabs האט אויפגעהויבן 80 מיליאן דאלאר Series B.
- **26 יאנואר:** Hugging Face און Qdrant – Hugging Face האט זיך פאראייניגט מיט Google Cloud. Qdrant האט אויפגעהויבן 28 מיליאן דאלאר Series A.
- **29 יאנואר:** Voltron און Tachyum – Voltron Data האט געקויפט Tachyum. Claypot טענהט אז איר Prodigy פראצעסאר קען קאנקורירן מיט צענדליגער Nvidia H200.
- **31 יאנואר:** Shortwave – Shortwave האט ארויסגעלאזט AI אויטאמאטישע דערגאנצונג פאר אימעילס.

רעגירונגען, געזעצן און זיכערהייט

- **4 יאנואר:** דיגיטאלע אונטערשריפטן אין קאמערעס – Canon, Sony און Nikon לייגן אריין דיגיטאלע אונטערשריפטן אין בילדער, כדי מען זאל זיי קענען אונטערשיידן פון AI פעלשונגען.

- **8 יאנואר:** קאפירעכט קלאגעס – שרייבער האבן געקלאגט Microsoft און OpenAI און פאדערן ביז 150,000 דאלאר פער ווערק. OpenAI האט געענטפערט אויף דער New York Times קלאגע אז טריינירן אויף עפנטליכן אינהאלט איז "fair use". די Times פאדערט ביליאנען דאלאר און דאס צעשטערן פון מאדעלן.
- **11 יאנואר:** Valve – Steam ערלויבט רוב שפילן מיט AI, אויב די אנטוויקלער זאגן אן אז זיי ניצן AI.
- **15 יאנואר:** מיליטערישע באנוץ – OpenAI האט געטוישט אירע כללים, אזוי אז מעגליך מעג מען איצט ניצן אירע טעכנאלאגיעס פאר מיליטערישע צוועקן.
- **16 יאנואר:** וואלן – OpenAI גרייט זיך צו די וואלן אין 2024 דורך פארמיידן אויסנוץ און מער דורכזיכטיגקייט. זי שאפט אויך א Collective Alignment מאנשאפט פאר עפנטליכע אינפוט.
- **18 יאנואר:** באליידיגונג קלאגע – א ריכטער אין Georgia האט נישט אוועקגעווארפן 'Mark Walters' קלאגע אקעגן OpenAI איבער פאלשע טענות וואס ChatGPT האט אויסגעטראכט וועגן אים.
- **29 יאנואר:** FTC אויספארשונג – די FTC פארשט די אינוועסטירונגען פון Microsoft, Amazon און Google אין OpenAI און Anthropic.
- **30 יאנואר:** איטאליע – איטאליע'ס פריוואטקייט-אויטאריטעט טענהט אז ChatGPT ברעכט אייראפעאישע געזעצן, און האט געגעבן OpenAI 30 טעג צו ענטפערן.

אנאליזן און מיינונגען

- **2 יאנואר:** אליינמענט און באוואוסטזיין – א ארטיקל אויף LessWrong טענהט אז אליינמענט איז שווער. 19 פארשער האבן געשאפן א רשימה פון סימנים און באשלאסן אז קיין היינטיגע AI איז נישט באוואוסטזיניג. לויט Crunchbase האבן רוב אינוועסטירער אין 2023 געווארט צו זען וואס עס וועט ווערן.
- **3 יאנואר:** איבערבליק פון 2023 – Simon Willison שרייבט אז מאדעלן זענען לייכט צו בויען, אבער נאך נישט פארלעסליך. antirez זאגט אז זיי העלפן אסאך מיט Python און מיט נודנע ארבעט. ChatGPT האט דערגרייכט 100 מיליאן באניצער א חודש אין יאנואר 2023.
- **8 יאנואר:** 2,778 פארשער – אן אנקעטע ווייזט אז די פארשער זענען צעטיילט. איבער האלב זענען אפטימיסטיש, אבער עס בלייבן זארגן וועגן אויסנוץ.
- **10 יאנואר:** Hanson און "Anatomy Framework" – Joe Carlsmith ענטפערט אויף Robin Hanson'ס טענה אז מורא האבן פון AI אלס "פרעמד" איז א פארורטייל. אן אנדער ארטיקל שטעלט פאר א פינף-שיכטיגן ראם צו מעסטן חידושים.
- **11 יאנואר:** GPT-4 Turbo – אין מעסטונגען האט GPT-4 געלייזט 70% פון די פראגראמירונג אויפגאבן, און GPT-4 Turbo 68.8%.
- **16 יאנואר:** IMF – לויט IMF קען AI באאיינפלוסן כמעט 40% פון אלע ארבעטס-שטעלעס אויף דער וועלט.
- **23 יאנואר:** AGI צייטליניעס – Alex Irpan שאצט 10% שאנס פאר AGI ביז 2028 און 50% ביז 2045. Larry Page האט אנגערופן Musk א "speciesist".
- **29 יאנואר:** אפן-סאורס – א ארטיקל פרעגט צי AI זיכערהייט ארגאניזאציעס ווילן פארבאטן אפן-סאורס.

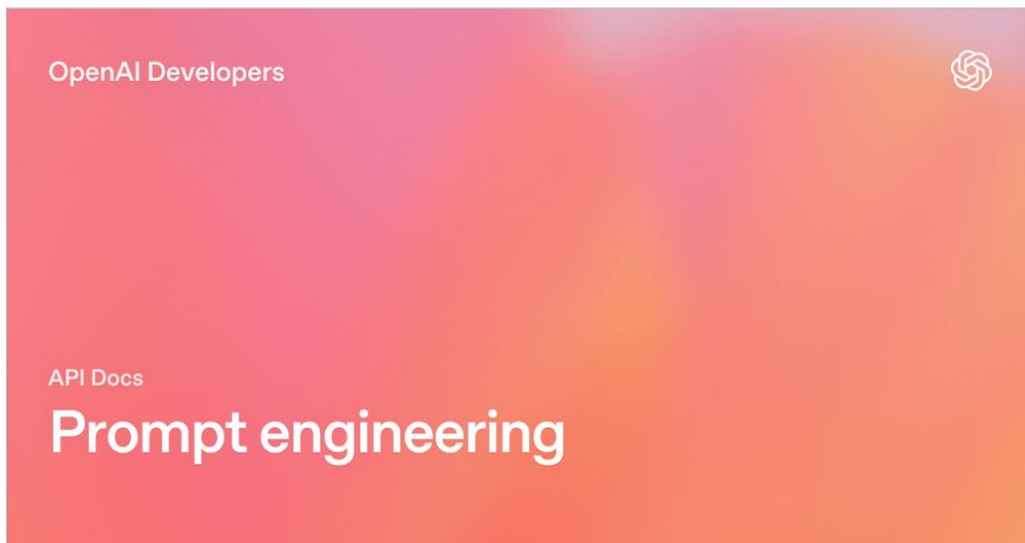
• **31 יאנואר: Altman** – א ארטיקל טענהט אז Sam Altman'ס סותר'דיגע מעלדונגען, למשל אז GPT-5 וועט זיין "okay", זענען א סטראטעגיע.

ווייטערע ידיעות

• **15 יאנואר: אנטיביאטיקע** – MIT פארשער האבן מיט deep learning געפונען חומרים וואס טייטן MRSA.

• **2 יאנואר: Steamboat Willie** – א דאטא-סעט פון יעדן בילד פון Steamboat Willie, איצט וואס Mickey איז ארויס פון קאפירעכט.

• **5 יאנואר: OpenAI פירער און Dropbox** – OpenAI האט ארויסגעגעבן א פירער פאר prompt engineering. Dropbox באניצער האבן מורא אז זייער דאטא ווערט גענוצט צו טריינירן.



OpenAI פירער און Dropbox · 5 יאנואר 2024

דעם פערברואר 2024 זענען ארויסגעקומען א ריי גרויסע מאדעלן. Google האט לאנסירט Gemini Advanced מיט Gemini Ultra 1.0, דערנאך Gemini 1.5 Pro מיט א קאנטעקסט פון ביז 1 מיליאן טאקעס, און די אפענע Gemma מאדעלן. OpenAI האט געוויזן Sora, א מאדעל וואס שאפט ווידעאס אין א שטוינענדיגער קוואליטעט, און האט אנגעהויבן אויספרובירן "זכרון" אין ChatGPT. Stability האט געמעלדן Stable Diffusion 3, און Mistral האט ארויסגעלאזט Mistral Large צוזאמען מיט א שותפות מיט Azure. אויפן געלט-פראנט איז OpenAI געשאצט געווארן איבער 80 ביליאן דאלאר, איר יערליכע איינקונפט איז אריבער 2 ביליאן, און Sam Altman רעדט וועגן אויפנעמען ביז 7 טריליאן דאלאר פאר טשיפ-פאבריקן. Nvidia איז געקומען צו א ווערט פון 1.7 טריליאן. אין פארשונג האט מען שטארק געארבעט אויף קלענערע און שפארעוודיגערע מאדעלן, ביז צו מאדעלן מיט בלויז 1.58 ביטס פאר יעדן פאראמעטער.

מאדעלן און נייע לאנסירונגען

- **1 פערברואר:** לייק פון **miqu-1-70b** – אויף Hugging Face און 4chan איז ארויסגעליקט געווארן א מאדעל "miqu-1-70b", וואס קומט נאנט צו GPT-4 און איז פארבונדן מיט Mistral 8x7b's Mixtral. דער הויפט פון Mistral האט באשטעטיגט אז א "באגייסטערטער קונה" האט עס ארויסגעליקט. דער פאל ווייזט ווי שטארק אפן-סאורס AI איז שוין געווארן.
- **1 פערברואר:** **Llava 1.6** און **פארוואנדטע** – Llava 1.6, א מאדעל וואס פארשטייט בילדער און טעקסט, איז בעסער געווארן אין לייענען טעקסט פון בילדער (OCR), אין טראכטן און אין וועלט-וויסן. אויף געוויסע אויפגאבעס איז ער גלייך מיט Gemini, און די גרופע פלאנירט צו פארעפנטלעכן די דאטא, דעם קאוד און דעם מאדעל. FireLlava איז אן אפענער מאדעל וואס מען מעג נוצן פאר מסחר. ער איז גלייך מיט דעם ערשטן Llava, אבער נאך נישט מיט Llava 1.5. Ollama שטיצט שוין "זעענדיגע" מאדעלן, אריינגערעכנט Llava 1.6, סיי אין Python און סיי אין JavaScript.
- **2 פערברואר:** **Bard מאכט בילדער** – Bard איז צוגעגליך אין מער שפראכן און קען יעצט אויך שאפן בילדער.
- **2 פערברואר:** **OLMo פון AI2** – AI2 האט ארויסגעלאזט א אינגאנצן אפענעם מאדעל און סיסטעם פאר טריינירן. מען באקומט די גאנצע טריינינג-דאטא, די וועיטס, דעם קאוד פאר טריינירן און אפשאצן, און איבער 500 צווישן-שטאפלען (checkpoints) פאר יעדן מאדעל. אלעס איז אונטער דער Apache 2.0 לייסענס.
- **2 פערברואר:** **Rufus' Amazon** – Amazon האט לאנסירט Rufus, א AI העלפער פאר איינקויפן אין איר אפליקאציע. ער נוצט Amazon'ס קאטאלאג און אינפארמאציע פון וועב, און איז דערווייל צוגעגליך פאר א טייל באניצער אין אמעריקע.
- **5 פערברואר:** **Hugging Chat Assistants** – Hugging Face האט לאנסירט Hugging Chat Assistants. דערמיט קען מען בחינם שאפן טשעטבאטס אויף אפן-סאורס מאדעלן, אלס קאנקורענץ צו OpenAI'ס Custom GPTs, פאר וועלכע מען דארף צאלן א מאנאטליכן אבאנעמענט.

- **6 פֿעברואַר:** Qwen – Qwen 1.5 האט ארויסגעלאזט א סעריע פון מאדעלן אין אסאך שפראכן, פון 0.5B ביז 72B פאראמעטערס. דער קלענסטער איז דער ערשטער ערנסטער מאדעל מיט אונטער 1B. די מאדעלן זענען טריינירט אויף 3T טאָקענס און קומען אויך אין טשעט-ווערסיעס און אין צוזאמגעפרעסטע ווערסיעס (quants). זיי זענען באזונדערס שטארק אין טראכטן, מאטעמאטיק און קאוד.
- **7 פֿעברואַר:** Microsoft Copilot – Microsoft's Copilot for Sales און Copilot for Service זענען יעצט צוגעגליך פאר אלעמען. זיי פארבינדן זיך מיט CRM סיסטעמען ווי Salesforce און מאכן אויטאמאטיש איבערגעחזר'טע ארבעט. Copilot האט אויך באקומען א נייעם אויסזען, שאפן און רעדאגירן בילדער, און א נייעם מאדעל מיטן נאמען Deucalion.
- **7 פֿעברואַר:** Meta Voice – א קליינער אפן-סאורס מאדעל מיט 1.3B פאראמעטערס, וואס מאכט פון טעקסט רייז און קען אויך נאכמאכן א קול.



Meta Voice · 7 פֿעברואַר 2024

- **8 פֿעברואַר:** Smaug-72B – Abacus AI האט אנטוויקלט Smaug-72B אויפן יסוד פון Alibaba's Qwen-72B. ער שטייט יעצט אויבנאן אויף Hugging Face'ס רשימה פון די בעסטע מאדעלן. אין עטליכע בענטשמארקס איז ער בעסער פון GPT-3.5 און Mistral Medium, מיט א דורכשניט פון איבער 80 אויף די הויפט-טעסטן.
- **9 פֿעברואַר:** Gemini Advanced – Google האט לאנסירט Gemini Advanced, א באצאלטער מאנאטליכער סערוויס וואס גיט צוטריט צו Gemini Ultra 1.0, צוזאמען מיט אן אפליקאציע פארן טעלעפאן. לויט Ethan Mollick'ס אנאליז איז ער אויף דער מדרגה פון GPT-4, און שטארק אין ערקלערן זאכן און אין פארבינדן בילדער מיט זוכן. Google ברענגט אויך די פלענער "Gemini Business" און "Gemini Enterprise" פאר Workspace. די פלענער האבן פארשטארקטע שוץ פאר דאטא, און די שמועסן פון די ארבעטער ווערן נישט גענוצט צו טריינירן.
- **12 פֿעברואַר:** BUD-E פון Laion – Laion האט ארויסגעלאזט אן אפענעם העלפער מיט א נאטירליך קול. ער לויפט אויף שטארק אפטימיזירטע מאדעלן, אזוי אז מען קען אים נוצן אפילו אויף א לעפטאפ פאר שפילן.

- **13 פעברואר: Apple MGIE – Apple** האט לאנסירט MGIE, אן אפן-סאורס מאדעל וואס רעדאגירט בילדער לויט באפעלן אין טעקסט. ער נוצט מולטימאדאלע שפראך-מאדעלן כדי צו פארשטיין די אנווייזונגען, און Apple האט אויך פארעפנטלעכט דעם קאוד.
- **13 פעברואר: Reka's 21B מאדעל** – דער סטארטאפ Reka האט געמעלדן א מאדעל מיט 21B פאראמעטערס, טריינירט פון אנהייב אן, וואס פארשטייט פון יסוד אן אסאך סארטן אינפארמאציע. ער איז בעסער פון Llama 70B און Mixtral אין וויכטיגע בענטשמארקס, און גלייך מיט Gemini Pro. מען קען אים אויספרובירן אויף Reka's פלאטפארמע.
- **14 פעברואר: זכרון אין ChatGPT – OpenAI** פרובירט אויס א מעגליכקייט אז ChatGPT זאל געדענקען פאקטן פון איין שמועס צום אנדערן, כדי די ענטפערס זאלן זיין מער צוגעפאסט צום באניצער. מען קען דאס אויסלעשן.
- **14 פעברואר: Stable Cascade – Stability** האט צוזאמען מיט rstchenüW ארויסגעלאזט Stable Cascade, א מאדעל וואס מאכט בילדער פון טעקסט, דערווייל נאר פאר פארשונג. ער איז בעסער פון SDXL אין שיינקייט, אין פאלגן דעם באפעל און אין דער קוואליטעט פון טעקסט אינעם בילד. ער איז געבויט אין עטליכע שטאפלען, וואס מאכט גרינג צו פאסן אים צו.
- **15 פעברואר: Nvidia Chat with RTX – Nvidia** האט ארויסגעלאזט א דעמא פון Chat with RTX, א טשעטבאט וואס לויפט גאנץ אויפן אייגענעם קאמפיוטער, אן א וואלקן (cloud) און אן אינטערנעט. ער אנאליזירט און פאסט צוזאם פערזענליכע טעקעס אין פארשידענע פארמאטן, און קען אויך אריינברענגען אינהאלט פון YouTube רשימות. דערווייל ארבעט ער נאר אויף Windows מיט געוויסע Nvidia קארטלעך.
- **16 פעברואר: OpenAI Sora – OpenAI** האט געוויזן Sora, א מאדעל וואס שאפט ווידיאס אין א שטוינענדיגער קוואליטעט. דערווייל איז ער נאר צוגענגליך פאר "red teamers", וואס פרובירן אים אויס פאר סכנות. מען האט שוין צונויפגעשטעלט א רשימה פון 30 פארשונג-ארבעטן וואס זענען פארבונדן מיט Sora.
- **16 פעברואר: Meta's V-JEPA – Meta** האט פארעפנטלעכט V-JEPA, מאדעלן וואס זענען טריינירט אויף ווידיא כדי צו פארשטיין געוויסע זייטן פון דער וועלט. לויט Yann LeCun איז דאס א שריט צו מאדעלן וואס לערנען זיך אליין און דארפן ווייניגער דאטא. די מאדעלן זענען צוגענגליך פארן עולם.
- **16 פעברואר: Aya פון Cohere – Cohere** ס'נישט-פאר-פראפיט פארשונג-לאבאראטאריע C4AI האט ארויסגעלאזט Aya, אן אפן-סאורס מאדעל וואס דעקט 101 שפראכן, אריינגערעכנט איבער 50 שפראכן וואס האבן ביז יעצט נישט באקומען גענוג אכטונג.
- **19 פעברואר: Gemini 1.5 Pro – Google** האט ארויסגעלאזט Gemini 1.5 Pro, א MoE מאדעל (געבויט פון עטליכע "מומחים"). ער איז גלייך מיט Gemini 1.0 Ultra, אבער קען האנדלען מיט ביז 1 מיליאן טאקענס אויף איינמאל, און נוצט ווייניגער רעכנ-קראפט ווייל ער איז קלענער. ער פארשטייט פון יסוד אן אסאך סארטן אינפארמאציע.
- **21 פעברואר: GPT-4 ביז דעצעמבער 2023 – OpenAI** לויט OpenAI'ס נייע דאקומענטאציע נעמט די טריינינג-דאטא פון GPT-4 מאדעלן יעצט אריין אינפארמאציע ביז דעצעמבער 2023.

• **21 פעברואר: Amazon TTS BASE** – די AGI גרופע פון Amazon האט געבויט BASE TTS, א מאדעל וואס מאכט פון טעקסט רייד. ער שמועסט אויף א זייער נאטירליכן אופן, און ווייזט לויט די פארשער שפראך-פעאייגקייטן וואס מען האט אים נישט ספעציעל געלערנט ("emergent abilities").

• **22 פעברואר: Gemma פון Google** – Google האט פארעפנטלעכט די וועיטס פון Gemma 2B און 7B אויף Hugging Face. די מאדעלן זענען טריינירט אויף 2T און 6T טאָקענס, און זענען אסאך בעסער פון Llama 2 אין א ברייטן קרייז פון בענטשמארקס. זיי קומען אין א יסוד-ווערסיע און אין א ווערסיע וואס פאלגט אנווייזונגען. Google'ס אינווינירן האבן אויך ארויסגעלאזט א לייכטן C++ מאָטאָר צו לויפן Gemma אויף א CPU, ענליך צו llama.cpp.

• **23 פעברואר: Stable Diffusion 3** – Stability AI האט געמעלדן Stable Diffusion 3, א "Diffusion Transformer" ענליך צו Sora. די מאדעלן גייען פון 800M ביז 8B פאראמעטערס, אסאך גרעסער ווי די פריערדיגע בילד-מאדעלן. מען וועט זיי ארויסלאזן נאך א תקופה פון פארשונג.

• **23 פעברואר: Phind 70B** – Phind'ס מאדעל פאר קאוד מיט 70B פאראמעטערס באקומט 83% אויף HumanEval, גלייך מיט GPT-4. ער איז ווייניגער "פויל" און זאגט ווייניגער אפט אפ צו ענטפערן. דער 34B מאדעל וועט באלד ארויסקומען, און דער 70B שפעטער.

• **27 פעברואר: Mistral Large** – Mistral האט ארויסגעלאזט Mistral Large דורך איר API, און געמעלדן א שותפות מיט Azure. דער מאדעל איז שטארק אין אסאך שפראכן און שטייט בלויז עטליכע פראצענט הינטער GPT-4 אויף MMLU. אינעם API איז אויך צוגעקומען א נייער mistral-small.

• **29 פעברואר: Llama 3 אין יולי** – לויט א באריכט פלאנירט Meta ארויסצולאזן Llama 3 אין יולי. די גרופע ארבעט נאך אז דער מאדעל זאל ענטפערן מער פראגעס וואס זענען אין אמת נישט קיין שטרייט-ענינים.

• **29 פעברואר: Adobe Acrobat AI** – Adobe האט אריינגעשטעלט א AI העלפער אין Acrobat. ער ענטפערט פראגעס איבער דאקומענטן און ווייזט פון וואו די ענטפער קומט, מאכט צוזאמפאסונגען, שלאגט פאר פראגעס, און העלפט שרייבן טעקסט.

פארשונג און טעכניק

• **1 פעברואר: LASER פון Microsoft Research** – Microsoft Research האט פארגעשטעלט LASER, א טעכניק וואס פארבעסערט די גענויקייט פון שפראך-מאדעלן. מען פארבייט א וועיט-מאטריץ מיט א קלענערער צוגענענטערטער ווערסיע.

• **1 פעברואר: Weaver פאר דערציילן** – Weaver איז א סעריע פון מאדעלן טריינירט ספעציעל צו דערציילן מעשיות. דער גרעסטער (34B) איז בעסער פון GPT-4 אויף א בענטשמארק פון דערציילן.

• **1 פעברואר: טעקסט אין בילדער** – GSDM איז א מאדעל וואס ראטעוועט טעקסט אין בילדער וואס זענען צעשטערט. די פארשער האבן אויך צוגעגרייט א נייעם בענטשמארק און צוויי דאטאסעטס, און אנגעוויזן א שטארקע פארבעסערונג. Hi-SAM טיילט צו טעקסט אויף אלע שטאפלען, פון א שטריך ביז א גאנצן פאראגראף, און אנאליזירט אויך דעם אויסשטעל פון א בלאט. CAM פארבעסערט דערקענען טעקסט אין קאמפליצירטע סצענעס.

• **1 פעברואר:** פארשטיין ווידיאו – Norton פארבעסערט דאס פארשטיין פון לאנגע ווידיאו. SNP לערנט ווידיאו און טעקסט צוזאמען אויף א מער שפארעוודיגן אופן. TDViT פארבעסערט אנאליז פון יעדן איינציגן בילד אין א ווידיאו. א באזונדערער מאדעל נוצט "ring attention" מיט א 7B מאדעל כדי צו ענטפערן גענויע פראגעס איבער ווידיאו פון א שעה, מיט א מיליאן טאקעס. ער איז בעסער פון קאמערציעלע מאדעלן.

• **2 פעברואר:** מאדעלן פאר טעלעפאנען – Google's MobileDiffusion מאכט בילדער אויף א טעלעפאן אין ווייניגער ווי א סעקונדע. א פארשונג-ארבעט שטודירט ווי צו מאכן קליינטשיגע שפראך-מאדעלן פאר טעלעפאנען, און MobileVLM V2 ווייזט שטארקע פארבעסערונגען אין מאדעלן וואס פארשטייען בילד און טעקסט אויף א טעלעפאן. Meta's MobileLLM מיט 350M פאראמעטערס קומט נאנט צו Llama 7B אין רופן פונקציעס, אבער ער איז נאך נישט ארויסגעלאזט. MobiLlama (0.5B) איז אינגאנצן אפן-סאורס, און איז אפילו בעסער פון געוויסע גרעסערע מאדעלן.

• **5 פעברואר:** רעקאמענדאציעס – PAP-REC שאפט אויטאמאטיש פערזענליכע באפעלן (prompts) פאר מאדעלן וואס גיבן רעקאמענדאציעס. DreamRec שאפט אן "אידעאלן" נעקסטן חפץ פארן באניצער, אנשטאט צו קלאסיפצירן פון א מישמאש פון ביישפילן. PromptMM פארבעסערט רעקאמענדאציעס אויף פלאטפארמעס ווי Amazon און TikTok דורך "knowledge distillation" פון בילד, טעקסט און קלאנג.

• **5 פעברואר:** פאלשע נייעס און קלאנג – א מאדעל BERT וואס איז ספעציעל צוגעפאסט, דערקענט פאלשע נייעס בעסער ווי GPT-3.5-turbo אזוי ווי ער איז. PAM שאצט אפ די קוואליטעט פון קלאנג און קיין פארגלייך-אפנעמונג און אן ספעציעלן טריינירן.

• **5 פעברואר:** שאפן ווידיאו – AnimateLCM מאכט שנעל ווידיאו פון הויכער קוואליטעט. Boximator לאזט פירן א ווידיאו מיט קעסטלעך און פיילן, אבער די קאנטראל איז נאך נישט פולקום. InteractiveVideo לאזט ענדערן דעם ווידיאו בשעת מען שאפט אים, מיט טעקסט, בילדער, מאלן און שלעפן. SEINE פארלענגערט קורצע קליפס צו לאנגע סצענעס מיט פליסיגע איבערגאנגען. Google's Lumiere נוצט א "space-time UNet". Snap 'ס מאדעל קומט צו דער פריערדיגער שפיץ-מדרגה (אחוץ Sora) און לויפט 3 מאל שנעלער. WoVoGen שאפט ווידיאו פון גאסן פאר אויטאנאמע מאשינען.

• **6 פעברואר:** פעלערן אין מולטימאדאלע מאדעלן – א פארשונג-ארבעט ווייזט אז מאדעלן פון בילד און שפראך באשרייבן בילדער פאלש ("הזיות") צומאל צוליב "semantic shift bias", באזונדערס ביי איבערגאנגען צווישן פאראגראפן. MHaluBench איז א נייער בענטשמארק צו שאצן ווי גוט מען כאפט אזעלכע פעלערן. CorrelationQA ווייזט אז GPT-4V און ענליכע מאדעלן טועה'ן ווען א בילד פירט אין א טעות אדער סותר'ט דעם טעקסט. ChartX טעסט ווי גוט זיי פארשטייען טשארטס. דער Q-Pathway דאטאסעט האט געזאמלט מיינונגען פון 58,000 מענטשן איבער 18,973 בילדער, וועגן קלארקייט, פארבן און ליכטיגקייט.

• **6 פעברואר:** אויסגלייכן מאדעלן (alignment) – KTO נוצט די "prospect theory" פון Kahneman און Tversky, און מאקסימירט נוצן אנשטאט נאר די ווארשיינליכקייט פון א פרעפערענץ. א פארשונג-ארבעט געפינט א שטארקע פשוטע מעטאדע: נעם די 1,000 לענגסטע ביישפילן. אן אנדערע ארבעט ווייזט אז טריינירן מיט א שטערקערן מאדעל ווי GPT-4 קען זיין בעסער ווי די קאמפליצירטע צוויי-שטאפלדיגע מעטאדע (SFT און דערנאך RL). נאך אן ארבעט ווייזט אז די פשוטע REINFORCE איז בעסער פון PPO. Gemini נוצט REINFORCE, און מען גלייבט אז GPT-4 נוצט PPO.

- **7 פֿעברואר: CLIP פארבעסערונגען** – מען קען טריינירן א שטארקן CLIP מאדעל מיט 30 מיליאן בילדער וואס זענען אינגאנצן קינסטליך געשאפן. דאס איז וויכטיג ווייל דאטאסעטס ווי LAION זענען אראפגענומען געווארן, און עס זענען דא ריכטיגע זארגן וועגן קאפירעכט. אן אלטע מעטאדע (GDA) פארבעסערט CLIP אין קלאסיפֿיצירן בילדער אן נאך טריינירן. SpLiCE מאכט CLIP'ס אינעווייניגסטע פארשטענדליך. א נייע מעטאדע שיצט מאדעלן ווי OpenFlamingo און LLaVA פון אטאקעס מיט בילדער, דורכן צופאסן דעם CLIP טייל אן אויפזיכט.
- **8 פֿעברואר: קול-מאדעלן פון Stability AI** – Stability AI האט טריינירט מאדעלן פון רייד וואס מען קען פירן מיט גענויע באפעלן אין טעקסט. ווייל עס איז נישטא קיין גרויסער דאטאסעט מיט באשרייבונגען פון קלאנג, האט מען קינסטליך באשריבן א גרויסן קארפוס פון רייד.
- **8 פֿעברואר: מוזיק** – Google'ס MusicLM גרופע האט גענוצט 300 טויזנט רעאקציעס פון מענטשן פאר RL. דאס רעזולטאט געפעלט מענטשן מער ווי דער יסוד-מאדעל. ChatMusician ווייזט אז מען קען שטארק פארבעסערן דאס פארשטיין פון מוזיק, אן שאדן צו די שפראך-פעאיגקייטן, ווען מען שרייבט די מוזיק אין ABC נאטאציע. Google האט אויך אפגעהאלטן א רעדע איבער "לייענען" מוזיק פונעם מוח.
- **8 פֿעברואר: SELF-DISCOVER** – DeepMind און USC האבן אנטוויקלט SELF-DISCOVER. דער מאדעל קלייבט און באַלייגט אליין "מאדולן" פון טראכטן צו א סטרוקטור פאסיג פאר יעדע פראבלעם, און פארבעסערט די רעזולטאטן ביז 32% ביי מאדעלן ווי GPT-4. עס איז שוין דא אן אומאפיציעלע אימפלעמענטאציע.
- **9 פֿעברואר: FunSearch** – DeepMind'ס FunSearch, באשריבן אין Nature, פארבינדט א שפראך-מאדעל מיט א פראגראם וואס אפשאצט די רעזולטאטן. ער האט געמאכט באווייזבארע אנטדעקונגען אין מאטעמאטיק, ביי דער "cap set" פראבלעם און ביי "bin-packing". די לעזונגען זענען פארשטענדליך פאר מענטשן.
- **9 פֿעברואר: אגענטן און שפילן** – א פארשונג-ארבעט שטעלט אריין אנווייזונגען פון א שפיל אין א "decision transformer", כדי אן אגענט זאל קענען לייענען די כללים און שפילן. Skill Set Optimization זוכט א סעט פון פעאיגקייטן וואס מען קען נוצן אין פארשידענע סביבות. MetaTree איז א decision tree געבויט אויף א טראנספארמער.
- **9 פֿעברואר: ראבאטן** – 1X ווייזט א ראבאט וואס טוט אסאך אויפגאבעס, אינגאנצן געפירט פון ניראלע נעצן וואס שיקן 10 באפעלן א סעקונדע לויטן ווידיאו. "Box o Flows" איז א סיסטעם צו טעסטן RL מיט פליסיגקייטן אין דער אמת'ער וועלט. "Learning to Fly" טריינירט קאנטראל פאר דראָנען אין א סימולאציע, אין סעקונדעס אויף א געוויינליכן לעפטאפ. 3D Diffuser Actor פארבעסערט ווי ראבאטן נוצן חפצים. פארשער פון UCSD האבן טריינירט מענטש-ענליכע ראבאטן צו זיין מער אויסדריקליך, און GeneOH Diffusion פארריכט פעלערן אין ווי הענט באגלייטן זיך מיט חפצים.
- **12 פֿעברואר: מענטשליכע דאטא** – א בלאג-ארטיקל באהאנדלט די ראלע פון מענטשן אין שאפן דאטא: זאמלען פרעפערענצן, אנאטירן און באצייכענען.
- **13 פֿעברואר: WebLinx** – א דאטאסעט פון 100 טויזנט אקציעס אויף וועב אין שמועס-פארעם, צו שטודירן ווי מאדעלן פירן זיך אויפן וועב.
- **14 פֿעברואר: מאטעמאטיק** – DeepSeekMath 7B איז טריינירט אויף א ריזיגער מאסע מאטעמאטישן אינהאלט. א נייע טעכניק לאזט יסוד-מאדעלן אליין אויסקלייבן גוטן מאטעמאטישן אינהאלט.

OpenMathInstruct-1 האט 1.8 מיליאן פראבלעמען מיט לעזונגען, און ברענגט אפן-סאורס מאדעלן צו דער מדרגה פון פארמאכטע. ConceptMath טעסט מאטעמאטיק אויף ענגליש און כינעזיש, צעטיילט לויט באזונדערע באגריפן.

• **14 פעברואר: פריוואטקייט** – א פארשונג-ארבעט שטעלט פאר א סיסטעמאטישן אופן צו שאצן ווי גוט מאדעלן וואס נוצן דרויסנדיגע געצייג שיצן פריוואטע אינפארמאציע. אן אנדערע ארבעט האט געפונען אז פראוון אויסצוגעפינען צי א געוויסע דאטא איז גענוצט געווארן צום טריינירן, זענען מערסטנס נישט בעסער ווי טרעפן אויף גורל.

• **14 פעברואר: איבערזעצן** – GenTranslate נוצט שפראך-מאדעלן צו פארבעסערן איבערזעצונגען, און איז בעסער פון די פירנדיגע מאדעלן. ZeroSwot איבערזעצט רייד, כאטש מען טריינירט בלויז אויף דאטא פון דערקענען רייד. VSP-LLM דערקענט און איבערזעצט רייד פון ליפן אין ווידיאו מיט א שפראך-מאדעל.

• **15 פעברואר: בילד-פארשונג** – א מעטאדע רעדאגירט בילדער לויט טעקסט אן קיין מאסקע. א נייע טעכניק טיילט צו בילדער מיט בלויז פשוטע לאבעלס. Yolo-World דערקענט אין רעאלע צייט וועלכע חפצים מען וויל, אויך אזעלכע וואס זענען נישט באשטימט געווארן פארן טריינירן. CogCoM "באארבעט" דאס בילד בשעת'ן טראכטן. ChatterBox איז שטארק אין שמועסן איבער בילדער. א מעטאדע דערקענט חפצים ווי פאראויסזאגן דעם נעקסטן טאָקען, און FiT שאפט בילדער אין יעדער גרייס און פראפארציע.

• **16 פעברואר: צוזאמפרעסן ווידיאו** – נייע מעטאדעס נוצן "diffusion" מאדעלן צו צוזאמפרעסן ווידיאו. איינע דערגרייכט 0.02 ביטס פאר פיקסל. דער טריק איז צו מעסטן ווי ענליך עס זעט אויס, און שיקן אן ארגינעלן בילד ווען נויטיג.

• **16 פעברואר: Meta טעסטן מיט AI** – Meta האט גענוצט שפראך-מאדעלן צו שרייבן טעסטן פאר איר קאוד, און געפונען א שטארקע פארבעסערונג אין ווי פיל פונעם קאוד ווערט געטעסט און אין דער קוואליטעט פונעם קאוד.

• **16 פעברואר: מער שפארעוודיגע מאדעלן** – SLEB שניידט אויס איבעריגע בלאקן פון א טראנספארמער. RoCo איז א פאליסי פאר דעם "cache" ביי ווייניג רעסורסן. KV Quant פארמינערט זכרון ביים לויפן, און ברענגט ביז 10M טאָקענס. אן אנדערע מעטאדע מאכט MoE מאדעלן קלענער אן פארלירן קוואליטעט. IR-QLoRA פארבעסערט צוזאמגעפרעסטע מאדעלן, און LoRA+ נוצט פארשידענע לערן-שנעלקייטן כדי צו פארבעסערן און פארשנעלערן דאס צופאסן.

• **16 פעברואר: CLIP-MUSED** – א מעטאדע וואס אנטשלוסלט וויזועלע אינפארמאציע פונעם מוח ביי פארשידענע מענטשן.

• **21 פעברואר: טעקסט און embeddings** – Contextual האט טריינירט א מאדעל וואס שאפט סיי טעקסט און סיי embeddings, און איז אסאך בעסער פון באזונדערע ספעציאליסטישע מאדעלן. "Echo embeddings" איבערחזרט דעם טעקסט צוויי מאל, כדי צו נוצן אינפארמאציע פון שפעטערע טאָקענס.

• **21 פעברואר: Mamba מיט Triton** – Sasha Rush האט ארויסגעלאזט א לערנונג מיט ערקלערונגען ווי צו פארשנעלערן Mamba מיט Triton. עס וואקסט נאך נישט צו גרויסע מאסן צוליב א פעלער אינעם Triton קאמפיילער.

• **22 פעברואר: בענטשמארקס** – Nicholas Carlini האט ארויסגעלאזט זיין אייגענעם בענטשמארק, אין וועלכן GPT-4 שטייט נאך מער פאראויס ווי ביי אנדערע. Stanford האט ארויסגעלאזט HELM-Instruct,

- פאר פאלגן אנווייזונגען. GTBench טעסט סטראטעגיע אין שפילן. די מאדעלן זענען גוט אין שפילן פון גורל, אבער שוואך אין שפילן וואס האבן קיין גורל נישט. Berkeley האט א רשימה פאר רופן פונקציעס, INSTRUCTIR איז פאר זוך-מאשינען, TIER שאצט אפ בילדער וואס AI האט געשאפן, Anime Bench פאר פאקטן, און BoCoEL שאצט אפ מאדעלן מיט Bayesian אפטימיזאציע.
- **23 פעברואר: זיכערע מאדעלן** – GradSafe כאפט געפערליכע באפעלן דורכן אנאליזירן "gradients", אן פיל טריינירן. ShieldLM איז א דעטעקטאר פון זיכערהייט וואס מען קען צופאסן, און ער ערקלערט זיינע באשלוסן. א פראיעקט שטעלט פאר מעטאדעס אז אויטאנאמע אגענטן זאלן בלייבן פארלעסליך.
 - **26 פעברואר: Anthropic פארשטיין מאדעלן** – Anthropic'ס פארשער ארבעטן מיט "Circuits", צו געפינען וועלכע טיילן פון א מאדעל ווערן גענוצט פאר געוויסע אויפגאבעס. זיי האבן ארויסגעגעבן א מאנאטליכן באריכט איבער זייערע עקספערימענטן.
 - **26 פעברואר: פראטאיינען און רפואות** – מען לייגט צו "flow matching" צו AlphaFold, וואס פארשטארקט שטארק די מעגליכקייט צו מאדעלירן די גאנצע וועלט פון פראטאיינען. "Graph diffusion policy optimization" פארבעסערט שאפן גרעפס פאר פלאנירן רפואות.
 - **28 פעברואר: MegaScale** – Bytedance'ס MegaScale האט טריינירט א 175B מאדעל אויף 12,288 GPUs, מיט 55.2% MFU (ווי גוט מען נוצט די רעכן-קראפט), א זייער שטארק רעזולטאט. Bytedance פלאנירט צו עפענען טיילן פונעם קאוד.
 - **29 פעברואר: 1-ביט מאדעלן און לאנגע טעקסטן** – א פארשונג-ארבעט שלאגט פאר צו נוצן 1.58 ביטס פאר פאראמעטער (יעדער איז 0, -1, אדער 1), און ווייזט די זעלבע רעזולטאטן ביז 3B פאראמעטערס. קיין קאוד איז נאך נישט ארויס. Dual Chunk Attention לאזט Llama2 70B האנדלען מיט איבער 100 טויזנט טאקענס אן נאך טריינירן.
 - **29 פעברואר: ווידיאו ווי א שפראך** – א פארשונג-ארבעט ווייזט ווי ווידיאו-מאדעלן קענען דינען אלס פלאנירער, אגענטן, רעכן-מאשינען און סימולאטארן פון א סביבה.
 - **29 פעברואר: 3D און מענטשן** – GPAvatar שאפט 3D קעפ פון בילדער. 3D topia מאכט 3D פון טעקסט אין צוויי שטאפלען. ImplicitDeepfake1 פארבינדט NeRFs און Gaussian Splatting פאר 3D אוואטארן. CapHuman שאפט פארשידענע בילדער פון א מענטש פון איין פאטא. PTV3 באארבעט 3D פונקטן שנעלער. א פראיעקט מאכט 3D פון בלויז 4 בילדער, MVDiffusion++ פאריינציגט פארשידענע בליקן אויף א חפץ צו 3D, אן אנדער פראיעקט בויט איבער 3D פון עטליכע בליקן אן צו וויסן פריער די טיפקייט, און FlowMDM שאפט לאנגע באוועגונגען פון א מענטש פון טעקסט.
 - **29 פעברואר: אויטאנאמע מאשינען** – MF-MOS דערקענט חפצים וואס באוועגן זיך מיט LiDAR. EMIFF דערקענט 3D חפצים מיט קאמערעס פון סיי די מאשין און סיי דעם וועג. "Spiking Neural Networks" דערגרייכן הויכע רעזולטאטן מיט ביז 85% ווייניגער ענערגיע.
 - **29 פעברואר: ווייטערע פארשונג** – Lag-Llama איז א יסוד-מאדעל פאר פאראויסזאגן "time series". Video Annotator נעמט אריין מומחים אינעם באצייכענען ווידיאו. ListT5 פארבעסערט די רייע פון זוך-רעזולטאטן, און HiGPT לערנט פון פארשידענארטיגע גרעפס. BirdCast נוצט ראדאר און מאשין-לערנען צו פאראויסזאגן ווען פייגל וועלן וואנדערן. Google האט ארויסגעלאזט Dinosaur, דעם קאוד וואס איז דער יסוד פון DeepMind'ס וועטער-מאדעלן.

געצייג פאר פראגראמירער און אפן-סאורס

- **1 פעברואר: אגענטן און RAG** – עס איז דא אן אימפלעמענטאציע פון MCTS (דער זוך-אלגאריטם פון AlphaGo) אין Jax. uAgents איז א ביבליאטעק פאר אויטאנאמע אגענטן אין Python. GPTNewspaper איז 6 אגענטן וואס בויען צוזאמען א צייטונג. RAGs איז אן אפליקאציע וואס בויט RAG סיסטעמען לויט אן באשרייבונג אין פשוטער שפראך.
- **2 פעברואר: דאטא און embeddings** – Nous Research האט ארויסגעלאזט דעם OpenHermes-2.5 דאטאסעט מיט איבער 1 מיליאן ביישפילן. BGE-M3 שטיצט איבער 100 שפראכן און טעקסטן ביז 8192 טאקענס. Nomic-Embed-Text-V1 איז אינגאנצן אפן, מיט 235 מיליאן פארן טעקסטן.
- **5 פעברואר: עטליכע געצייג** – Aphrodite – א מאָטאָר וואס פארמינערט VRAM און פארגרעסערט די שנעלקייט ביים לויפן מאדעלן. Diffusers 0.26 – ווידיאו און עטליכע IP-adapters אויף איינמאל. ChatLLM – אן אפענע אלטערנאטיווע צו Copilot אין VS Code, מיט שטיצע פאר Ollama. א פראיעקט מאכט מוזיק פון א בילד.
- **6 פעברואר: עטליכע פראיעקטן** – מען בויט אן אפן-סאורס ווערסיע פון Rabbit. NaturalSQL פון ChatDB איבערזעצט שפראך צו Meta's RL ביבליאטעק Pearls, וואס איז שוין אין גאנג ביי ליציטאציעס און רעקאמענדאציעס, איז גרייט פאר נוצן, און עס איז דא א לערנונג דערצו.
- **7 פעברואר: עטליכע ארויסגאבעס** – Bria AI האט געעפנט איר מאדעל צו אראפנעמען דעם הינטערגרונט פון בילדער (נישט פאר מסחר). Laxta זענען מאדעלן פאר באסקיש, אויף Llama 2. Fabric איז א סיסטעם צו פארשטארקן מענטשן מיט AI. א Chrome צוגאב פילטערט סאציאלע מעדיע מיט א לאקאלן מאדעל.
- **9 פעברואר: אפשאצן און צופאסן** – Hugging Face האט ארויסגעלאזט lichteval פאר אפשאצן מאדעלן. א "קאכבוך" ווייזט ווי צו בויען א RAG אויפן אייגענעם קאמפיוטער מיט Ollama און pgvector. HelixML ערקלערט ווי זיי האבן הצלחה געהאט צו צופאסן Mistral-7B, מיט פראגעס און ענטפערס פון פארשידענע בליקן.
- **12 פעברואר: ראמען** – Llmware איז א ראם פאר שפראך-מאדעל אפליקאציעס. SPIN לאזט א מאדעל שאפן זיין אייגענע טריינינג-דאטא, אבער דער קאוד איז שווער צו נוצן. Phidata איז פאר בויען העלפער וואס רופן פונקציעס.
- **13 פעברואר: Axolotl און packing** – Axolotl שטיצט יעצט Mac M טשיפס, און עס איז דא א Colab ווערסיע. א פראיעקט פאקט ריכטיג צוזאם ביישפילן ביים טריינירן אויף Hugging Face, כדי זיי זאלן זיך נישט "פארמישן".
- **14 פעברואר: עטליכע געצייג** – ZLUDA לאזט לויפן CUDA אויף AMD טשיפס (אלפא). Design2Code מאכט פון סקיצעס און Figma אן SGLang HTML/CSS/JS איז א שפראך פאר סטרוקטורירטע ענטפערס. Reor איז א נאטיץ-אפליקאציע מיט לאקאלע מאדעלן.
- **15 פעברואר: MaGNET פון Meta** – Meta האט ארויסגעלאזט אן אפן-סאורס מאדעל פאר קלאנג און מוזיק. ער איז 7 מאל שנעלער פון אנדערע, אן פארלירן קוואליטעט.

- **16 פעברואר:** קליינע פראיעקטן – Neural Flow צייכנט אויס די רעזולטאטן פון יעדן שיכט פון Mistral 7B. GPTScript איז א שפראך פאר אויטאמאטישע שמועסן מיט OpenAI מאדעלן. Andrej Karpathy האט ארויסגעלאזט minbpe, א ריינע אימפלעמענטאציע פון "byte pair encoding".
- **20 פעברואר:** טריינירן און באגלייטן – ControlVectors זענען ביליגער ווי LoRA צו פירן א מאדעל. OpenRLHF מיט Ray איז פאר מאדעלן אין Llama סטיל. א ארטיקל ערקלערט ווי צו צופאסן LR און WD אין AdamW. OpenLLMetry-JS גיט פולע איבערזיכט פון אפליקאציעס. עס איז דא א רשימה פון GPU קלאסטערס צו דינגען פאר שעה.
- **22 פעברואר:** LoRAs און אגענטן – Predibase האט ארויסגעלאזט 25 LoRAs אויף Mistral 7B, וואס יעדער איינער איז בעסער פון GPT-4 אויף זיין אויפגאבע. Lorax איז א סערווער וואס לאזט וועקסלען LoRAs אויף איין יסוד-מאדעל. Meta האט ארויסגעלאזט MMCSG, שמועסן רעקארדירט מיט Aria ברילן. CrewAI איז א ביבליאטעק צו פירן אגענטן מיט LangChain.
- **23 פעברואר:** מעדיצין און אפשאצן – MMedLM 2 איז טריינירט אויף 25.5 ביליאן טאקעס אין זעקס שפראכן, און איז בעסער פון GPT-4 אין מעדיצינישע פראגעס. Mozilla ערקלערט איר ראלע אלס אפשאצער אין דער NeurIPS געיעג פאר שפארעוודיגקייט. Parea ערקלערט מאסן פאר אפשאצן מאדעלן, מיט Python.
- **26 פעברואר:** פאראייניגטע מאדעלן – OpenCodeInterpreter, געבויט אויף CodeLlama און DeepSeek Coder, האט באקומען איבער 85% אויף HumanEval. AlphaMonarch איז א צונויפגעשמאלצענער מאדעל פאר לאנגע שמועסן. FuseChat פאראייניגט די שטארקייטן פון עטליכע מאדעלן. Simulatrex איז פאר סימולאציעס מיט אגענטן. עס איז דא א רשימה פון פארשונג-ארבעטן איבער באצייכענען מיט מאדעלן.
- **29 פעברואר:** StarCoder v2 און מער – Big Code האט ארויסגעלאזט StarCoder v2, מיט א קאנטעקסט פון 16k און טריינירט אויף 4T טאקעס. ער איז שטארק, אבער נאך אונטער DeepSeek coder. Distilabel האלט זאמלען און רייניגן דאטא פאר RLHF און DPO. Modal שטיצט websockets פאר "streaming". Character AI ניצט א סטיל פון שרייבן קאוד וואס מאכט די פארמען פון טענזארן קלארער, און עס איז דא א קורצער קאוד מיט TRL צו מאכן א מאדעל מער באשרייבנדיג.

פירמעס, געלט און ביזנעס

- **1 פעברואר:** ServiceNow – ServiceNow (150 ביליאן מארקעט ווערט) מעלדט אז אירע AI פראדוקטן האבן געברענגט דעם גרעסטן נייעם איינקונפט פון אלע איר נייע פראדוקטן אין א ערשטן קווארטאל.
- **2 פעברואר:** a16z רשימה – a16z האט צונויפגעשטעלט 50 סטארטאפס וואס נוצן AI אין פארטיידיגונג, ענערגיע, טראנספארט און פאבריקאציע.
- **5 פעברואר:** קורצע ידיעות – Sakana האט באקומען צוטריט צו א יאפאנעזישן סופער-קאמפיוטער. The Browser Company בויט אן אגענט אין Arc וואס זוכט אליין אויפן וועב. Banana מאכט צו איר serverless GPU ביזנעס.
- **6 פעברואר:** Roblox איבערזעצער – Roblox האט א מאדעל וואס איבערזעצט שמועסן אין רעאלע צייט צווישן 16 שפראכן, מיט א MoE סטרוקטור.

- **8 פעברואר: AI מכשירים** – Rabbit R1 פאר 199 דאלאר האט באקומען פאר 10 מיליאן דאלאר פאר-באשטעלונגען, א טייל דאנק דעם דיזיין מיט Teenage Engineering. Brilliant Labs פון סינגאפור האט באקומען אינוועסטירונג פון דעם שאפער פון Pokemon GO פאר Frame, ברילן מיט א AI העלפער. Shortwave האט ארויסגעלאזט א העלפער פאר אימעילס אויפן טעלעפאן.
- **12 פעברואר: OpenAI איינקונפט** – OpenAI איז אריבער 2 ביליאן דאלאר יערליכע איינקונפט, און האפט עס מער ווי צו פארטאפלען אין 2025.
- **12 פעברואר: AI קאמפיוטערס** – מען רעכנט אז היי-יאר וועלן ארויסגעשיקט ווערן כמעט 50 מיליאן AI קאמפיוטערס, און ביז 2027 נאנט צו 60% פון אלע קאמפיוטערס.
- **14 פעברואר: Altman'ס 7 טריליאן** – Sam Altman פירט שמועסן מיט אינוועסטארן, ווי פאנדן פון רעגירונגען אין UAE, צו אויפנעמען ביז 7 טריליאן דאלאר פאר טשיפ-פאבריקן. דער פלאן קען אויפרייצן זארגן וועגן גיאפאליטיק, און ער זוכט יעצט א הסכמה פון וואשינגטאן.
- **14 פעברואר: אייגענע טשיפס** – Meta וועט היי-יאר אריינשטעלן אייגענע AI טשיפס אין דאטא צענטערס, צו ווייניגער אנגעוויזן זיין אויף Nvidia. Microsoft בויט אויך אייגענע סערווערס.
- **14 פעברואר: YouTube און Google** – YouTube'ס הויפט האט באשריבן 4 וועטעס פאר 2024. די באניצונג פון קאנאל-מעמבערשיפס איז געוואקסן מיט Google. 50% האט אנגעהויבן א פראגראם פון 25 מיליאן אייראָ פאר לערנען AI אין אייראפע.
- **15 פעברואר: Karpathy און Nvidia** – Andrej Karpathy פארלאזט OpenAI כדי צו ארבעטן אויף אייגענע פראיעקטן, אן קיין דראמע. Nvidia'ס ווערט איז דערגרייכט 1.7 טריליאן דאלאר, אזוי פיל ווי אלע כינעזישע פירמעס אויף דער האנג קאנג בערזע.
- **16 פעברואר: געלט-רונדעס** – Lambda Labs האט אויפגענומען 320 מיליאן דאלאר פאר א GPU וואלקן. Upstash האט אויפגענומען 10 מיליאן פון a16z, און לאנסירט א serverless וועקטאר דאטאבאזע. Sam Altman איז דער איינציגער באזיצער פון OpenAI'ס אינוועסטירונג-פאנד.
- **19 פעברואר: OpenAI אגענטן און זוכן** – OpenAI אנטוויקלט לויט באריכטן צוויי סארטן אגענטן: איינער נעמט איבער דעם מכשיר פונעם באניצער, און דער אנדערער טוט ארבעט אויפן וועב. זי ארבעט אויך אויף א זוך-סערוויס צו קאנקורירן מיט Google.
- **20 פעברואר: OpenAI 80 ביליאן, Magic** – OpenAI איז געשאצט איבער 80 ביליאן דאלאר אין א דיעל מיט Thrive Capital, כמעט 3 מאל מער אין ווייניגער ווי 10 חדשים. דער דיעל לאזט ארבעטער פארקויפן זייערע שערס. Magic, וואס בויט א AI פראגראמירער, האט אויפגענומען 117 מיליאן.
- **21 פעברואר: SoftBank און Scribe** – SoftBank'ס Masayoshi Son וויל אויפנעמען 100 ביליאן דאלאר פאר Izanagi, א טשיפ-פירמע צו קאנקורירן מיט Nvidia. Scribe האט אויפגענומען 25 מיליאן. מער ווי א מיליאן גרופעס נוצן עס, אריינגערעכנט 97% פון Fortune 100.
- **22 פעברואר: Reddit און Apple** – Reddit האט א הסכם פון בערך 60 מיליאן דאלאר א יאר מיט א גרויסער AI פירמע, וועמענס נאמען האט נישט אויסגעזאגט, פאר צוטריט צו אירע אינהאלטן. Apple אנטוויקלט א AI געצייג אין Xcode צו שרייבן קאוד.
- **23 פעברואר: Sierra און Marqo** – Bret Taylor'ס פירמע Sierra בויט AI אגענטן פאר באדינען קונים. Marqo האט אויפגענומען 12.5 מיליאן פאר וועקטאר-זוכן.

- **27 פֿעברואר:** Datology AI – SF Compute האט אויפגענומען 11 מיליאן פאר אויסקלייבן און רייניגן דאטא. San Francisco Compute באט אן די ביליגסטע H100 קלאסטערס פאר קורצע צייט. דער ערשטער, Angel Island, ארבעט שוין.
- **28 פֿעברואר:** Apple – Apple מאכט צו איר פראיעקט פון א זעלבסט-פארנדיגן אויטא, נאך א יארצענדליג און ביליאנען דאלארן, און ארבעטער גייען איבער צו AI. Tim Cook זאגט אז Apple וועט אין 2024 "ברעכן נייע ערד" אין AI.

רעגירונגען, געזעצן און זיכערהייט

- **2 פֿעברואר:** IBM בלאקטשעיין – IBM Consulting און Casper Labs בויען א געצייג אויף בלאקטשעיין צו באאויפזיכטיגן AI. עס וועט האבן ווערסיעס, אויפזיכט און קאנטראל ווער עס מעג נוצן די מאדעלן.
- **6 פֿעברואר:** OpenAI "דעמאקראטיש" – א טרעף מיט Colin Megill, וועגן זיין פלאטפארמע Polis, האט אנגעצונדן OpenAI'ס פראגראם "Democratic Inputs to AI". דער פראגראם האט געגעבן 1 מיליאן דאלאר אין גרענטס, צו אנטדעקן ווי דער עולם קען מיטערעדן אין פירן AI.
- **7 פֿעברואר:** סימנים אויף AI בילדער – OpenAI לייגט אריין C2PA סימנים אין DALL-E 3 בילדער: אומזעבארע מעטאדאטא און א סימבאל CR וואס מען זעט. Meta וועט באצייכענען בילדער פון AI אויף אירע פלאטפארמעס, אפילו ווען קאנקורענטן האבן זיי געשאפן.
- **8 פֿעברואר:** EU AI Act – דאס אייראפעאישע AI געזעץ האט דורכגעגאנגען א וויכטיגע שטימונג פון די פארשטייער פון די מדינות, וועלכע האבן באשטעטיגט דעם ענדגילטיגן טעקסט.
- **8 פֿעברואר:** קינדער-זיכערהייט – OpenAI האט געגרינדעט א גרופע צו פארמיידן אז קינדער זאלן מיסברויכן אירע געצייג.
- **15 פֿעברואר:** מדינה-שפילער – OpenAI האט געפונען און פארמאכט קאנטעס פון גרופעס פארבונדן מיט מדינות, וואס האבן גענוצט אירע מאדעלן פאר שלעכטע צוועקן.
- **16 פֿעברואר:** קיין פאטענטן פאר AI – דער אמעריקאנער פאטענט אפיס זאגט אז נאר מענטשן וואס שטייערן באדייטנד צו צו אן אויפטו קענען זיין "ממציאים", נישט AI.
- **19 פֿעברואר:** Anthropic און וואלן – Anthropic פרובירט אויס Prompt Shield, וואס שיקט באניצער פון Claude אין אמעריקע וואס זוכן אינפארמאציע וועגן וואלן צו פארלעסליכע מקורות ווי TurboVote.
- **26 פֿעברואר:** Gemini בילדער – Google האט מודה געווען אז Gemini האט געשאפן היסטארישע בילדער וואס זענען נישט געווען ריכטיג, צוליב פראבלעמען אין דער טריינינג-דאטא. זי וועט עס פארבעסערן, אבער האט זיך נישט אינגאנצן אנטשולדיגט.
- **27 פֿעברואר:** Scale AI – Pentagon האט באקומען א קאנטראקט פאר א יאר צו בויען א סיסטעם צו טעסטן שפראך-מאדעלן פארן Pentagon, מיט ספעציעלע דאטאסעטס.

אנאליזן און מיינונגען

- **1 פֿעברואר:** ארבעט מיט AI – Ethan Mollick שרייבט אז AI קען יעצט טון אסאך אויפגאבעס אין ווייניגער ווי א מינוט, און פירמעס דארפן זיך באמיען אז מען זאל עס נוצן פאר גוטס. אן אנדערער ארטיקל טענה'ט אז מען איז נישט פארשפעטיגט מיט AI, אזוי ווי ביים אנהייב פון אינטערנעט.

- **6 פברואר: ארבעט-מארקעט** – א פארשונג-ארבעט פון MIT און IBM געפינט אז נאר 23% פון אויפגאבעס וואס פאדערן זען, לוינט זיך צו אויטאמאטיזירן. די פארשער זאגן פאראויס א שטופנווייזע השפעה אויף ארבעט.
- **7 פברואר: עקאנאמיע** – אן ארטיקל שאצט ווי GPT-5 קען ענדערן מארקעטן, רעאל-עסטעיט, ענערגיע און גיאפאליטיק. אן אנדערער זאגט אז AI וועט מאכן פראצעסן ביליגער, און דערמיט עפענען ביזנעסער וואס האבן פריער נישט געלוינט. א דריטער טענה'ט אז די אינדוסטריע גייט צו קלענערע, שפארעוודיגע מאדעלן ווי Phi 2, כדי צו ווערן פראפיטאבל.
- **9 פברואר: אינוועסטירן** – אן "angel" אינוועסטאר צעטיילט AI אין שיכטן: אינפראסטרוקטור, מאדעלן און פראדוקטן. אן ארטיקל זאגט אז אנגענטן ווערן א נייער וועג צו פארקויפן, און פירמעס וועלן דארפן צופאסן זייערע פראדוקטן פאר AI. אן אנדערער ארטיקל ווארנט אז רוב סטארטאפס וועלן פארלירן קעגן גרויסע פירמעס וואס האבן שוין דאטא און סיסטעמען.
- **12 פברואר: קונסט און שאפן** – USV שרייבט אז AI וועט ענדערן ווי מען שאפט קונסט. אן אנדערער ארטיקל זאגט אז סטארטאפס וואס זענען געבוירן מיט AI וועלן האבן א גרויסן פארטייל.
- **13 פברואר: ארבעט און מעדל-קלאס** – א ארטיקל שלאגט פאר "Massed Muddler Intelligence", א סיסטעם פון צעשפרייטע אנגענטן אנשטאט איין ריזיגן מאדעל. Noema טענה'ט, אנטקעגן Elon Musk און Geoffrey Hinton, אז ווייל עס פעלן ארבעטער, קען AI העלפן אויפבויען די מיטל-קלאס.
- **14 פברואר: LeCun אינטערוויי** – Yann LeCun רעדט מיט Time איבער AGI, אפן-סאורס און סכנות.
- **15 פברואר: טשעט איבער דאטא** – מען שרייבט אז צו בויען טשעטבאטס איבער אייגענע דאטא איז שווערער ווי מען מיינט, ווייל עס פאדערט גוטע זוכן און קאמפליצירטע פראגעס.
- **19 פברואר: VC און דאטא** – איבער 75% פון האנדל אין בערזעס ווערט געפירט דורך אלגאריטמען. לויט פאראויסזאגן וועלן איבער 75% פון VC איבערקוקן נוצן AI ביז 2025.
- **20 פברואר: אינעווייניג OpenAI** – Logan Kilpatrick ערקלערט OpenAI'ס קולטור פון שנעלקייט און אחריות, און ווי זי באשליסט צו לאנסירן פראדוקטן.
- **21 פברואר: פארשנעלערטע צוקונפט** – Ethan Mollick שרייבט אז Gemini'ס מיליאן טאקענס און Groq'ס כמעט מידיגע ענטפערס ווייזן א גרויסן שפּרונג. פירער מוזן שנעל לערנען זיך צופאסן.
- **22 פברואר: AI וועט פארבייטן** – אן ארטיקל טענה'ט אז דער געדאנק אז AI וועט נאר העלפן ארבעטער און קיינמאל נישט פארבייטן איז נישט אמת. ער פארגלייכט מיט וואס טעכנאלאגיע האט געטון צו פאבריקן און אגריקולטור.
- **26 פברואר: Altman'ס חשבון** – Astral Codex Ten ערקלערט אז די 7 טריליאן ווייזן ווי שנעל די קאסטן וואקסן פאר יעדע נייע ווערסיע פון GPT, און פרעגט צי די וועלט איז גרייט פאר אזא שנעלקייט.
- **28 פברואר: צען לעקציעס** – גרינדער פון Databricks, Anyscale און Microsoft טיילן זייערע מיינונגען איבער AGI, שפראך-מאדעלן און ווי גרינדער זאלן נוצן AI אין זייערע פראדוקטן.

ווייטערע ידיעות

- **1 פעברואר: געצייג פאר ביזנעס** – Brainner – קוקט איבער קורות-חיים. AI Client Finder – געפינט קאנטאקטן פון קונים. SellScale – גיט באמערקונגען אויף פארקויף-אימעילס. Matrices – א שפרעדשיט וואס פילט זיך אליין. Weave – ארבעט-פראצעסן אן קאוד. OpenWidget – א ChatGPT ווידזשעט פאר וועבזייטלעך.
- **12 פעברואר: Neuralink** – לויט Elon Musk האט Neuralink דורכגעפירט איר ערשטע אפעראציע צו אריינשטעלן א מכשיר אין מוח.
- **14 פעברואר: נאך געצייג** – Galileo AI – שאפט דיזיינס פאר אפליקאציעס. AI Form Roast – פארבעסערט פארמען אויפן וועב. PressPulse – ברענגט דערמאנונגען אין פרעסע. Autotab – לערנט טון ארבעט אין בראוזער. Unspam – העלפט אז אימעילס זאלן נישט פאלן אין ספאם. AdGen AI – שאפט איבער 100 ווערסיעס פון א רעקלאמע פון איין לינק. Persuva – שרייבט רעקלאמעס. Mindy – א "Chief of Staff" דורך אימעיל. BookPecker – 14,509 ספרים צוזאמגעפאסט אין 5 פונקטן יעדער.

מעריץ 2024 איז געווען א חודש פון גרויסע מאדעל-לאנסירונגען: Anthropic האט ארויסגעלאזט די Claude 3 פאמיליע וואס ווערט פארגליכן מיט xAI, GPT-4, האט געעפנט Grok-1 מיט 314 ביליאן פאראמעטערס, און Databricks און AI21 האבן ארויסגעגעבן DBRX און Elon Musk האט געקלאגט OpenAI, און OpenAI האט געענטפערט מיט אלטע אימעילס. אין ביזנעס האט Microsoft אריינגענומען די גרינדער פון Inflection, Amazon האט פארענדיגט א 4 ביליאן דאלאר אינוועסטירונג אין Anthropic, און ראבאטיק-פירמעס ווי Figure און Physical Intelligence האבן אויפגענומען גרויסע סומעס. דאס אייראפעאישע פארלאמענט האט באשטעטיגט דעם ערשטן גרויסן AI געזעץ, און Cognition האט באוויזן Devin, אן AI ווייכווארג-אינושענער.

מאדעלן און נייע לאנסירונגען

- **1 מעריץ: Google DeepMind's Genie – Genie** איז א וועלט-מאדעל וואס איז טריינירט געווארן אויף אינטערנעט-ווידיעס. ער קען שאפן אן אומענדליכע צאל וועלטן וואו מען קען שפילן און קאנטראלירן די אקציעס, פון סינטעטישע בילדער, פאטאגראפיעס און אפילו פון סקיצן.
- **1 מעריץ: Mist Rime – Rime** האט לאנסירט א נייעם מאדעל וואס פארוואנדלט טעקסט אין רייז. ער קען גוט שמועסן, און קען אריינלייגן "אום" און נאטירליכע פויזעס, אנדערש ווי פריערדיגע מאדעלן.
- **5 מעריץ: Anthropic's Claude 3** Anthropic האט טריינירט דריי נייע מאדעלן אין דער Claude 3 פאמיליע. דער שטארקסטער פון זיי גלייכט זיך מיט GPT-4'ס געמאלדענע בענטשמארק-רעזולטאטן. די מאדעלן ארבעטן אויך מיט בילדער, און קענען שרייבן קאוד אסאך בעסער ווי פריער. Claude 3 Opus האט באוויזן צו פארוואנדלען א לאנג ווידעא פון Andrej Karpathy אין א שיינעם בלאג-ארטיקל, מיט עפעס הילף צו צוגרייטן די דאטא. א ארטיקל אין every.to באשרייבט Claude 3 אלס דעם "מערסט מענטשליכן" AI אין שעפערישע ארבעט. שפעטער איז ארויסגעקומען Claude 3 Haiku, דער שנעלסטער און ביליגסטער מאדעל פאר פירמעס, און Anthropic האט ארויסגעגעבן א ביבליאטעק פון פראמפטס וואס ארבעטן גוט מיט די מאדעלן.

Prompting best practices

2024 מערץ 5 · Claude 3 Anthropic

- **5 מערץ:** ChatGPT לייענט הויך – OpenAI האט צוגעלייגט א "Read Aloud" אפציע. ChatGPT קען יעצט ענטפערן מיט א שטימע אין 37 שפראכן, מיט פינף סארטן שטימען, סיי אויפן וועב און סיי אויף מאביל.
- **5 מערץ:** Adobe מוזיק-געצייג – Adobe's Project Music GenAI Control שאפט מוזיק און לאזט זי אנפאסן, לויט א טעקסט-באשרייבונג אדער לויט א ניגון וואס מען גיט אים אלס ביישפיל.
- **6 מערץ:** Stable Diffusion 3 – די פארשונג-ארבעט פון Stable Diffusion 3 באשרייבט א נייע ארכיטעקטור, Multimodal Diffusion Transformer, וואס האט באזונדערע וואגן פאר טעקסט און פאר בילדער. דערמיט פארשטייט דער מאדעל בעסער דעם פראמפט און שרייבט בעסער אותיות. SD3-Turbo פארקלענערט די צאל דיפוזיע-טריט ביז 4, און די קוואליטעט בלייבט די זעלבע.
- **6 מערץ:** Mistral AI – Le Chat האט לאנסירט "le Chat Mistral", א טשעטבאט וואס רעדט אסאך שפראכן, מיט א באזונדערע ווערסיע פאר פירמעס.
- **7 מערץ:** Moondream 2 – א קליינער מאדעל וואס פארשטייט בילדער, געבויט אויף Phi-2 און SigLIP. די נייע ווערסיע האט א קאמערציעלע ליצענץ און בעסערע בענטשמארקס, און פאסט פאר מאשינען מיט ווייניג כח.
- **8 מערץ:** Inflection 2.5 – Inflection's נייער מאדעל פאר זיין אסיסטענט Pi ערגרייכט 94% פון GPT-4. די פירמע זאגט אז ער האט געדארפט נאר 40% פונעם רעכנ-כח צו טריינירן. א דורכשניטליכער שמועס מיט Pi דויערט 33 מינוט.
- **12 מערץ:** xAI עפנט Grok – Elon Musk האט אנגעזאגט אז Grok וועט ווערן אפן-סאורס, און xAI האט טאקע ארויסגעגעבן די וואגן פון Grok-1, א Mixture-of-Experts מאדעל מיט 314 ביליאן פאראמעטערס, געשריבן אין JAX. Musk האט אויך געזאגט אז אלע Premium באניצער אויף X וועלן באקומען צוטריט צום טשעטבאט.
- **12 מערץ:** Covariant RFM-1 – Covariant האט באוויזן RFM-1, א גרויסער שפראך-מאדעל פאר ראבאטן, געבויט אויף א ריזיגע זאמלונג דאטא פון איר Brain AI פלאטפארמע. דער ציל איז צו פארבעסערן ווי ראבאטן נעמען באשלוסן.

- **12 מערץ:** Cohere's Command-R For AI האט געשאפן א מאדעל מיט מער ווי 30 ביליאן פאראמעטערס, וואס איז שטארק אין לאגיק, צוזאמפאסונגען און ענטפערן אויף פראגעס אין 10 שפראכן. דער מאדעל איז אויסגעשטעלט פאר RAG, און וועט צוקומען צום Nvidia API קאטאלאג.



2024 מערץ 12 · Cohere's Command-R

- **13 מערץ:** Devin's Cognition – Devin האט באוויזן Devin, אן "AI וויכווארג-אינזשעניר". ער האט באקומען 14% אויף SWE-Bench, קעגן 1.7% פאר GPT-4.
- **13 מערץ:** Gemini פאסט אן ענטפערס – אין דער Gemini וועב-אפליקאציע קענען באניצער יעצט אנפאסן ענטפערס מיט אפציעס ווי ווידעו-שאפן, קירצער, לענגער און אויסמעקן.
- **14 מערץ:** DeepMind's SIMA – SIMA איז אן אלגעמיינער AI אגענט פאר 3D שפילן, וואס פאלגט באפעלן אין פשוטער שפראך אין פארשידענע ווידעו-שפילן.
- **14 מערץ:** Sora – OpenAI's CTO זאגט אז Sora וועט ווערן ארויסגעלאזט פאר'ן עולם דאס יאר, נאכן טעסטן און אויפשטעלן זיכערהייט-מאסנאמען. OpenAI ווייזט שוין Sora פאר Hollywood סטודיאס און אגענטורן, און האט פארעפנטלעכט א זאמלונג ווידעאס וואס קינסטלער און פילם-מאכער האבן געשאפן דערמיט.
- **15 מערץ:** Figure 01 רעדט – מיט OpenAI's טעכנאלאגיע קענען Figure's ראבאטן פירן פולע שמועסן. OpenAI's מאדעלן גיבן דאס זען און פארשטיין שפראך, און Figure's נעצן פירן אויס די פייע באוועגונגען.
- **15 מערץ:** Tesla FSD v12.3 – דער אפדעיט, וואס Musk רופט "א גרויסע ארויסגאבע", ניצט end-to-end נעצן צו פירן דעם קאר, אנשטאט קאנטראל וואס פראגראמירער האבן אויסגעשריבן.
- **18 מערץ:** GPT-4.5 Turbo דורכגעלאזט? – א בלאג-ארטיקל פון OpenAI איז ארויס בטעות, איז אינדעקסירט געווארן אין Bing און DuckDuckGo, און דערנאך אראפגענומען געווארן. ער האט אנגעזאגט GPT-4.5 Turbo מיט א קאנטעקסט פון 256 טויזנט טאקענס. OpenAI האט נישט קאמענטירט.
- **19 מערץ:** Apple, Stability, Google – Apple פארשער האבן אנטוויקלט מעטאדן צו טריינירן מאדעלן אויף טעקסט און בילדער צוזאמען, מיט רעזולטאטן אויפן העכסטן שטאפל. Stability AI האט

- ארויסגעלאזט Stable Video 3D, וואס שאפט 3D מאדעלן פון איין בילד. Google's VLOGGER מאכט פון איין פאטא א ווידעא פון א מענטש וואס רעדט און באוועגט זיך.
- **19 מערץ:** GPT-4 Turbo אין Copilot – Microsoft האט צוגעלייגט GPT-4 Turbo צו דער פרייער ווערסיע פון Copilot.
 - **20 מערץ:** IBM און NASA – IBM און NASA האבן אנטוויקלט מאדעלן וואס זענען טריינירט אויף וויסנשאפטליכע ליטעראטור. די מאדעלן זענען אפן-סאורס אויף Hugging Face.
 - **21 מערץ:** Copilot4D's Waabi – א גרונט-מאדעל פאר זעלבסט-פארנדיגע קארס, וואס ניצט LiDAR דאטא צו פארויסזאגן ווי די 3D סביבה וועט זיך טוישן.
 - **22 מערץ:** Japan – Sakana AI – Sakana AI האט לאנסירט מאדעלן פאר יאפאניש, פאר בילדער און פאר שאפן בילדער. זי האט אויך ארויסגעגעבן אן "עוואלוציאנערע מערדזש" מעטאד, וואס לאזט אנטוויקלען גרונט-מאדעלן אן טייערע פאר-טריינירונג.
 - **22 מערץ:** Meta's SceneScript – א מאדעל פון 70 מיליאן פאראמעטערס, וואס מאכט פון בילדער א 3D פארשטעלונג פון א סצענע. ער לויפט שנעל אויפן מאשין אליין, און איז זייער סטאביל.
 - **25 מערץ:** GitHub פאראכטן שוואכקייטן – GitHub האט אנגעהויבן א בעטא פון code-scanning autofix, וואס פארבינדט Copilot מיט CodeQL. דער געצייג טרעפט און פאראכט זיכערהייט-שוואכקייטן אין Python, JavaScript, TypeScript, Java און Python.
 - **25 מערץ:** DeepGram Aura – א API וואס פארוואנדלט טעקסט אין רייד אין ווייניגער ווי 250 מילי-סעקונדעס.
 - **26 מערץ:** Open Interpreter 01 Light – א טראגבארע שטימע-מאשין וואס קאנטראלירט דעם קאמפיוטער אינדערהיים. זי זעט דעם עקראן, ניצט די אפליקאציעס און לערנט זיך נייע זאכן. זי איז אפן-סאורס.
 - **26 מערץ:** Character Voice – Character.AI האט צוגעלייגט שטימען צו די איינס-אויף-איינס שמועסן, אלס ערשטער טריט צו א מולטימאדאלן אינטערפעיס.
 - **28 מערץ:** Databricks's DBRX – Databricks און Mosaic האבן טריינירט א MoE מאדעל מיט 132 ביליאן פאראמעטערס אויף H100 3,000 טשיפס. די וואגן זענען ארויסגעגעבן געווארן, און דער מאדעל איז אויך בנמצא דורך Databricks's API.
 - **29 מערץ:** AI21's Jamba – א מאדעל וואס מישט Mamba מיט MoE שיכטן. ער לויפט 1,600 טאקענס א סעקונדע, מיט א קאנטעקסט פון 128 טויזנט טאקענס, און ערגרייכט 67% אויף MMLU. די וואגן זענען בנמצא.
 - **29 מערץ:** Hume's EVI – Hume האט ארויסגעלאזט Empathic Voice Interface, וואס די פירמע באשרייבט אלס דעם ערשטן שמועס-AI מיט עמאציאנעלע אינטעליגענץ.
 - **29 מערץ:** Google און Meta אין פראדוקטן – Google טעסט AI צוזאמפאסונגען גלייך אין די הויפט-זוכן, ביי א טייל באניצער אין אמעריקע. Meta לייגט צו AI צו די Ray-Ban ברילן, פאר איבערזעצונג און צו דערקענען זאכן, חיות און מאנומענטן. Pixel 8 וועט באקומען Gemini Nano אין א פרי-ווערסיע פאר אנטוויקלער.

פארשונג און טעכניק

- **1 מערץ:** קוואנטיזירונג און שנעלקייט – א שטודיע ווייזט אז קוואנטיזירונג נאכן טריינירן מאכט מאדעלן ווי OPT און LLaMA2 ביליגער אין זכרון און רעכענען. א סטאטישער KV Cache קען צוויי מאל שנעלער מאכן Llama אויף DeepSpeed. Hugging Face שטיצט יעצט 6 ביטס פאר יעדן פאראמעטער, מיט מער ווי דאפלטע שנעלקייט. א נייע ארבעט ווייזט 1-ביט קוואנטיזירונג פון די Linear שיכטן אן פארלירן קוואליטעט, אזוי אז א 70B מאדעל קען לויפן אויף געווענליכע GPUs.
- **1 מערץ:** דער פאראדאקס פון דיפוזיע-דיסטילאציע – א ארטיקל וואס דערקלערט פארוואס מען שטרעבט צו שאפן א בילד אין איין טריט, כאטש דיפוזיע איז אויפגעבויט אויף טיילן דעם פראבלעם אין אסאך קליינע טריט. ער ברענגט אויך די פארשידענע וועגן פון פארשונג אויף דעם.
- **4 מערץ:** לאנגע קאנטעקסט – מען גלייבט, און עס איז אומאפיציעל באשטעטיגט, אז Gemini'ס קאנטעקסט פון מער ווי א מיליאן טאקענס ניצט HyperAttention. Resonance RoPE העלפט מאדעלן ארבעטן מיט לענגערע טעקסטן ווי זיי זענען טריינירט געווארן, אן נאך רעכן-כח.
- **4 מערץ:** אגענטן – GPT-4 פירט אויס נאר 14% פון אגענט-אויפגאבעס, וואו מענטשן געלינגט עס איבער 70% פון מאל. לויט דער ארבעט וועט מען מעגליך דארפן באזונדער טריינירטע מאדעלן.
- **4 מערץ:** טעאריע פון טריינירן – א ארבעט דערקלערט פארוואס MuP איבערטראגט היפער-פאראמעטערס: דער גרעסטער eigenvalue פונעם Hessian הענגט נישט אפ אין דער ברייט און טיפקייט פונעם נעץ. א צווייטע ארבעט ווייזט אז איין self-attention שיכט צעטיילט דעם פראבלעם אין "שווערע צוריקנעמען" און "ווייכע צוזאמשטעלן". א צאל עקספערימענטן מיט 16 טאקעניזערס ווייזט אז די גרייס פונעם ווערטערבוך מאכט נישט אויס אזוי פיל ביי גרויסע מאדעלן.
- **5 מערץ:** GPQA – א בענטשמארק מיט 448 שווערע פראגעס אין ביאלאגיע, פיזיק און כעמיע. עקספערטן באקומען בערך 65%, נישט-עקספערטן 34%, און GPT-4 נאר 39%.
- **7 מערץ:** מאטעמאטיק מיט סינטעטישע דאטא – MathScale האט געשאפן 2 מיליאן סינטעטישע פראבלעמען און דערמיט טריינירט א 7B מאדעל וואס שטייט זיך גוט קעגן די בעסטע מאדעלן. Microsoft האט ארויסגעגעבן דעם Orca דאטאסעט פון מאטעמאטישע ווערטער-פראבלעמען.
- **7 מערץ:** Design2Code – א בענטשמארק און א 18B מאדעל וואס מאכן קאוד פון בילדער פון דיזיין. לויט דער ארבעט האלט מען שוין נאענט ביי פשוטע דיזייןס, און אמאל האט מען בעסער ליב GPT-4V'ס קאוד ווי קאוד וואס מענטשן האבן געשריבן.
- **7 מערץ:** פלאנירן – KEPP ניצט א פראבאביליסטישן וויסן-גראף צו סדר'ן אקציעס ביז'ן ציל. KnowAgent לייגט צו קלארע וויסן וועגן אקציעס, כדי מאדעלן זאלן בעסער פלאנירן.
- **8 מערץ:** tinyBenchmarks – מען קען פארלעסליך אפשאצן א מאדעל מיט נאר 100 ביישפילן, אנשטאט בענטשמארקס מיט איבער 14 טויזנט.
- **8 מערץ:** טריינירן אויף גרויס – א לערנונג וועגן די סארטן פאראלעליזם. Meta און IBM האבן ערגרייכט 57% MFU מיט PyTorch FSDP. Databricks Mosaic דערציילט וועגן טריינירן מיט fp8.
- **8 מערץ:** מאדעלן ארבעטן צוזאמען – א מעטאד וואו עטליכע מאדעלן שרייבן אויף איבערבייט, טאקען נאך טאקען, און דאס פארבעסערט די רעזולטאטן.

- **11 מערץ: Yi - 01** AI האט ארויסגעגעבן א ארבעט מיט אסאך פרטים ווי זי האט געזאמלט דאטא און טריינירט די Yi מאדעלן.
- **11 מערץ: זיכערהייט און פראמפטס** – מען ניצט איבערזעצונג צו פארמינערן שעדליכן אינהאלט אין שפראכן וואו עס פעלט דאטא. מעטאהעויריסטיק, מער ווי 100 אפטימיזאציע-מעטאדן, פארבעסערן פראמפטס.
- **13 מערץ: גנב'ענען א מאדעל** – מען קען גנב'ענען א טייל פון א פארמאכטן מאדעל, למשל די embeddings שיכט, נאר דורך זיין פובליק API, פאר ווייניגער ווי 2,000 דאלאר.
- **13 מערץ: Mamba און SSM** – Caduceus איז א Mamba מאדעל פאר DNA. VideoMamba פארשטייט ווידעא. SSM מאדעלן שאפן לענגערע ווידעאס מיט לינעארן זכרון. DiS ניצט SSM אנשטאט U-Net פאר בילדער. VMRNN פארבינדט Mamba מיט LSTM פאר פארויסזאגן.
- **15 מערץ: מישן עקספערטן** – Branch-Train-MiX ווייזט אז מען קען באזונדער טריינירן מאדעלן און זיי צונויפגיסן אין איין MoE. MoE אדאפטערס לאזן זיך ווייטער לערנען און פארגעסן וואס מען האט שוין געלערנט.
- **15 מערץ: LiveCodeBench** – א וועג צו מעסטן קאוד-מאדעלן און "פארשמוצונג", ווייל א טייל אפענע מאדעלן זענען צו פיל צוגעפאסט צו HumanEval.
- **15 מערץ: "Evil Geniuses"** – א שטודיע געפינט אז AI אנענטן שטייען שוואכער קעגן שלעכטע אטאקעס, און אז עס איז שווערער צו כאפן זייערע אומפאסיגע ענטפערס.
- **18 מערץ: Cappy** – Google'ס קליינער מאדעל וואס גיט א ציון ווי גוט אן ענטפער פאלגט די אנווייזונגען. ער איז בעסער ווי אסאך גרעסערע מאדעלן.
- **20 מערץ: RAG 2.0** – Contextual AI שרייבט אז "פארפרוירענע RAG" ארבעט שוואך אויסער דעמאס, און שלאגט פאר אז מען זאל אנפאסן די גאנצע מערכת צו דער אויפגאבע.
- **20 מערץ: Quiet-STaR** – א מעטאד וואס גיט מאדעלן א "אינערליכן מאנאלאג", און פארבעסערט שטארק Mistral 7B אויף לאגיק-טעסטן. דער קאוד איז ארויסגעגעבן געווארן.
- **20 מערץ: ווי Sora ארבעט** – א ארטיקל פון א געוועזענעם OpenAI וויסנשאפטלער וועגן Sora, און ווי פיל עס קען קאסטן צו שאפן ווידעאס.
- **21 מערץ: מאטריצן-מולטיפליקאציע** – פארשער פון Tsinghua און UC Berkeley האבן געמאכט דעם גרעסטן פארשריט אין פארמינערן די קאמפלעקסיטעט פון מאטריצן-מולטיפליקאציע.
- **22 מערץ: PuzzleVQA** – א דאטאסעט וואס טעסט אבסטראקטע לאגיק ביי מאדעלן ווי GPT-4V.
- **28 מערץ: בינארע וועקטארן** – אז מען פארבייט fp32 צאלן מיט 0 אדער 1, און מען ניצט KNN און א reranker, קען מען פארמינערן דעם זכרון 30 מאל און פארלירן קוואליטעט.
- **29 מערץ: לאנגע פאקטן** – Google האט ארויסגעגעבן א דאטאסעט און בענטשמארקס פאר פאקטישע גענויקייט. מאדעלן האבן געוואונען קעגן מענטשליכע אנאטאטארן אין רוב פעלער.
- **1 מערץ: ווייטערע פארשונג אין בילד און ווידעא** – UniVS: איינהייטליכע ווידעא-סעגמענטאציע. CFLD: בילדער פון מענטשן לויט פאזע. ליכטיגע 3D מאפע-נעצן. Drive-WM: ווידעאס צו פארויסזאגן סצענארן פאר זעלבסט-פארנדיגע קארס. CricaVPR: דערקענען ערטער. DSNIE: פלאכן-נארמאלן. ASMV2:

באציאונגען צווישן זאכן אין א בילד. SURE: פארלעסליכע אומזיכערקייט-פארויסזאגונגען. Vision-RWKV: בילדער אין הויכע רעזאלוציע. SSD: טיפקייט אין פינסטערניש און רעגן. VideoElevator: בעסערע טעקסט-צו-ווידעא. 3D אביעקטן פון איין בילד. MoAI: צוגעלייגטע וויזועלע אינפארמאציע. DreamDA: דאטא-פארמערונג. Chain-of-Spot: קאנצענטרירן זיך אויף וויכטיגע געגנטן. StableVITON: אנפרובירן קליידער. OVAM: אטענשאן-מאפעס פאר יעדעס ווארט. HETAL: פריוואטקייט דורך האמאמארפישע ענקריפציע. שאטנס, דערקענען מענטשן פון דראנס, בילדער אין פלעקסיבלע גרייסן, UFineBench, אולטראסאונד, SegRefiner, AniPortrait, א איבערבליק פון דיפפעי-דערקענונג, און CoDA פאר צופאסן מאדעלן צו נייע סביבות.

געצייג פאר פראגראמירער און אפן-סאורס

- **1 מערק:** Modular MAX - Modular MAX'ס, וואס פארפשוטערט און פארשנעלערט AI אינסטאלאציעס, איז יעצט בנמצא.
- **1 מערק:** דאטאסעטן פאר אנווייזונגען - Bonito, אן אפן-סאורס מאדעל, פארוואנדלט טעקסט אן אנאטאציעס אין טריינינג-דאטא. Genstruct'ס Nous Research שאפט אנווייזונגען פון טעקסט.
- **4 מערק:** טריינירן 70B אינדערהיים - Unsloth האט שנעלער געמאכט דאס צופאסן פון Gemma מיט 2.4 מאל. Answer AI האט ארויסגעגעבן אן אפן-סאורס געצייג מיט FSDP און QLoRA, וואס לאזט טריינירן 70B מאדעלן אויף געווענליכע גיימינג GPUs. Axolotl ווייזט ווי אזוי צו טריינירן Mixtral אויף אזעלכע מאשינען.
- **5 מערק:** Open Sora Plan - א פראיעקט וואס וויל איבערמאכן Sora, און זוכט אפן-סאורס מיטארבעטער.
- **6 מערק:** MobileClip - לויפט 7 מאל שנעלער ווי CLIP מיט דער זעלבער קוואליטעט, און קען ארבעטן אויפן מאשין אליין.
- **7 מערק:** Hugging Face, MLX, EasyLM - א געצייג פון Hugging Face צו גרופירן טעקסטן, וואס פארארבעט 100 טויזנט ביישפילן אין 10 מינוט. Apple'ס MLX Swift, ביישפילן, אריינגערעכנט טריינירן אן MNIST מאדעל אויף אן iPhone. EasyLM פאר JAX.
- **13 מערק:** Transformer Debugger - א געצייג פון OpenAI'ס Superalignment גרופע צו פארשן ווי קליינע מאדעלן פירן זיך. ער ניצט sparse autoencoders.
- **18 מערק:** 900 אפן-סאורס געצייג - א אנאליז פון GitHub שטערן ווייזט אז רוב אפן-סאורס AI געצייג זענען אינפראסטרוקטור און אפליקאציעס.
- **22 מערק:** TorchTune און Mistral - Meta האט ארויסגעגעבן TorchTune, א PyTorch ביבליאטעק צו צופאסן מאדעלן. Mistral האט אנגעזאגט v0.2 פון איר 7B מאדעל, מיט ריינעם קאוד צו צופאסן אים.
- **26 מערק:** Cerebras CS3 - Cerebras'ס נייער טשיפ, א גאנצער וועיפער, קען טריינירן מאדעלן מיט 24 טריליאן פאראמעטערס, און שטיצט PyTorch.
- **1 מערק:** עטליכע נייע געצייג - ווידעא-פארשטאנד מיט LLMs. מאדעלן צולייגן צו Ollama. 3D בילדער (GLB). Smooth Diffusion. Emissary: בענטשמארקס פאר פראמפטס. בילד-רעסטאוראציע. Fructose: טייפירטע אינטערפעיסן. אטענשאן-אימפלעמענטאציעס מיט ביז 10 מאל אונטערשייד אין

שנעלקייט. ViewFusion. SplattingAvatar. Face2Diffusion. AICI. TripoSR: 3D פון איין בילד. superopenai: לאגס און קעש. Skyvern: אויטאמאטיזירן דעם בראווער. רירענקינג. Lamini. Clarity Upscaler. ViT-CoMer. AMD GPUs. א דאטאסעט פון פראגעס און ענטפערס פון kapa.ai. MusicLang. earnings calls. פאר דאקומענטאציע. LLM4Decompile. LlamaGym. Triton Puzzles. FRESCO. GCTMs. Fed3DGS. CLIP מעדיצינישע אנאמאליעס. Codel. PSALM. Optimal Transport. AdaIR. Street Fighter. OPTIN. AID/PAID. Superpipe. CT-RATE/CT- דורך בענטשמארקס. Saliency DETR. CLIP. backtracing. MRA. 200 מיליאן OCR זייטן ביכער פונעם Internet Archive.

פירמעס, געלט און ביזנעס

- **1 מערץ:** Figure – Figure האט אויפגענומען 675 מיליאן דאלאר, אריינגערעכנט פון OpenAI, מער ווי 150 מיליאן העכער ווי זי האט געזוכט. זי האט אויך געשלאסן אן אפמאך מיט OpenAI אויף מאדעלן פאר אירע ראבאטן.
- **4 מערץ:** ראבאטיק – MagicLab האט באוויזן MagicBot, וואס טאנצט, לייגט צוזאם קליידער און קען מאכן סאלטאס. Amazon'ס פאנד פון 1 ביליאן דאלאר וועט מער אינוועסטירן אין פירמעס וואס פארבינדן AI מיט ראבאטן.
- **4 מערץ:** Cohere – Cohere האט געעפנט אן אפיס אין New York, און ארבעט צוזאמען מיט Accenture צו ברענגען אירע embeddings צו מער ווי 9,000 פירמעס.
- **6 מערץ:** Perplexity – Perplexity רעדט וועגן א געלט-רונדע ביי א ווערט העכער ווי 520 מיליאן דאלאר, און האט אריינגעלייגט Yelp'ס דאטא אין אירע ענטפערס.
- **6 מערץ:** Reka – Yi Tay באשרייבט ווי שווער עס איז צו טריינירן מאדעלן אויסער Google, צוליב אומסטאבילע קלאסטערס און ווייכווארג וואס איז נאך נישט דערוואקסן.
- **7 מערץ:** Mistral – Mistral Large קומט צו Microsoft Snowflake Cortex. האט אינוועסטירט 15 מיליאן אייראס אין Mistral, אלס טייל פון איר סטראטעגיע צו מאכן Azure א "גארטן" פון מאדעלן.
- **7 מערץ:** Waymo אין Austin – Waymo טעסט אירע קארס אן א שאפער. די ערשטע פארן זענען פאר אירע ארבעטער.
- **7 מערץ:** Stack Overflow און Brave – OverflowAPI ברענגט Stack Overflow'ס וויסן צו Gemini אויף Google Cloud. Leo איז יעצט בנמצא אויף Android.
- **8 מערץ:** Microsoft – Copilot פאר פינאנץ אין Excel און Outlook. Surface Pro 10 און Laptop 6 מיט NPUs, צו ווערן באוויזן דעם 21טן מערץ. א AI אונטערנעמונג דעם 20טן מאי מיט Satya Nadella.
- **11 מערץ:** טשיפס – Nvidia שטעלט אפ די GTX קארטלעך און גייט אריבער נאר צו RTX. Google Tensor G4 קומט אין Pixel 9.
- **11 מערץ:** OpenAI'ס באורד – דריי נייע מיטגלידער, און Sam Altman קומט צוריק אין באורד.
- **13 מערץ:** Meta'ס 24 טויזנט H100 – Meta באשרייבט די אינפראסטרוקטור פאר טריינירן Llama 3, אריינגערעכנט סטארעדזש, נעץ-פארבינדונגען און PyTorch.

- **13 מערץ: Physical Intelligence** – די פירמע איז ארויס פון באהעלטעניש מיט 70 מיליאן דאלאר seed געלט, צו בויען גרונט-מאדעלן פאר ראבאטן.
- **14 מערץ: געלט-רונדעס** – Applied Intuition: 250 מיליאן ביי א ווערט פון 6 ביליאן, א שפרונג פון 67%. Together AI: 106 מיליאן, און וואקסט 3 מאל יעדן חודש. Assort Health: 3.5 מיליאן. NLX: 15 מיליאן.
- **15 מערץ: OpenAI און צייטונגען** – אפמאכן מיט Le Monde און Prisa Media צו ברענגען זייער אינהאלט אין ChatGPT.
- **20 מערץ: Microsoft AI** – Mustafa Suleyman און n Simonyan פארלאזן Inflection צו גרינדן Microsoft AI, און Sean White ווערט דער נייער CEO. א ארטיקל באהאנדלט די זארג פון קאנצענטראציע אין די הענט פון ריזן-פירמעס.
- **20 מערץ: Lilac** – Databricks האט אויפגעקויפט Lilac.
- **21 מערץ: Stability AI** – Robin Rombach פארלאזט די פירמע, און Emad Mostaque האט רעזיגנירט אלס CEO צו ארבעטן אויף דעצענטראליזירטן AI. דער COO און דער CTO פירן יעצט צוזאמען.
- **22 מערץ: Apple** – Apple רעדט מיט Google וועגן Gemini אין iPhone, און באטראכט אויך ChatGPT און Baidu. לויט באריכטן וועט iOS 18 נישט האבן אן אייגענעם טשעטבאט. אויף WWDC קען קומען אן AI App Store.
- **26 מערץ: Anyscale און Nvidia** – א שותפות וואס פארבינדט Ray מיט Triton און TensorRT-LLM.
- **27 מערץ: GPT Store** – דער GPT Store פילט זיך אן מיט ספאם און מיט GPTs וואס מעגליך פארלעצן קאפירעכט. OpenAI טעסט טיילן פארדינסטן מיט א קליינע גרופע בויער אין אמעריקע.
- **28 מערץ: Amazon אין Anthropic** – Amazon האט צוגעלייגט נאך 2.75 ביליאן דאלאר, און דערמיט פארענדיגט א 4 ביליאן דאלאר אינוועסטירונג, איר גרעסטע ביז יעצט.
- **28 מערץ: MLPerf** – Nvidia'ס GPUs, בעיקר H200, האבן געפירט אין די MLPerf בענטשמארקס.
- **5 מערץ: קליינע אפליקאציעס** – PitchLeague: א קאוטש פאר פיטש-דעקס. Superhuman: אימעיל מיט AI. Kopia: אנפרובירן קליידער. JARS: פראגראמען מיט פריינט. Toolify: א קאטאלאג פון AI געצייג. Creatie: פראדוקט-דיזיין. Coframe: A/B טעסטן פון בילדער. Summarizeing: צוזאמפאסונגען פון YouTube. Copy.ai האט דערגרייכט 10 מיליאן באניצער אין פיר יאר.

רעגירונגען, געזעצן און זיכערהייט

- **4 מערץ: Musk קעגן OpenAI** – Elon Musk קלאגט OpenAI, און טענה'ט אז GPT-4 איז שוין AGI און דערפאר נישט אונטער'ן אפמאך מיט Microsoft. OpenAI האט פארעפנטלעכט אימעילס וואס ווייזן אז Musk האט געהאלטן אז OpenAI'ס שאנס צו געלינגען איז 0%, און אז ער האט מסכים געווען צו פארמאכן געוויסע מאדעלן. Gary Marcus רופט עס OpenAI'ס "אייגענער גאל".
- **6 מערץ: ליצענצן פאר AI** – א ארטיקל טענה'ט אז מען קען נישט רעגולירן AI ווי א פיזישע זאך, נאר מען דארף ציל'ן אויף ספעציפישע שלעכטע באנוצן.
- **8 מערץ: SEC** – די SEC אויספארשט צי Sam Altman האט פארפירט אינוועסטירער.

- **12 מערץ:** סייבער-זיכערהייט – האקערס האבן פארדינט 50 טויזנט דאלאר ביי Google's bugSWAT. פארשער האבן געשאפן Morris II, א "וואָרעם" וואס פארשפרייט זיך צווישן AI סיסטעמען. א side-channel אטאקע קען לייענען פארשלאסענע שמועסן מיט AI אסיסטענטן, אחוץ Gemini.
- **12 מערץ:** קאפירעכט – Microsoft בעט צו פארווארפן די קלאגע פון New York Times. Midjourney האט פארבאטן אלע Stability AI ארבעטער, און זאגט אז זיי האבן געפרוואווט אראפשלעפן אירע דאטא. א געריכט אין Beijing האט פסק'נט אז א בילד וואס AI האט געשאפן איז באשיצט מיט קאפירעכט.
- **13 מערץ:** Obama – Barack Obama האט גערעדט וועגן רעגולירן AI, און גערופן טעכנאלאגיע-עקספערטן צו גיין ארבעטן אין רעגירונג.
- **14 מערץ:** Google קעגן ספאם – Google's מערץ אפדעיט וועט ווייזן ווייניגער רעזולטאטן וואס זענען געשריבן פאר זוך-מאשינען אנשטאט פאר מענטשן.
- **15 מערץ:** אייראפעאישער AI געזעץ – דאס אייראפעאישע פארלאמענט האט באשטעטיגט דעם ערשטן פולן AI געזעץ, וואס צעטיילט AI לויט סכנה-שטאפלען. דאס געזעץ וועט אריין אין קראפט סטאפנווייז פון 2025.
- **22 מערץ:** Intel – אמעריקע'ס Department of Commerce שלאגט פאר צו געבן Intel ביז 8.5 ביליאן דאלאר אונטער'ן CHIPS Act. Intel פלאנט צו אינוועסטירן 100 ביליאן דאלאר.
- **28 מערץ:** דער קאמף צו קאנטראלירן AI – Politico שרייבט אז לאנד-פירער און טעכנאלאגיע-שעפן זענען צעטיילט, און עס איז נישטא קיין הסכמה ווי צו רעגולירן AI.

אנאליזן און מיינונגען

- **5 מערץ:** Stratechery און סינטעטישע דאטא – Stratechery שרייבט אז AI וואס גיט איין ענטפער שטעלט אין סכנה דאס ביזנעס-מאדעל פון אגרעגאטארן ווי Google און Meta. א צווייטער ארטיקל שאצט אז קוואליטעט-דאטא וועט אויסגיין ביז 2027, און זעט סינטעטישע דאטא אלס א לעזונג.
- **6 מערץ:** Sora און די צוקונפט – Vanity Fair שרייבט ווי Sora קען איבערקערן Hollywood. איין ארטיקל זעט אין Sora "איין טריט פון דער Matrix", און א צווייטער ווייזט אויף אירע חסרונות. Stratechery זעט אין Sora און Groq א טריט צו ווירטועלע וועלטן אין עכטע צייט.
- **7 מערץ:** Waymo פארברענט – נאכדעם וואס מען האט אונטערגעצונדן א Waymo קאר אין San Francisco, טענה'ט Newcomer אז זעלבסט-פארנדיגע קארס קענען שטארק פארמינערן טויטע אויף די וועגן.
- **8 מערץ:** סטראטעגיע פאר סטארטאפס – NFX פארגלייכט AI מיט וואסער אין פלעשלעך, וואו דער אונטערשייד קומט פון מארקעטינג און באניצער-דערפארונג. א ארטיקל אנאליזירט 70 פירמעס וואס האבן אויפגענומען איבער 7 מיליאן דאלאר. A Smart Bear שרייבט אז די גרויסע פירמעס האבן א מעלה קעגן סטארטאפס. Glean's CEO רעדט וועגן ווערט פאר פירמעס. Tom Tunguz זאגט פאר אז AI SaaS פירמעס וועלן זיין מער פראפיטאבל.
- **11 מערץ:** דער GPT-4 שלאקבוים – Simon Willison שרייבט אז פיר מאדעלן האבן ערגרייכט GPT-4, אבער קיינער פון זיי איז נישטא אפן אדער דורכזיכטיג וועגן זיין טריינינג-דאטא.

- **12 מערץ: קריטיק** – Rachel פון fast.ai טענה'ט אז "AI וועט היילן ראק" פארשטייט נישט סיי AI און סיי מעדיצין. א צווייטער מחבר שלאגט פאר צו ניצן ווערטער ווי "שעדליך" און "פעאיג" אנשטאט "גוט" און "שלעכט". א דריטער רופט צו קעמפן קעגן AI אין שעפערשקייט.
- **14 מערץ: אינטעליגענץ-טעסטן** – GPT-4 און Gemini Advanced פאלן דורך וויזועלע IQ טעסטן. Strange Loop Canon טענה'ט אז Claude-3 פארמאגט א "ספעציעלע אינטעליגענץ" און נישט קיין אלגעמיינע. Tomas Pueyo דערמאנט אז ביי Klarna האט AI פארביטן 700 ארבעטער.
- **19 מערץ: פראמפט-אינווענירינג** – לויט IEEE Spectrum פארמאגן פראמפטס וואס אן אלגאריטם שאפט אמאל א מאדנע פארעם, און זיי זענען בעסער ווי פראמפטס פון מענטשן. Ethan Mollick באטאנט אז מען דארף צולייגן קאנטעקסט און chain-of-thought.
- **20 מערץ: סקעפטיציזם און רעקלאמעס** – א סעריע ארטיקלען טענה'ט אז AI וואס פילטערט רעקלאמעס קען שעדיגן א 1 טריליאן דאלאר אינדוסטריע. Mind Prison טענה'ט אז מען קומט צו צו די גרענעצן פון וואקסן. Gary Marcus בלייבט ביי זיין מיינונג אז נאר מער דאטא און מער רעכנ-כח וועט נישט ברענגען AGI.
- **14 מערץ: פראקטישע אנאליזן** – a16z איבער 40% פון די טאפ וועב-פראדוקטן זענען נייע. א ארטיקל וועגן נייע UX מוסטערן. דאס "ליכט-גיכקייט" לימיט פון אינפערענס ווערט באשטימט דורך די ברייט פונעם זכרון. Python, Go און Rust אין פראדוקציע. The Verge: AI זוך-מאשינען קענען נישט ערזעצן Google. צופאסן GPT-4 גיט 3 מאל בעסערע רעזולטאטן פאר 1.5 מאל דעם פרייז. Simon Willison שרייבט C ערווייטערונגען פאר SQLite מיט ChatGPT.
- **29 מערץ: Kurzweil און ארבעט** – Ray Kurzweil רעדט וועגן עקספאנענציעלן וואוקס און "longevity escape velocity". Noahpinion טענה'ט אז מענטשן קענען בלייבן מיט גוט-באצאלטע ארבעט, צוליב גרענעצן אין רעכנ-כח, אבער דאס איז נישט קיין גאראנטיע.

ווייטערע ידיעות

- **12 מערץ: פויזע-ענערגיע** – Princeton פארשער האבן טריינירט אן AI וואס פארויסזאגט און פארמיידט אומסטאבילקייט אין פלאזמע אין א טאקאמאק, דורך אנפאסן מאגנעטישע פעלדער אין עכטע צייט.
- **13 מערץ: פאעזיע** – אין א פרובע האבן מענטשן געשריבן בעסער ווי Bard, ChatGPT-4 און Claude. ענגליש-עקספערטן האבן בעסער דערקענט וועלכע לידער AI האט געשריבן.
- **13 מערץ: צוטריט פאר באהינדערטע** – Microsoft האט באוויזן "Speak For Me" פאר מענטשן מיט רייד-שוועריגקייטן, און נייע Azure AI מעגליכקייטן.
- **15 מערץ: Curtis Priem** – Nvidia'ס מיט-גרינדער האט געגעבן 275 מיליאן דאלאר צו RPI, און האט אוועקגעגעבן זיינע אקציעס.
- **19 מערץ: KAIST טשיפ** – דער C-Transformer טשיפ ניצט אסאך ווייניגער ענערגיע ווי Nvidia'ס A100.
- **20 מערץ: Fitbit און Minecraft** – Fitbit בויט א Gemini מאדעל פאר פערזענליכע עצות וועגן געזונט. MineDreamer פאלגט קאמפליצירטע באפעלן אין Minecraft.
- **21 מערץ: פאזערס** – Claude קען שאפן עפעקטיווע פאזערס פאר פארסינג-קאוד. אויך: א לערנונג ווי אזוי צו שאפן פיקסל-קונסט אין JavaScript.

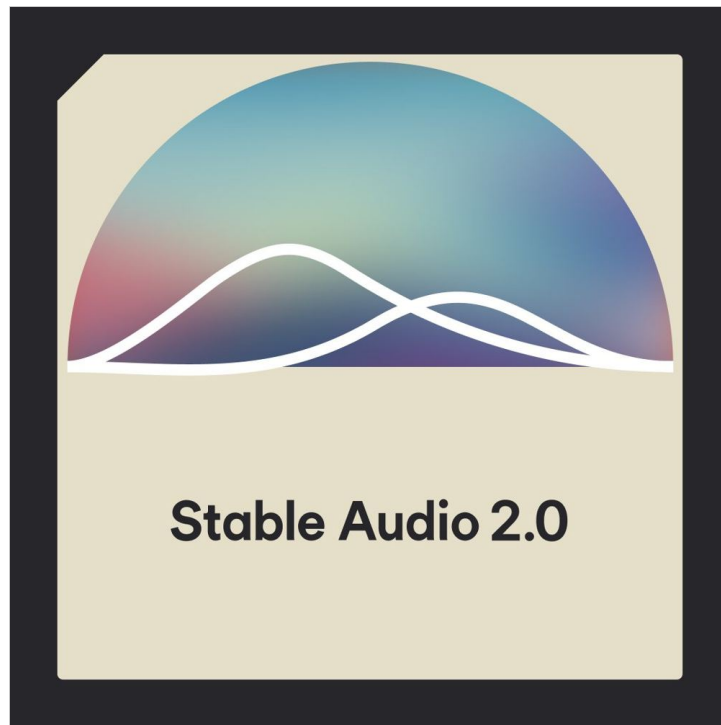
- **21 מערץ: GPT-5** – לויט Indian Express קען GPT-5 ארויסקומען אין זומער.
- **25 מערץ: מעדיצין** – וויסנשאפטלער האבן געשאפן נייע אנטיקערפער מיט AI. CLASSIX פון Manchester האט פארארבעט 5.7 מיליאן קאויד-סעקווענצן אין 1 ביז 2 טעג אויף א געווענליכן לעפטאפ. Insilico Medicine'ס קאנדידאט גייט פאראויס אין קלינישע פרובעס.
- **26 מערץ: באנען אין China – AI** האט פארמינערט קליינע פעלערן אויף די שינעס פון China'ס שנעל-באנען מיט 80%.

אפריל 2024 איז געווען א חודש פון גרויסע מאדעל-לאנסירונגען, ספעציעל Llama 3 פון Meta (8B און 70B, מיט א 400B מאדעל נאך אין טריינירן), Mixtral 8x22B פון Mistral, Phi 3, Grok-1.5 און די עפענטלעכע פרוויז פון Gemini 1.5 Pro. אין דער סאמע צייט איז די אפן-סאורס וועלט שטארק געוואקסן מיט אסאך קליינע מאדעלן און געצייג, און Claude 3 Opus איז צום ערשטן מאל ארויף איבער GPT-4 אויפן Chatbot Arena. אויפן ביזנעס-פראנט האבן זיך געשאפן נייע טשיפס פון Intel, Google און Meta, Nvidia האט אפגעקויפט Run:ai פאר בערך 700 מיליאן דאלאר, און Microsoft מיט OpenAI פלאנירן לויט באריכטן א סופער-קאמפיוטער פון 100 ביליאן דאלאר. פארשער האבן אויך פיל געארבעט אויף לאנגע קאנטעקסטן, אויף שנעלערע אינפערענץ און אויף זיכערהייט פון מאדעלן.

מאדעלן און נייע לאנסירונגען

- **1 אפריל: Grok-1.5 פון xAI – xAI** האט אנאנסירט Grok-1.5 מיט א קאנטעקסט פון 128k און בעסערע ריזאנירונג. ער איז שטארק אין רעטירינג און פראגראמירן, און ווייזט קאנקורענטפעקע בענטשמארק-רעזולטאטן קעגן GPT-4 אין מאטעמאטיק און קאוד. דער מאדעל איז צוגענגליך פאר Premium+ באניצער פון X, און עס איז פלאנירט אים צו ברענגען אויך צו געווענליכע Premium באניצער. שפעטער איז ארויסגעקומען Grok-1.5 Vision Preview, וואס האט לויט xAI ראיה-פעיקייטן גלייך מיט די בעסטע מאדעלן, אין געוויסע פעלער אפילו בעסער.
- **1 אפריל: Claude 3 איבער GPT-4** פון Anthropic איז צום ערשטן מאל ארויף איבער GPT-4 אויפן Chatbot Arena, א טאבעלע וואס ווערט געפירט דורך LMSYS און ווערט אויסגערעכנט פון באניצער-שאצונגען. דער ארטיקל באמערקט אבער אז GPT-4 איז שוין מער ווי א יאר אלט.
- **1 אפריל: Qwen מאדעלן – Qwen MoE** ווייזט די לייסטונג פון א שטארקן 7B מאדעל, אבער מיט נאר א דריטל אקטיווירטע פאראמעטערס. דאס Qwen טים האט אויך ארויסגעלאזט Qwen1.5-32B, וואס פאסט אריין אין סיסטעמען מיט בעסקידענעם זכרון. דערצו איז געקומען Code Qwen 1.5, א סעריע 7B מאדעלן געטריינירט אויף 3T טאָקענס קאוד-מאטעריאל, מיט גוטע רעזולטאטן אויף HumanEval און לאנגן קאנטעקסט ביז 64k.
- **2 אפריל: ReALM'ס Apple און Ferret-UI** – פארשער ביי Apple האבן אנטוויקלט ReALM, א סיסטעם וואס העלפט Siri פארשטיין קאנטעקסט פון דעם וואס שטייט אויפן עקראן, פונעם שמועס און פון הינטערגרונט-זאכן. לויט בענטשמארקס איז עס בעסער ווי ChatGPT 4.0. אן אנדער מולטימאדאל מאדעל, Ferret-UI, קען פארשטיין די באניצער-אינטערפייסעס פון מאביל-עקראנען.
- **2 אפריל: OpenAI'ס Voice Engine – Voice Engine** פון OpenAI קען נאכמאכן א מענטש'ס שטימע פון בלויז א 15-סעקונדיקן אויפנאם. מען קען עס ניצן פאר לערן-הילף, איבערזעצונגען און הילף פאר מענטשן וואס קענען נישט רעדן. צוליב דער סכנה פון מיסברויך גייט OpenAI אבער פארזיכטיג מיטן ארויסלאזן.
- **2 אפריל: ChatGPT אן א חשבון** – מען קען יעצט ניצן ChatGPT אן עפענען א חשבון. פאר אזעלכע באניצער זענען דא נאך שוץ-מיטלען, און זיי קענען אפזאגן אז זייערע דאטא זאל ניצן ווערן פאר טריינירן.

- **4 אפריל:** Stability AI – **Stable Audio 2.0** האט לאנסירט Stable Audio 2.0, געטריינירט אויף ריכטיג לייסענסירטע מוזיק. עס קען שאפן ביז 3 מינוט הויך-קוואליטעט מוזיק, און עס האט אויך audio-to-audio. שפעטער איז ארויסגעקומען א פארשונג-ארבעט וואס באשרייבט דעם diffusion transformer מאדעל.



2024 אפריל 4 · Stable Audio 2.0

- **4 אפריל:** **Opera** מיט **לאקאלע מאדעלן** – דער בלעטערער Opera לאזט יעצט אראפלאדן און לויפן מאדעלן לאקאל אויפן קאמפיוטער, מיט איבער 150 מאדעלן פון מער ווי 50 פאמיליעס.
- **5 אפריל:** **Diffuse** פון **Higgsfield AI**, געגרינדעט פון א געוועזענעם עקזעקוטיוו ביי Snap, האט לאנסירט Diffuse. דאס איז א מאביל-אפליקאציע צו שאפן און רעדאגירן ווידעאס, געצילט פאר שאפער און מארקעטער אויף סאשעל מידיא.
- **8 אפריל:** **Coherer's Command R+ און Rerank 3** – Cohere האט ארויסגעלאזט Command R+, א מאדעל פאר ביזנעסער מיט RAG מיט ציטאטן, 10 הויפט-שפראכן און ניצן פון געצייג. נאכדעם איז געקומען Rerank 3, א מאדעל צו פארבעסערן ביזנעס-זוכן און RAG. ער ארבעט מיט יעדע דאטאבאזע, און קען פארבעסערן זוך-רעזולטאטן אדער אראפברענגען די קאסטן כמעט אן פארשפעטיקונג.
- **8 אפריל:** **DALL-E רעדאגירן אין ChatGPT** – באניצער קענען יעצט רעדאגירן DALL-E בילדער גלייך אין ChatGPT, אויפן וועב, iOS און Android.
- **10 אפריל:** **נייע Gemma מאדעלן** – Google האט געטריינירט CodeGemma און RecurrentGemma, מיט קאנקורענטפעיקער לייסטונג און FIM פעיקייטן. דער רעקורענטער מאדעל איז אסאך שנעלער און ניצט ווייניגער זכרון.
- **10 אפריל:** **Llama 3 פון Meta** – Meta האט באשטעטיגט אז Llama 3 קומט, און עס זענען געווען באריכטן אז OpenAI און Meta גרייטן מאדעלן מיט בעסערע ריזאנירונג און פלאנירונג. די טעכנאלאגיע-וועלט איז אבער געבליבן סקעפטיש. אום 19טן אפריל איז ארויסגעקומען Llama 3 אין

גרייסן 8B און 70B, מיט א שטארקער פארבעסערונג אין ריזאנירונג, קאנטעקסט-לענג און קאוד. א 400B מאדעל ווערט נאך געטריינירט, און ער וועט לויט Meta זיין גלייך מיט Opus; דאס זענען יעצט די שטארקסטע אפענע מאדעלן. Llama 3 זאגט אויך אסאך ווייניגער אפ ביי פראגעס וועלכע עס האט נישט געדארפט אפזאגן ווי די פריערדיגע Llama מאדעלן.

• **11 אפריל: Gemini 1.5 Pro** – פרייווי – Google האט געמאכט Gemini 1.5 Pro צוגענגליך אין געפנטלעכער פרייווי אויף Vertex AI, מיט א קאנטעקסט פון ביז 1 מיליאן טאקענס.

• **12 אפריל: Mistral 8x22B** – Mistral האט ארויסגעלאזט א נייעם 8x22B מאדעל דורך א magnet link, און די ערשטע בענטשמארקס פון דער געמיינדע ווייזן 77 אויף MMLU. שפעטער איז ארויסגעקומען א באריכט, צוזאמען מיט אן אינסטרוקציע-מאדעל אונטער א פרייער לייסענס. לויט Mistral האט ער שטארקע רעזולטאטן אויף MMLU און HumanEval, ער ארבעט גוט אין ענגליש, פראנצויזיש, איטאליעניש, דייטש און שפאניש, ער שטיצט function calling, און ער האט א קאנטעקסט פון 64k.

• **12 אפריל: Spotify AI פלעיליסטס** – Spotify האט ארויסגעלאזט א פונקציע וואס שאפט פערזענלעכע פלעיליסטס לויט א געשריבענער בקשה.

• **16 אפריל: Adobe's ווידעא AI** – Adobe ברענגט AI ווידעא-געצייג אריין אין Premiere Pro, באזירט אויף איר Firefly ווידעא-מאדעל. די געצייג קענען צולייגן אדער אראפנעמען אביעקטן, פארלענגערן שאַטס און שאפן ווידעא פון טעקסט. Adobe ארבעט אויך מיט דריטע פירמעס, און באהאנדלט די זארג פון פאלשע ווידעאס דורך Content Credentials.

• **17 אפריל: Apple's iOS 18 אויפן מכשיר** – לויט שמועסן וועלן די ערשטע AI פונקציעס אין iOS 18 ארבעטן אינגאנצן אויפן מכשיר אן אינטערנעט, מיט Apple's אייגענעם מאדעל וואס הייסט אינערליך "Ajax".

• **18 אפריל: Stable Diffusion 3 API** – Stability AI האט געמאכט Stable Diffusion 3 צוגענגליך פאר געוויסע פראגראמירער דורך API און דורך דער נייער פלאטפארמע Stable Assistant Beta. דער מאדעל איז נאך אין פרייווי.

• **18 אפריל: OLMo פארבעסערט** – דער אינגאנצן אפענער 7B מאדעל OLMo האט זיך פארבעסערט מיט 24 פונקטן אויף MMLU, דורכן טוישן די מישונג פון דאטא. דאס טים האט פובליצירט די גענויע קוואלן און פראצענטן.

• **19 אפריל: DALL-E 2 ווערט אפגעשטעלט** – OpenAI ענדיגט אפ DALL-E 2. דער ארטיקל דערמאנט ווי די לאנסירונג אין 2022 האט אויפגעוועקט סיי באגייסטערונג סיי עטישע פראגעס וועגן טריינירן אויף קינסטלערישע ארבעט אן רשות.

• **22 אפריל: Brave Search מיט AI** – Brave Search גיט יעצט איבער די גאנצע וועלט ענטפערס וואס ווערן צוזאמגעשטעלט דורך AI. Brave AI זאגט אז זי וועט נאכקוקן די השפעה אויף וועבזייטלעך.

• **24 אפריל: Ray-Ban Meta ברילן** – די Ray-Ban Meta ברילן האבן באקומען מולטימאדאל AI פאר אידענטיפיצירן און איבערזעצן דורך שטימע-באפעלן. די פעיקייטן זענען אבער באגרעניצט, און אמאל נישט פינקטלעך.

- **24 אפריל: Phi 3 – Phi 3** איז א סעריע מאדעלן פון 3B ביז 14B. לויט די באריכטן איז דער 3B מאדעל בעסער ווי דער ארגינעלער ChatGPT. די וועיטס זענען ארויסגעגעבן, און עס איז דא א ווערסיע מיט 128k קאנטעקסט.
- **25 אפריל: Apple's OpenELM – Apple** האט ארויסגעלאזט OpenELM, אכט אפן-סאורס מאדעלן פון 270 מיליאן ביז 3 ביליאן פאראמעטערס, געמאכט צו לויפן אויף איין מכשיר. צוזאמען דערמיט איז ארויסגעקומען CoreNet, א ביבליאטעק פאר טריינירן.
- **25 אפריל: Sakana AI – EvoSDXL-JP** האט ארויסגעלאזט EvoSDXL-JP, א בילד-מאדעל פאר יאפאנישע בקשות, געמאכט דורך evolutionary model merging. לויט Sakana איז ער צען מאל שנעלער ביי אינפערענץ.
- **26 אפריל: Snowflake Arctic – Snowflake** האט לאנסירט Arctic, א מאדעל פאר ביזנעסער. לויט Snowflake האט ער שפיץ-לייסטונג אין SQL, קאוד און נאכפאלגן אינסטרוקציעס, פאר א בראכטייל פון די געווענליכע קאסטן. ער איז אפן-סאורס און צוגעגלייך אויף Hugging Face.
- **26 אפריל: Bunny מיט Llama3** – די Bunny סעריע מולטימאדאלע מאדעלן האט ארויסגעלאזט אירע ערשטע מאדעלן באזירט אויף Llama3 8B, מיט גוטע רעזולטאטן אויף MMMU פאר זייער גרייס.
- **30 אפריל: GitHub Copilot Workspace – GitHub** האט אנאנסירט א פראיעקט וואו מאדעלן קענען האלב-אליין ענדיגן פונקציעס און פארריכטן באגס.
- **30 אפריל: זכרון אין ChatGPT – OpenAI** האט לאנסירט "Memory" פאר ChatGPT Plus באניצער. מיט דעם געדענקט ChatGPT אינפארמאציע פון איין שמועס צום צווייטן, און מען קען עס קאנטראלירן אין די סעטינגס.
- **30 אפריל: StarCoder2 15B Instruct** – דער מאדעל איז געטריינירט אויף סינטעטישע און מענטשליכע דאטא, ער האט א פרייע לייסענס, און ער קריגט 72.6 אויף HumanEval. די אנטוויקלער ערקלערן ווי אזוי מען קען איבערחזרן זייער דאטא-פרעצעס.

פארשונג און טעכניק

- **1 אפריל: אויטאמאטיש פארן און דרויסנדיגע סצענעס** – א נייער צוגאנג באשרייבט דרויסנדיגע סביבות און באגייט די פראבלעמען פון טוישנדיגע באדינגונגען און פעלנדיגע דאטא. א קאנטראל-סיסטעם העלפט זיכער טוישן ליינס אין געדיכטן טראפיק. WcDT פארבינדט diffusion מיט transformers צו שאפן וועגן פאר אויטאמאטישע קארס. אויטאמאטישע רעיסעס, ווי דער Indy Autonomous Challenge, שטופן פאראויס די פארשונג פון זיך-פארנדיגע קארס.
- **1 אפריל: evals פאר ביזנעס** – דאטא, evals און קאמפיוט זענען יסודות פאר א גוטער AI. evals קענען זיין א שוץ-גראבן (moat) פאר פירמעס.
- **2 אפריל: אומלעזבארע פראבלעמען** – א פארשונג-ארבעט שטעלט פאר א בענטשמארק (UPD) צו פרובירן צי ראי-ה-מאדעלן ווי GPT-4V דערקענען ווען א פראגע אויף א בילד האט נישט קיין ענטפער, און זי ווייזט אויך מעגליכע פארבעסערונגען.

- **2 אפריל: Transformer-Lite אויף טעלעפאנען** – מיט quantization, אראפנעמען דעם kv cache און נאך אפטימיזאציעס קריגט מען 30 טאָקענס א סעקונדע פאר Gemma 2B אויף א טעלעפאן. דאס איז בערך 3 מאל שנעלער ווי אנדערע סיסטעמען.
- **3 אפריל: דאטא-סעטס** – עס איז ארויסגעקומען א דאטא-סעט פון 26 מיליאן זייטן און 18B טאָקענס, הויך-קוואליטעט OCR פון אינדוסטריעלע PDF's. HuggingFace ווייזט ווי אזוי צו שאפן סינטעטישע דאטא אויף א גרויסן פארנעם פאר Gretel. pre-training. האט ארויסגעגעבן 23 מיליאן טאָקענס פון text-to-SQL דאטא.
- **3 אפריל: many-shot jailbreaking** – Anthropic האט אנטדעקט א טעכניק צו jailbreak'ן מאדעלן מיט לאנגן קאנטעקסט. זי האט עס מיטגעטויילט מיט אנדערע ארגאניזאציעס, און זי האט איינגעפירט שוין-מיטלען דערקעגן.
- **4 אפריל: AI לערנט אן AI** – פארשער האבן געבויט א נעץ וואו איין AI לערנט א צווייטן אויספירן א ארבעט דורך נאטירליכער שפראך, מיט א מאדעל וואס הייסט S-Bert. דאס קען העלפן אין ראבאטיק.
- **4 אפריל: שנעלערע טראנספארמערס** – DiJiang פארוואנדלט עקזיסטירנדיגע טראנספארמערס אין שנעלערע מאדעלן אן א שווער איבער-טריינירן. LASP באהאנדלט לאנגע סעקווענצן מיט linear attention. Mixture of Depths פון DeepMind לאזט אויסגעבן ווייניגער רעכענונג אויף גרינגערע טאָקענס, מיט 50% ווייניגער flops ביי גלייכער לייסטונג. SqueezeAttention רעדוצירט דעם KV cache זכרון מיט 30%-70% און טאפלט די דורכפלוס.
- **5 אפריל: ענטפערס און פעלערן** – MGFiD פארבעסערט ענטפערן פראגעס דורך אויסקלייבן ראיות אויף עטליכע שטאפלען. ReaLMistake איז א בענטשמארק צו דערקענען פעלערן אין מאדעל-ענטפערס. Ada-LEval פרובירט מאדעלן אויף לאנגע און זייער לאנגע דאקומענטן.
- **4 אפריל: RealKIE דאטא-סעטס** – פינף נייע בענטשמארק דאטא-סעטס צו מעסטן ארויסציען אינפארמאציע פאר RAG ביי ביזנעסער.
- **5 אפריל: בילד-שאפונג** – InstantStyle טיילט אפ סטיל פון אינהאלט פאר קאנסיסטענטע סטיליזאציע און טונינג. TGATE טיילט די שאפונג אין פלאנירן און פארפיינערן, און פארבעסערט די קוואליטעט. PAE פארפיינערט אויטאמאטיש בקשות מיט UniFL reinforcement learning. diffusion מאדעלן דורך א קייט פון פידבעק. SwapAnything קען אויסטוישן סיי וועלכן אביעקט אין א בילד און לאזן דאס איבעריגע ווי עס איז.
- **8 אפריל: שוין פאר בילדער** – "Anything Unsegmentable" שיצט בילדער פון ווערן צוטיילט דורך AI מאדעלן, צוליב קאפירעכט און פריוואטקייט.
- **8 אפריל: אפטימיזירט און ReFT** – Meta פארשער האבן ארויסגעגעבן קאוד פאר א schedule-free אפטימיזירטער, וואס דארף נישט וויסן פון פאראויס די צאל טריינירונג-טריט. ReFT איז א פיין-טונינג מעטאד וואס איז אסאך ביליגער אפילו ווי PeFT.
- **9 אפריל: 3D** – Seg-NN מאכט גרינגער 3D צוטיילונג אן DreamView pre-training. שאפט 3D אביעקטן פון טעקסט. Hash3D פארשנעלערט 3D שאפונג מיט InstantMesh hashing. מאכט א 3D mesh פון איין בילד אין סעקונדעס. SplatPose דערקענט אנאמאליעס מיט Any2Point Gaussian splatting. ברענגט פעיקייטן פון אנדערע מאדעלן צו 3D. 6Img-to-3D. שאפט 3D פון זעקס בילדער.

- **9 אפריל:** געזונט-מאדעלן – אן איבערבליק איבער Healthcare Foundation Models, וואס קענען זיך צופאסן צו פארשידענע מעדיצינישע ארבעטן.
- **11 אפריל:** איבעררעדן – א פארשונג פון Anthropic האט געפונען אז Claude 3 Opus איז כמעט אזוי איבעררעדנדיג ווי מענטשן. מענטשן זענען געווען אביסל בעסער, אבער נישט סטאטיסטיש באדייטנד, און גרעסערע מאדעלן ווערן מער איבערצייגנדיג.
- **12 אפריל:** ראיה און חושים – א weak-to-strong סיסטעם פאר 3D דערקענונג מיט עטליכע קאמערעס; PoLoPCAC פאר לאסלעס קאמפרעסיע פון point clouds; MoCha-Stereo; וואס היט אפ געאמעטרישע דעטאלן; HomoFusion צו דערקענען פארדעקטע ליניעס אויפן וועג; MambaMOS פאר באוועגנדיגע אביעקטן אין LiDAR; און MIM4D פאר אויטאמאטיש פארן.
- **12 אפריל:** באגריפן אין שיכטן – א פארשונג געפינט אז פשוטע ארבעטן ווערן באהאנדלט אין פריערדיגע שיכטן, און קאמפליצירטע אין טיפערע.
- **15 אפריל:** מאניפולירן פראדוקטן – א פארשונג געפינט אז געוויסע טעקסט-סעקווענצן אין פראדוקט-באשרייבונגען קענען מאניפולירן מאדעלן אין זוך-מאשינען צו בעפארצוגן געוויסע פראדוקטן.
- **15 אפריל:** פארשידענע מעטאדן – MindBridge דעקאדירט מוח-סיגנאלן פון עטליכע מענטשן מיט איין מאדעל. DGMamba באנוצט Mamba פאר MPPE domain generalization. zero-shot פארבעסערט dialogue state tracking.
- **16 אפריל:** Pile-T5 – Eleuther האט איבערגעטריינירט T5 מיט א מאדערנעם טאקעניזער, און לענגער, און באקומען א שטארק פארבעסערטן מאדעל פאר encoding.
- **16 אפריל:** ווידעא און פאנאראמע – אן ארטיקל ערקלערט ווי צו טריינירן diffusion מאדעלן פאר ווידעא. PanFusion שאפט 360-גראד בילדער פון טעקסט. PLLaVA פאסט צו בילד-מאדעלן פאר ווידעא. TrackGPT פאלגט נאך אביעקטן לויט אינסטרוקציעס.
- **17 אפריל:** אומענדליכער קאנטעקסט – Google's Infini-attention לאזט מאדעלן ארבעטן מיט טעקסט אן א סוף, מיט א פעסטן זכרון און רעכענונג. Megalodon איז בעסער ווי Llama 2 אויף 2T טאקענס און קען אויך האבן אומענדליכן קאנטעקסט. TransformerFAM לייגט צו פידבעק און רעקורענץ.
- **17 אפריל:** קאמפרעסיע און אינטעליגענץ – א פארשונג-ארבעט ווייזט אז בענטשמארק-רעזולטאטן שטימען שטארק מיט ווי גוט א מאדעל קען קאמפרעסירן נייעם טעקסט.
- **17 אפריל:** quantization – 1-bit מאדעלן, נאכן פיין-טונען בלויז 0.65% פון די פאראמעטערס, זענען בעסער ווי 2-bit Quip#. DecoupleQ פון ByteDance פארבעסערט די פינקטלעכקייט ביי זייער נידריגע ביטס. א פארשונג פרובירט Llama3 מיט נידריגע ביטס. FP6-LLM ערמעגליכט פולע tensor core ניצונג אן אומסטאביליטעט.
- **17 אפריל:** ראיה-שפראך מאדעלן – א נייער פראגע-פארמאט מיט semantic segmentation; MoVA קלייבט אויס פארשידענע vision encoders לויטן קאנטעקסט; SEED-X פארשטייט בילדער פון יעדע גרייס; LaDiC פאר בילד-באשרייבונג דורך diffusion.
- **19 אפריל:** DeepMind און מאטעמאטיק – מאטעמאטיקער מיט DeepMind האבן גענוצט מאדעלן צו געפינען גרויסע זאמלונגען אן געוויסע מוסטערן, וואס העלפט שטודירן ערגסטע מצבים ווי צוואמפאלן פון סערווערס.

- **19 אפריל:** נאך מעטאדן – FedPFT פאסט צו מאדעלן און היט אפ פריוואטקייט. א מעטאד פארבעסערט In-Context Learning דורך state vectors. Component modeling צעלייגט א פראגנאז אויף אירע באזונדערע טיילן.
- **4 אפריל:** ספארט – ASTRA דערקענט וויכטיגע מאמענטן אין פוסבאל, און SoccerNet-GSR איז א דאטא-סעט צו רעקאנסטרוקטירן דעם מצב פונעם שפיל.
- **22 אפריל:** DeepMind alignment – Google DeepMind האט ארויסגעגעבן alignment ארבעט, בעיקר sparse autoencoders אויף Gemini Ultra, א גרויסער שפרונג אין דער גרייס פון אינטערפרעטאציע.
- **22 אפריל:** Q-function און ריזאנירונג-טאָקענס – א טעכנישע RL פארשונג-ארבעט ווייזט דעם טעארעטישן יסוד פון reward מאדעלן. "Reasoning tokens" לערנען מאדעלן צו טראכטן פאראויס; ביי GPT-2 האט עס געברענגט א 35% ירידה אין loss.
- **22 אפריל:** נאך פארשונג – א קאמבינאציע פון segmentation און optical flow פאר ווידעא; פארלענגערן דעם קאנטעקסט פון embedding מאדעלן ביז 32,000 טאָקענס אן טריינירן; און, FineWeb, 15T טאָקענס פון Common Crawl, אויסגערייניגט פון דופליקאטן.
- **24 אפריל:** instruction hierarchy – OpenAI האט פובליצירט ווי אזוי צו געבן מער וואג צו system prompts. דאס שטארקט שטארק די שוץ קעגן jailbreaks.
- **25 אפריל:** דערקענען sleeper agents – Anthropic ווייזט אז פשוטע linear probes מיט דער פראגע "וועסטו טון עפעס געפערליך?" דערקענען זייער פארלעסליך מאדעלן וואס זענען געטריינירט צו טון ביזע זאכן ביי געוויסע ווערטער.
- **29 אפריל:** "dot by dot" – א פארשונג-ארבעט ווייזט אז chain-of-thought קען מען באהאלטן מיט "...". טאָקענס, אויב מען טריינירט דעם מאדעל דערויף. דאס הייסט אז מען קען נישט שטענדיג פארשטיין די טריט פונעם טראכטן.
- **30 אפריל:** JAT – א טים פארבונדן מיט Hugging Face האט געטריינירט JAT, א נאכמאך פון DeepMind's GATO, וואס ארבעט גאנץ גוט אויף א ברייטע ריי ארבעטן.
- **23 אפריל:** קורצע פארשונגען – MAexp פאר multi-agent; TF-GPH פאר בילד-הארמאניזאציע; FinLangNet פאר קרעדיט-ריזיקע; MultiBooth פאר בילדער מיט עטליכע באגריפן; קאנטראל פון כאראקטערן אין רעאל-צייט; CutDiffusion פאר הויכע רעזאלוציע; CFExplainer פאר זיכערהייט-שוואכקייטן אין קאוד; WCPS פאר טוישנדיגע דאטא; אראפנעמען דראטן פון ווידעאס; GNN פאר NAS; און AAPL צו דערקענען נישט-געזעענע אביעקטן.

געצייג פאר פראגראמירער און אפן-סאורס

- **1 אפריל:** קול און קאוד – VoiceCraft איז א 700m מאדעל פאר zero-shot שטימע-קלאנירונג און רייד. EvoEval איז א שטרענגערער בענטשמארק פאר פראגראמירן. "Interrupting Cow" איז א הילף וואס כאפט אריין און רעדט ארויס ווען זי פאראויסזאגט גענוג טאָקענס.
- **2 אפריל:** Gemma, OpenUI, TensorRT – OpenChat ווייזט אז מען קען טונען Gemma ביז דער לייסטונג פון Mistral. Wandb איז א באשרייבונג אין React אדער Svelte.

- Nvidia's TensorRT דערגרייכט ביז 28 מאל שנעלער, און Llama 2 לויפט מיט 1,200 טאקעס א סעקונדע.
- **2 אפריל: Microsoft קורס** – ווערסיע 2 פון Microsoft 'ס קורס פאר אנהייבער, מיט 18 לעקציעס אויף GitHub.
 - **3 אפריל: SWE-agent** – דער אפן-סאורס SWE-agent ניצט GPT-4 צו פארריכטן באגס אין GitHub, און לייזט 12.29% אויף SWE-bench, ענליך צו Devin.
 - **3 אפריל: נאך געצייג** – Chug איז א מולטימאדאל דאטא-לאדער. SPHINX-V איז א מאדעל וואס פארשטייט ראיה-אנווייזונגען. AutoQuant עקספארטירט מאדעלן אין quantization 5 פארמאטן.
 - **8 אפריל: אנגענטן און מידלווער** – LLocalSearch איז א לאקאלע זוך-מאשין מיט AI אנגענטן. AIOS בויט אריין א מאדעל ווי דער "מוח" פון אן אפערירונג-סיסטעם. FlyFlow איז API מידלווער.
 - **9 אפריל: TGI צוריק אפן** – Hugging Face האט צוריקגעבויט TGI צו א פרייע לייסענס. דער ניט-קאמערציעלער לייסענס האט נישט געהאלפן אין ביזנעס, און עס האט פארמינערט די באטייליגונג פון דער געמיינדע.
 - **9 אפריל: llm.c און Attorch** – Andrej Karpathy 'ס llm.c טריינירט מאדעלן אין פשוטן C/CUDA און PyTorch. Attorch שרייבט איבער טיילן פון PyTorch nn אין Triton.
 - **9 אפריל: Claude ניצט געצייג** – Anthropic האט לאנסירט א בעטא וואו Claude קען ניצן געצייג פון די באניצער, דורך JSON schemas. Haiku איז בעסער ווי GPT-4 Turbo אין 50% פון די פעלער אויפן Berkeley Function Calling בענטשמארק.
 - **10 אפריל: בענטשמארק און קאוד** – א כעמיע-בענטשמארק פאר מאדעלן; קאוד פאר "next-resolution prediction" בילד-שאפונג; און טריינירן איבער 200k קאנטעקסט אויף H100.
 - **11 אפריל: קליינע מאדעלן און RL** – Policy-guided diffusion פאר offline RL. א קליינער ראיה-מאדעל באזירט אויף Gemma און SigLIP. Moondream באזירט אויף Phi-2, לויפט אויף א טעלעפאן; עס איז דא א llamafile ווערסיע פון Moondream 2.
 - **12 אפריל: JetMoE** – א MoE מאדעל געטריינירט פאר 100 טויזנט דאלאר, וואס גלייכט זיך מיט Llama 7B.
 - **15 אפריל: נאך געצייג** – Jina AI reader מאכט וועבזייטלעך אין markdown פאר מאדעלן; א ריינע implementation פון discrete diffusion; דער BabyLM טשאלענדזש מיט בלויז ~10M טאקעס; און Github2File.
 - **17 אפריל: Storm און Ray** – Stanford 'ס Storm שרייבט באריכטן מיט ציטאטן. Ray 2.10 שטיצט Intel Gaudi 3.
 - **18 אפריל: Effort, Luminal** – Effort לאזט טוישן אין רעאל-צייט וויפיל רעכענונג ווערט געטון ביי אינפערענץ אויף Apple Silicon, און ווערט אזוי שנעלער. דערצו זענען דא א snippet צו פרובירן DDP וועיטס, און Luminal, א שנעלע deep learning ביבליאטעק.
 - **19 אפריל: AI Gateway** – AI Gateway פארייניגט API בקשות צו פארשידענע פראוויידערס מיט load balancing. דערצו א וועגווייזער צו Sentence Transformers, און Tabnine.

- **22 אפריל: Llama 3 פיינ-טונינג** – וועגווייזער פאר Q-Lora, ORPO, FSDP, און unsloth מיט 6 מאל לענגערן קאנטעקסט. עס ווייזט זיך אויך ווי אראפצונעמען די ווייניגסט-נוצליכע שיכטן, און עס איז דא א פשוטער jailbreak.
- **22 אפריל: Mistral Common און Cauldron** – Mistral האט ארויסגעגעבן אירע טאָקענייזער-געצייג. The Cauldron פארייניגט 50 ראיה/שפראך דאטא-סעטס.
- **24 אפריל: Instructor, BitBLAS, MaxText** – Instructor פאר סטרוקטורירטע ענטפערס; Microsoft's BitBLAS קערנעלס פאר BitNet; און Google's MaxText אין Jax פאר TPUs.
- **25 אפריל: Cohere און Anthropic** – Cohere האט אפן-סאורס געמאכט איר טשעטבאט-אינטערפייס, און Anthropic האט א ביבליאטעק פון prompts.
- **29 אפריל: torchtitan און llamafile** – Meta's torchtitan פאר pre-training אין PyTorch; OpenLIT פאר observability; FrugalGPT צו אראפברענגען API קאסטן; Llamafile איז 2 מאל שנעלער; און AI town לאקאל מיט Llama 3.
- **30 אפריל: Whisper און Mistral.rs** – א Whisper CLI פון די שאפער פון Ruff; Mistral.rs, א שנעלע אינפערענץ-פלאטפארמע; Memary פאר זכרון פון אגענטן; און א Jax וועגווייזער.

פירמעס, געלט און ביזנעס

- **1 אפריל: Stargate** – Microsoft און OpenAI פלאנירן לויט באריכטן א דאטא צענטער פראיעקט פון ביז 100 ביליאן דאלאר, וואס זאל ענדיגן מיט א סופער-קאמפיוטער "Stargate" ביז 2028.
- **1 אפריל: Microsoft Copilot** – Copilot וועט לויפן לאקאל אויף קאמפיוטערס, מיט NPU פון איבער 40 Microsoft TOPS. פרובירט אויך אן Xbox טשעטבאט פאר סופארט.
- **3 אפריל: OpenAI Startup Fund** – Sam Altman האט איבערגעגעבן דעם קאנטראל פונעם פאנד צו Ian Hathaway.
- **3 אפריל: YC, Replit, Samsung** – YC – 25 סטארטאפס האבן געטריינירט אייגענע מאדעלן אין דריי חדשים. Replit לאנסירט Replit Teams מיט אן AI אגענט וואס פארריכט פעלערן. Samsung באטראכט צו פארבעסערן Bixby.
- **4 אפריל: Google SGE** – Google באטראכט צו מאכן SGE א באצאלטע פונקציע אין Google One AI Premium.
- **5 אפריל: Lambda** – Lambda האט באקומען 500 מיליאן דאלאר אין חוב-פינאנצירונג פאר GPU קלאוד, נאך 230 מיליאן אין Series C.
- **5 אפריל: OpenAI Custom Model** – OpenAI פארברייטערט איר פראגראם פאר געמאכטע מאדעלן פאר ביזנעסער.
- **5 אפריל: טאלאנט** – Tesla הייבט אויף די געהאלט פון AI אינזשעניערן קעגן OpenAI. לויט אן אנאליז פירט Microsoft אין דער מלחמה אויף טאלאנט.
- **8 אפריל: xAI** – xAI פירט שמועסן צו אויפנעמען 3 ביליאן דאלאר ביי א ווערט פון 18 ביליאן.
- **8 אפריל: AMD** – AMD וועט אפן-סאורס מאכן די firmware פון Radeon.

- **8 אפריל: טשיפס** – אין MLPerf האבן Nvidia'ס סיסטעמען מיט H200 געוואונען קעגן Intel און Qualcomm. Intel Gaudi 3 איז לויט ביז 1.7 מאל שנעלער ווי H100 אין טריינירן, און ביליגער. Meta'ס נייער טשיפ האט 128GB אויף 5nm און Google TPU v5p איז כמעט דריי מאל שנעלער ווי v4, און עס איז דא א נייע Axion CPU.
- **8 אפריל: Rabbit מיט ElevenLabs** – Rabbit ניצט ElevenLabs פאר שטימע-באפעלן אויף r1. ElevenLabs האט אויפגענומען 80 מיליאן.
- **9 אפריל: Altman מיט Jony Ive** – Sam Altman און Jony Ive זוכן ביז 1 ביליאן דאלאר פאר א פערזענלעך AI מכשיר וואס וועט נישט אויסזען ווי א סמארטפאון.
- **9 אפריל: Yahoo און Artifact** – Yahoo קויפט די טעכנאלאגיע פון Artifact. Re-Up האט איינגעפירט א ראבאט וואס פרעגלט עסן.
- **11 אפריל: Amazon** – Amazon האט אינוועסטירט נאך 2.75 ביליאן אין Anthropic, און ארבעט אויף איר אייגענעם מאדעל Olympus.
- **12 אפריל: Waymo מיט UberEats** – Waymo קארס ברענגען עסן פאר UberEats אין Phoenix.
- **15 אפריל: OpenAI און Amazon** – OpenAI האט לויט באריכטן ארויסגעווארפן צוויי פארשער צוליב ארויסלאזן סודות. Andrew Ng גייט אריין אין Amazon'ס באורד.
- **16 אפריל: OpenAI Japan** – OpenAI האט געעפנט א ביורא אין יאפאן. Daikin, Rakuten און TOYOTA Connected ניצן ChatGPT Enterprise.
- **16 אפריל: Google Code Assist** – Gemini Code Assist האט 1 מיליאן טאָקענס קאנטעקסט. TikTok פרובירט AI אוואטארן פאר רעקלאמעס, אבער זיי זענען נאך נישט אזוי ווירקזאם ווי מענטשן.
- **18 אפריל: Mistral** – Mistral זוכט עטליכע הונדערט מיליאן ביי א ווערט פון 5 ביליאן.
- **18 אפריל: Sanctuary AI** – Magna וועט פרובירן דעם ראבאט Phoenix אין א פאבריק.
- **19 אפריל: Stability AI** – האט אפגעזאגט בערך 20 ארבעטער, ארום 10%.
- **22 אפריל: Nvidia און Intel** – Nvidia וועט מיט Fujitsu באזארגן יאפאן'ס קוואנטום סופער-קאמפיוטער ABCI-Q מיט Intel H100 2,000. האט פארענדיגט צו צוזאמשטעלן דעם ערשטן High-NA EUV מאשין פון ASML.
- **24 אפריל: OpenAI ביזנעס** – נייע פונקציעס פאר API: זיכערהייט, Assistants API, Projects קאסטן-קאנטראל.
- **24 אפריל: Profluent** – דער סטארטאפ Profluent האט מיט AI געשאפן OpenCRISPR-1 און דערמיט רעדאגירט מענטשליכע DNA. ער וועט אפן-סאורס מאכן די גענען-רעדאקטארן, אבער נישט די AI טעכנאלאגיע.
- **25 אפריל: Augment און FlexAI** – Augment, א קאנקורענט פון Copilot, איז ארויס מיט 252 מיליאן ביי א ווערט נאענט צו 977 מיליאן. FlexAI האט אויפגענומען 30 מיליאן seed.
- **25 אפריל: Gemini Nano** – Google וועט האבן ווערסיע 2 פאר Galaxy S25.

- **26 אפריל:** אפקויפן – Apple האט געקויפט Datalab פון פאריז. Nvidia קויפט Run:ai פאר בערך 700 מיליאן.
- **29 אפריל:** Apple, Cohere, Home Assistant – פירט שמועסן מיט OpenAI און Google פאר Cohere iOS 18 האט לאנסירט א Toolkit. Home Assistant איז אריין אין Open Home Foundation.
- **30 אפריל:** Financial Times – OpenAI וועט טריינירן אויף דעם אינהאלט פון Financial Times.
- **1 אפריל:** קליינע אפליקאציעס – Bezi, Airtable AI, 3D, Bluedot פאר אימעילס, א ליסטע פון 100 געצייג, DataMotto, Butternut AI, AI Flashcards, Sonnet AI, PhoneScreen.AI, Limitless, Panjaya פאר ווידעא-לאקאליזאציע.

רעגירונגען, געזעצן און זיכערהייט

- **2 אפריל:** זיכערהייט-טעסטן – METR ארבעט מיט OpenAI און Anthropic אויף טעסטן פאר אויטאנאמיע און זיך-קאפירן, אבער מען גיבט צו אז די טעסטן זענען נאך אין אנהייב און קענען נישט גארנטירן זיכערהייט.
- **4 אפריל:** רעגולירן – אן ארטיקל באשרייבט די קאמפליצירטע אינטערנאציאנאלע באמאונגען צו רעגולירן AI.
- **5 אפריל:** YouTube קעגן Sora – YouTube'ס CEO זאגט אז טריינירן Sora אויף YouTube ווידעאס וואלט זיין א בראך פון די כללים, הגם ער ווייסט נישט צי דאס איז געשען.
- **9 אפריל:** קאנאדע – Trudeau האט אנאנסירט 2.4 ביליאן דאלאר פאר AI אין קאנאדע.
- **12 אפריל:** Meta לעיבלס – Meta פארברייטערט "Made with AI" לעיבלס פון מאי.
- **22 אפריל:** GPT-4 און שוואכקייטן – GPT-4 קען אויסנוצן שוואכקייטן לויט CVE באריכטן מיט 87% הצלחה.
- **23 אפריל:** AI דאגפייט – א דזשעט וואס ווערט געפירט דורך AI, פון DARPA, האט מצליח געווען אין א דאגפייט קעגן א מענטשליכן פילאט.
- **29 אפריל:** DHS באורד – Homeland Security האט געשאפן א AI באורד, מיט Jensen Huang און Sam Altman.
- **30 אפריל:** OMB פאליסי – קריטיקער זארגן אז די נייע OMB פאליסי וועט אפשטעלן AI אין רעגירונג.

אנאליזן און מיינונגען

- **1 אפריל:** כינע טאלאנט – כינע פראדוצירט כמעט האלב פון די בעסטע AI פארשער. 38% פון די שפיץ-פראפעסיאנאלן אין אמעריקע זענען געבוירן אין כינע.
- **1 אפריל:** AI מיט א גוף – Meta פארשער גלייבן אז אמת'ע אינטעליגענץ פאדערט א גוף. AI Habitat 3.0 האט יעצט מענטשליכע אוואטארן.

- **2 אפריל: RPA און סכנות** – נאר 3% פירמעס האבן מצליח געווען צו פארברייטערן RPA, און Foundation Capital זאגט פאראויס א פארגרעסערונג פון 10 מאל. אן אנדער ארטיקל טענה'ט אז מען דארף זיך קאנצענטרירן אויף ארבעט-פלעצער און פאלשע אינפארמאציע.
- **3 אפריל: מארקעט מאפע** – Matt Turck'ס צענטע מאפע ציילט מער ווי צוויי טויזנט פירמעס. Instacart און Faire ערקלערן ווי AI פארבעסערט זייערע מארקעטס.
- **4 אפריל: Gary Marcus** – Gary Marcus מיינט אז אויב קיין GPT-5 קומט נישט ביז סוף 2024, וועט דער הייפ פארשווינדן, און אז סקילינג וועט קיינמאל נישט ברענגען AGI.
- **4 אפריל: אסטראטעגיע** – אנאליזן וועגן AI אינפראסטרוקטור, ווערטיקאלע AI, meta-moats, און אז דער שוץ-גראבן קומט פון ספעציעלע דאטא און נישט פון די מאדעלן.
- **8 אפריל: דאטא שטערונג** – הויך-קוואליטעט דאטא קען זיך ענדיגן ביז 2026. די פירמעס ניצן YouTube טראנסקריפטן, טוישן פריוואטקייט-פאליסיס און באטראכטן סינטעטישע דאטא.
- **9 אפריל: Pew** – ניצן פון ChatGPT אין אמעריקע איז געוואקסן פון 18% צו 23%, אבער 40% גלייבן נישט די אינפארמאציע וועגן וואלן.
- **10 אפריל: Hassabis און שטימען** – א פראגע צי Demis Hassabis קען ראטעווען Google. שטימע-אקטיארן ביי ווידעא-שפילן זארגן וועגן גנבה פון זייערע שטימען.
- **11 אפריל: סטאטוס** – a16z'ס ליסטע פון די 100 פאפולערסטע אפליקאציעס; א באריכט וועגן דעם מצב פון AI אין 2024; און Tom Tunguz מיינט אז מען דארף נאך מיקרא-מענדזשן AI.
- **16 אפריל: Humane Ai Pin** – רעצענזיעס זאגן אז דער Ai Pin האט אסאך טעכנישע פראבלעמען, בפרט די איבערזעצונג.
- **17 אפריל: Stanford AI Index** – דער זיבעטער יערליכער באריכט באהאנדלט מולטימאדאלע מאדעלן, אינוועסטירונגען, בענטשמארקס און רעגולאציעס.
- **18 אפריל: אויסער Attention און CUDA** – נייע ארכיטעקטורן קאנקורירן ביי קליינע גרייסן. CUDA בלייבט א ריזיגער שוץ פאר Nvidia קעגן AMD'ס ROCm.
- **22 אפריל: נוצן** – Ben Evans זוכט פראקטישע נוצן. Foundation Capital זעט 4.6 טריליאן דאלאר אין "Service-as-Software". אן ארטיקל ערקלערט ווי צו בויען פארלעסליכע סיסטעמען פון נישט-פארלעסליכע אגענטן.
- **24 אפריל: ווי ChatGPT ארבעט** – Evan Morikawa פון OpenAI ערקלערט טאקעניזאציע, embeddings און ווי דער מאדעל קלייבט דאס נעקסטע ווארט.
- **25 אפריל: אפן-סאורס וואך** – די לעצטע וואך פון מערץ האט געברענגט DBRX, Jamba און Samba-CoE, און דער אונטערשייד צווישן אפענע און פארמאכטע מאדעלן ווערט קלענער.
- **26 אפריל: קריטיק** – רופן צו איבערמאכן Siri; א טענה אז האלוצינאציעס קען מען נישט אינגאנצן לייזן; און אן ארטיקל אז AI פירט נישט אויס די הבטחות.

ווייטערע ידיעות

- **2 אפריל: ראבאט Emo** – א ראבאט וואס פאראויסזאגט און מאכט נאך פנים-אויסדרוקן אין רעאל-צייט.

- **4 אפריל: Worldcoin Orb** – Worldcoin האט אפן-סאורס געמאכט די ווייכווארג פונעם Orb , וואס ניצט NVIDIA Jetson, Rust און Orb OS.
- **5 אפריל: AI און אמונה** – א שטודיע פון Joshua Conrad Jackson און Adam Waytz געפינט א פארבינדונג צווישן AI/ראבאטן און א ירידה אין רעליגיעזע גלויבן.
- **5 אפריל: TypeScript און YC ווידעא** – TypeScript פון Microsoft מאכט OpenAPI קירצער. א YC פירמע האט געשאפן איר Demo Day ווידעא מיט AI.
- **10 אפריל: אוויאציע האקאטאן** – 200 אינוועניערן האבן געבויט געצייג ווי AI קאנטראלירער פאר לופט-פארקער.
- **11 אפריל: נאטור-אומגליקן** – Microsoft מיט OpenAI האבן א פאטענט צו דערקענען ערדציטערנישן און איינברוכן דורך קלאנגען.
- **17 אפריל: ראבאטן און פארן** – ALOHA Unleashed פון DeepMind ;Wayve ס Lingo-2 פארט מיט ערקלערונגען; און Google Maps ניצט AI פאר EV לאדערס.
- **19 אפריל: מעגליכע מוחות** – אן עסיי וועגן ווי AI פארשטייט פרעגט וואס עס הייסט צו זיין א מענטש.
- **24 אפריל: Los Angeles** – א פיילאט ניצט AI צו פאראויסזאגן היימלאזיגקייט; 86% פון די באטייליגטע זענען געבליבן אין א דירה, און כמעט 800 הויזגעזינד האבן באקומען הילף.
- **26 אפריל: Drake** – Drake'ס ליד ניצט AI שטימען פון Tupac און Snoop Dogg.

דער גרעסטער נייעס פון חודש מאי איז געווען OpenAI's נייער מאדעל GPT-4o, וואס פארשטייט טעקסט, בילדער און קול אינאיינעם. באלד דערנאך האט Google ביי I/O באוויזן Gemini Flash, Veo, און Jan Leike און Ilya Sutskever ביי OpenAI אליין איז געווען א טומל: Project Astra. דער superalignment-טיים איז צעלייגט געווארן, און מען האט געקריטיקירט די שווייג-אפמאכן פאר געוועזענע ארבעטער. דערביי האט OpenAI אונטערגעשריבן א ריי אפמאכן מיט צייטונגען און וועבזייטלעך, בשעת אכט צייטונגען האבן זי פארקלאגט. אויף דער געלט-זייט האבן xAI (\$6 ביליאן), Scale AI (\$1 ביליאן) און Coreweave (\$1.1 ביליאן) אויפגעשטעלט ריזיגע סומעס. Anthropic האט געמאפירט מיליאנען באגריפן אינעווייניג אין Claude Sonnet, און Microsoft האט לאנסירט Copilot+ קאמפיוטערס.

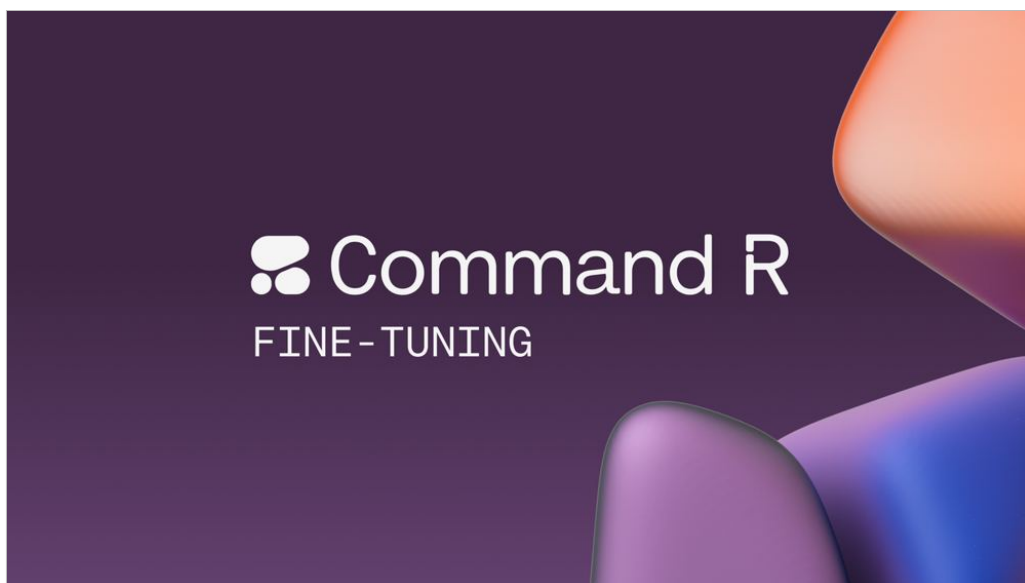
מאדעלן און נייע לאנסירונגען

- **2 מיי:** א מיסטעריעזער **gpt2-chatbot** – אויף lmsys.org איז אויפגעטויכט א מאדעל מיטן נאמען gpt2-chatbot, וואס האט געוויזן פעאיגקייטן בערך ווי GPT-4.5. מען האט געמיינט אז דאס איז א שטילער טעסט פון OpenAI. שפעטער איז באשטעטיגט געווארן אז עס איז טאקע אן OpenAI מאדעל, ווייל א 429 רעיט-לימיט טעות האט ארויסגעוויזן פרטים פון OpenAI's API.
- **2 מיי:** **Amazon Q** – AWS האט לאנסירט Amazon Q, א גענעראטיוון AI געהילף וואס ניצט א פירמע'ס אינעווייניגסטע דאטא. פאר פראגראמירער העלפט עס מיט קאוד, טעסטן און אפליקאציעס, און פאר ביזנעס-באניצער גיט עס ענטפערס אויף פשוטע שפראך. מיט Amazon Q Apps קען מען אויך בויען אייגענע AI אפליקאציעס אן קאוד.
- **7 מיי:** **DeepSeek-V2** – DeepSeek האט ארויסגעלאזט א מאדעל מיט איבער 200B פאראמעטערס, פון וועלכע 21B זענען אקטיוו אויף איינמאל. עס ארבעט זייער גוט ביי קאוד און לאגישן טראכטן. מען ווייסט נאך נישט צי עס איז אינגאנצן בעסער ווי Llama 3 70B, אבער עס איז א ווילקאמער צוגאב צו די אפענע מאדעלן.



2024 מיי 7 · DeepSeek-V2

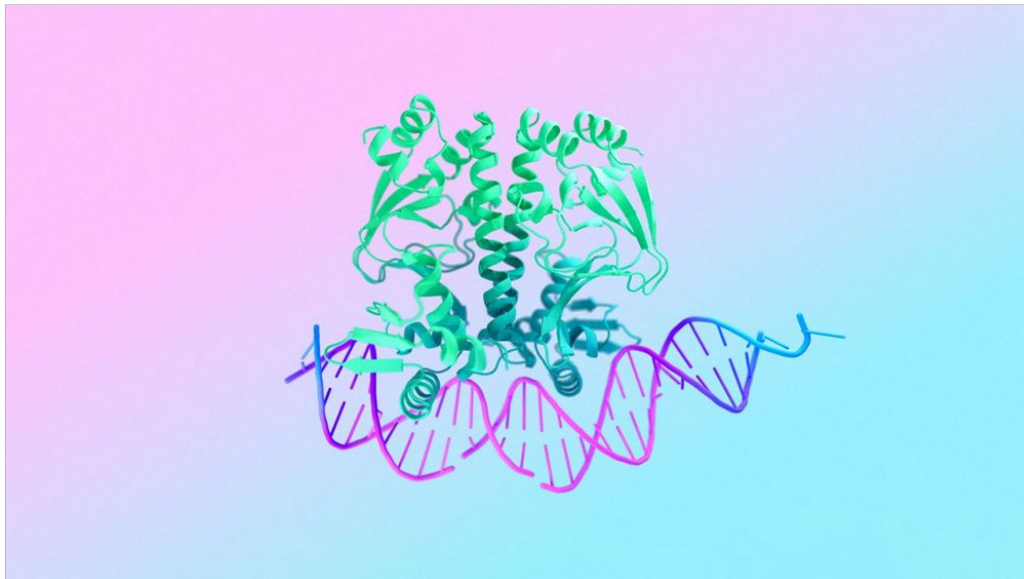
- **7 מיי: Cohere Command R** – די Command R און R+ מאדעלן, וואס זענען צוגעפאסט פאר לאנגע טעקסטן און וויכטיגע ביזנעס-ארבעטן, זענען יעצט צוטריטלעך אויף Amazon Bedrock. שפעטער האט Cohere לאנסירט fine-tuning פאר Command R. לויט Cohere איז עס ביליגער און שלאגט גרעסערע מאדעלן, און עס איז צוטריטלעך אויף איר פלאטפארמע און אויף Amazon Sagemaker.



2024 מיי 7 · Cohere Command R

- **9 מיי: Microsoft MAI-1** – Microsoft בויט אן אייגענעם מאדעל מיטן נאמען MAI-1. לויט ידיעות האט ער בערך 500 ביליאן פאראמעטערס, און ער צילט צו איבערשטייגן די מאדעלן פון Google און OpenAI.
- **9 מיי: Apple M4 טשיפ** – Apple האט באוויזן דעם M4, איר ערשטן טשיפ וואס איז ספעציעל געבויט פאר AI. ער ניצט 3 נאנאמעטער טעכנאלאגיע און דעבייטירט אינעם 2024 iPad Pro. זיין 10-קאר CPU איז 50% שנעלער ווי דער M2, און דער נייער Neural Engine טוט 38 טריליאן אפעראציעס א סעקונדע. מען האט פריער ערווארט אז Tim Cook וועט ביי דער "Let Loose" אסיפה נאר אנדייטן אויף AI פיטשערס, מיט מער פרטים ביי WWDC אין יוני.

- **9 מיי: AlphaFold 3** – Google DeepMind און Isomorphic Labs האבן ארויסגעלאזט די דריטע דור פון AlphaFold, וואס זאגט פאראויס ווי פראטעיןען לייגן זיך צוזאם. ער איז 50% גענויער ווי די פריערדיגע דורות. ער האט ריכטיג פאראויסגעזאגט די סטרוקטור פונעם spike פראטעין פונעם Coronavirus OC43. דערצו איז לאנסירט געווארן AlphaFold Server, א פרייער וועג צו ניצן דעם מאדעל.



AlphaFold 3 · 9 מיי 2024

- **10 מיי: Stable Artisan** – Stability AI האט באוויזן Stable Artisan, א Discord באט וואס פארבינדט אלע אירע מאדעלן דורך איר API. מען זעט דאס ווי א דירעקטע קאנקורענץ מיט Midjourney.
- **10 מיי: ElevenLabs מוזיק** – ElevenLabs ווייזט פאר א נייעם מאדעל וואס מאכט פון א פראמפט א ליד מיט ווערטער. דער פראמאציע-שטייגער איז ענליך צו וויאזוי OpenAI האט פראמאטעד Sora. איינער פון זייערע פארשער האט ארויסגעלאזט אן audio-diffusion רעפאזיטאריע אין PyTorch, וואס קען דינען אלס אנהייב-פונקט צו נאכמאכן אזעלכע רעזולטאטן.
- **10 מיי: Alibaba Qwen2.5** – Alibaba Cloud האט ארויסגעלאזט Tongyi Qianwen Qwen2.5. עס איז א שטארקע פארבעסערונג לגבי Qwen2.0 אין לאגישן טראכטן, פארשטיין קאוד און פארשטיין טעקסט.
- **13 מיי: BLIP 3** – Salesforce האט ארויסגעלאזט די דריטע ווערסיע פון BLIP, נישט פאר קאמערציעלן באנוץ. די מאדעלן פארבינדן בילד און שפראך און ווערן מערסטנס גענוצט צו פארשטיין בילדער און זיי באשרייבן.



2024 מיי 13 · BLIP 3

- **14 מיי:** OpenAI – GPT-4o האט באוויזן GPT-4o (דער "o" שטייט פאר omni), א מאדעל וואס איז פון זיך אליין מולטימאדאל. ביי טעקסט איז ער בעסער ווי GPT-4, און ביי אנדערע פארמען (בילד, קול) איז ער אויפן העכסטן שטאפל. מיט אים זענען געקומען א דעסקטאפ אפליקאציע, א קול-אינטערפעיס כמעט אין רעאלער צייט, און פארבעסערטע לאגישע פעאייגקייטן. ChatGPT האט אויך פארבעסערט GPT-4o'ס פעאייגקייט צו אנאליזירן דאטא: ער ווייזט טאבעלעס, ווייזט ענדערונגען און מאכט פון זיי גראפיקן. א טעסט מיט א מאטעמאטיק פראבלעם האט געפונען אז א פשוטער טעקסט פראמפט מיט Chain-of-Thought גיט די בעסטע רעזולטאטן, אבער דער מאדעל פארשטייט נאך נישט גוט קליינע פרטים אין בילדער. די macOS אפליקאציע איז שוין צוטריטלעך פאר Plus אבאנענטן.
- **14 מיי:** IBM Granite קאוד מאדעלן – IBM עפנט די Granite קאוד מאדעלן אלס אפן-סאורס. זיי זענען פון 3 ביז 34 ביליאן פאראמעטערס און דעקן א ריי פראגראמיר-ארבעטן.
- **15 מיי:** Google I/O – Google האט באוויזן Gemini Flash, א לייכטער מולטימאדאלער מאדעל מיט א פענסטער ביז א מיליאן טאקענס. דערצו זענען געקומען Veo פאר מאכן ווידעא, Imagen 3, און Project Astra, Google'ס נייער געהילף. אנדערע אנאנסירונגען: א קאנטעקסט פון 2 מיליאן טאקענס, פיל ביליגערע מאדעלן, Firebase Genkit פאר JavaScript/TypeScript, און LearnLM. מען האט אויך ארויסגעלאזט PaliGemma, אן אפן-סאורס בילד-שפראך מאדעל געבויט אויף SigLIP. ער איז גרינג צו טריינירן פאר באזונדערע ארבעטן און שלאגט אנדערע מאדעלן אין דערקענען און צעטיילן אביעקטן און אין OCR. מיט GPT-Fast קען מען אים לויפן אין רעאלער צייט.
- **16 מיי:** Nous Hermes Theta – Theta איז א צונויפגיסונג פון Hermes 2 Pro מיט Llama 3 8B instruct. עס רופט גוט אויף פונקציעס און שניידט גוט אפ אויף בענטשמארקס.
- **16 מיי:** GPT קראם פאר אלעמען – OpenAI'ס קראם פון אייגענע GPTs איז יעצט אפן פאר אלע באניצער אומזיסט.
- **20 מיי:** Grok-1 אפן – xAI האט ארויסגעלאזט די וואגן און ארכיטעקטור פון Grok-1, א Mixture-of-Experts מאדעל מיט 314 ביליאן פאראמעטערס. עס איז א רוייער באזע-מאדעל פון דער ערשטער טריינירונג, נישט צוגעפאסט פאר קיין באזונדערע ארבעט. xAI האט אים טריינירט פון אנהייב מיט אן אייגענער מערכה.

- **20 מיי: Meta Chameleon – Meta** האט באוויזן Chameleon, א מאדעל וואס ארבעט פון זיך אליין מיט בילדער און טעקסט צוגלייך, ענליך צו GPT-4o, און ער שלאגט אסאך אנדערע מאדעלן. לויט דער ידיעה איז Meta'ס טים אינעווייניג שוין ווייט איבערגעשטיגן די ארבעט.
- **22 מיי: Phi-3** – די קליינע Phi-3 סעריע האט באקומען א 14B ווערסיע, וואס איז אויף איין שטאפל מיט Command R (א 104B מאדעל). Microsoft האט אויך מודיע געווען אז Phi-3 איז צוטריטלעך פאר אלעמען. דערצו האט זי באוויזן Phi-3-Silica, מיט 3.3 ביליאן פאראמעטערס, צוגעפאסט צו לויפן לאקאל אויף די NPU טשיפס פון Copilot+ קאמפיוטערס.
- **22 מיי: Copilot+ קאמפיוטערס** – Microsoft וויל מאכן Windows פאר אן AI אפערירונג-סיסטעם, און לאנסירט Copilot Plus קאמפיוטערס. צווישן די נייע פֿיטשערס איז Recall, וואס העלפט געפינען אפליקאציעס, פיילס און אנדערע זאכן וואס מען האט שוין אמאל געזען.
- **22 מיי: Grok מולטימאדאל** – xAI ארבעט אז Grok זאל קענען נעמען בילדער: מען וועט קענען ארויפלייגן א בילד און באקומען א טעקסט-ענטפער.
- **22 מיי: ElevenLabs Reader** – ElevenLabs האט לאנסירט די Reader אפליקאציע, וואס לייענט אויף א קול טעקסט פון פארשידענע דאקומענטן מיט 11 פארשידענע קולער.
- **23 מיי: Aurora וועטער-מאדעל** – Microsoft האט טריינירט א יסוד-מאדעל פאר פאראויסזאגן די אטמאספערע. ער האט געשטעלט א נייעם רעקארד אויף טעסטן פון 5 און 10 טעג גלאבאלע וועטער פאראויסזאגונג.
- **23 מיי: MiniCPM-V אויף Llama 3** – א נייע 8B ווערסיע פון MiniCPM-V, געבויט אויף Llama 3, שלאגט אסאך פארמאכטע מאדעלן. זי ארבעט אין 30 שפראכן און איז שטארק אין OCR און אין ענטפערן פראגעס אויף בילדער.
- **23 מיי: Mistral 7B v3** – Mistral האט ארויסגעלאזט די נעקסטע ווערסיע פון איר 7B מאדעל, מיט א לענגערן קאנטעקסט און בעסערער פערפארמענס.
- **24 מיי: Cohere Aya** – דער Aya פראיעקט האט דריי מאדעלן פון פארשידענע גרייסן, וואס רעדן 101 שפראכן, פון זיי אסאך מיט זייער ווייניג דאטא. עס איז א גרויסער שריט פארן אפענעם פארשונג-געמיינדע.
- **29 מיי: Copilot אנגענטן** – ביי Build 2024 האט Microsoft באוויזן Team Copilot פאר ארבעטן אין א גרופע, אייגענע AI אנגענטן צו אויטאמאטיזירן ארבעט-גאנגען, און Copilot Extensions און Connectors. זיי זענען יעצט אין א פריוואטער פרובירונג, מיט א ברייטערער פארשפרייטונג ערווארט שפעטער אין 2024.
- **30 מיי: Cartesia Sonic** – Cartesia, די פירמע פון די שאפער פון SSMs און Mamba, האט טריינירט Sonic, א בליץ-שנעלער און רעאליסטישער קול-מאדעל. דאס ווייזט אז זיי ווילן אריינגיין אין דעם געביט פון AI געהילפן.
- **30 מיי: Codestral** – Mistral האט ארויסגעלאזט א 22B קאוד מאדעל מיט 32k קאנטעקסט. ער איז שנעל און שטארק אין אסאך פראגראמיר-שפראכן, מיט אפענע וואגן, און איז צוטריטלעך דורך Mistral'ס פלאטפארמע.
- **31 מיי: Perplexity Pages** – Perplexity האט באוויזן א פראדוקט צו שאפן מיט AI פארשונג-ארבעטן וואס בלייבן און וואס מען קען טיילן מיט אנדערע.

- **31 מיי:** Anthropic – **Claude tool use** פירט איין "tool use" פאר Claude. באניצער קענען בויען געהילפן וואס פארבינדן זיך מיט סיי וועלכע דרויסנדיגע API. זיי קענען אנאליזירן דאטא, רעקאמענדירן פראדוקטן, נאכפאלגן באשטעלונגען, געבן טעכנישע הילף און באארבעטן בילדער.
- **31 מיי:** Google – **Chromebook AI** האט צוגעלייגט AI פיטשערס צו Chromebook Plus: א שרייב-געהילף, א מכשיר צו שאפן וואלפעפער, און גרינגן צוטריט צו Gemini.

פארשונג און טעכניק

- **1 מיי:** Mamba פאר בילדער – עטליכע ארבעטן האבן אונטערגעזוכט די Mamba ארכיטעקטור פאר בילדער. Mamba באהאנדלט לאנגע סעקווענצן אן די חסרונות פון Transformers, און א ווערסיע טריינירט מיט CLIP איז שטארק אין zero-shot קלאסיפיקאציע. אויך איז ארויס א פולשטענדיגע איבערבליק-ארבעט. אבער פארשער האבן געפונען אז Mamba איז נישט אפעקטיוו פאר בילד-קלאסיפיקאציע, נאר אויסזיכטפול פאר דערקענען און צעטיילן אביעקטן. אן אנדערע ארבעט האט געוויזן אז א גוט-אויסגעשטעלטער CNN שלאגט Mamba אויף א ריי בילד-ארבעטן.
- **1 מיי:** פריע ארבעטן – Seismic ארגאניזירט די אינדעקסן פאר טעקסט-זוכן שנעלער און גענויער. CIT העלפט גראף נעצן זיך צופאסן צו ענדערנדע סטרוקטורן. Meta-prompting צעטיילט א קאמפליצירטע ארבעט צווישן ספעציאליזירטע ווערסיעס פונעם זעלבן מאדעל, און דאס פארבעסערט די רעזולטאטן.
- **2 מיי:** KAN נעצן – (Kolmogorov-Arnold Networks) KAN ניצט געלערנטע אקטיוואציע-פונקציעס אויף די קאנטן אנשטאט פעסטע, און פארבייט די וואגן מיט splines. Wav-KAN ניצט wavelet פונקציעס פאר שנעלערע טריינירונג, מער פעסטקייט און בעסער פארשטיין וואס דער מאדעל טוט.
- **2 מיי:** Reka Vibe-Eval – Reka, וואס האט כמעט דערגרייכט די גרעסטע שפילער מיט א קליינעם טייל פונעם געלט, האט ארויסגעלאזט א טייל פון אירע אינעווייניגסטע טעסטן. דער בענטשמארק Vibe-Eval האט 269 פראמפטס צו טעסטן בילד-פארשטאנד, פון זיי 100 באזונדערס שווערע.
- **3 מיי:** קאנטעקסט ביו 128K – א פארשונג ווייזט וויאזוי צו פארלענגערן דעם קאנטעקסט ביו 128K, און באטאנט אז ס'איז וויכטיג סיי וויפיל טריינירונג-דאטא מען האט און סיי ווי פארשידנארטיג זי איז.
- **3 מיי:** בילד און קלאנג פארשונג – FFDConv פארבעסערט דערקענען קלאנג-געשעענישן. אן ארבעט ווייזט וויאזוי מען מאכט Vision Transformers מער עפעקטיוו דורך quantization און צוגעפאסטע הארדוועיר. Spikformer V2 פארבינדט self-attention מיט ספייקינג נעצן, וואס שפארן ענערגיע. ASAM פארבעסערט SAM'ס דורך SUNDAAE adversarial tuning שפארט זכרון ביי 3D רענדערינג. InstructDr איז שטארק אין פארשטיין דאקומענטן פון בילדער.
- **6 מיי:** א מאדעל פון א קינד'ס אויגן – ד"ר Brenden Lake פון NYU רעקארדירט וואס זיין טאכטער זעט, כדי צו בויען א שפראך-מאדעל וואס לערנט ווי א קליין קינד, אנשטאט פון ריזיגע דאטא-סעטן.
- **6 מיי:** נאך פארשונג – FeNNol ערלייכטערט מאלעקולארע סימולאציעס מיט היברידע נעצן. Spider פארשטייט זאכן וואס האבן א טייטש נאר לויטן קאנטעקסט, ווי באהאלטענע אביעקטן אדער מעדיצינישע וואונדן. א נייער דאטא-סעט און מאדעל פאלגן אנווייזונגען איבער עטליכע בילדער. RaffesDG פארבעסערט מעדיצינישע בילד-מאדעלן ווען די דאטא קומט פון פארשידענע מקורות. Google האט ארויסגעלאזט 15 טויזנט בילדער מיט זייער פרטימדיגע לעיבלס.

- **9 מיי:** בילד-באשרייבונגען און **quantization** Image In Words – ניצט צוויי רונדעס פון VLMs צו שאפן בילד-באשרייבונגען מיט מער פרטים ווי סיי וועלכער פריערדיגער דאטא-סעט. א דיפוזיע-מאדעל מיט unidirectional block attention שאפט בילדער מיט זייער הויכע רעזאלוציע. QoQ פארשנעלערט מאדעלן מיט 4-ביט וואגן, 8-ביט אקטיוואציעס און א 4-ביט KV cache. ChemFlow נאויגירט דעם כעמישן רוים מיט גענעראטיווע מאדעלן.
- **10 מיי:** YOCO און **Consistency LLMs** – YOCO איז א decoder-decoder ארכיטעקטור וואס פארמינערט שטארק דעם GPU זכרון און די ווארט-צייט, און האלט אבער גלאבאלע attention. Consistency Language Models ניצן פאראלעלע דעקאדירונג צו גענערירן מער ווי איין טאקען אויף איינמאל, מיט א פריע פארשנעלערונג פון 3x, גלייך ווי speculative decoding.
- **10 מיי:** דאטא און יושר – DiffMatch ניצט VLMs צו מאכן לעיבלס פאר באמערקן ענדערונגען אין בילדער. Buzz איז א טריינירונג-דאטא-סעט וואס נעמט אריין דאטא וועגן וואס מענטשן האבן בעסער ליב, און די מאדעלן וואס זענען דערמיט טריינירט שניידן גוט אפ אויף אזעלכע טעסטן. אן אלגאריטם מיט א "fairness cost" רעגולירט נאכדעם די רעזולטאטן כדי זיי זאלן זיין יושרדיג צו פארשידענע גרופעס. א דיסקוסיע איבער Gemma מיט 10M קאנטעקסט ווייזט פארשידענע וועגן צו פארלענגערן דעם קאנטעקסט.
- **13 מיי:** טאקענס און קערנעלס – א פארשונג ווייזט אז אפילו שטארקע מאדעלן האבן טאקענס וואס זענען קיינמאל נישט טריינירט געווארן, און ווי צו געפינען זיי. ThunderKittens איז א נייע שפראך פאר CUDA קערנעלס; זיין flash attention איז 30% שנעלער און נעמט נאר 100 שורות קאוד. Mirage שאפט זייער שנעלע קערנעלס, און E2URec לאזט רעקאמענדאציע-סיסטעמען פארגעסן א באניצער'ס דאטא.
- **13 מיי:** Lumina – Lumina וויל זיין איין מאדעל וואס מאכט פון טעקסט סיי וואס. ער איז טריינירט אויף בילדער, ווידעא, קלאנג און טעקסט צוזאמענגעמישט.
- **14 מיי:** פארשידענע מאדעלן – MatterSim סימולירט אטאמען פון פארשידענע עלעמענטן ביי פארשידענע טעמפעראטורן און דרוקן. א טראנספארמער-מעטאד מאכט שנעלער גראפן פון א בילד'ס סצענע. InternLM-XComposer2 שאפט און פארשטייט טעקסט מיט בילדער, מיט א TEFN. Partial LoRA פארבעסערט לאנג-טערמיניגע פאראויסזאגונגען, MRSegmentator דערקענט 40 ארגאנען אין MRI בילדער, און MoonDream האט מיט COYO צוגעגרייט 5 מיליאן נייע בילד-באשרייבונגען.
- **15 מיי:** state-space און **ESDNet** – א נייער state-space מאדעל האט אן אלגאריטם וואס לויפט פאראלעל אן א "state". ESDNet איז א ספייקינג נעץ וואס נעמט אראפ רעגן פון בילדער.
- **16 מיי:** בילד און ווידעא – Hunyuan-DiT שטעלט א נייעם שטאפל אין מאכן בילדער פון טעקסט, באזונדערס אויף ענגליש און כינעזיש. Coin3D לאזט דעם באניצער אריינגעבן א גראבע פארם בעפאר מען שאפט 3D. PAG פארבעסערט בילדער פון דיפוזיע-מאדעלן אן נאך טריינירונג. IC-Relight טוישט די ליכטיגקייט אין בילדער. EfficientTrain++ פארקירצט די טריינירונג פון ResNet און Swin ביז דריי מאל. SqueezeTime איז א לייכטע נעץ צו פארשטיין ווידעא אויף טעלעפאנען, און Hopfield Boosting דערקענט דאטא וואס איז אנדערש פון דער טריינירונג.
- **17 מיי:** Cursor'ס שנעלע רעדאקטירונג – פארשער האבן גענוצט אן אייגענעם Llama 3 70B מיט speculative editing צו רעדאקטירן פיילס מיט 1,000 טאקענס א סעקונדע, כמעט מיטאמאל, און דאס אן diffs.

- **17 מיי:** בילד-פארשונג – א נייער וועג צו שאצן א אביעקט'ס פאזיציע פון RGB-D בילדער ניצט דענסע קארעספאנדענץ. CDFormer פארבעסערט די רעזאלוציע פון בילדער. Xmodel-VLM איז א 1B בילד-שפראך מאדעל פאר געווענליכע GPUs. א טעכניק פולט אויס לאכער אין LiDAR סקענס פאר אויטאנאמע קארס, און MasterWeaver פארבעסערט פערזענליכע בילדער.
- **20 מיי:** LoRA און Gemini באריכט – א פארשונג געפינט אז LoRA לערנט ווייניגער ווי פולע fine-tuning, אבער פארגעסט אויך ווייניגער פון דעם וואס דער מאדעל האט פריער געקענט. Google'ס נייער טעכנישער באריכט פאר Gemini 1.5 זאגט ווייניג וועגן די מאדעלן און דאטא, אבער פיל וועגן טעסטן און זיכערהייט.
- **21 מיי:** 3D און זיכערהייט – LeMeViT פארמינערט די קאסטן פון Vision Transformers. Art 3D קינסטלערישע 3D סצענעס פון טעקסט. אנדערע ארבעטן פארשטייען וואו זאכן זענען אין 3D (און מיט PSG-4D אויך איבער צייט). SWARM איז א נייער אטאקע וואס ניצט א "switch token" צו אקטיווירן געהיים א הינטער-טיר אין Vision Transformers.
- **22 מיי:** Claude'ס אינעווייניג – Anthropic האט געמאפירט מיליאנען באגריפן ("features") אינעווייניג אין Claude 3 Sonnet. דורך פארשטארקן א feature קען מען פירן דעם מאדעל'ס ענטפערס; למשל האבן די פארשער אים געמאכט גלייבן אז ער איז דער Golden Gate בריק. דער ציל איז צו מאכן AI זיכערער.
- **22 מיי:** בילד-שפראך – א Smart Expert System ניצט LLMs פאר טעקסט-קלאסיפיקאציע מיט ווייניגער צוגרייטונג. CSTA פארבעסערט ווידעא-סיכומים. אן ארבעט פארמינערט די נייגונג פון בילד-שפראך מאדעלן זיך צו פארלאזן אויף שפראך אנשטאט אויפן בילד. DeepSeek-VL איז אן אפן-סאורס בילד-שפראך מאדעל פאר פראקטישן באנוץ.
- **23 מיי:** מאטעמאטיק און מעדיצין – MathBench איז א פולשטענדיגער בענטשמארק פאר מאטעמאטיק. MedLFQA און OLAPH פארבעסערן די גענויקייט פון לאנגע מעדיצינישע ענטפערס.
- **24 מיי:** נאך ארבעטן – GLSCL פארבעסערט זוכן ווידעא לויט טעקסט ביי נידריגערע קאסטן. ProtT3 פארבינדט פראטעין-מאדעלן מיט שפראך-מאדעלן צו באשרייבן פראטעיןען. SirLLM העלפט מאדעלן געדענקען לענגער אין לאנגע שמועסן, און א דיפוזיע-מאדעל באשרייבט ענדערונגען אין סאטעליט-בילדער.
- **27 מיי:** Lyft – Lyft האט גענוצט online reinforcement learning, באלוינט לויט די צוקונפטיגע פארדינסטן פון די דרייווערס, צו צופאסן דרייווערס מיט פאסאזשירן. לויט דער שאצונג ברענגט דאס בערך \$30 מיליאן א יאר מער.
- **27 מיי:** עוואלוציע און מעטאדן – אן ארבעט ווייזט קלארע, איבערחזר'בארע קריטעריעס צו טעסטן שפראך-מאדעלן. אנדערע ארבעטן: בילדער וואס האלטן א מענטש'ס אידענטיטעט אן נאך טריינירונג, LoRA-Ensemble, modular norm פאר טריינירונג פון נעצן פון אלע גרייסן, MobileNet-V4 וואס לויפט שנעל אויף קליינע מכשירים, און א פארשונג צי מאדעלן ניצן מערדימענסיאנאלע features.
- **28 מיי:** בילד-צעטיילונג – Scribble2Scene פארמינערט די לעיבלס וואס מען דארף צו פארענדיגן 3D סצענעס. HDC און SSA פארבעסערן צעטיילן בילדער. RSBuilding פארבינדט ארויסנעמען געביידעס און באמערקן ענדערונגען אין סאטעליט-בילדער. Meteor פארבעסערט בילד-שפראך מאדעלן, און InstructAvatar שאפט 2D אוואטארן לויט טעקסט אנווייזונגען.

• **29 מיי:** סקעילינג און אפטימיזער – אן ארבעט ניצט דעם gzip קאמפרעסיע-פראפארציע צו פאראויסזאגן סקעילינג געזעצן לויט דער דאטא. נייע embeddings, וואס מערקן יעדע ציפער'ס פאזיציע, לערנען טראנספארמערס רעכענען. Meta's Schedule Free אפטימיזער איז אן אלטערנאטיוו צו Adam, באזונדערס ווען מען ווייט נישט פון פאראויס וויפיל שריט מען וועט טריינרן. Vista איז א וועלט-מאדעל פאר אויטאנאמע דרייווינג.

• **30 מיי:** PRISM – PRISM איז א יסוד-מאדעל טריינירט אויף 1.2 ביליאן מאסע-ספעקטן פון קליינע מאלעקולן, צו פארשנעלערן געפינען נייע מעדיצינען. דערצו: DMPlug פאר inverse פראבלעמען, PatchScaler פאר שנעלערע העכערע רעזאלוציע, און Reason3D צו פארשטיין 3D סביבות.

• **31 מיי:** Yuan 2.0-M32 – MoE מאדעל מיט 40B פאראמעטערס, פון זיי 3.7B אקטיוו, וואס קומט נאנט צו Llama 3 70B מיט 1/19 פונעם רעכנ-כח. ער איז טריינירט אויף 2T טאקענס. אן אנדערע ארבעט שלאגט פאר א קאנסטאנטע learning rate מיט אן אפקילונג צום סוף, כדי ביליגער צו שאפן סקעילינג געזעצן.

• **31 מיי:** בילד און ווידעא – 4K-Rain13k איז א דאטא-סעט פון 13,000 פארן 4K בילדער פאר אראפנעמען רעגן. SDXL Scribble ControlNet לאזט צייכענען ליניעס צו פירן דאס בילד. EasyAnimate מאכט ווידעא מיט טראנספארמערס. SELM פארבעסערט preference optimization, און Text2Video Turbo פארמינערט די שריט צו מאכן ווידעא.

געצייג פאר פראגראמירער און אפן-סאורס

• **1 מיי:** ExecuTorch – Meta's ExecuTorch לאזט לויפן Llama מאדעלן אויף iPhone און Galaxy טעלעפאנען. ביי 7B מאדעלן דערגרייכט עס ביז 11 טאקענס א סעקונדע.

• **1 מיי:** פרייע געצייג – Nvidia's CUDA Checkpointing שפייכלט דעם CUDA צושטאנד אויף צו איבערפירן און ווידער אנהייבן, נוצבאר פאר צעשפרייטע טריינירונג. א רעפאזיטאריע ווייזט ווי צו מאכן גוטע גראפיקן פון רעזולטאטן. QUICK איז אפטימיזירטע CUDA קערנעלס פאר NVIDIA GPUs. FlowTestAI איז אן אפן-סאורס IDE פאר API ארבעט-גאנגען, און Phospho אנאליזירט טעקסט פון LLM אפליקאציעס.

• **2 מיי:** לערנען און 3D – א בלאג-ארטיקל דערקלערט דעם טראנספארמער דורך Gemma, מיט קאוד און בילדער. דערצו: אן איבערבליק פון RAG מאדעלן, Lightplane צו שפארן זכרון ביי 3D, און MicroDreamer וואס פארשנעלערט 3D גענעראציע.

• **3 מיי:** StreamMultiDiffusion – א געצייג צו שאפן בילדער פון טעקסט לויט געגנטן, אין רעאלער צייט.

• **6 מיי:** Bonito און Prometheus – Bonito מאכט אנווייזונג-דאטא-סעטן פון רויען טעקסט. Prometheus, געבויט אויף Mistral, איז א מאדעל וואס שאצט אפ די ענטפערס פון אנדערע מאדעלן. Penzai איז א JAX ביבליאטעק צו זען און באהאנדלען מאדעלן. דערצו: א טוטאריאל צו ציילן קארס אין ווידעא, און PeopleGPT פאר זוכן ארבעטער.

• **7 מיי:** Unsloth און Hugging Face – א ווידעא פונעם גרינדער פון Unsloth ווייזט ווי זיי ניצן PyTorch און שרייבן קערנעלס. Hugging Face האט ארויסגעלאזט א ביבליאטעק פאר ראבאטיק, און Morph-Tokens פארבעסערן פארשטיין און שאפן בילדער.

- **9 מיי:** **DocRes** און **Copilot Workspace** – DocRes פאררעכט דאקומענט-בילדער (גלייכמאכן, אראפנעמען שאטנס און טונקלקייט, אא"וו) אינעם זעלבן סיסטעם. GitHub Copilot Workspace, יעצט אין א באגרענעצטער פרובירונג, לאזט פראגראמירן אויף פשוטע ענגליש, פון פלאנירן ביז אויספירן, אינעם בראווער.
- **13 מיי:** **AnythingLLM** און **Bedrock Studio** – AnythingLLM איז א פולער סיסטעם צו לויפן LLMs מיט RAG. Cleanlab האט אויפגעשטעלט א פלעץ צו פרובירן איר Trustworthy Language Model. Amazon האט לאנסירט Bedrock Studio צו פרובירן, טריינירן און ארויסשטעלן מאדעלן.
- **14 מיי:** **Pipecat** און **WebLlama** – Pipecat איז א סיסטעם צו בויען קול-אגענטן. WebLlama איז א מאדעל וואס זוכט אויפן וועב און ענטפערט פראגעס. Eraser AI מאכט דיאגראמען און דאקומענטן לויט א פראמפט.
- **15 מיי:** **Genkit** – **Firestore** Genkit איז א נייער סעט געצייג פאר Firestore צו בויען גענעראטיווע פראדוקטן, און קען לויפן סערווערס פאר אפן-סאורס מאדעלן.
- **16 מיי:** **Project IDX** – Google's AI-IDE אינעם בראווער איז יעצט אין אן אפענער beta, נאכדעם וואס איבער 100,000 פראגראמירער האבן עס שוין פרובירט. מען האט צוגעלייגט פארבינדונגען מיט Google Maps Platform, Chrome Dev Tools און Lighthouse.
- **17 מיי:** **Llama3.np** – א פשוטע ימפלעמענטאציע פון Llama 3 נאר אין NumPy.
- **17 מיי:** **Hugging Face GPUs** – Hugging Face שענקט \$10 מיליאן אין פרייע GPUs פאר פראגראמירער, אקאדעמיקער און סטארטאפס, קעגן דעם אז די גרויסע פירמעס האלטן אלעס ביי זיך.
- **20 מיי:** **Model Explorer** און **נאך** – Google's Model Explorer ווייזט דעם רעכן-גראף פון א מאדעל. Experts פארבינדט OpenAI Assistants אין א "פאנעל פון עקספערטן". garple.com, א זייטל וואס פארקויפט קורצע com. דאמענס, איז אינגאנצן געשריבן דורך אן AI אגענט. Julius AI האט לאנסירט דעם ערשטן AI קאוד אינטערפרעטער פאר R, און MoonDream לויפט אינעם בראווער מיט WebGPU.
- **21 מיי:** **Devon** און **Llama 3** – Devon איז אן אפן-סאורס פראגראמיר-שותף. א טוטאריאל ווייזט יעדן שריט פון אויפבויען Llama 3 פון אנהייב און לויפן די וואגן.
- **22 מיי:** **Auto Wiki** – אויטאמאטישע דאקומענטאציע פאר סיי וועלכע רעפאזיטאריע.
- **23 מיי:** **Tarsier** – **Reworkd**'s Tarsier מערקט צו עלעמענטן אויף וועבזייטלעך, אזוי אז א מאדעל אן ראיא פארשטייט דאס זייטל. דאס שלאגט בילד-שפראך מאדעלן אויף בענטשמארקס.
- **24 מיי:** **Llama.cpp** **צעשפרייט** – Llama.cpp שטיצט יעצט צו לויפן א מאדעל צעטיילט איבער עטליכע מאשינען, דערווייל נאר אין FP16. דערצו: א וועגווייזער צו מאכן אינטעראקטיווע embedding בילדער, Krea AI פאר ווידעא פון שליסל-בילדער, און Glyph ByT5 וואס פארבעסערט טעקסט אינעווייניג אין בילדער.
- **27 מיי:** **Mistral Finetune** און **Llamafile** – Mistral האט ארויסגעלאזט אן אפיציעלע רעפאזיטאריע צו טריינירן אירע מאדעלן. דאנק Mozilla's Llamafile לויפט Mixtral 8x22b מיט 20 טאקענס א סעקונדע אויף א געווענליכן CPU \$400.

- **28 מיי:** מעסטן און דיאגראמען – א געצייג גיט א סך מאסן צו מעסטן ווי גוט דאס זוכן ארבעט אין א RAG סיסטעם, און א וועגווייזער ווייזט ווי צו צייכענען דיאגראמען פון נייראנישע קרייזן.
- **29 מיי:** Ilm.c – Andrej Karpathy'ס Ilm.c קען יעצט טריינירן דעם קלענסטן GPT-2 (124M) אין 90 מינוט פאר \$20, מיט ווייניג אפהענגיקייטן. Marigold פאר טיפקייט איז יעצט אין diffusers, און LaVague איז אן אפן-סאורס סיסטעם פאר וועב אגענטן.

פירמעס, געלט און ביזנעס

- **1 מיי:** Apple'ס AI פלענער – Apple האט אריבערגעצויגן AI עקספערטן פון Google און געעפנט א געהיימע לאבאראטאריע אין ציריך, "Vision Lab", וואס בויט מאדעלן וואס פארבינדן טעקסט מיט בילד. Safari 18 וועט האבן Intelligent Search וואס פארקירצט אינהאלט, און א "Web Eraser". Apple איז נאנט צו אן אפמאך מיט OpenAI צו ברענגען ChatGPT אויף iPhone אין iOS 18. איר אסטראטעגיע איז פראקטישע פיטשערס (Project Greymatter), און קלאוד-דאטא וועט ווערן באארבעט אין א פארזיכערטן "black box".
- **1 מיי:** Diddo און Cognition – Diddo האט אויפגעשטעלט געלט פאר אן API וואס לאזט קויפן גלייך פון א סטרימינג פראגראם, אן QR קאודס. Cognition Labs זוכט א ווערט פון \$2 ביליאן פאר Devin, איר AI וואס שרייבט קאוד.
- **2 מיי:** Coreweave – Coreweave, וואס פארדינגט GPU קלאסטערס, האט אויפגעשטעלט \$1.1 ביליאן ביי א ווערט פון \$19 ביליאן.
- **2 מיי:** Field AI – Field AI, געגרינדעט דורך א געוועזענעם פירער ביי NASA JPL, בויט ראבאטן וואס ארבעטן אן מאפעס, GPS אדער השגחה, פאר אינדוסטריע און בויען.
- **2 מיי:** סטארטאפס אין שוועריגקייטן – Anthropic און Inflection AI, Stability AI האבן געלט-דרוק פון די הויכע קאסטן, בשעת OpenAI'ס איינקונפט וואקסט. Stability AI האט גערעדט מיט מעגליכע קונים צו פארקויפן די פירמע. Inflection האט א נייע פירערשאפט אונטער Reid Hoffman, וואס זאגט אז די פירמע האט נאך געלט אויף 18 חדשים.
- **3 מיי:** Suleyman קעגן Hassabis – Mustafa Suleyman, מיט-גרינדער פון DeepMind, איז יעצט CEO פון Microsoft AI. זיין געוועזענער שותף Demis Hassabis פירט די AI פארשונג ביי Google.
- **3 מיי:** דאטא צענטערס – Meta בויט א דאטא צענטער פאר \$800 מיליאן אין Montgomery, וואס זאל זיין גרייט ביזן סוף 2026. Microsoft אינוועסטירט \$3.3 ביליאן אין וויסקאנסין.
- **6 מיי:** געלט-רונדעס – Lamini האט באקומען \$25 מיליאן (Series A) פאר אן AI פלאטפארמע פאר פירמעס. DatologyAI האט באקומען \$46 מיליאן פאר אויסקלויבן דאטא, און Yellow האט באקומען \$5 מיליאן פאר 3D פיגורן.
- **6 מיי:** OpenAI'ס זוכעכץ – OpenAI בויט א זוכ-מאשין פאר ChatGPT, מעגליך אויף Bing'ס אינפראסטרוקטור, וואס קען קאנקורירן מיט Google. מען האט ערווארט אז די אסיפה פון 13טן מאי וועט ברענגען דערהייבונגען צו ChatGPT און GPT-4.
- **7 מיי:** OpenAI'ס צייטונגען-אפמאכן – OpenAI האט אונטערגעשריבן שותפותן מיט Stack Overflow, Reddit (צוטריט צו זייער Data API), News Corp, The Atlantic, און Vox Media, צו ניצן זייער

- אינהאלט מיט אנגעבן דעם מקור. א דורכגעזיקערטע פרעזענטאציע ווייזט אז דער Preferred Publishers Program גיט צייטונגען געלט און מער זעבארקייט אין ChatGPT.
- **7 מיי: Grok נייעס** – Elon Musk פלאנט אז Grok זאל צונויפלייגן נייעס פון די שמועסן אויף X, מיט מקורות. עס זענען דא שוועריגקייטן מיט אנגעבן מקורות און געזעצליכע פראגעס.
 - **10 מיי: Mistral** – Mistral שטעלט אויף געלט ביי א ווערט פון \$6 ביליאן, דריי מאל מער ווי אין דעצעמבער. זי האט זיך פארבונדן מיט Harvey, וואס וועט ווארשיינליך מיינען א מאדעל פאר אדוואקאטן. זי האט אויך אריינגעפירט א נייע לייסענס, MNPL, נעבן Apache 2.0.
 - **13 מיי: Altera** – Altera האט באקומען \$9 מיליאן פאר AI אגענטן וואס שפילן שפילן מיט א מענטש.
 - **15 מיי: Anthropic אין אייראפע** – Claude איז יעצט צוטריטלעך אין אייראפע אויפן וועבזייטל, אויף iOS, און פאר טימס, און Anthropic הייבט אן אויפשטעלן מער געלט.
 - **15 מיי: xAI** – xAI פירט פארהאנדלונגען צו דינגען סערווערס פון Oracle פאר \$10 ביליאן, און האט מודיע געווען א Series B פון \$6 ביליאן פאר אינפראסטרוקטור, פראדוקטן און פארשונג.
 - **16 מיי: AI הוצאות** – לויט Ramp זענען AI הוצאות ביי אירע קונים געוואקסן 293% אין א יאר. אויך פירמעס אויסער טעכנאלאגיע, ווי געזונט און פינאנצן, ניצן עס.
 - **17 מיי: Meta אויערן** – Meta ארבעט לויט ידיעות אויף "Camerabuds", אויערן מיט קאמערעס וואס דערקענען אביעקטן און איבערזעצן שפראכן.
 - **20 מיי: Unitree G1** – Unitree G1 הומאנאיד ראבאט קאסט \$16,000. ער האט 3D LiDAR און א באטעריע פון 9,000 mAh, וועגט בערך 35 קילא, און לייגט זיך צוזאם קליין.
 - **20 מיי: געלט-רונדעס** – Gamma האט באקומען \$12 מיליאן; 18 מיליאן באניצער האבן שוין געשאפן איבער 60 מיליאן פרעזענטאציעס. Scale AI האט באקומען \$1 מיליאן (Series F). Suno האט באקומען \$125 מיליאן פאר מוזיק, און H, געגרינדעט דורך DeepMind פארשער, האט באקומען \$220 מיליאן אין א seed רונדע. Patronus AI האט באקומען \$17 מיליאן פאר זיכערהייט פון מאדעלן, און Bootloader Studio האט באקומען \$5 מיליאן.
 - **21 מיי: שותפותן** – SoundHound ניצט Perplexity פאר דערהיינטיגטע ענטפערס, און Opera האט אריינגעשטעלט Gemini אין Aria.
 - **27 מיי: Character.ai** – xAI און Meta שלאגן זיך ווער עס וועט זיך פארבינדן מיט Character.ai, געגרינדעט דורך Noam Shazeer.
 - **28 מיי: Humane צו פארקויפן** – Humane זוכט צו פארקויפן די פירמע פאר \$750 מיליאן ביז \$1 ביליאן, נאך א שוואכן אנהייב פונעם AI Pin.
 - **30 מיי: PwC און Midjourney** – PwC איז OpenAI'ס גרעסטער פירמע-קונה, מיט 100,000 באניצער, און דער ערשטער שותף וואס פארקויפט OpenAI'ס ענטערפרייז פראדוקט ווייטער. Midjourney האט צונויפגעשטעלט א הארדוועיר טים.
 - **31 מיי: Wonder Dynamics** – Autodesk האט געקויפט Wonder Dynamics, א פירמע פאר AI עפעקטן אין פילמען, פאר א נישט-אויסגעזאגטע סומע.

רעגירונגען, געזעצן און זיכערהייט

- **3 מיי:** צייטונגען קלאגן – אכט צייטונגען וואס געהערן צו Alden Global Capital קלאגן OpenAI און Microsoft אויף קאפירעכט, נעבן דעם פריערדיגן קלאג פון New York Times.
- **3 מיי:** Microsoft און פאליציי – Microsoft פארבאט אמעריקאנער פאליציי צו ניצן Azure OpenAI פאר דערקענען פנימער.
- **6 מיי:** AI פליגער – די אמעריקאנער לופט-מאכט צילט צו האבן איבער 1,000 פליגער אן א פילאט ביז 2028.
- **7 מיי:** Microsoft באריכט – אין איר ערשטן באריכט פון אחריות'דיגן AI זאגט Microsoft אז זי האט אין 2023 געשאפן 30 AI געצייג און אריינגעשטעלט זיכערהייט-שריט.
- **9 מיי:** OpenAI Model Spec – OpenAI האט ארויסגעגעבן א דאקומענט וואס באשרייבט ווי זי וויל אז אירע מאדעלן זאלן זיך פירן אין פארשידענע מצבים.
- **13 מיי:** קאליפארניע SB1047 – דער געזעץ-פארשלאג וועט רעגולירן מאדעלן מיט איבער 10^{26} FLOPs. ער פאדערט א זיכערע סביבה, א מעגליכקייט אפצושטעלן דעם מאדעל שנעל, און טעסטן קעגן שלעכטן באנוץ.
- **14 מיי:** DALL-E דערקענונג – OpenAI האט לאנסירט א געצייג וואס דערקענט בילדער פון DALL-E 3, און ניצט C2PA סטאנדארטן.
- **16 מיי:** טומל ביי OpenAI – Ilya Sutskever גייט אוועק, און Jakub Pachocki ווערט דער נייער הויפט-פארשער. נאכן אוועקגיין פון Sutskever און Jan Leike איז דער superalignment טים צעלייגט געווארן. ס'איז ארויסגעקומען אז געוועזענע ארבעטער קענען פארלירן זייערע אקציעס אויב זיי קריטיקירן די פירמע. Sam Altman האט צוגעזאגט עס צו פאררעכטן, און OpenAI האט שפעטער באפרייט די געוועזענע ארבעטער פון די אפמאכן. לויט Vox האבן די טים-מיטגלידער פארלוירן צוטרוי אין Altman.
- **21 מיי:** Sky קול – OpenAI האט אפגעשטעלט דעם "Sky" קול נאכן טענה אז עס איז נאכגעמאכט פון א שוישפילערין'ס קול אן איר רשות. זי זוכט געזעצליכע עצה.
- **21 מיי:** זיכערהייט-פלענער – Google DeepMind'ס Frontier Safety Framework באשטימט שטאפלען פון געפערליכע פעאיקייטן, טעסט די מאדעלן קעגן זיי, און נעמט שריט ווען מען דערגרייכט זיי. Anthropic ציילט אויף וואס זי האט זיך געלערנט פון איר Responsible Scaling Policy.
- **23 מיי:** שקיפות אינדעקס – Stanford HAI האט אפגעשאצט 14 גרויסע AI פירמעס. זיי האבן ארויסגעגעבן מער אינפארמאציע ווי פריער, אבער די דורכשניטליכע ציפער איז בלויז 58 פון 100.
- **24 מיי:** Prove AI – Casper Labs מיט IBM האבן געשאפן Prove AI צו קענען נאכקוקן AI ביי פירמעס.
- **29 מיי:** זיכערהייט-קאמיטעט – OpenAI האט געשאפן א קאמיטעט פאר זיכערהייט און באוואונטגעמאכט אז זי טריינירט א נייעם מאדעל. Jan Leike איז אריבער צו Anthropic צו פירן א נייעם superalignment טים.

אנאליזן און מיינונגען

- **3 מיי:** NYC און פרייזן – Lux Capital זעט אז ניו יארק וואקסט אלס AI צענטער, מיט א \$400 מיליאן פארשונג-קאנסארציום. אן אנאליז ווייזט אז רוב AI אפליקאציעס רעכענען א מאנאטליכן אבאנעמענט פער באניצער, בשעת נייע פירמעס פרובירן רעכענען לויטן רעזולטאט.
- **6 מיי:** AI הארדוועיר – Humane Pin און Rabbit R1 זענען קריטיקירט געווארן, אבער לויט Every איז דא נאך אסאך אינוועסטירונגס-געלט אין דעם פעלד, און Sam Altman קוקט אויף גרויסע אינוועסטירונגען.
- **7 מיי:** פראמפטס און ראיה – IEEE Spectrum זאגט אז דאס האנט-שרייבן פון פראמפטס וועט ווערן איבריג. א פארשונג ווייזט אז GPT-4V און ענליכע מאדעלן האבן שוועריגקייטן מיט שווערע בילד-רעטענישן. Stanford האט ארויסגעגעבן א באריכט מיט 13 גראפיקן.
- **9 מיי:** אויסגעמוטשעטע אינזשענירן – AI אינזשענירן דערציילן פון אויסמוטשונג און פון ארבעט וואס ווערט ארויסגעלאזט אין א יאגעניש, מער פאר די אינוועסטארן ווי פאר די באניצער.
- **13 מיי:** AGI – Ethan Mollick דערמאנט פאראויסזאגונגען אז AGI וועט קומען צווישן 2027 און 2047, און זאגט אז דערווייל איז AI בעסטנס אלס מיטהעלפער. Ben Evans זאגט אז מען ווייסט נישט, ווייל ס'איז נישטא קיין קלארע טעאריע פון אינטעליגענץ.
- **14 מיי:** פארשונג און Copilot – לויט Liang et al. איז ביז 18% פון פארשונג-ארבעטן אין געוויסע פעלדער געשריבן מיט AI הילף. Microsoft לייגט צו אויטאמאטישע פראמפט-פארענדיגונג צו Copilot.
- **16 מיי:** פראגראמירן און שכירות – לערער לייגן ווייניגער געוויכט אויף סינטאקס און מער אויף העכערע פעאיקייטן. אן ארטיקל זאגט אז AI מאכט אינטערוויוס שווערער, און שלאגט פאר אינטערוויוס פנים אל פנים. אן אנדערער ארטיקל ווארנט אז AI וועט אראפברענגען די שכירות פון קוואליפיצירטע ארבעטער.
- **17 מיי:** מאכט פון ריזן – אויף Hacker News קלאגן פראגראמירער אז א האנטפול ריזן-פירמעס האלטן אלע קארטן. אן ארטיקל זאגט אז Meta'ס אפענע Llama 3, מיט א 405B ווערסיע וואס ווערט נאך טריינירט, שטעלט אין געפאר OpenAI'ס ביזנעס-מאדעל.
- **21 מיי:** Llama 3 און Google – לויט LMSYS איז Llama 3-70B שטארק ביי שאפערישע ארבעטן, אבער שוואכער ווען די פראגעס ווערן שווערער. CEO Aravind Srinivas א'Perplexity זאגט אז Google'ס AI סיכומים קענען שעדיגן Google'ס ביזנעס פון רעקלאמעס. Gemini קען נאך אלץ נישט מאכן בילדער פון מענטשן, און א שרייבער קלאגט אז Google'ס AI זוכעכץ לייגט צו דריי סעקונדעס.
- **22 מיי:** Claude קעגן GPT – אויף Hacker News האבן רוב פראגראמירער בעסער Claude Opus פאר קאוד און שרייבן. Scale'ס פריוואטער רייע ווייזט אז Llama 3 70B פאלגט אנווייזונגען אפט בעסער ווי Claude Opus.
- **23 מיי:** OpenAI'ס ביבליאטעק – OpenAI'ס הויפטקווארטיר איז געבויט ווי א ביבליאטעק, ווי א סימבאל, ווי די New York Times באמערקט, צו די ויכוחים איבער קאפירעכט.
- **27 מיי:** נאך מיינונגען – א שרייבער זאגט אז LLMs זענען נישט גוט פאר געדאנקען-שטורעמען, ווייל זיי ענדיגן ביי קאנסענווס. Hypergrowth Partners ראטן AI סטארטאפס צו פירן מיט מארקעטינג.

- **29 מיי:** רעגולאציע – אן ארטיקל לייגט פאר אז מען זאל רעגולירן דאס פירונג (conduct) מיט די עקזיסטירנדע געזעצן, אנשטאט רעגולירן יעדן באנוץ, און קריטיקירט דעם EU AI Act.

ווייטערע ידיעות

- **9 מיי:** טיניידזשערס מיט טשעטבאטס – יונגע לייט ניצן Character.AI פאר געזעלשאפט און שטיצע; דאס העלפט זיי, אבער מען ווערט צוגעוואוינט, און מען זארגט וועגן שלעכטע עצות.
- **22 מיי:** Apple צוגענגליכקייט – Apple ברענגט Eye Tracking צו iPad און iPhone, Vocal Shortcuts פאר Siri, Music Haptics, און דערקענען אומגעווענליכע רייד.
- **31 מיי:** מוח-טשיפ רעקארד – Precision Neuroscience זאגט אז זי האט געשטעלט א רעקארד מיט 4,096 עלעקטראדן אויף א לעבעדיגן מענטשנס מוח.

דער חודש יוני 2024 איז געווען פול מיט גרויסע לאנסירונגען. Apple האט אויף WWDC אנטפלעקט "Apple Intelligence" צוזאמען מיט א שותפות מיט OpenAI, און Anthropic האט ארויסגעלאזט Claude 3.5 Sonnet, וואס איז שטארקער ווי Opus און קאסט א פינפטל. Nvidia איז געווארן די ווערטפולסטע פובליק-פירמע אויף דער וועלט מיט 3.34 טריליאן דאלאר. Ilya Sutskever האט געגרינדעט א נייע פירמע פאר זיכערע סופער-אינטעליגענץ, און OpenAI האט פארדאפלט אירע איינקונפטן און אנגעהויבן קויפן פירמעס. דערצו זענען ארויסגעקומען א ריי שטארקע אפן-סאורס מאדעלן (Qwen 2, Nemotron-4, DeepSeek Coder V2, Gemma 2), און א גרופע אנגעשטעלטע האבן געפאדערט שוץ פאר די וואס ווארענען פון סכנות.

מאדעלן און נייע לאנסירונגען

- **3 יוני: Google – Google AI Overviews** האט ערקלערט אז זי האט איינגעפירט מער ווי א טון טעכנישע פארבעסערונגען נאך טעותדיגע רעזולטאטן אין AI Overviews, למשל בעסער דערקענען אומזיניגע פראגעס און ווייניגער אינהאלט פון באניצער. די ענטפערס ווייזן זיך יעצט ארויס אין בערך 11% פון די זוכונגען, אראפ פון 27% ביים לאנסירן. מערסטנס קומען זיי ארויס ביי געזונט-פראגעס, הגם די רעקלאמע האט באטאנט איינקויפן, רעסטוראנטן און רייזעס. BrightEdge האט פובליצירט דאטא וועלכע אינדוסטריעס ווערן מערסטנס באריכט.
- **4 יוני: Claude 3 Opus אויף Google Cloud – Vertex** גיט יעצט Claude 3 Opus מיט "tool use", צוזאמען מיט די קלענערע מאדעלן, אלס טייל פון Vertex AI.

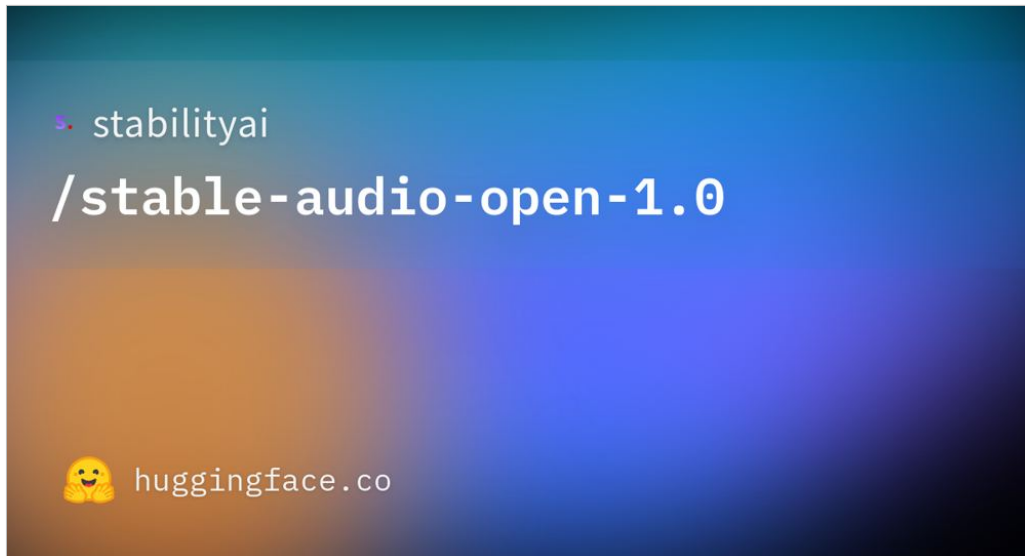


2024 יוני 4 · Vertex אויף Claude 3 Opus

- **4 יוני: Mamba-2** – די מאנשאפט הינטער דעם "state space" מאדעל Mamba האט ארויסגעלאזט א צווייטע ווערסיע מיט א לאנגער ערקלערונג פון די פארבעסערונגען. שפעטער האט Nvidia ארויסגעלאזט א היברידן Mamba מאדעל פון אירע פארשונג-ארבעט.

• **6 יוני: Mistral פיין-טונינג** – Mistral האט לאנסירט צופאסונג פון אירע מאדעלן דורך איר פלאטפארמע API.

• **6 יוני: Stable Audio Open** – Stability AI האט ארויסגעגעבן די וואגן פון Stable Audio, וואס איז טריינירט געווארן צו שאפן קלאנג-עפעקטן אויף אודיא מיט פרייע ליצענץ.



2024 יוני 6 · Stable Audio Open

• **7 יוני: Qwen 2** – Alibaba Cloud האט ארויסגעלאזט Qwen 2, איר נייער הויפט-מאדעל. אויף ענגלישע בענטשמארקס איז ער א ביסל בעסער ווי Llama 3 70B, און ער איז שטארקער אין פילע שפראכן.

• **10 יוני: Apple Intelligence** – Apple האט אויף WWDC 2024 אנטפלעקט "Apple Intelligence" אין iOS 18, iPadOS 18, און macOS Sequoia. עס כולל סיכומים פון מעלדונגען, קלוגע ענטפערס אין Messages און Mail, סארטירן בליצבריוו, "Genmoji", פאטא-רעדאגירונג, טראנסקריפציעס און א שמועסדיגער Siri. ChatGPT (GPT-4o) ווערט אריינגעבויט פאר קאמפליצירטע פראגעס, און שפעטער זאל עס אויך ארבעטן מיט קול. די פונקציעס וועלן זיין נאר אויף נייערע מכשירים, אויף ענגליש, דעם האַרבסט. Apple באטאנט פריוואטקייט דורך ארבעט אויפן מכשיר און "Private Cloud Compute".

• **10 יוני: Copilot אויף Telegram** – Microsoft האט לאנסירט Copilot, באזירט אויף GPT, אויף Telegram.

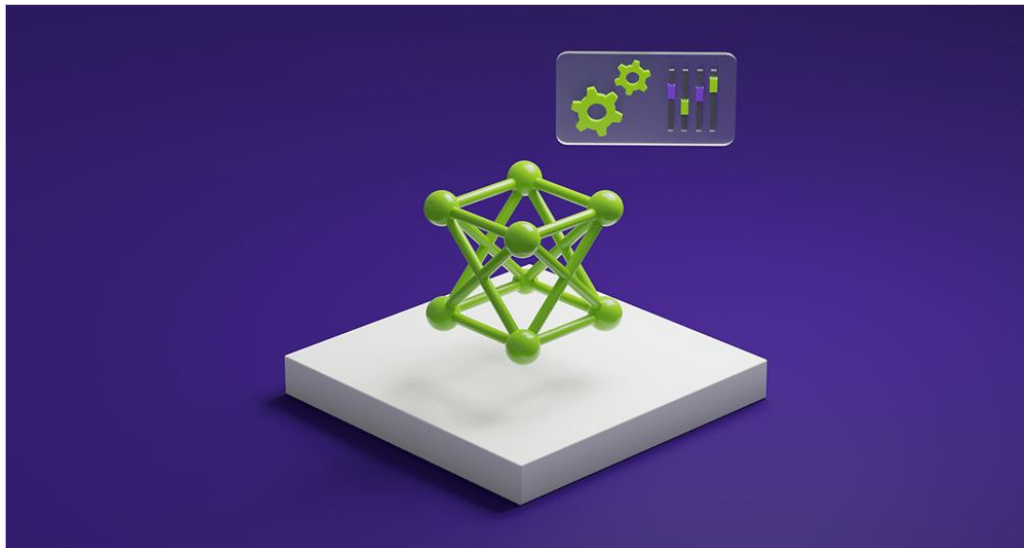
• **11 יוני: GLM 4 9B** – Tsinghua KEM גרופע האט ארויסגעלאזט א מאדעל טריינירט אויף 10 טריליאן טאקענס, מיט שטיצע פאר 26 שפראכן.

• **13 יוני: Stable Diffusion 3** – Stability AI האט ארויסגעגעבן די וואגן פון SD 3 פאר נישט-קאמערציעלע באנוץ. קאמערציעלע באנוץ איז מעגליך אויף בקשה.

• **14 יוני: Gemini 1.5 אפיציעל** – Google האט ארויסגעלאזט די סטאבילע ווערסיעס פון Gemini 1.5 Pro און Flash 1.5, מיט העכערע API גרעניצן, טונינג פאר JSON schema, Flash, און שטיצע פאר מאביל.

• **14 יוני: ElevenLabs קלאנג-עפעקטן** – ElevenLabs האט אריינגעפירט א מאדעל וואס שאפט קלאנג-עפעקטן, טרעקס און שטימען פון טעקסט. ער שטיצט זיך אויף דער אודיא-ביבליאטעק פון Shutterstock.

• **17 יוני: Nvidia Nemotron-4 340B** – Nvidia האט ארויסגעלאזט א פאמיליע אפענע מאדעלן צו שאפן סינטעטישע דאטא פאר טריינירן קאמערציעלע מאדעלן. דער "reward" מאדעל קומט גלייך צום ערשטן GPT-4 און לויפט אויף Nvidia H100 8 האט אויך פובליצירט HelpSteer2, א דאטאסעט מיט א רעצעפט פאר טריינירן אזעלכע מאדעלן.



2024 יוני 17 · Nvidia Nemotron-4 340B

• **18 יוני: Runway Gen 3** – Runway האט טריינירט א זייער שטארקן נייעם ווידעא-מאדעל, וואס וועט שטיין הינטער פילע פונקציעס אויף איר פלאטפארמע.

• **18 יוני: DeepMind V2A** – DeepMind אנטוויקלט V2A, וואס שאפט סינכראניזירטע מוזיק, קלאנג-עפעקטן און דיאלאג פאר ווידעאס. עס ניצט "diffusion" מאדעלן טריינירט אויף אודיא, טראנסקריפטן און ווידעא.

• **18 יוני: DeepSeek Coder V2** – דער קאוד-מאדעל קריגט איבער 90 אויף HumanEval און גלייכט זיך מיט GPT-4 Turbo אויף פילע בענטשמארקס. ער איז צוגענגליך דורך API און פריי פאר קאמערציעלע באנוץ.

• **18 יוני: Depth Anything V2** – דער נייער מאדעל פאר מעסטן טיפקייט איז טריינירט מערסטנס אויף סינטעטישע דאטא, און ער איז פיל בעסער ביי קאמפליצירטע סצענעס.

• **19 יוני: Meta Chameleon** – Meta האט ארויסגעלאזט די טשעקפוינטס און קאוד פון Chameleon, א מאדעל וואס ארבעט גלייכצייטיג מיט בילד- און טעקסט-טאקענס.

• **19 יוני: Granite און FireFunction** – IBM האט ארויסגעלאזט אפענע Granite קאוד-מאדעלן, וואס זענען אויך גוט אין פאררעכטן, ערקלערן און אפשאצן קאוד. FireFunction-v2, געבויט אויף Llama 3 70B, קומט גלייך צו GPT-4o ביי "function calling".

• **20 יוני: Microsoft Florence-2** – Microsoft האט ארויסגעלאזט קליינע וויזע-מאדעלן אונטער MIT ליצענץ, וואס זענען פיל בעסער ווי גרעסערע מאדעלן ביי באשרייבן בילדער, באגרעניצן אביעקטן און קלאסיפיצירן. מען קען זיי פיין-טונען אין PyTorch, כאטש זיי זענען נישט אזוי שטארק ווי PaliGemma.

• **21 יוני: Claude 3.5 Sonnet** – Anthropic האט לאנסירט Claude 3.5 Sonnet. ער איז שטארקער ווי Opus און קאסט א פינפטל, און ער איז דער בעסטער וויזע-מאדעל אויפן מארק. אויף עטליכע

- בענטשמארקס איז ער בעסער ווי GPT-4o און Gemini, און ער איז שנעלער. ער ברענגט מיט זיך די נייע "Artifacts" פונקציע, און Anthropic וויל אז Claude זאל ווערן א צענטראל געצייג פאר ביזנעסער.
- **21 יוני: Apple אפן-סאורס** – Apple האט צוגעלייגט 20 Core ML מאדעלן אויף Hugging Face, פאר קלאסיפֿיצירן בילדער און טיפקייט-סעגמענטאציע. דאס קומט נאך Ferret און די פיר OpenELM מאדעלן.
 - **24 יוני: Snapchat AI לענוועס** – Snap האט איינגעפירט א מאדעל אויפן מכשיר וואס שאפט AR לענוועס פון טעקסט. Lens Studio 5.0 ברענגט א GenAI סוויט צו שאפן 3D אביעקטן, פנים-מאסקעס און טעקסטורן.
 - **25 יוני: Toucan TTS** – Toucan האט אויסגעברייטערט אירע טעקסט-צו-רייד מאדעלן אויף יעדע ISO-639-3 שפראך, בערך 7000 שפראכן.
 - **26 יוני: Imbue 70B** – Imbue האט טריינירט און ארויסגעלאזט א 70B מאדעל וואס קומט גלייך צו GPT-4. ער איז טריינירט מיט איר אייגענעם "optimizer" און גוטע דאטא-פילטערונג, אן קיין שום "loss" "spikes".
 - **26 יוני: ChatGPT נייעס** – OpenAI האט אפגעשטופט די פארגעשריטענע קול-מאָד, וואס וועט ערשט סוף יולי ארויסקומען אין "alpha" פאר געוויסע Plus באניצער. די ChatGPT אפליקאציע פאר macOS איז יעצט צוגענגליך פאר אלע.
 - **27 יוני: מעדיצינישער 7B** – א 7B מאדעל, טריינירט אויף מעדיצינישע דאטא, איז געקומען גלייך מיט GPT-4 לויט דאקטוירים וואס האבן אפגעשאצט די ענטפערס.
 - **28 יוני: CriticGPT** – OpenAI האט טריינירט א מאדעל באזירט אויף GPT-4 וואס כאפט טעותים פון GPT-4 און העלפט מענטשן זיי באמערקן אינעם RLHF פראצעס.
 - **28 יוני: Meta LLM Compiler** – Meta האט ארויסגעלאזט צוויי מאדעלן וואס קאמפילירן קאוד צו "assembly" און דעקאמפילירן צו LLVM IR. זיי זענען טריינירט אויף 546 ביליאן טאקענס. זיי דערגרייכן 77% פון אפטימיזירטע "assembly" לייסטונג, און 45% ביי א פולער "round trip".
 - **28 יוני: אנדערע לאנסירונגען** – Instagram פרובירט אויס אין אמעריקע אן "AI Studio", וואו שאפער קענען מאכן טשעטבאט-ווערסיעס פון זיך. Google האט צוגענגליך געמאכט Gemma 2 אין גרייסן 9B און 27B. ElevenLabs האט לאנסירט איר ערשטע iOS אפליקאציע, וואס לייענט פאר טעקסט און וועב-לינקס.

פארשונג און טעכניק

- **3 יוני: פרי-יוני פארשונג** – Random Color Erasing מאכט ווייזע-מאדעלן שטארקער דורך אויסמעקן פארבן ביים טריינירן. א פארשונג-ארבעט ווייזט ווי אזוי צו ניצן muP פאר ספארס מאדעלן, וואס פארבעסערט טריינירן און שפארט קאמפיוטינג. Conifer פארבעסערט די אויספירונג פון קאמפליצירטע אנווייזונגען. FABRIC פארבעסערט בילד-קוואליטעט אן טריינירן. Mora פרובירט צו פארמעסטן זיך מיט Sora דורך ניצן עטליכע אנגענטן.
- **3 יוני: Mozilla alt טעקסט** – Mozilla האט טריינירט א מאדעל וואס לויפט אינעם בלעטערער און שרייבט אויטאמאטיש באשרייבונגען פאר בילדער.

- **4 יוני: מער אַרבעטן** – TrainAllInfAttn העלפט אין געביטן מיט ווייניג דאטא, און InvariantSelectPR צופאסן מולטימאדאלע מאדעלן צו מעדיצין. MegActor שאפט רעאליסטישע פארטרעט-אנימאציע. MeshXL שאפט הויך-קוואליטעט 3D "meshes". ICAL פארבעסערט דערקענען האנטגעשריבענע מאטעמאטיק.
- **4 יוני: FineWeb** – Hugging Face האט פובליצירט אן אויספירליכן ארטיקל וועגן ריינמאכן און פילטערן וועב-דאטא פאר טריינירן שפראך-מאדעלן, צוזאמען מיט שפיץ-דאטאסעטן.
- **4 יוני: וועטער-פאָרויסזאגונג** – דער WeatherMesh מאדעל פון WindBorne ניצט ERA5 דאטא און איז בעסער ווי טראדיציאנעלע מאדעלן, מיט פיל ווייניגער קאמפיוטינג.
- **5 יוני: MMLU Pro** – א נייער, שווערערער און ריינערער בענטשמארק פאר שכל-פעאיקייט. ער ווערט געשאפן ווייל מאדעלן האבן זיך שוין "צוגעפאסט" צום אלטן MMLU.
- **5 יוני: בילד-טעכניק** – Tree Diffusion "דיפוזירט" קאוד וואס שאפט בילדער, און מען קען עס קאמבינירן מיט זוכן. Control-GIC איז פאר קאמפרעסירן בילדער. Omot, פונעם שאפער פון ControlNet, שרייבט איבער דעם "prompt" אלס קאוד וואס מען קען רעדאגירן. אן אנדער מעטאד הייבט אויף די ראם-ראטע פון ווידעא אן טריינירן.
- **5 יוני: הינט-בילן** – פארשער פון University of Michigan, צוזאמען מיט INAOE אין מעקסיקע, אנאליזירן הינט-בילן צו דערקענען שפילעריי אדער אגרעסיע, און אויך די ראסע, עלטער און מין פונעם הונט.
- **6 יוני: Google שיפן און מער** – Google האט פארבעסערט די מאטעמאטישע אפטימיזאציע פון שיפן-רוטן, מיט 13% ווייניגער ברענשטאף און קאסטן. אנדערע ארבעטן: LlamaCare פאר מעדיצין, DreamMat פאר 3D מאטעריאלן, מאדעלן אן מאטרייז-מולטיפליקאציע וואס ארבעטן גוט אפילו ביי ביליאנען פאראמעטערס, XRec פאר ערקלערבארע רעקאמענדאציעס, GenS פאר רעקאנסטרוקציע פון אויבערפלאכן, און CV-VAE פאר ווידעא.
- **7 יוני: GPT-4 קאנצעפטן** – OpenAI האט געפונען 16 מיליאן פארשטענדליכע "features" אין GPT-4, ווי פרייז-העכערונגען און אלגעברעאישע רינגען, און האט צוגעשטעלט דעם קאוד. Together האט באוויזן DragonFly פאר זייער הויכע רעזאלוציע. אנדערע ארבעטן: FFNification, וויזועלע געא-לאקאליזאציע, GrootVL, און ReLU פאר INRs.
- **10 יוני: Claude'ס כאראקטער** – Anthropic ערקלערט ווי אזוי זי פורעמט Claude'ס פערזענליכקייט דורך א כאראקטער-ווערסיע פון Constitutional AI, אזוי אז ער זאל נישט נאר צושטימען צום באניצער.
- **10 יוני: וויזיע פארשונג** – YOLO-World ברענגט "open-vocabulary" דערקענונג פון אביעקטן. SDA ניצט סינטעטישע דאטא, און ReNO אפטימיזירט דעם אנהייב-ראש ביי טעקסט-צו-בילד. אנדערע ארבעטן: קאמערע-לאקאליזאציע וואס שוינט די פריוואטקייט, א 6D דאטאסעט מיט איבער 700 טויזנט בילדער, און Spatial Reasoning GPT פאר גרייסן פון אביעקטן.
- **11 יוני: BitsFusion** – Snap Research האט פארקלענערט דעם Stable Diffusion UNet פון 1.72 GB צו 219MB, און די לייסטונג איז דערביי נאך בעסער געווארן. Google'ס Gboard ניצט PaLM2-XS אויפן מכשיר צו פאררעכטן טעותים מיט איין דריק. אנדערע ארבעטן: OVMR, PDF פאר מולטימאדאל טריינירן, און טיפקייט-באמערקן מיט געווענליכע קאמערעס.

- **12 יוני: Together – Mixture of Agents** MoA קאמבינירט עטליכע מאדעלן אין שיכטן און קריגט 65.1% אויף AlpacaEval 2.0, מער ווי GPT-4o, אבער עס איז פאמעלעכער. אנדערע ארבעטן: סטרוקטורירטע מאטריצן אנשטאט געדיכטע שיכטן, PHATGOOSE פאר "zero-shot", CARES פאר מעדיצינישע מאדעלן, LE3D פאר נאכט-בילדער, HGM פאר 3D מענטשן, און דערקענען דייעכטיק-שאדנס אין די אויגן.
- **12 יוני: בענטשמארק-איבערפאסונג** – Scale AI SEAL האט געשאפן GSM1k און געפונען אז Mistral און Phi זענען "overfit" אויף GSM8k, אבער GPT, Claude, Gemini און Llama נישט.
- **13 יוני: TextGrad און מאטעמאטיק** – TextGrad באהאנדלט מאדעלן וואס פארבעסערן טעקסט אזוי ווי "backpropagation". DeepMind האט אויטאמאטיזירט "process supervision" און דערמיט שטארק פארבעסערט מאטעמאטיק ביי Gemini Pro. אנדערע ארבעטן: BERT קען לערנען אינעם קאנטעקסט, LlamaGen שאפט בילדער "autoregressive" מיט א נייעם FID רעקארד, AsyncDiff פאר פאראלעלע דיפוזיע, און לינעארע "attention" מיט "speculative decoding".
- **14 יוני: Sakana evolution** – Sakana AI האט אנטדעקט, דורך עוואלוציע איבער בערך 100 דורות פון קאוד געשריבן דורך מאדעלן, אפטימיזאציע-מעטאדן שטארקער ווי DPO. אנדערע ארבעטן: איבערבאשרייבן ביליאנען בילדער מיט Llama 3 פארבעסערט מאדעלן ווי CLIP. Microsoft, Samba, א Mamba מאדעל אויף Phi 3 דאטא, איז פיל בעסער ווי טראנספארמער אויף MMLU, HumanEval און "retrieval". אויך: DiffRIR פאר אקוסטיק, און "masked diffusion" פאר שפראך.
- **17 יוני: MCTS מאטעמאטיק** – Llama 3 8B מיט MCTS זוכן קומט גלייך צו GPT-4 אין מאטעמאטיק. אנדערע ארבעטן: MLKV שפארט זכרון, א דיפוזיע שפראך-מאדעל וואס שרייבט נישט אויף די ריי, VideoGPT+, און סאטעליט-סצענע גראפן.
- **18 יוני: Cerebras טשיפ** – דער צווייטער דור "wafer-scale" טשיפ פון Cerebras איז שנעלער ווי דער שנעלסטער סופערקאמפיוטער ביי מאלעקולארע דינאמיק, און קען לויפן ספארס מאדעלן מיט א דריטל פון די ענערגיע.
- **18 יוני: נאך ארבעטן** – א פארשונג-ארבעט פארשט פארוואס בילד-מאדעלן "כאלוצינירן". CPO פארבעסערט לאגישע טראכטונג דורך ToT בוימער. ERASE דערהיינטיקט וויסן-באזעס. אויך: Duoduo, CLIP, CAMixerSR, OGNI-DC, און "subobject" טאקעניזאציע.
- **19 יוני: MLPerf** – סיסטעמען מיט Nvidia Hopper האבן געוועלטיגט ביי צוויי נייע MLPerf טעסטן, פיין-טונינג פון מאדעלן און גראף נעצן.
- **20 יוני: Logit Prisms** – א נייע מעטאד צוטיילט די רעזולטאטן פון א מאדעל לויט די באזונדערע טיילן, כדי צו פארשטיין ווי gemma-2 ארבעט. אויך: Decision QA, און TroL מאדעלן פון 1.8B ביז 7B.
- **21 יוני: Character AI אינפערענץ** – Character AI באדינט 20,000 פראגעס א סעקונדע, 20% פון Google'ס זוכונגען, מיט א ריי נייע עפעקטיווע מעטאדן. אויך: LayerMerge, ChangeViT, CIFAR-10, TimeSieve, און 94% אין 3.29 סעקונדעס.
- **24 יוני: איבערצייגונג** – א פארשונג-ארבעט געפינט שוואכע ראיות אז גרעסערע מאדעלן זענען נישט באדייטנד מער איבערצייגנד, כאטש מען קען זיי מעגליך טונען דערצו. אויך: MacroHFT פאר קריפטא האנדל, און QMIX פארבעסערונגען.

- **25 יוני:** CVPR **און מער** – א ליסטע פון 40 אויסגעקליבענע CVPR ארבעטן. אויך: PE-Rank, Mixture of Attention, און CDMask.
- **26 יוני:** RAG **פיינ-טונינג** – Together און Morph Labs ווייזן ווי אזוי צו טונען מאדעלן פאר RAG מיט סינטעטישע דאטא. LLaMini's Memory Tuning רעדוצירט כאלוצינאציעס מיט PyTorch 95%. האט קערנעלס פאר 2:4 ספארס טריינירן. אויך: EvalAlign, FreeTraj, און GeoMFormer.
- **27 יוני:** DeepMind **דאטא** – DeepMind ווייזט ווי אזוי צו דערגרייכן שפיץ-רעזולטאטן מיט בלויז 10% פון די FLOPs דורך אויסקלייבן ביישפילן. אויך: א קלארע ערקלערונג פון SAEs מיט קאוד, RA-IT, און Director3D.
- **28 יוני:** קליינע מאדעלן – Apple און Microsoft לייגן דעם טראָפּ אויף קלענערע מאדעלן מיט גלייכע אדער בעסערע רעזולטאטן. לויט דעם CEO פון OpenAI גייען מיר שוין ווייטער פון דער תקופה פון גרויסע מאדעלן. אויך: ALPBench, א סורויי וועגן GenIR, Point-SAM, און MatText.

געצייג פאר פראגראמירער און אפן-סאורס

- **4 יוני:** לאקאלער קאוד-געהילף – מיט Ollama און Continue קען מען לויפן "autocomplete" און שמועסן גאנץ אויפן מכשיר.
- **7 יוני:** Google Mesop – א פיור Python קאנקורענט צו Gradio. ער לאזט זיך בעסער צוזאמענשטעלן, אבער ער איז ווייניגער אויסגעצייטיגט.
- **10 יוני:** עטליכע נייע געצייג – Inspectus פאר וויזואליזירן מאדעלן. א GPT-2 מאדעל אינגאנצן אין א "spreadsheet". Thread, א Jupyter נאטבוק אין סטיל פון "code interpreter".
- **12 יוני:** Mistral.rs **און מער** – Mistral.rs אין Rust איז שנעלער געווארן און שטיצט GGUF. PowerInfer-2 לויפט א 47B MoE מאדעל אויפן טעלעפאן. AXLearn איז א ביבליאטעק אויף JAX.
- **13 יוני:** DeepMind Torax – DeepMind האט געעפנט איר "tokamak" סימולאטאר אין JAX, און אויך Splash Attention, א קערנעל פאר ספארס "attention".
- **14 יוני:** כלים – DiffusionKit פאר Apple MLX, א פארט פון Karpathy's LLM.c. LARS לויפט מאדעלן לאקאל אויף אייגענע דאקומענטן. Gaianet-Node פאר אן אייגענעם אנענט. TokenCost רעכנט אויס API קאסטן.
- **18 יוני:** טעכנישע לערנונגען – א ערקלערונג ווי אזוי צו שרייבן "differentiable" SVG. א רעדע פונעם שאפער פון RAGatouille וועגן פארבעסערן RAG און COLBERT.
- **19 יוני:** Open Interpreter – Local III קאנטראלירט דעם קאמפיוטער גאנץ לאקאל אן אינטערנעט. LlamaIndex האט איינגעפירט Property Graph Index פאר וויסן-גראפן.
- **20 יוני:** Argilla **און מער** – Argilla איז א פלאטפארמע פאר מיטארבעט אויף הויך-קוואליטעט דאטא. Paramount מעסט די פינקטלעכקייט פון אנענטן. א ליסטע פון פראיעקטן פאר אנהייבער.
- **21 יוני:** WebCanvas – א פריימווערק צו פרובירן וועב-אגענטן אויף לעבעדיגע זייטלעך, און א נייער בענטשמארק פאר טיפקייט.

- **24 יוני: Karpathy קורס** – Andrej Karpathy גרייט צו א קורס צו בויען א מאדעל וואס דערציילט מעשיות. אויך: AutoCodeRover פאר פאררעכטן קאוד, ReaLHF פאר RLHF, NLUX פאר וועב UI, און Claudette פאר Claude.
- **25 יוני: MLX שנעלקייט** – "batch" KV קעש אין MLX ברענגט 1300 טאקענס א סעקונדע אויף Mac. TRL שטיצט יעצט וויזיע-מאדעלן ווי LLaVA. א Llama מאדעל לויפט אין א פאנט-פייל. June איז א לאקאלער קול-טשעטבאט.
- **26 יוני: לאקאלער קול-געהילף** – ווי אזוי מען מאכט א קול-געהילף שנעלער מיט RAG. LLM Radar שיקט מעלדונגען וועגן נייע מאדעלן. OpenGlass מאכט יעדע ברילן "קלוג".
- **27 יוני: RES-Q און Slack** – RES-Q איז א בענטשמארק פאר רעדאגירן קאוד-רעפאזיטאריעס. א פראמפט-געצייג איז כמעט אינגאנצן געשריבן דורך Slack. Sonnet 3.5 האט איבערגעשטעלט איבער 15,000 טעסטן פון Enzyme צו React Testing Library מיט Claude 2.1, מיט 80% הצלחה.
- **28 יוני: AuraSR** – FAL האט געעפנט אן "upscaler" וואס פארגרעסערט בילדער 4 מאל אין איין גאנג. אויך: א פשוטע אימפלעמענטאציע פון Mamba 2.

פירמעס, געלט און ביזנעס

- **3 יוני: OpenAI ראבאטיק** – OpenAI באלעבט ווידער איר ראבאטיק גרופע נאך דריי יאר.
- **3 יוני: טשיפס** – לויט באריכטן גרייט Nvidia א PC טשיפ מיט Arm Cortex-X5 און Blackwell. AMD האט אנטפלעקט MI325X, וואס קומט Q4 2024, און דערנאך MI350 אין 2025, מיט 35 מאל בעסערע אינפערענץ, און MI400 אין Raspberry Pi 2026. לאנסירט אן AI טשיפ.
- **3 יוני: OpenAI בארד און Musk** – א געוועזענער בארד-מיטגליד פון OpenAI זאגט אז Altman האט פארשווינדן אינפארמאציע און איז געווען נישט ערליך מיט דעם בארד. די בארד האט זיך דערוואוסט פון ChatGPT דורך Musk. Twitter האט צוריקגעצויגן זיין קלאגע קעגן OpenAI, אבער ער קען זי ווידער אריינגעבן.
- **5 יוני: Altman'ס אינוועסטירונגען** – Sam Altman פארוואלטעט אינוועסטירונגען ווערט כאטש 2.8 ביליאן דאלאר.
- **6 יוני: קליינע געלט-רונדעס** – Cartwheel האט אויפגענומען 5.6 מיליאן דאלאר פאר 3D אנימאציע. CaseMark 1.7 מיליאן פאר אדוואקאטן. Ability AI 1.1 מיליאן פאר מארקעטינג-אגענטן.
- **7 יוני: Musk'ס GPUs** – Musk האט איבערגעפירט 12,000 H100, וואס Tesla האט באשטעלט, צו X.
- **7 יוני: Nvidia שווי** – Nvidia איז ערשט געווארן ווערט 3.01 טריליאן, מער ווי Apple, און דערנאך 3.34 טריליאן, מער ווי Microsoft. דערמיט איז זי די ווערטפולסטע פובליק-פירמע, נאך א שטייג פון איבער 170% דעם יאר.
- **10 יוני: Databricks** – Databricks האט געקויפט Tabular פאר אן אפענעם דאטא-סטאנדארט, און בויט מיט Shutterstock א מאדעל צו שאפן בילדער.
- **12 יוני: Mistral** – Mistral האט אויפגענומען 600 מיליאן ייראָ אין א Series B, אויף איר ערשטן געבורטסטאג.

- **12 יוני: OpenAI אקציעס** – נאך באזארגטקייט האט OpenAI געטוישט די פאליסי, אזוי אז אלע אנגעשטעלטע, יעצטיגע און געוועזענע, קענען גלייך פארקויפן אקציעס, און זי האט אפגעשאפט געצוואונגענע צוריקקויפונגען.
- **13 יוני: OpenAI איינקונפטן** – OpenAI'ס יערליכע איינקונפטן האבן זיך מער ווי פארדאפלט צו 3.4 ביליאן דאלאר, מערסטנס פון אירע פראדוקטן.
- **13 יוני: Perplexity** – Perplexity פלאנירט צו טיילן די איינקונפטן מיט פובליצירער, און גיט אין יאפאן א פרייע יאר Pro דורך SoftBank. איר שווי איז 1 ביליאן דאלאר.
- **14 יוני: Microsoft** – Microsoft בויט אן אייגענע AI מאנשאפט אונטער Mustafa Suleyman און האט אינוועסטירט 1.5 ביליאן אין א פירמע פון אבו דאבי.
- **14 יוני: OpenAI-Apple אפמאך** – Apple צאלט נישט קיין געלט. זי באטראכט דעם אויסשטעל פון OpenAI'ס בראנד אויף אירע מכשירים אלס די צאלונג. דער אפמאך איז נישט עקסקלוסיוו, און Apple רעדט מיט Google וועגן Gemini. Musk האט געסטראשעט צו פארבאטן Apple מכשירים ביי זיינע פירמעס.
- **14 יוני: נייע קויפונגען און באשטימונגען** – Hugging Face האט געקויפט OpenAI. Argilla האט באשטימט א נייעם CFO, און Kevin Weil אלס CPO.
- **17 יוני: Cohere און Amazon** – Aidan Gomez פון Cohere רעדט וועגן דעם וועג צו פראפיט. Amazon וועט אוועקלייגן ביז 230 מיליאן דאלאר פאר סטארטאפס.
- **17 יוני: נאציאנאלע AI** – Nvidia ערווארט כמעט 10 ביליאן דאלאר דעם יאר פון רעגירונגען וואס בויען דאטא צענטערס פאר AI אין זייערע אייגענע שפראכן.
- **13 יוני: Meta דאטא צענטערס** – Meta וואקסט צו 600,000 GPUs פאר טריינירן.
- **19 יוני: Hinton און Decagon** – Geoff Hinton און Max Welling בויען א סטארטאפ פון 30 מיליאן דאלאר פאר מאטעריאלן צו כאפן קוילנשטאף. Decagon האט אויפגענומען 35 מיליאן פאר קונה-שטיצע אגענטן.
- **20 יוני: Ilya Sutskever** – Ilya Sutskever האט געגרינדעט Safe Superintelligence Inc (SSI), וואס וועט זיך נאר פארנעמען מיט א זיכערער, שטארקער סיסטעם, אן קאמערציעלן דרוק.
- **20 יוני: Roblox 4D** – Roblox בויט צו 4D AI, וואס ברענגט באוועגונג און אינטעראקציע.
- **21 יוני: HeyGen און Factory** – HeyGen האט אויפגענומען 15 מיליאן דאלאר פון Sequoia; אירע Droids האבן א רעקארד אויף HeyGen. SWE-bench האט אויפגענומען 60 מיליאן, מיט 35 מיליאן יערליכע איינקונפטן.
- **24 יוני: OpenAI קויפט** – OpenAI האט געקויפט Rockset פאר זוך-אינפראסטרוקטור און Multi פאר דעסקטאפ-טעכנאלאגיע.
- **24 יוני: Butterflies AI** – א סאציאלע אפליקאציע מיט AI פיגורן האט אויפגענומען 4.8 מיליאן דאלאר.
- **25 יוני: Amazon Metis** – Amazon אנטוויקלט א טשעטבאט "Metis" קעגן ChatGPT, מיט א נייעם מאדעל.

- **26 יוני:** Etched – Stability האט אויפגענומען 120 מיליאן דאלאר פאר טראנספארמער-טשיפס ביי TSMC (4 נאנאמעטער). זי זאגט אז זיי קענען דערגרייכן 500 טויזנט טאקענס א סעקונדע, און אז איין 8xSohu קען פארבייטן 160 H100. Stability AI האט באקומען א נייעם CEO און א שטארקע געלט-אינוועסטירונג.
- **27 יוני:** Evolutionary Scale – Evolutionary Scale האט אויפגענומען 142 מיליאן דאלאר און ארויסגעלאזט ESM 3, א ביאלאגיע-מאדעל מיט 98B פאראמעטערס, טריינירט אויף 771 ביליאן טאקענס. דער מאדעל האט אנטדעקט א נייעם גרינעם פלואַרעסצירנדן פראטעין.
- **27 יוני:** YouTube – Waymo האט פירט פארהאנדלונגען מיט רעקארד-פירמעס פאר ליצענצן צו מוזיק פאר AI. Waymo One איז יעצט אפן פאר יעדן אין San Francisco.
- **28 יוני:** Dust – Dust האט אויפגענומען 16 מיליאן דאלאר נאכן דורכגיין 1 מיליאן יערליכע איינקונפטן.

רעגירונגען, געזעצן און זיכערהייט

- **5 יוני:** רעכט צו ווארענען – יעצטיגע און געוועזענע אנגעשטעלטע פאדערן שוץ פאר די וואס הייבן אויף סכנות, און אז מען זאל אפשאפן אפמאכן וואס פארבאטן קריטיק. Daniel Kokotajlo זאגט אז ער האט פארלאזט OpenAI ווייל ער האט פארלוירן צוטרוי אז זי וועט זיך פירן פאראנטווארטליך.
- **5 יוני:** אטאקעס – פארשער האבן פארבעסערט GCG "jailbreaking" און אנדערע מעטאדן קעגן מאדעלן. Haize Labs האט אויטאמאטיזירט "red-teaming", און א פארשונג-ארבעט ווייזט ווי שוואך ווייזע-אגענטן זענען קעגן אטאקעס.
- **7 יוני:** OpenAI זיכערהייט – OpenAI באשרייבט ווי אזוי זי שיצט די וואגן פון אירע מאדעלן אויף Azure און Kubernetes.
- **10 יוני:** ChatGPT שטים – OpenAI זאגט אז די שטים "Sky" איז נישט די שטים פון א באקאנטער אקטריסע, און זי האט אפגעשטעלט דעם באנוץ דערפון.
- **14 יוני:** Nakasone – OpenAI האט צוגעלייגט צו איר בארד און צו דער זיכערהייט-קאמיטע דעם געוועזענעם NSA הויפט Paul Nakasone.

אנאליזן און מיינונגען

- **3 יוני:** Edge AI און KL – אן אנאליז וועגן AI וואס ציט זיך צו מולטימאדאל און צו די מכשירים. אן ערקלערונג פון KL divergence. a16z גיט 7 סטראטעגיעס צו האלטן באניצער.
- **4 יוני:** קאמוניקאציע – לויט אן אנאליז ברענגט GPT-4o אן איבערקערעניש אין קאמוניקאציע צווישן מענטש און AI.
- **5 יוני:** ראבאטיק און דזשאבס – א דעבאטע צי גרייסער-מאכן (scaling) וועט לייען ראבאטיק. א ארטיקל וועגן "speed of light" אינפערענץ. Chip Huyen האט איבערגעקוקט 900 אפן-סאורס געצייג. Noah Smith טענהט אז מענטשן וועלן ווייטער האבן גוט באצאלטע ארבעט צוליב "comparative advantage".
- **6 יוני:** OpenAI און צוגעמאכטע מאדעלן – איינער איז פעסימיסטיש אויף OpenAI צוליב דעם ווען מענטשן פארלאזן די פירמע. אן אנדערער שרייבט אז די צוקונפט געהערט צו צוגעמאכטע מאדעלן.

- **7 יוני:** קול-אגענטן און הייפ – a16z מאפט אויס דעם מארק פון קול-אגענטן. Glyph טענהט אז AI הייפ-ציקלען גייען אין א "sigmoid" פארעם.
- **10 יוני:** פראדוקטן – Ben Evans שרייבט אז מאדעלן זענען נישט קיין דאטאבאזעס און דארפן אריינגעבויט ווערן אין ספעציפישע געצייג. אן ערקלערונג ווי בילד-מאדעלן ארבעטן.
- **18 יוני:** אלגעמיינע אינטעליגענץ – א פארשער ביי OpenAI שרייבט אז מיר זענען בלויז א פאר יאר אוועק פון אן אלגעמיינע אינטעליגענטן אגענט.
- **20 יוני:** קאפירעכט – O'Reilly שלאגט פאר ווי אזוי צו באצאלן שאפער פאר זייער אינהאלט. אן אנאליז פון די מאי-אנאנסירונגען: GPT-4o איז צוויי מאל שנעלער און האלב אזוי טייער, און Gemini 1.5 Flash איז שנעלער און ביליגער.
- **21 יוני:** אייגענע Llama – אליין האלטן Llama-3 8B קאסט בערך 17 דאלאר פאר א מיליאן טאקענס, קעגן 1 דאלאר ביי ChatGPT. מען קען עס אראפברענגען אונטער 0.01 דאלאר, אבער עס נעמט בערך 5.5 יאר ביז מען קריגט צוריק די הוצאות.
- **24 יוני:** לאנגער בליק – אן אנאליז זאגט אז AI קען שטארק פארגיכערן דעם וואוקס. swyx רעדט וועגן געלעגנהייטן פאר פראגראמירער.
- **25 יוני:** Apple סטראטעגיע – Ben Evans שרייבט אז Apple צוטיילט צווישן "קאנטעקסט" און "וועלט" מאדעלן, און אז OpenAI'ס פאזיציע איז נישט שטארק באשיצט. Daring Fireball לויבט Apple'ס ווייכקייט אויף פריוואטקייט.
- **27 יוני:** אגענטן און רייזעס – לאנג-צייטיגע פלאנירונג און סיסטעם-שטארקייט פאר אגענטן. ווי AI קען איבערבויען די רייזע-אינדוסטריע. צוויי מיינונגען וועגן "roll-ups" מיט AI.

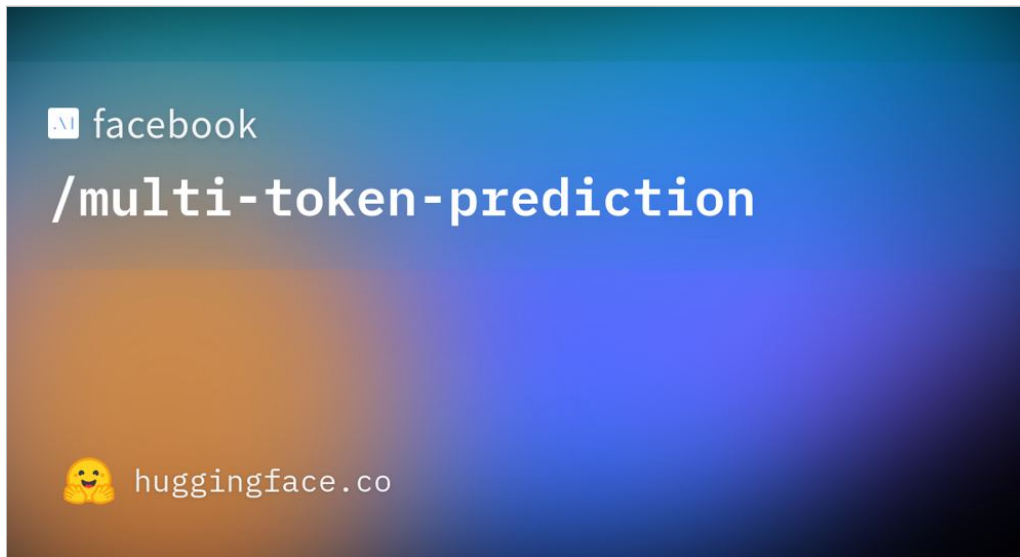
ווייטערע ידיעות

- **3 יוני:** Sakana און Tribeca – Sakana AI שטיצט א "LLM Merging" פארמעסט ביי NeurIPS. דער Tribeca פעסטיוואל וועט ווייזן פינף קורצע פילמען געשאפן מיט Sora.
- **4 יוני:** פראדוקטן – Bestever פאר רעקלאמעס. Amazon'ס Project PI געפינט חסרונות אין סחורה פארן שיקן. Facia פאר פנים-דערקענונג קעגן שווינדל.
- **6 יוני:** אפליקאציעס – Captions פאר ווידעאס, Synthflow פאר קול-געהילף, ChainGPT פאר קריפטא, MindMatrix פאר מיינד-מאפס.
- **11 יוני:** Asana – Asana האט איינגעפירט "AI טיממעטס" וואס פירן פראיעקטן אונטער מענטשליכער אויפזיכט.
- **12 יוני:** Mozilla Builders – Mozilla'ס ערשטער "accelerator" לייגט דעם טראָפּ אויף לאקאלע AI אויף פערזענליכע מכשירים.
- **19 יוני:** HubSpot און Khroma – HubSpot האט AI פונקציעס, און Khroma שאפט פארבן-פאלעטן.
- **24 יוני:** AI Steve – א ביזנעסמאן פון Sussex שטעלט "AI Steve" אויף אלס קאנדידאט אין די בריטישע וואלן, און זאגט צו אז די וויילער וועלן זיין שטענדיג פארבונדן מיט אים.

יולי 2024 איז געווען א חודש פון גרויסע אפן-סאורס לאנסירונגען: Meta האט ארויסגעלאזט Llama 3.1 מיט'ן 405B מאדעל, Mistral האט געענטפערט מיט Mistral Large 2 און Mistral Nemo, און Apple האט ארויסגעגעבן אלעס פון איר DCLM מאדעל. OpenAI האט לאנסירט GPT-4o Mini, אנגעהויבן פרובירן SearchGPT און א נייעם קול-טשעטבאט, אבער אויך געהאט פראבלעמען: א העקער, קלאגעס איבער NDA-אפמאכן, און באריכטן אז די פירמע קען פארלירן 5 ביליאן דאלאר. DeepMind's AlphaProof האט דערגרייכט א זילבערנעם מעדאל-שטאפל ביי דער מאטעמאטיק אלימפיאדע. ביי די פירמעס האבן מען געזען גרויסע אינוועסטירונגען (Cohere ביי 5.5 ביליאן, Harvey, Hebbia), די גרויסע פירמעס וואס "שלינגען" סטארטאפס ווי Adept, און א ברייטע וויכוח צי AI איז א בלאז.

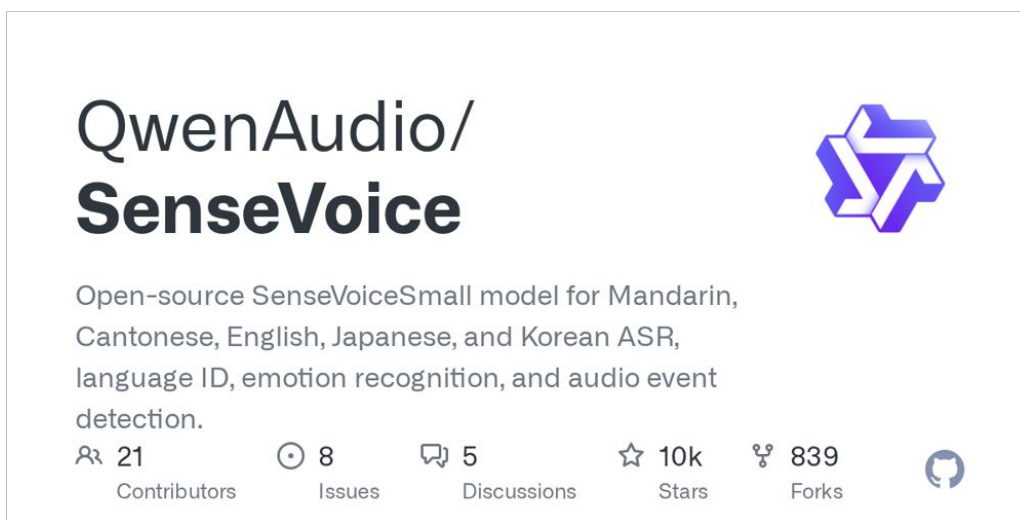
מאדעלן און נייע לאנסירונגען

- **2 יולי: Figma AI – Figma** האט לאנסירט א סעריע AI געצייג פאר דיזיינער: וויזועלע זוכן, זוכן פון אסעטס, רעדאקטירן טעקסט און בילדער, פראטאטייפונג, אומבאנעמען לעיערס און מאכן דיזיינס. זיי לויפן אויף מאדעלן פון דריטע פירמעס און זענען בחינם בשעת'ן בעטא.
- **4 יולי: Grok 2 אין אויגוסט – xAI** קומענדיגער מאדעל Grok 2 איז געפלאנט פאר אויגוסט. לויט באריכטן האט מען אים טריינירט מיט 20 טויזנט H100 טשיפס. Grok 3 דערווארט מען לויט קלאנגען ביז סוף יאר, און ער קען דארפן 100 טויזנט H100.
- **4 יולי: Kyutai Moshi** – דער פראנצויזישער אפענער פארשונג-לאב Kyutai האט טריינירט א ריינעם קול-LLM מיט גאר א קליינע פארשפעטיגונג. דער דעמא איז זייער אימפרעסיוו, און מען וועט אים מאכן אפן-סאורס אין די קומענדיגע חדשים.
- **4 יולי: Meta מולטי-טאקען מאדעלן – Meta** האט ארויסגעלאזט פאר פארשונג פריי-טריינירטע מאדעלן וואס זאגן פאראויס עטליכע קומענדיגע ווערטער אויף אמאל. דאס צייגט שטארקע רעזולטאטן און פארשפרעכט צו פארקירצן דאס טריינירן אויף א גרויסן פארנעם.



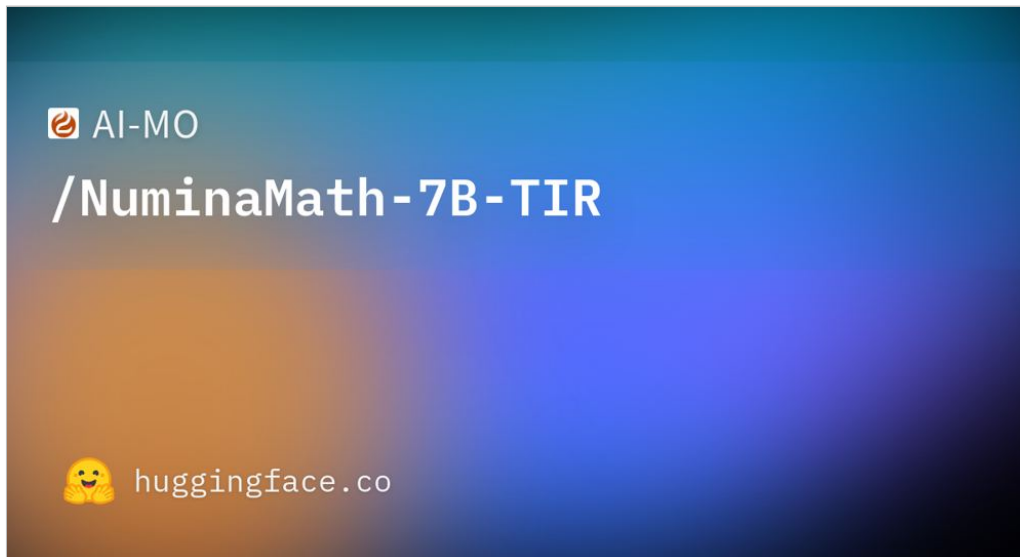
Meta מולטי-טאקען מאדעלן · 4 יולי 2024

- **8 יולי: ElevenLabs Reader** – ElevenLabs האט לאנסירט א Reader אפליקאציע מיט דיגיטאלע קולות פון פארשטארבענע סעלעברייטס, צוזאמען מיט זייערע עסטעיטס. דאס האט אויפגעהויבן פראגעס פון מאראל און געזעץ.
- **8 יולי: Tongyi קול-מאדעלן** – די Tongyi קול-גרופע האט ארויסגעלאזט זייער שטארקע מאדעלן פאר פארשטיין און שאפן אודיא. דערצו איז געקומען SenseVoice, א יסוד-מאדעל וואס דערקענט רייד, די שפראך, געפילן אין קול און אודיא-געשעענישן.



Tongyi קול-מאדעלן · 8 יולי 2024

- **10 יולי: Poe Previews** – Poe's Quora האט צוגעלייגט "Previews", וואס לאזט באניצער שאפן און טיילן אינטעראקטיווע אפליקאציעס, ווי דאטא-וויזואליזאציעס און שפילן, גלייך אינעם שמועס מיט א טשעטבאט.
- **11 יולי: מאטעמאטיק אלימפיאדע מאדעל** – Hugging Face האט ארויסגעלאזט א זייער שטארקן אפענעם מאטעמאטיק-מאדעל.



מאטעמאטיק אלימפיאדע מאדעל · 11 יולי 2024

- **11 יולי:** Claude פראמפט-שפילפלאץ – Anthropic האט צוגעלייגט פאר אנטוויקלער געצייג וואס אויטאמאטיזירן דאס שרייבן פראמפטס: זיי שאפן, פרובירן און פארבעסערן פראמפטס מיט א שנעלן אפרוף.
- **12 יולי:** Samsung Bixby – Samsung וועט נאך דעם יאר לאנסירן א פארבעסערטע Bixby, געבויט אויף אירע אייגענע AI מאדעלן.
- **12 יולי:** Gemini פאר ראבאטן – Google זאגט אז Gemini 1.5 Pro קען לערנען ראבאטן זיך צו באוועגן און אויספירן אויפגאבעס דורך ווידעא-טורן און אנווייזונגען אין פשוטע שפראך. דאס האט דערגרייכט 90% הצלחה.
- **15 יולי:** AuraFlow v0.1 – Fal האט טריינירט AuraFlow, אן אפענעם בילד-מאדעל מיט 5.8 ביליאן פאראמעטערס, מיט דער muP מעטאד.
- **15 יולי:** Lynx – א מאדעל פון Patronus וואס כאפט אסאך בעסער ווי פריערדיגע מעטאדן ווען א שפראך-מאדעל "כאלוצינירט".
- **17 יולי:** Codestral Mamba און Mathstral – Mistral האט טריינירט א Mamba 2 קאוד-מאדעל וואס ארבעט זייער גוט מיט לאנגע קאנטעקסטן און באקומט 75% אויף HumanEval. דערצו האט Mistral ארויסגעלאזט Mathstral, א מאטעמאטיק-מאדעל וואס שניידט גוט אפ אויף פארשידענע בענטשמארקס.
- **17 יולי:** SmolLM – Hugging Face האט ארויסגעלאזט א סעריע קליינע שפראך-מאדעלן וואס שלאגן אסאך גרעסערע. דער סוד ליגט אינעם פארזיכטיגן אויסקלויבן פון דער דאטא.
- **17 יולי:** Claude אויף Android – Anthropic האט לאנסירט די Claude אפליקאציע פאר Android, מיט אנאליז פון בילדער, איבערזעצונגען און סינכראניזירן צווישן מכשירים. אין API האט מען צוויי מאל פארגרעסערט דעם מאקסימום אויסגאנג פאר Claude 3.5 Sonnet, פון 4096 ביז 8192 טאקענס.
- **18 יולי:** Google Vids – Google Vids איז אן AI אפליקאציע צו שאפן ווידעאס פאר ארבעט, טיף איינגעבויט אין Workspace.

- **19 יולי:** Mistral – Mistral Nemo 12B האבן ארויסגעלאזט Mistral Nemo, א פילשפראכיקער מאדעל מיט 12 ביליאן פאראמעטערס, טריינירט מיט א נייעם טאקעניזער. ער ווייזט שטארקע רעזולטאטן אין ענגליש און אנדערע שפראכן, און שטיצט 128 טויזנט טאקענס קאנטעקסט.
- **19 יולי:** OpenAI – GPT-4o Mini האט לאנסירט GPT-4o Mini, א קלענערער און ביליגערער מאדעל וואס זאל פארבייטן GPT-3.5. ער באקומט 82 אויף MMLU, וואס איז גוט פאר א ביליגן מאדעל. דער מאדעל ניצט א נייע טעכניק "אינסטרוקציע-היעראכיע", וואס גיט פארצוג די אנווייזונגען פונעם אנטוויקלער, כדי צו פארשליסן דעם "איגנאָריר אלע פריערדיגע אנווייזונגען" טריק. אזוי וועלן AI אגענטן זיך נישט לאזן אפנארן. OpenAI האט אויך אנגעבאטן בחינם fine-tuning פאר GPT-4o Mini ביז 23סטן סעפטעמבער, ערשט פאר Tier 4 און 5 באניצער.
- **19 יולי:** Flow Studio – Flow Studio שאפט פולשטענדיגע ווידעאס פון 3 מינוט אין פילם-קוואליטעט, מיט א מעשה, קאנסיסטענטע פיגורן און צוגעפאסטע מוזיק און קלאנגען.
- **22 יולי:** Apple – Apple DCLM האט ארויסגעלאזט אלעס, אריינגערעכנט קאוד און דאטא, פון איר 7B מאדעל, וואס קען זיך פארמעסטן מיט Mistral.
- **23 יולי:** ElevenLabs Turbo 2.5 – דער נייער מאדעל ברענגט שמועס-AI פון הויכער קוואליטעט מיט קליינער פארשפעיטונג פאר כמעט 80% פון דער וועלט, אין 31 שפראכן. צום ערשטן מאל שטיצט ער וועטנאמעזיש, אונגאריש און נארוועגיש, און ענגליש איז 25% שנעלער ווי ביי Turbo v2.
- **24 יולי:** Meta – Llama 3.1 405B האט ארויסגעלאזט איר ערשטן אפענעם "פראנטיר" מאדעל, צוזאמען מיט פרטים איבער'ן טריינירן. ער איז גלייך מיט די בעסטע געשלאסענע מאדעלן, מיט 128 טויזנט טאקענס קאנטעקסט און טשעט-מאדעלן. Mark Zuckerberg האט ערקלערט פארוואס אפן-סאורס איז דער שליסל צו Meta'ס סטראטעגיע. Meta האט אויך ארויסגעגעבן א ביישפיל-סיסטעם ווי אזוי צו ניצן Llama ווי AI אגענטן.
- **25 יולי:** Mistral – Mistral Large 2 האט ארויסגעלאזט א מאדעל מיט 123 ביליאן פאראמעטערס וואס איז גלייך מיט Llama 3 405B. ער קומט ארויס אונטער א פריידיגער פארשונג-ליצענץ.
- **25 יולי:** Stability AI – Stable Video 4D'ס נייער מאדעל מאכט פון איין ווידעא פון א חפץ נייע ווידעאס פון 8 פארשידענע ווינקלען. מיט איין דורכלויף שאפט ער 5 פרעיס אין 8 ווינקלען אין בערך 40 סעקונדעס, און דער באניצער קען אויסקלייבן די קאמערע-ווינקלען.



2024 יולי 25 · Stable Video 4D

- **26 יולי:** OpenAI – SearchGPT פרובירט א פראטאטייפ פון א זוך-סיסטעם.
- **26 יולי:** DeepMind – AlphaProof האט פאראייניגט א שפראך-מאדעל אין Gemini-סטיל מיט אן RL אלגאריטם אין AlphaGo-סטיל. דער סיסטעם האט געלייזט 4 פון 6 פראבלעמען אין דעם יאר'ס אינטערנאציאנאלע מאטעמאטיק אלימפיאדע (IMO), אויפ'ן שטאפל פון א זילבערנעם מעדאל.
- **26 יולי:** Bing ג'ענערעטיוו זוכן – Microsoft האט אריינגעשטעלט אין Bing ענטפערס געשריבן דורך AI ווי די הויפט רעזולטאטן, מיט לינקס צו די מקורות. דאס איז נאך אין פראווא-שטאפל פאר געוויסע פראגעס, נאכדעם וואס Google האט צוגעלייגט AI Overviews.
- **29 יולי:** Gemini 1.5 Flash – Google האט פארבעסערט Gemini 1.5 Flash: ער ענטפערט שנעלער, נעמט אריין מער אינהאלט אין פראמפטס, און איז בעסער אין באטראכטן און פארשטיין בילדער.
- **30 יולי:** SAM 2 – Meta האט ארויסגעלאזט SAM 2, וואס צעטיילט בילדער און ווידעאס לויט א פראמפט אין רעאלער צייט.
- **30 יולי:** Apple's מאדעלן – Apple האט באשריבן די יסודות פון אירע שפראך-מאדעלן פאר Apple Intelligence. זי האט אויך געזאגט אז מען האט זיי פריי-טריינירט אויף טשיפס וואס Google האט דיזינט.
- **31 יולי:** Meta AI Studio – Meta's נייע AI Studio וועט לאזן באניצער אן טעכנישע וויסן שאפן אייגענע טשעטבאטס פאר Instagram, Messenger און WhatsApp, מיט פולע קאנטראל איבער אויטאמאטישע ענטפערס.
- **31 יולי:** ChatGPT קול-מאדע – Sam Altman האט באשטעטיגט אז דער נייער קול-אינטערפעיס קומט פאר ChatGPT Plus אבאנענטן. OpenAI לאנסירט א קול-טשעטבאט געבויט אויף GPT-4o, מיט נאטירליכע, פליסיגע שמועסן, ערשט פאר געוויסע Plus באניצער.

• **31 יולי:** Azure AI נייעס – Microsoft האט צוגעלייגט serverless fine-tuning פאר Phi-3 מאדעלן און פארבעסערט Phi-3-mini, GPT-4o mini און Llama 3.1 זענען איצט איינגעבויט אין Azure AI.

פארשונג און טעכניק

• **1 יולי:** בילד און טריינירן – Lambda Labs האט טריינירט דעם Open Sora ווידעא-מאדעל צו מאכן Lego פילמען. די Mosaic גרופע האט מיט PyTorch באשריבן ווי אזוי זיי וואקסן MoE מאדעלן ביז טויזנטער GPUs. MUMU איז א בילד-מאדעל וואס מישט טעקסט און בילדער אין דער אנווייזונג, און שניידט בעסער אפ ווי מאדעלן וואס ניצן נאר טעקסט.

• **1 יולי:** שרייבן לויט אנווייזונגען – Suri איז א דאטאסעט פון 20 טויזנט לאנגע טעקסטן מיט קאמפליצירטע אנווייזונגען, צוזאמען מיט דער I-ORPO מעטאד וואס ניצט קינסטלעך פארדארבענע אנווייזונגען. אן אנדער ארבעט ווייזט ווי אזוי צו לערנען מאדעלן צו האלטן זיך ביי א געבעטענער לענג, ווייל רוב מאדעלן טוען דאס נישט.

• **1 יולי:** Cambrian-1 און DEX-TTS – NYU's Cambrian-1 איז א פולשטענדיג אפענער, שטארקער וויזועלער מאדעל, מיט ברייטע עקספערימענטן אויף טעקסט-ענקאדערס און דאטא-מישונגען. DEX-TTS איז אן אויסדרוקספולער טעקסט-צו-קול מאדעל וואס ניצט א ביישפיל-קול.

• **2 יולי:** muP און פערסאנעס – א צוטריטלעכער ארטיקל ערקלערט ווי אזוי צו שטעלן היפער-פאראמעטערס ווען מען וואקסט. אן ארבעט באטראכט וועב-טעקסט ווי עס קומט פון א "פערסאנע", און ניצט א ביליאן פערסאנעס צו שאפן סינטעטישע דאטא. דאס האט געגעבן א שפּרונג פון 20 פונקטן אויף MATH.

• **2 יולי:** נאך פארשונג – א נייער פראצעס פארבעסערט Web2Code צו שאפן קאוד פון וועבזייטלעך. UnSAM צעטיילט בילדער אן מענטשליכע באצייכענונגען. Block Transformer פארקלענערט שטארק דעם KV cache, אבער עס איז נאך נישט באוויזן אויף א גרויסן פארנעם. דערצו קומט א נייע אויפגאבע צו געפינען "מאדנע" חפצים אין א סצענע.

• **3 יולי:** Meta 3D Gen און מער – Meta האט טריינירט א מאדעל צו שאפן 3D חפצים מיט PBR טעקסטורן, און ניצט דערביי איר אייגענעם 2D מאדעל צו שאפן סינטעטישע דאטא. אן ארבעט לאזט קאנטראלירן א מאדעל'ס אויסגאנג מיט לאגישע באדינגונגען אין דער צייט פון אינפערענס. PDL ווערט אויסגעברייטערט צו קלאסיפיקאציע.

• **3 יולי:** ראבאטן און מולטימאדאל – LLaRA ניצט שמועס-פארמאט אנווייזונגען צו פארבעסערן ראבאט-אקציעס. MM-Instruct איז א גרויסער דאטאסעט פאר מולטימאדאלע מאדעלן. RSCaMa באשרייבט מיט ווערטער ענדערונגען אין סאטעליט-בילדער, און InstantStyle+ טראגט איבער סטיל מיט דיפוזיע-מאדעלן.

• **4 יולי:** דיפוזיע-שפראך און מעדיצין – Sasha Rush האט געגעבן א לעקציע איבער ניצן דיפוזיע פאר שפראך-מאדעלן. MInference פארשנעלערט די באהאנדלונג פון לאנגע פראמפטס. נייע מעטאדן שאפן 3D סצענעס פון א פארשווימענעם בילד, דיאגנאזירן Alzheimer פון MRI אויף א מער ערקלערבארן אופן, און כאפן א שלאג (stroke) אין סעקונדעס דורכן אנאליזירן דאס פנים.

• **5 יולי:** ווייזע און דאטאסעטס – CELLO איז א דאטאסעט פון 14,094 פראגעס איבער סיבה און תוצאה. UIST פארבעסערט די פאראויסזאגונג פון קליקס. Google's Magic Insert שטעלט אריין חפצים אין

בילדער. OpenStreetView-5M האט מער ווי 5 מיליאן גאס-בילדער פון 225 לענדער. FlexiFilm שאפט ווידעאס לענגער ווי 30 סעקונדעס. דערצו קומען מעטאדן פאר קאמערע-טראקינג, שארפקייט פון סמארטפאון-בילדער, און קלארע רייד (SEMamba).

• **8 יולי: קליינע מאדעלן** – Meta's פארשונג-ארבעט ווייזט ווי אזוי צו טריינירן מאדעלן אונטער א ביליאן פאראמעטערס זאלן זיין ווי שטארקסט מעגליך (צוזאמען מיט MobileLLM, 24071006). Craiyon's CapPa ווייזט טאפ רעזולטאטן אויף וויזועלע בענטשמארקס, און SDXL איז ווידער טריינירט פון אנהייב מיט א פולן LLM, פאר בעסערן טעקסט-פארשטאנד.

• **8 יולי: מעדיצינישע בילדער** – א מאדעל איבערזעצט MRI בילדער צווישן סעקווענצן אן פאר-ביישפילן. EGIInet פארפולשטענדיגט פוינט-קלאודס, און א נייע מעטאד בויט אויף 3D סצענעס פון אפעראציעס.

• **8 יולי: Google Smart Paste** – Google's גרופע פאר קאוד-געצייג האט טריינירט א מאדעל וואס פאסט צו אריינגעקלעפטן קאוד לויטן קאנטעקסט. 42% פון די פארשלאגן ווערן אנגענומען, ביי 6% פון אלעם קאפירטן קאוד.

• **9 יולי: DeepMind ראץ** – DeepMind און Harvard האבן געבויט א ווירטועלע ראץ מיט נייראל נעצן טריינירט אויף ריכטיגע באוועגונגען, צו פארשן ווי דער מוח פירט באוועגונגען. דאס ברענגט אויך ידיעות פאר ראבאטיק. Neel Nanda פון DeepMind האט פארעפנטלעכט זיין ליסטע פון בעסטע ארבעטן אין mechanistic interpretability.

• **9 יולי: אגענטן און 3D** – AriGraph בויט פאר AI אגענטן א זכרון-גראף. PartCraft לאזט אויסקלייבן טיילן פון א בילד. PlatoNeRF פון MIT און Meta ניצט שאטנס און lidar צו בויען 3D סצענעס, אריינגערעכנט פארשטעלטע ערטער. דערצו קומען AnySR און א 3D פאזע-מעטאד אן מענטשליכער אויפזיכט.

• **10 יולי: System 2 און TTT** – מען קען דיסטילירן דאס "System 2" באטראכטן (ווי Chain of Thought) אריין אין א שנעלערן "System 1" מאדעל, מיט כמעט די זעלבע פינקטלעכקייט. א נייער RNN וואס לערנט אין דער צייט פון טעסט שלאגט Mamba, ווייל זיין דערהיינטיקונגס-פונקציע איז אליין אן ML מאדעל. אנדערע ארבעטן: Video-STaR, טראפיק-סיגנאלן מיט LLM, און OccSora.

• **11 יולי: Anole און Paints UNDO** – Anole לייגט צוריק בילד-שאפונג צו Meta's Chameleon. Paints UNDO, פונעם שעפער פון ControlNet, מאלט א בילד מיט שטריכן, און קען צוריקנעמען שטריכן און טוישן דעם צוגאנג, ענליך צו א מענטשלעכן קינסטלער. דערצו קומען SuperFlow, PerlDiff, פאר אויטאנאמע קארס, און פערזענליכע בילד-מאדעלן.

• **12 יולי: FlashAttention 3 און PaliGemma** – FlashAttention 3 ניצט 75% פון H100's קראפט, אנשטאט 35% פריער. דער טעכנישער באריכט פון PaliGemma (און SigLIP) ערקלערט די באשלוסן איבער ארכיטעקטור און דאטא.

• **12 יולי: נאך ארבעטן** – OV-DINO און OVFormer זענען פאר אפענעם-ווערטערבוך דעטעקציע. Lookback Lens כאפט כאלוצינאציעס. UCE ציט ארויס עטליכע באגריפן פון איין בילד. דערצו קומען סאטעליט-ענדערונגען און PosFormer פאר האנטגעשריבענע מאטעמאטיק.

- **15 יולי:** **OpenDiLoCo** – Prime Intellect האט איבערגעמאכט DeepMind's DiLoCo, וואס לאזט טריינירן איבער פארשידענע דאטא צענטערס און דאך פול ניצן די GPUs. דערצו קומען MARS פאר בילדער, ViTime פאר צייט-סעריעס, און GBRL.
- **16 יולי:** **SpreadsheetLLM** – א פארשונג-ארבעט ווייזט ווי אזוי צו פארוואנדלען א ספרעדשיט אין א פארם וואס א LLM פארשטייט, פאר פראגעס און פארמאטירן. דערצו קומען LAPT, Hyper-3DG, ProDepth, LightenDiffusion, און Open-Canopy, א דאטאסעט פון בוים-הייך איבער פראנקרייך.
- **17 יולי:** **SciCode** – HumanEval איז אין עיקר שוין געלייזט, און SciCode איז א נייער בענטשמארק מיט שווערע וויסנשאפטליכע פראגראמיר-פראבלעמען. DeRTa פארבעסערט זיכערהייט דורך פארריכטן די פאזיציע פון אפואגן. Yi Tay האט געשריבן איבער ענקאודערס און דעם גורל פון BERT.
- **18 יולי:** **Prover-Verifier** – OpenAI האט טריינירט א שטארקן מאדעל צו שרייבן טעקסט וואס א שוואכער מאדעל קען ריכטיג אפשאצן, און דאס האט אלגעמיין פארבעסערט ווי לייענבאר דער טעקסט איז. אנדערע ארבעטן: SFPNet פאר LiDAR, ReasonVOS, און נאך.
- **19 יולי:** **RAG מיט שפראך-מאפעס** – Mutable AI בויט א Wikipedia-ענליכע "שפראך-מאפע" פון א קאוד-באזע, און דאס פארבעסערט שטארק די ענטפערס. RECE מעקט אויס אומפאסיגן אינהאלט פון דיפוזיע-מאדעלן אין 3 סעקונדעס. דערצו קומען patch-level טריינירן, E5-V, GraphMuse, און SPIQA פאר וויסנשאפטליכע ארבעטן.
- **22 יולי:** **ווערטערבוך-גרייס** – א פארשונג ווייזט אז גרעסערע מאדעלן דארפן א גרעסערן ווערטערבוך. דערצו קומען MoME פאר מולטימאדאלע מאדעלן און CerberusDet.
- **23 יולי:** **AssistantBench** – א נייער בענטשמארק פאר די אגענט-פעאיגקייטן פון א מאדעל. אנדערע ארבעטן: PlacidDreamer פאר טעקסט-צו-3D, SurvReLU, און TRG.
- **24 יולי:** **LocoTrack** – LocoTrack פאלגט נאך פונקטן אין ווידעא, בערך 6 מאל שנעלער ווי די בעסטע מאדעלן. Skoltech און Sber זאגן פאראויס דערטן א יאר פאראויס, און דער גרעסטער באנק אין רוסלאנד וועט דאס איינבויען. דערצו קומען Odyssey פאר Minecraft, STAMP, Artist, און די "star operation".
- **25 יולי:** **KAN אדער MLP** – א יושרדיגער פארגלייך געפינט אז KAN איז בעסער ווי MLP נאר ביי סימבאלישע פארמלען, און ערגער ביי אלע אנדערע אויפגאבעס. Gretel האט פארגליכן סינטעטישע דאטא מיט דאטא פון עקספערטן. דערצו קומען DeepRank-GNN-esm פאר פראטאינען, INF-LLaVA, MOMAland, און Change-Agent.
- **26 יולי:** **u-muP** – א ווערסיע פון muP וואס לאזט אויך איבערטראגן היפער-פאראמעטערס צו קוואנטיזירטע מאדעלן. דערצו קומען LoFormer, Raindrop Clarity מיט 15,186 בילדער, Trajectory2Pose, און א סקירה פון 3D Gaussian Splatting.
- **29 יולי:** **MINT-1T** – Salesforce האט ארויסגעלאזט MINT-1T, א מולטימאדאלן דאטאסעט וואס איז 10 מאל גרעסער ווי פריער. דערצו קומען ALDI, LKCell, RefMask3D, StreamMOS, BetterDepth, און מאביל-פאטאס מיט ספעקטראמעטערס.
- **30 יולי:** **קליין אבער אפט** – אויב מען ניצט א קלענערן מאדעל עטליכע מאל, קען מען בעסער שרייבן קאוד פאר כמעט די זעלבע הוצאה. צוויי מאדעלן פאר מאלעקולן, טריינירט אויף 100 מיליאן מאלעקולן, האבן

געבראכט א פארבעסערונג פון 8%. א דאטאסעט פון מער ווי 300 טויזנט סינטעטישע שמועסן איז ארויס פאר אפגענוע מאדעלן. דערצו קומען AccDiffusion, HYBRIDDEPTH, VSSD.

• **31 יולי: Theia און SaulLM** – Theia פאראייניגט עטליכע וויזועלע מאדעלן פאר ראבאט-לערנען מיט ווייניגער דאטא. SaulLM-54B און SaulLM-141B זענען מאדעלן פאר משפט, געבויט אויף Mixtral. דערצו קומען LightGODE, פראגעס אויף טשארטס, בעסערע פנים-בילדער, און DGFNet.

געצייג פאר פראגראמירער און אפן-סאורס

• **1 יולי: Llama Agents און מער** – llama-agents איז א פריימווערק צו בויען סיסטעמען מיט עטליכע AI אגענטן. דערצו קומען א Laplace אפעראטאר פאר PDEs, א פירער פאר דיבאגן PyTorch, און Vision Agent, וואס שרייבט קאוד פאר וויזועלע אויפגאבעס.

• **2 יולי: Composio און Screen Pipe** – Composio גיט AI אגענטן געצייג, און Screen Pipe פארוואנדלט וואס איז אויפ'ן עקראן אין אקציעס.

• **3 יולי: Mutahunter; אוועק פון LangChain** – Mutahunter איז אן אפן-סאורס געצייג פאר mutation testing. Octomind האט פארלאזט LangChain, ווייל זיינע אבסטראקציעס זענען נישט געווען בייגעווייגט, און איז אריבער צו פשוטערע בויסטיינער, וואס האט פארבעסערט דעם קאוד און די פראדוקטיוויטעט.

• **4 יולי: Agentless** – א סיסטעם לייזט מער ווי 24% פון SWE-bench נאר מיט א שפראך-מאדעל, אן אגענטן. דערצו קומען א מאנועל פאר torch.compile און Maestro פאר Claude Opus.

• **5 יולי: LLM101n** – Andrej Karpathy האט אנאנסירט א קורס צו בויען א "Storyteller" LLM פון אנהייב, מיט Python, C און CUDA. דערצו קומען PTQ4SAM, און Florence 2 וואס לויפט אינגאנצן אינעם בלעטערער מיט WebGPU.

• **8 יולי: Quality prompts** – א ביבליאטעק וואס זאמלט פראמפט-טריקס און ווענדט זיי אויטאמאטיש אן.

• **9 יולי: Minference און Micro Agent** – Microsoft's Minference פארשנעלערט אינפערענס שטארק. Micro Agent שרייבט און פארריכט קאוד.

• **10 יולי: GraphRAG און Tinygrad** – GraphRAG פון Microsoft פארשט אומגעארדנטן טעקסט מיט שפראך-מאדעלן און לויפט אויף Azure. דערצו קומען נאטיצן איבער Tinygrad.

• **11 יולי: minRF און RouteLLM** – minRF איז א מינימאלע אימפלעמענטאציע פון Rectified Flow Transformers, וואס ווערן גענוצט אין RouteLLM. Stable Diffusion 3 איז פאר LLM ראוטערס. א טריק פארשנעלערט 30 מאל דאס אנהייבן פון א מאדעל אין HF Transformers.

• **15 יולי: GPU CPP און Answer AI** – Answer AI האט ארויסגעלאזט א לייכטע ביבליאטעק פאר GPU חשבונות, געבויט אויף WebGPU.

• **16 יולי: Crawlee** – א געצייג צו זאמלען וועב-דאטא אויף א גרויסן פארנעם.

• **18 יולי: PraisnAI; GraphRAG אנאליז** – PraisnAI איז א low-code פריימווערק פאר סיסטעמען מיט עטליכע אגענטן. אן אנאליז געפינט אז Neo4j איז נישט שטארק בעסער ווי FAISS ביים ארויסזוכן קאנטעקסט, הגם זיין אינדעקס פארבעסערט ווי טריי די ענטפערס זענען.

- **19 יולי: LLM2sh; Abliteration** – LLM2sh איבערזעצט בקשות אין shell באפעלן. "Abliteration" נעמט ארויס דעם אפזאג-מעכאניזם פון Llama מאדעלן אן טריינירן פון ניי.
- **22 יולי: Mantle און Endia** – Mantle האט גענוצט Gemini 1.0 Pro מיט א קאנטעקסט פון א מיליאן טאקענס צו איבערמאכן א פראטאטייפ אין א פראדוקציע-קאוד, און דאס האט געשפארט חדשים ארבעט. דערצו קומען Endia פאר OpenOCR, Txtai, Mojo, און א CLI מיט Claude Sonnet 3.5 וואס פארשלאגט קאוד-ענדערונגען.
- **23 יולי: Orch** – Orch איז א ביבליאטעק אין Rust פאר LLM אפליקאציעס. דערצו קומען א ארטיקל וויאזוי צו ניצן LLMs צו אפשאצן אנדערע מאדעלן, און שנעלקייט פון סטרוקטורירטער גענעראציע.
- **25 יולי: Rule Based Rewards** – OpenAI האט ארויסגעגעבן קאוד און דאטא פון איר פראיעקט פון כללים-באזירטע באלינונגען פאר זיכערהייט.
- **26 יולי: dlordinal** – dlordinal איז א Python ביבליאטעק פאר אַרדינאלע קלאסיפיקאציע. דערצו קומט א ארטיקל וויאזוי צו בויען א ג'ענעראטיוו AI פלאטפארמע.
- **29 יולי: פילשטאפלדיגע LLM אפליקאציעס** – טאקטיקן מיט קאוד-ביישפילן.
- **30 יולי: AppWorld** – א סביבה מיט 9 אפליקאציעס און 457 APIs צו טעסטן AI אגענטן. דערצו קומט א קורס איבער MLPs פאר גראפיק.

פירמעס, געלט און ביזנעס

- **1 יולי: Amazon שלינגט Adept** – Amazon האט אנגעשטעלט Adept'ס גרינדער און רוב מאנשאפט, און ליצענצירט אירע טעכנאלאגיע. The Verge זעט דאס ווי א מוסטער, ענליך צו Microsoft מיט Inflection, ווי די גרויסע פירמעס נעמען AI טאלאנט אן א פארמעלן קויף, כדי צו אויסמיינדן אנטי-טראסט אויפזיכט.
- **1 יולי: OpenAI און Time** – Time גיט OpenAI רשות צו טריינירן אויף 101 יאר פון אירע ארטיקלען און צו ניצן אירע אקטועלע נייעס. אין אויסטויש וועט OpenAI ציטירן Time און לינקן צו איר.
- **1 יולי: Apple, OpenAI און Gemini** – Apple פלאנט צו מעלדן דעם הערבסט א אפמאך פאר Gemini אויף אירע מכשירים. Phil Schiller האט געזאלט באקומען אן אבזערוואַטאר-פלאץ אויף OpenAI'ס באורד, אבער צוליב רעגולאטארישע אויפזיכט האט Apple זיך צוריקגעצויגן און Microsoft האט אוועקגעלייגט איר אבזערוואַטאר-ראלע.
- **4 יולי: Nvidia אסיפה** – Jensen Huang זאגט אז Nvidia האלט מער ווי 80% פונעם AI טשיפ-מארק צוליב אן אינוועסטירונג פון א צענדליג יאר צוריק, און אז די פירמע איז געווארן א דאטא-צענטער פירמע.
- **4 יולי: Persona AI** – Jerry Pratt (פריער CTO פון Figure) און Nic Radford האבן געגרינדעט Persona AI, א נייע פירמע פאר מענטש-ענליכע ראבאטן.
- **5 יולי: Intel אפטישע טשיפ** – Intel האט געוויזן דעם ערשטן פולשטענדיג איינגעבויטן אפטישן OCI chiplet, צו פארגרעסערן די דאטא-שנעלקייט פאר AI.
- **8 יולי: Figma ציט צוריק** – Figma האט צייטווייליג אראפגענומען "Make Designs" ווייל עס האט געשאפן דיזיינס צו ענליך צו Apple'ס אפליקאציעס. דער CEO Dylan Field באשולדיגט די דריטע-פירמע

- מאדעלן (GPT-4, Amazon Titan) און א דיזיין-סיסטעם, און נישט טריינירן אויף Figma'ס אינהאלט. מען וועט עס צוריקשטעלן מיט בעסערע קאנטראלן.
- **8 יולי:** **Apple M5** און **Siri** – לויט באריכטן וועט M5 ניצן SoIC טעכנאלאגיע פאר Macs און AI סערווערס. א בעסערע Siri מיט Apple Intelligence קומט מיט iOS 18.4 אין פרילינג 2025.
 - **10 יולי:** **a16z Oxygen** – Andreessen Horowitz האט פארזיכערט טויזנטער טשיפס, אריינגערעכנט H100, צו פארטיילן צו אירע פירמעס אין אויסטויש פאר אקציעס.
 - **10 יולי:** **xAI און Oracle** – xAI האט פארלאזט א אפמאך מיט Oracle פאר 24 טויזנט H100, און וועט בויען איר אייגענעם דאטא צענטער נאכ'ן טריינירן Grok 2.
 - **12 יולי:** **Snapdragon לעפטאפס** – Qualcomm'ס Snapdragon X Elite און X Plus פארמעסטן זיך מיט Apple M3, Intel און AMD אין כח און באטעריע, און זענען ביליגער.
 - **15 יולי:** **אגענט אינושעניר** – Bret Taylor'ס Sierra האט אריינגעפירט א נייע ראלע: Agent Engineer.
 - **15 יולי:** **OpenAI'ס פינאנצן** – OpenAI'ס איינקונפט ווערט געשאצט אויף 3.4 ביליאן דאלאר. א באריכט זאגט אז די הוצאות פאר טריינירן און אינפערענס קענען דערגרייכן 7 ביליאן אין 2024, און די פירמע קען פארלירן 5 ביליאן און דארפן נייע געלט אין א יאר. זי ניצט דאס גלייכע פון 350 טויזנט A100 סערווערס.
 - **16 יולי:** **Hebbia** – Hebbia האט אויפגענומען 130 מיליאן דאלאר ביי א ווערט פון 700 מיליאן, מיט 13 מיליאן ARR. איר פראדוקט Matrix ווערט גענוצט פון 30% פון די asset managers.
 - **16 יולי:** **Pixel 9 Pro** – קויפער וועלן באקומען א יאר Gemini Advanced בחינם.
 - **16 יולי:** **Lattice** – די HR פירמע האט אפגעזאגט א פלאן צו געבן AI "רעכטן" ווי דיגיטאלע ארבעטער, נאך א שטארקע קעגנרעאקציע.
 - **17 יולי:** **Eureka Labs** – Andrej Karpathy'ס נייע פירמע פאר AI אין חינוך.
 - **17 יולי:** **Redis** – לויט Redis'ס אייגענע בענטשמארקס איז זי 62% שנעלער ביי פשוטע דאטאסעטס און 21% ביי הויך-דימענסיאנאלע, קעגן Qdrant, Milvus און Weaviate.
 - **18 יולי:** **Anthology Fund** – Anthropic מיט Menlo Ventures לאנסירן א פאנד פון 100 מיליאן דאלאר, מיט מינדסטנס 100 טויזנט פאר יעדן סטארטאפ.
 - **18 יולי:** **Thrive AI Health** – OpenAI און Thrive Global פינאנצירן א פערזענליכן AI געזונט-קאוטש פאר שלאף, עסן און דאס גלייכן.
 - **18 יולי:** **aiXplain** – 6.5 מיליאן דאלאר pre-Series A פון Wa'ed Ventures פון Saudi Aramco.
 - **19 יולי:** **Cohere** – Cohere און Fujitsu גייען צוזאמען פאר AI אין יאפאניש. Cohere האט אויפגענומען 500 מיליאן דאלאר אין א Series D ביי א ווערט פון 5.5 ביליאן, מיט AMD, Fujitsu, Cisco, PSP און EDC Ventures.
 - **19 יולי:** **OpenAI און Broadcom** – די צוויי פירמעס האבן גערעדט וועגן בויען א נייעם AI סערווער טשיפ.

- **22 יולי: Nvidia B20 פאר כינע** – Nvidia אנטוויקלט א B20 טשיפ וואס פאסט צו די אמעריקאנער עקספארט-געזעצן. די H20 קען דעם יאר פארקויפן מער ווי א מיליאן שטיק אין כינע, ווערט 12 ביליאן דאלאר.
- **22 יולי: Artificial Agency** – געוועזענע DeepMind פארשער האבן אויפגענומען 16 מיליאן דאלאר פאר AI פיגורן אין ווידעא-שפילן.
- **24 יולי: Harvey און Gumloop** – Harvey, א סטארטאפ פאר אדוואקאטן, האט אויפגענומען 100 מיליאן דאלאר אין א Series C און איז געווארן א "יוניקארן". Gumloop האט אויפגענומען 3.1 מיליאן אין א seed רונדע פאר no-code AI אויטאמאציע.
- **24 יולי: Tenstorrent; Samsung** – Tenstorrent פארקויפט Wormhole קארטלעך פאר 999 און 1,399 דאלאר, און סטאנציעס פאר 12 און 15 טויזנט. Samsung לייגט אריין דעם לייב'ס חלק פון איר פארשונג אין "AI פאונס".
- **26 יולי: Neuralink** – געוויסע מיטארבעטער גרייטן זיך צו פארקויפן אקציעס נאכן שפּרונג אין ווערט.
- **29 יולי: Musk פירמעס** – Elon Musk וועט "דיסקוטירן" מיט Tesla איבער אינוועסטירן 5 ביליאן דאלאר אין xAI's robotaxi וועט ווערן געוויזן דעם 10טן אקטאבער. לויט קלאנגען קען Nvidia ארויסלאזן א 63% Titan GPU, שנעלער ווי RTX 4090.
- **31 יולי: Perplexity מודעות** – Perplexity וועט פארקויפן מודעות און טיילן די איינקונפט מיט פארלעגער וועמענס אינהאלט ווערט גענוצט.
- **31 יולי: Canva קויפט Leonardo** – Canva האט געקויפט Leonardo, א פירמע פאר בילד-שאפונג.

רעגירונגען, געזעצן און זיכערהייט

- **1 יולי: סקרעפינג קלאגעס** – AWS אויספארשט Perplexity איבער ארויסגעגעבן דאטא פון וועבזייטלעך אן רשות. די CEOs פון Freelancer.com און iFixit קלאגן אז Anthropic's ClaudeBot האט געשטערט זייערע זייטלעך. Anthropic זאגט אז זי האלט זיך ביי robots.txt, און וועט דאס אויספארשן.
- **2 יולי: דאטא פאר טריינירן** – Google און Meta טוישן זייערע תנאים צו ניצן באניצער-דאטא פאר טריינירן, און דאס רופט ארויס קעגנרעאקציעס. Anthropic, Nvidia, Apple און Salesforce האבן גענוצט אונטערשריפטן פון 173,536 YouTube ווידעאס פון 48 טויזנט קאנאלן. Taylor & Francis האט פארקויפט צוטריט צו פארשונג-ארבעטן צו Microsoft, אן צו פרעגן די שרייבער.
- **3 יולי: YouTube** – מען קען איצט פארלאנגען אראפצונעמען AI אינהאלט וואס נאכמאכט אייער קול אדער אויסזען.
- **5 יולי: OpenAI זיכערהייט** – די ChatGPT אפליקאציע פאר Mac האט געהאלטן שמועסן אין פשוטן טעקסט. א העקער איז אריין אין OpenAI'ס אינערליכע מעסעדזש-סיסטעם אין 2023, אבער נישט אין די AI סיסטעמען. מען האט נישט מודיע געווען דעם עולם, ווייל קיין קונה-דאטא איז נישט גע'גנב'עט געווארן.
- **12 יולי: שאנכאי ראבאטן-כללים** – די ערשטע כללים פאר מענטש-ענליכע ראבאטן, מיט פאקוס אויף מענטשליכער זיכערהייט.
- **15 יולי: OpenAI אין כינע** – OpenAI האט פארשפארט באניצער אין כינע, אפילו דורך VPN.

- **17 יולי: NDA קלאגס** – Whistleblowers בעטן ביי דער SEC אויסצופארשן OpenAI'ס NDAs, וואס זיי זאגן באגרענעצן אויף אן אומלעגאלן אופן קאמוניקאציע מיט רעגולאטארן.
- **17 יולי: Gray Swan AI** – א פירמע פאר AI זיכערהייט און ריזיקע-אפשאצונג.
- **18 יולי: Meta און EU** – Meta וועט נישט אנבאטן אירע קומענדיגע מולטימאדאלע מאדעלן אין EU צוליב רעגולאטארישע אומזיכערקייט, נאר א טעקסט-ווערסיע פון Llama 3.
- **22 יולי: UN** – די UN וויל שאפן א צענטראלן פארום פאר AI רעגירונג.
- **23 יולי: קאליפארניע געזעץ** – Scott Wiener'ס געזעץ פאדערט פון פירמעס וואס טריינירן מאדעלן פאר מער ווי 100 מיליאן דאלאר צו מאכן זיכערהייט-טעסטן און קענען אפשטעלן דעם מאדעל. די טעכנאלאגיע-אינדוסטריע קריטיקירט עס שטארק.
- **25 יולי: טראמפ און AI** – די רעפובליקאנער פלאטפארמע פארשפרעכט צו אנולירן Biden'ס עקזעקוטיוו-ארדער איבער AI. לויט קריטיקער (Tech Policy Press) וועט דאס שטארקן Big Tech.
- **31 יולי: Altman'ס פלאן** – Sam Altman רופט אז אמעריקע מיט אליאירטע זאלן פירן א גלאבאלע AI איניציאטיוו לויט דעמאקראטישע ווערטן: זיכערהייט, אינפראסטרוקטור, קאמערציעלע דיפלאמאטיע און געמיינזאמע נארמען, קעגן אויטאריטארע רעזשימען.

אנאליזן און מיינונגען

- **1 יולי: אינוועסטירן אין AI** – Eastwind זאגט אז טראץ דער "איפאריע" ווייזט דער מארק שוין שוואכקייטן, און אז דער מארק פאר אינפערענס איז באגרענעצט. VCs דארפן גלייבן אז ער וועט שטארק וואקסן.
- **2 יולי: Suleyman איבער Altman** – Mustafa Suleyman פון Microsoft AI האט געלויבט Sam Altman ווי ערנסט וועגן זיכערהייט, און גערופן צו דיאלאג און נישט זען AI ווי א געיעג, אריינגערעכנט מיט ארבעט מיט כינע.
- **2 יולי: Bias און סקעילינג** – Dan Luu באטראכט די וויכוח איבער AI בייאס. AI Snake Oil זאגט אז AGI דורך נאר סקעילינג איז נישט ווארשיינלעך, און אז גוטע דאטא ווערט שווערער.
- **3 יולי: Google AI Overviews** – לויט SE Ranking דערשיינען AI Overviews איצט נאר אין 8.71% פון זוכן, אראפ פון 64% פריער, און רוב לינקן צו די ערשטע 10 רעזולטאטן.
- **3 יולי: פינף שטאפלען** – Benjamin Bratton פארגלייכט די רעאקציע צו AI מיט א "קאפערניקאנישער טראומע" מיט פינף שטאפלען.
- **4 יולי: נייע ביזנעס-מאדעלן** – אנאליטיקער באשרייבן ווי AI דריקט אויף Salesforce און Zendesk. זיי רעדן פון "Service-as-a-software" און M-SaaS סטארטאפס, וואס הייבן אן מיט ארבעט פון מענטשן און ווערן מער אויטאמאטיש. Tanay Jaipuria טיילט אויף אין דריי מוסטערן: Copilots, Colleagues, AI-Native Services. מען גייט איבער פון פרייזן לויט זיץ צו פרייזן לויט ארבעט אדער רעזולטאט, און אדאפטירן אין פירמעס גייט נאך פאמעלעך.
- **5 יולי: Acemoglu** – Daron Acemoglu שאצט נאר 0.06% יערליכן TFP וואוקס פון AI. קריטיקער (Maximum Progress) האלטן אז ער אונטערשאצט.

- **8 יולי: 10 Superintelligence יאר** – א צענדליג יאר נאך Bostrom'ס בוך איז AI זיכערהייט געווארן א צענטראלע טעמע.
- **9 יולי: ChatGPT און קאוד** – הצלחה פון ווייניגער ווי 1% ביז 89%, לויט שוועריגקייט און פראגראמיר-שפראך.
- **10 יולי: AI זומער און Goldman** – Ben Evans באמערקט אז ChatGPT האט דערגרייכט 100 מיליאן אין צוויי חדשים, אבער אסאך באניצער האבן נישט געפונען א לאנגע נוצן. Goldman Sachs'ס באריכט פון 31 זייטן פרעגט צי מען גיט אויס צו פיל פאר צו ווייניג נוצן.
- **12 יולי: Augmented Intelligence** – א ארטיקל לייגט פאר צו זען AI ווי א געצייג וואס פארשטארקט דעם מענטש.
- **16 יולי: סקעילינג און פינף שטאפלען** – Kevin Scott פון Microsoft גלייבט אז די סקעילינג-געזעצן וועלן ווייטער האלטן. OpenAI האט געשטעלט פינף שטאפלען צו AGI, און האלט אז זי איז אויפ'ן ערשטן שטאפל, נאענט צום צווייטן.
- **16 יולי: באוואוסטזיין** – עס איז נאך נישטא קיין וועג צו וויסן צי AI קען האבן באוואוסטזיין. די טעאריעס זענען צעטיילט.
- **22 יולי: מודעות צאלן פאר AI** – מיט GPT-4o mini קאסט א בלאג-ארטיקל 0.00051525 דאלאר, און א מודעה ברענגט 0.0026, אזוי אז מען פארדינט 0.002 פאר יעדן בליק.
- **29 יולי: א בלאז?** – Kelvin Mu פון Translink Capital פארגלייכט מיט די 90ער יארן, און האלט אז מיר זענען נאך ווייט פון א בלאז.
- **31 יולי: פארוואס בחינם** – א ארטיקל זאגט אז Meta'ס Llama 3.1 קען מאכן LLMs פאר א פשוטע סחורה, און פארגרעסערן די נאכפראגע פאר סערווערס, אויפ'ן חשבון פון סטארטאפס.

ווייטערע ידיעות

- **11 יולי: ראווער און קארס** – ראווער אין סאן פראנציסקא האבן אריינגעגעבן כמעט 200 קלאגעס איבער Waymo און Cruise.
- **11 יולי: OpenAI און Los Alamos** – א שותפות צו פארשן זיכערע אנווענדונגען פון GPT-4o אין ביא-לאבאראטאריעס.
- **18 יולי: Khanmigo** – א שולע אין Newark ניצט Khan Academy'ס AI טוטאר, געשטיצט פון דער Gates פאונדעישאן.
- **24 יולי: AI יעדע מינוט** – יעדע מינוט גיט מען אויס כמעט 300 טויזנט דאלאר אויף AI, און מען שאפט כמעט 20 טויזנט בילדער.
- **31 יולי: AI אינטערוויס** – פירמעס ווי Meituan ניצן AI אוואטארן פאר ערשטע אינטערוויס. קריטיקער פאדערן מער דורכזיכטיגקייט.
- **9 יולי: עטליכע נייע געצייג** – Doti – געזונט-אסיסטענט; Krut AI – פאר e-commerce Pieces; Copilot+ – זכרון אויפ'ן מכשיר; AiEditor – טעקסט-רעדאקטאר; AI Web Designer; Project Atlas –

ביונעס-אגענטן; Shadow – מיטינג-אסיסטענט; HeyGen Interactive Avatar ; Friend – אפן-סאורס
האלדזבאנד; no-code – Airtable Cobuilder אפליקאציעס.

אויגוסט 2024 איז געווען א חודש פון א שטראם נייע מאדעלן: Grok 2, Flux פון Black Forest Labs, Ideogram 2.0. OpenAI און Imagen 3, Phi-3.5, Qwen2-Math, Hermes 3 צענטער. דריי גרינדער האבן פארלאזט די פירמע, Elon Musk האט ווידער אנגעהויבן זיין קלאגע, OpenAI האט לאנסירט Structured Outputs און פיין-טיונינג פאר GPT-4o, און עס פירט שמועסן איבער א געלט-רונדע וואס זאל זי אפשרצן העכער 100 ביליאן דאלאר. אויפן געזעץ-פראנט איז דער EU AI Act אריין אין קראפט, און קאליפארניע'ס געזעץ SB 1047 איז דורך די לעגיסלאטור. אין פארשונג האט מען שטארק געארבעט אויף נוצן מער רעכן-קראפט בשעת'ן ענטפערן, אויף קליינע מאדעלן דורך דיסטילאציע, און אויף Mamba-היברידן. אויך זענען געווען גרויסע אינוועסטירונגען, ווי Groq מיט 640 מיליאן, Cursor מיט 60 מיליאן און Magic מיט 320 מיליאן.

מאדעלן און נייע לאנסירונגען

- **1 אויגוסט: Grok און X'ס דאטא** – Elon Musk זאגט אז זיין AI פירמע טריינירט Grok אויף א שטארקן קלאסטער, מיטן ציל צו האבן "די שטארקסטע AI" ביז דעצעמבער. X נעמט די דאטא פון באניצער אויטאמאטיש פאר'ן טריינירן, און מ'קען דאס אפשרטעלן נאר אין דער וועב-אפליקאציע. נאך א לעגאלן שריט פון אירלאנד'ס דאטא-קאמישאנער וועט Musk אפשרטעלן דאס טריינירן אויף געוויסע EU דאטא. דעם 15טן האט xAI ארויסגעלאזט Grok 2, א פראנטיר-מאדעל פאר לאגיק, קאוד און מאטעמאטיק, און עס ארבעט צוזאמען מיט Black Forest Labs צו ברענגען FLUX צו X באניצער.
- **2 אויגוסט: Black Forest Labs און Flux** – די שאפער פון VQGAN און Stable Diffusion האבן געגרינדעט א נייע פירמע מיט איבער 30 מיליאן דאלאר און ארויסגעלאזט שטארקע בילדער-מאדעלן אין עטליכע מדרגות. Flux קען לויפן אויף א געווענליכן לעפטאפ. עס איז ספעציעל גוט אין צייכענען מענטשן און אין פאלגן דעם באפעל, און אין געוויסע זאכן גייט עס איבער Midjourney. עס קומט אין ווערסיעס Pro, Dev און Schnell, און מ'האט אנגעזאגט אויף אן אפן-סאורס טעקסט-צו-ווידעא מאדעל.
- **2 אויגוסט: Stable Fast 3D – Stability AI** האט ארויסגעלאזט א שנעלן און שטארקן 3D מאדעל, אונטער דער זעלבער קאמערציעלער ליצענץ ווי אירע פריערדיגע מאדעלן.



2024 אויגוסט 2 · Stable Fast 3D

- **5 אויגוסט:** נייעס צו Gemma 2 – Google לייגט צו צו דער Gemma 2 פאמיליע א נייעס 2B מאדעל, א מאדעל וואס סארטירט אינהאלט פאר זיכערהייט, און א געצייג צו פארשטיין ווי א מאדעל ארבעט.



2024 אויגוסט 5 · Gemma 2 נייעס צו

- **7 אויגוסט:** Figure – Figure האט ארויסגעלאזט איר נייעסטן הומאנאידן ראבאט, מיט מער קלוגשאפט און א בעסערן דיזיין.
- **7 אויגוסט:** Structured Outputs ביי OpenAI – א נייע מעגלעכקייט אין OpenAI'ס API גיט ענטפערס אין א פעסטן פארמאט, למשל JSON, וואס אנדערע פראגראמען קענען דירעקט נוצן.
- **7 אויגוסט:** Segment Anything 2 – Mark Zuckerberg האט צוזאמען מיט Jensen Huang פארגעשטעלט SA2, א מאדעל וואס צעטיילט ווידעאס אויף זייערע טיילן. דאס איז א שריט פאראויס פון דעם פאריגן מאדעל, וואס האט געארבעט נאר אויף בילדער.

- **8 אויגוסט:** **Gemini** אין **Google Home** – **Google** ברענגט **Gemini** צו **Nest** קאמערעס און רעדערס. עס שרייבט קלוגע באשרייבונגען פון ווידעא-בילדער, מ'קען שאפן רוטינען מיט פשוט'ע שפראך, און דער **Assistant** קלינגט מער נאטירליך. רוב פון די מעגלעכקייטן זענען הינטער אן **Nest Aware** אבאנעמענט.
- **9 אויגוסט:** **OpenAI – GPT-4o System Card** באשרייבט וויאזוי זי ארבעט צו מאכן דעם מולטי-מאדאלן קול-מאדעל זיכער און נוצבאר.
- **9 אויגוסט:** **Qwen2-Math** – די **Qwen** גרופע האט ארויסגעלאזט א **72B** מאטעמאטיק-מאדעל וואס שלאגט אלע אפענע און פארמאכטע מאדעלן אויף **MATH**. אויף עטליכע לאנג-בענטשמארקס גייט עס אויך איבער **Llama-3.1-405B**. דערווייל ארבעט עס נאר אויף ענגליש.
- **9 אויגוסט:** **Gemini** אין **Chrome** – **Google** ברענגט **Google Lens**, א פארגלייך פון טאבס פאר'ן איינקויפן, און זוכן אין דער בלעטער-היסטאריע מיט נאטירליכער שפראך צו **Chrome** אויף קאמפיוטערס.
- **12 אויגוסט:** **Mistral: La Plateforme** – פראגראמירער קענען יעצט צופאסן **Mistral Large 2** און **Codestral** מיט וויסן פון זייער אייגענעם פאך. מ'האט ארויסגעלאזט אן אלפא ווערסיע פון "Agents" פאר קאמפליצירטע ארבעט-גענג, און דער **SDK** פאר **Python** און **Typescript** איז אנגעקומען צו ווערסיע **1.0**.
- **13 אויגוסט:** **YouTube: Brainstorm with Gemini** – **YouTube** פראבירט א געצייג וואס העלפט אויסגעוויילטע שאפער טראכטן פון אידיען פאר ווידעאס און בילדלעך.
- **14 אויגוסט:** **Sakana's AI Scientist** – די יאפאנעזישע פירמע **Sakana** האט ארויסגעלאזט א סיסטעם וואס פירט אליין פארשונג. עס שלאגט פאר השערות, פירט דורך עקספערימענטן, שרייבט קאוד, און פאסט אלעס צוזאמען אין פארשונג-ארבעטן. מ'האט צוגעשטעלט ביישפילן, און עס איז דא אן אפן-סאורס ווערסיע.
- **14 אויגוסט:** **Gemini 1.5 Flash** ביליגער – דער פרייז פון אריינגעשיקטע טאקענס איז אראפ **78%**, און פון ארויסגעגעבענע טאקענס **71%**. דער **API** שטיצט יעצט איבער **100** שפראכן.
- **15 אויגוסט:** **Prompt Caching** ביי **Claude** – **Anthropic** לאזט פראגראמירער האלטן אין זכרון א טעקסט וואס מ'נוצט אפט, וואס שפארט שטארק אין קאסטן און אין צייט. פירמעס ווי **Notion** נוצן עס שוין.
- **15 אויגוסט:** **Nvidia's Minitron** מאדעלן – **Nvidia** האט דורך שניידן און דיסטילאציע געמאכט פון **Llama 3.1 8B** א **4B** מאדעל. עס האט באקומען **16%** בעסער אויף **MMLU** ווי טריינירן פון אנפאנג, מיט **40** מאל ווייניגער טאקענס. שפעטער איז געקומען **Mistral-NeMo-Minitron 8B**, א פארקלענערטע ווערסיע פונעם **12B** מאדעל, מיט א שפיץ-גענויקייט, וואס לויפט אויף **RTX** קאמפיוטערס.
- **15 אויגוסט:** נייער **GPT-4o** אין **ChatGPT** – **OpenAI** האט שטילערהייט ארויסגעלאזט א פארבעסערטן **GPT-4o** לויט באניצער-באמערקונגען. לויט **OpenAI** איז נישט געטוישט געווארן דער אופן ווי ער טראכט. די ווערסיע אין **ChatGPT** קען זיך אויך אונטערשיידן פון דער ווערסיע אין **API**.
- **15 אויגוסט:** **Pixel מיט AI** – דער **Pixel Watch 3** כאפט אויטאמאטיש אויף ווען מ'שלאפט, אנאליזירט די פארעם ביים לויפן, און רופט הילף ווען דער דופק פעלט. אין א רעצענזיע פון **Pixel 9** טעלעפאנען שטייט אז **Gemini** גיט פארשידענע מעגלעכקייטן, ווי פאסן צוזאם דאקומענטן, אבער אלס אסיסטענט איז עס געמישט, מיט פעלערס.
- **16 אויגוסט:** **Hermes 3** – **Nous Research** האט ארויסגעלאזט איר הויפט-מאדעל, געבויט אויף **Llama 3**, מיט שטארקע רעזולטאטן און א געשמאקע פערזענליכקייט.

- **19 אויגוסט:** **Imagen 3** און **Gems** – Google האט ארויסגעלאזט Imagen 3 דורך AI Test Kitchen און Vertex AI, מיט מער פרטים און בעסערע ליכטיגקייט, און מיט שוין-גרענעצן. דאך האבן באניצער געטראפן וועגן צו שאפן געשטאלטן וואס זענען ענליך צו באשיצטע מאטעריאלן. שפעטער זענען Gems און Imagen 3 געקומען צו Gemini Advanced. מיט Gems קען מען שאפן אייגענע ווערסיעס פון Gemini פאר ספעציפישע ארבעט.
- **19 אויגוסט:** **Runway Gen-3 Alpha Turbo** – עס שאפט ווידעאס 7 מאל שנעלער ווי דער פריערדיגער מאדעל, פאר האלב דעם פרייז.
- **19 אויגוסט:** **DeepSeek Prover** – DeepSeek האט געשאפן א שטארקן מאטעמאטיק-מאדעל, טריינירט מיט פראצעס-אויפזיכט, וואס שלאגט גרעסערע מאדעלן אויף MATH בענטשמארקס.
- **19 אויגוסט:** **Gemini Advanced** און **Gemini Live** – Gemini 1.5 Pro און Gemini Advanced איז פארבעסערט געווארן פאר לאגיק און קאוד. Gemini Live, Google'ס סיסטעם פאר רעדן מיטן קול, האט קריטיק באקומען פאר אויסגעטראכטע ענטפערס, און מ'קען עס ווייניגער צופאסן ווי OpenAI'ס Advanced Voice Mode.
- **21 אויגוסט:** **GPT-4o** – OpenAI לאזט פראגראמירער צופאסן GPT-4o מיט זייערע אייגענע דאטא, מיט 1 מיליאן טאקענס פריי א טאג ביז דעם 23טן סעפטעמבער.
- **22 אויגוסט:** **Ideogram 2.0** – א נייער מאדעל מיט זייער גוט שרייבן פון טעקסט אין בילדער, און שטארק אין שאפן בילדער.
- **22 אויגוסט:** **Zed AI** – א סערוויס וואס ברענגט AI-קאודירן אין דעם Zed רעדאקטאר, צוזאמען מיט Anthropic, צו רעדאקטירן שנעל מיט Claude.
- **22 אויגוסט:** **Phi-3.5** פון Microsoft – Microsoft האט ארויסגעלאזט דריי מאדעלן: Phi-3.5-mini, Phi-3.5-MoE און Phi-3.5-vision. זיי זענען אפן-סאורס אונטער MIT ליצענץ אויף Hugging Face, און קומען כמעט אן צום שפיץ-ניוואו כאטש זיי זענען קליינער. עס איז דא אן Unsloth סקריפט פאר שנעלן פיינ-טיונינג.
- **23 אויגוסט:** **Jamba Large** – AI21 האט ארויסגעלאזט נייע ווערסיעס פון איר היבריד-ארכיטעקטור פון Transformer און state space מאדעל.
- **23 אויגוסט:** **Gemini** אין **Gmail** – א נייע "Polish" מעגלעכקייט פארבעסערט שרייבן אין Gmail.
- **27 אויגוסט:** **xGen-VideoSyn-1** – Salesforce האט פארגעשטעלט א טעקסט-צו-ווידעא מאדעל וואס נוצט VidVAE צו פארקלענערן די ווידעא-דאטא, און א Diffusion Transformer פאר מער קאנסיסטענץ.
- **27 אויגוסט:** **D-ID** ווידעא-איבערזעצונג – א געצייג וואס קלאנירט דעם רעדנער'ס קול און פאסט צו די ליפן אין 30 שפראכן. פלענער הייבן זיך אן פון 56 דאלאר א יאר.
- **27 אויגוסט:** **Claude**: **סיסטעם-באפעלן** און **Artifacts** – Anthropic האט פובליקירט די סיסטעם-באפעלן פאר אלע אירע מאדעלן. Artifacts איז יעצט צוגענגליך פאר אלעמען, אויך אויף מאביל.

System prompts

Claude: סיסטעם-באפעלן און Artifacts · 27 אויגוסט 2024

- **28 אויגוסט: Cerebras Inference** – ווייל דער טשיפ פון Cerebras האט א ריזיגן פאראייניגטן זכרון, קען ער גיבן מאדעלן מיט טויזנטער טאקענס א סעקונדע.
- **28 אויגוסט: OpenAI's "Strawberry"** – לויט דער ידיעה פלאנט OpenAI ארויסצולאזן אין הארבסט א פראדוקט מיט שטארקע לאגיק, וואס זאל קענען ליזן אומבאקאנטע מאטעמאטיק-פראבלעמען און אויסארבעטן מארקעט-סטראטעגיעס.
- **28 אויגוסט: Higgs-Llama-3-70B-v2** – Boson AI's מאדעל איז שטארק אין Arena-Hard און AlpacaEval 2.0. אין פארגלייך מיט Claude 3.5 Sonnet פארמינערט עס ווי אפט מ'דארף איבערשאפן אן ענטפער מיט 21.6%, און פארגרעסערט ווי פיל באניצער קומען צוריק אויפן ערשטן טאג מיט 5.3%. אויפן ניוואו איז עס גלייך מיט Gemini 1.5 Pro. Boson האט אויך לאנסירט RP-Bench-Auto, אן אויטאמאטישער בענטשמארק פאר ראל-שפילן.
- **28 אויגוסט: Qwen 2 Audio און Qwen – VL** – Qwen האט ארויסגעלאזט מאדעלן וואס פארשטייען מוזיק, קולות און געזאנג, און באזונדער בילד-שפראך מאדעלן געבויט אויף Qwen 2.
- **29 אויגוסט: Cartesia** – די פירמע האט אנגעזאגט פארבעסערונגען אין אירע מאדעלן און סיסטעמען פאר AI אויפן מכשיר, און ארויסגעלאזט אן אפענעם היברידי state space מאדעל.
- **30 אויגוסט: Magic: 100 מיליאן טאקענס** – Magic Dev האט אנגעזאגט 100 מיליאן טאקענס קאנטעקסט פאר לאגיק-און פראגראמיר-אגענטן, און א שותפות צו בויען צוויי נייע סופער-קאמפיוטערס אויף Google Cloud. דאס קומט נאך א געלט-רונדע פון 320 מיליאן דאלאר.

פארשונג און טעכניק

- **1 אויגוסט: Sparse Autoencoders אויף Gemma** – א שיינער אפענער אינטערפעיס צו ערפארשן די אייגנשאפטן אין Gemma 2B. נאכדעם האט DeepMind פובליקירט די פארשונג-ארבעט צו Gemma Scope, וועגן דעם וויאזוי מ'האט טריינירט און וואס מ'האט געפונען.
- **1 אויגוסט: פארשידענע ארבעטן** – א איבערבליק איבער "אפלערנען" אין גענעראטיווע AI; Elysium, וואס נאכפאלגט חפצים אין ווידעאס; א צוויי-שטאפלדיגער Transformer וואס שפילט פיאנא לויט געפילן;

- DynaVol-S, א 3D מאדעל פון ווידעאס אן אנווייזונגען; און א בלאג איבער דער מאטעמאטיק פון flow matching ביי דיפיוזשאן מאדעלן.
- **2 אויגוסט: ARCLe און MMTrail** – א סביבה פאר reinforcement learning אויף ARC, און א דאטאסעט מיט איבער 20 מיליאן טרעילער-קליפס מיט מולטי-מאדאלע באשרייבונגען.
 - **5 אויגוסט: מער רעכן-קראפט בשעת'ן ענטפערן** – פארשער האבן געפונען אז ווען מ'נעמט עטליכע מאל ענטפערס פון קליינע מאדעלן און מ'האט א געצייג וואס באשטעטיגט זיי, פארבעסערט זיך שטארק דער בענטשמארק, און עס קאסט נאך אלץ 3 מאל ווייניגער ווי א גרעסערער מאדעל. נאך א ארבעט ווייזט אז מער רעכנען ביים ענטפערן קען זיין בעסער ווי מער פאראמעטער. Microsoft ווייזט אז זוכן אין א בוים (Mutual Reasoning) פארבעסערט שטארק קליינע מאדעלן אין מאטעמאטיק.
 - **5 אויגוסט: ווידעא, הענט און דאטא** – TSPM ענטפערט פראגעס אויף קול און בילד אין ווידעא; XHand ציילט אויס הענט אין עכטע צייט; SynthVLM שאפט בילדער פון באשרייבונגען פאר טריינירן; און א נעץ וואס פארקלענערט זיך אליין, ווייל מ'האט אריינגעלייגט זיין גרייס אין דער loss פונקציע.
 - **6 אויגוסט: פון ווידעא ביז אטאם-קראפט** – נאכפאלגן פונקטן אין ווידעאס מיט שפראך-embeddings; פארבעסערונגען אין line search פאר SGD; TexGen פאר טעקסטורן אויף 3D; א BYU פראפעסאר'ס AI אלגאריטם וואס קען פארקירצן דאס דיזיינען און ליצענצירן פון אטאם-רעאקטארן מיט א יארצענדליג, אין איין ביישפיל פון זעקס חדשים אויף צוויי טעג; און באלאנסירן די רעכן-לאסט ביים טריינירן בילד-שפראך מאדעלן.
 - **7 אויגוסט: רעצענזיעס און 3D** – אויטאמאטישע איבערבליקן פון וויסנשאפטליכע ליטעראטור מיט LLMs; MeshAnything V2 פאר קינסטלערישע meshes; CLEFT, א ביליגערער contrastive learning פאר מעדיצינישע בילדער; און RecDiffusion פאר צוזאמנייען בילדער.
 - **8 אויגוסט: 3D, LLaVA און muP** – 3D חפצים אלס 64x64 UV בילדער פאר א DiT מאדעל; LLaVA-OneVision, וואס מאכט גרינג צו טראנספערירן צווישן ארבעטן דורך אויסוואל פון דאטא און ארכיטעקטור; HARMONIC פאר סינטעטישע טאבעל-דאטא מיט פריוואטקייט; און א באווייז אז ABC invariance האלט ביי muP.
 - **9 אויגוסט: פאטאלאגיע, רעדן, 1.5-Pints** – SAM2-PATH פאר דיגיטאלע פאטאלאגיע; Speech-MASSIVE, א מולטי-לינגוואלער דאטאסעט פון גערעדטע שפראך; און 1.5-Pints, וואס מיט 57 ביליאן אויסגעקליבענע טאקענס שלאגט OpenELM און Phi אויף MT-Bench.
 - **12 אויגוסט: Tree Attention און מער** – BA-LoRA פארמינערט פארפולונג פון טריינינג-דאטא; Tree Attention, א פארלענגערונג פון Ring Attention פאר זייער לאנגן קאנטעקסט איבער פילע GPUs; MM-Forecast, וואס נוצט בילדער צו פאראויסזאגן געשעענישן; און trans-tokenization פאר שפראכן מיט ווייניג דאטא.
 - **13 אויגוסט: סענסארן, PDE, ECG** – סטערעא event קאמערע צוזאמען מיט LiDAR; UGrid, א נייראל-לייזער פאר לינעארע PDE; און ECG-FM, א יסוד-מאדעל פאר הארץ-אנאליז.
 - **14 אויגוסט: SWE-bench Verified און Agent Q** – OpenAI האט ארויסגעגעבן א גרינגערע טייל-זאמלונג פון SWE-bench. answerai-colbert-small האט נאר 33 מיליאן פאראמעטער און ווייזט שטארקע רעזולטאטן. Agent Q מ'Multion האט זיך פארבעסערט פון 18% אויף 81% אין וועב-ארבעטן

- ווי באשטעלן עסן, מיט MCTS און DPO, צוזאמען מיט Stanford. דערצו "lazy visual grounding" און click attention פאר צעטיילן בילדער.
- **15 אויגוסט:** קאוזאליטעט און לאנגע טעקסטן – Causal Agent פאר פראבלעמען פון סיבה און תוצאה; פארריכטן די פארפולונג אין CLIP; AgentWrite, וואס צעטיילט ארבעט אז מ'זאל קענען שרייבן איבער 20,000 ווערטער; און OpenResearcher פאר פארשונג-פראגעס.
 - **16 אויגוסט:** APO, מערדזשן און RAG – APO באנוצט אן אפגעשוואכטן ביישפיל אלס אנקער פאר העדפות; MetaSeg פאר צעטיילן בילדער; BadMerging, די ערשטע backdoor אטאקע אויף צוזאמגיין מאדעלן, און אן איבערבליק איבער מערדזשן; Databricks איבער לאנגן קאנטעקסט ביי RAG, וואס פאלט אויף מאדנע אבער פאראויסזעכארע אופנים; און PeriodWave פאר וועלן-סיגנאלן.
 - **19 אויגוסט:** BAM, FuseChat און זיכערהייט – BAM פארוואנדלט געדיכטע מאדעלן אין MoE, מיט גרויס שפארונג; DifuzCam ערוועצט א לינדז מיט א מאסקע און דיפיוזשאן; FuseChat גיסט צוזאם 7B מאדעלן מיט הויכע MT-Bench פונקטן; BAPLe, א backdoor אין מעדיצינישע מאדעלן; SigmaRL פאר אויטא-פלאנירונג; א בענטשמארק פאר 3D רעקאנסטרוקציע; און multi-view inpainting.
 - **20 אויגוסט:** BLIP-3 און אנדערע – אן איבערבליק איבער 180 בענטשמארקס פאר MLLMs; SAM2-UNet; BLIP-3, פיל פשוט'ער מיט אן autoregressive loss און א שפיץ-ניוואו אין באשרייבן בילדער; EasyRec פאר רעקאמענדאציעס; און שנעלע meshes פון gaussian splatting אויף איין GPU.
 - **21 אויגוסט:** Self-Taught Evaluator און ADAS – Meta FAIR טריינירט LLM-ריכטער אן מענטשליכע באצייכענונגען, מיט בעסערע רעזולטאטן אויף RewardBench. א פארשונג-ארבעט ווייזט ווי שוואך היינטיגע אגענט-סיסטעמען זענען, און שלאגט פאר אויטאמאטיש צו דיזיינען זיי. ShortCircuit, באזירט אויף AlphaZero, דיזיינט שאלטונגען מיט 84.6% דערפאלג. דערצו MeshFormer3D, רענטגן-באריכטן, TraDiffusion מיט מויז-שטריכן, און optical flow אויף שוואכע מכשירים.
 - **22 אויגוסט:** Transfusion און קאוד – Meta's Transfusion טוט סיי next-token און סיי דיפיוזשאן אויף טעקסט און בילדער צוזאמען, אויפן ניוואו פון Dalle 2 און Llama 2. Cohere באשטעטיגט אז טריינירן אויף קאוד העלפט לאגיק בכלל. דערצו optical flow ביז 70 מאל שנעלער, אפדעקן פאלשע בילדער, און PA-LLaVA פאר פאטאלאגיע.
 - **23 אויגוסט:** DeepMind זיכערהייט און Aurora – DeepMind's AGI זיכערהייט-גרופע האט באשריבן איר ארבעט: אויפזיכט, פראנטיר-זיכערהייט און מעכאנישע אינטערפערעטאציע. Microsoft האט עפנטליך ארויסגעגעבן Aurora, אן אטמאספערע-מאדעל. דערצו biorecap פאר bioRxiv, NuSegDG, Pano2Room, און אן איבערבליק איבער ראבאטיק.
 - **26 אויגוסט:** Inheritune און שאך – Inheritune בויט א 1.5B מאדעל פון א גרעסערן, אויף איין A6000 אין ווייניגער ווי א האלבן טאג, מיט 0.1% פון די דאטא. בעיס-מאדעלן שפילן שאך ארום 1800 elo, טשאט-מאדעלן פאלן אראפ, און מ'ווייזט ווי צו פארריכטן דאס. דערצו SAD, TSDE מיט spiking נעצן, GSAM, און GaussianOcc.
 - **27 אויגוסט:** Neural Assets – באהאנדלען 3D חפצים אין 2D סצענעס; T3M, אנימאציעס לויט טעקסט; FlexEdit; BiRefNet, א שפיץ-ניוואו אין אוועקנעמען הינטערגרונט; און Google's RB Modulation צו צופאסן דיפיוזשאן מאדעלן.

- **28 אויגוסט:** 2 ביטס פאר פאראמעטער – Fire-Flyer, DeepSeek, צוזאמען-דיזיין פון הארדוועיר און סאפטוועיר; Physics of Language Models 3.3 געפינט ארום 2 ביטס וויסן פאר יעדן פאראמעטער; Zyphra איבער טריינירן היברידע מאדעלן; Online Subspace Descent פאר שפארן זכרון; Llava BitNet, דער ערשטער טערנארער VLM; און Sketch2Scene פאר שפילבארע 3D סצענעס פון א סקיצע.
- **29 אויגוסט:** Generative Verifiers און דיסטילאציע – DeepMind נוצט ja/ logits אלס באלוינונג, און מיט CoT און ensembling איז דאס 16% בעסער. Aya's Cohere האט פארבעסערט די win rate דורך "מולטי-לינגוואלער ארביטראזש". מ'קען דיסטילירן Transformers אין לינעארע RNNs, און "Mamba in the Llama" ווייזט נייע וועגן צו פארשנעלערן היברידן. דערצו TAG, OctFusion (3D) אין 2.5 סעקונדעס אויף 4090), און דעבלורנינג מיט event קאמערע.
- **30 אויגוסט:** סקעדזשולנינג – ווען מ'לערנט צו ראגירן די לענג פון ענטפערס און מ'גייט די קורצע ערשט, ווערט דער דורכפלוס 6.5 מאל העכער. דערצו MTMamba++, RAW-Adapter, און SAM 2 פאר 3D מעדיצינישע בילדער.

געצייג פאר פראגראמירער און אפן-סאורס

- **1 אויגוסט:** Add-SD און Simply Draw – לייגן חפצים אריין אין בילדער מיט פשוט'ע באפעלן, און אן אפליקאציע וואס לערנט צייכענען מיט AI.
- **2 אויגוסט:** Mishax, Torch Chat, Engine Core – DeepMind'ס ביבליאטעק פאר אינטערפרעטאציע; PyTorch'ס ביבליאטעק צו שמועסן לאקאל מיט Llama 3.1; און א מוסטער פאר דינאמישע סיסטעם-באפעלן מיט געצייג.
- **5 אויגוסט:** GitHub Models – GitHub האט לאנסירט א קאנקורענט צו Hugging Face, מיט גרינגן צוטריט צו Llama 3.1 און GPT-4. Character AI האט עפנטליך ארויסגעגעבן Prompt Poet. Minigrid. לויפט יעצט אין JAX.
- **7 אויגוסט:** CogVideo און MiniCPM-V 2.6 – אן אפן-סאורס ווידעא-מאדעל נאענט צו די פארמאכטע מאדעלן, א בילד-שפראך מאדעל כמעט אין עכטע צייט, און X' Labs'ס ערשטע Canny controlnet פאר Flux-Dev.
- **9 אויגוסט:** FlexAttention – א נייער PyTorch API וואס קאמפילירט יעדע סארט attention צו Triton. דערצו Lighthouse פאר ווידעא-מאמענטן.
- **12 אויגוסט:** PufferLib און SQLite-Vec – א שנעלער פאראייניגטער ראם פאר RL, און א וועקטאר-דאטאבאזע אויף SQLite.
- **13 אויגוסט:** OCR מיט LLM – פארבעסערן OCR מיט LLMs, לאקאל אדער דורך API, און ProxyCLIP.
- **14 אויגוסט:** Triton קערנלס און UniBench – AI Toolkit, Flash Linear Attention, און UniBench וואס פאראייניגט איבער 50 בענטשמארקס.
- **15 אויגוסט:** קלאסטערן און Transformer Explainer – Together AI'ס וועגווייזער פאר גרויסע GPU קלאסטערן, א שיינע ווייזונג ווי א Transformer ארבעט, און Tavus פאר ווידעא-שמועסן מיט דיגיטאלע צווילינגען.

- **20 אויגוסט:** Storm און llamafile - Llama-3.1 Storm, די ערשטע צוגעפאסטע מאדעלן וואס פארבעסערן באמת די בעיס; llamafile מיט Whisper; און Hamming פאר טעסטן קול-אגענטן.
- **21 אויגוסט:** LossRider און SkyScript - אן אנימאציע פון loss קורוון, און א דאטאסעט פאר דראמע-סקריפטן.
- **22 אויגוסט:** TorchPruning - א ביבליאטעק מיט שפיץ-מעטאדן צו שניידן מאדעלן.
- **26 אויגוסט:** Liger און pgvectorscale - LinkedIn'ס Liger Kernel גיט 20% מער דורכפלוס און כמעט 60% ווייניגער זכרון פאר Llama, מיט 3 שורות קאוד. pgvectorscale איז ארום 28 מאל שנעלער. דערצו Meta איבער Llama 3.1, און א נייער SigLIP פון Google.
- **27 אויגוסט:** Infinite Dataset Hub - דאטא איבער יעדן נושא מיט phi-3-mini, ShellMate פאר'ן Mac טערמינאל, און Vyond פאר ווידעאס פאר פירמעס.
- **28 אויגוסט:** Llama Duo און LitServe - קליינע מאדעלן אלס רעזערוו ווען א פארמאכטער API פאלט, און א סערווירונג-מאשין אויף FastAPI.
- **29 אויגוסט:** Joy Caption און Hermes דאטאסעט - אן אפענער VLM פאר באשרייבן בילדער, Nous'ס דאטאסעט פאר פונקציע-רופן, און Wispr Flow, דיקטירן אויף 3 Mac מאל שנעלער ווי א קלאוויאטור.

פירמעס, געלט און ביזנעס

- **1 אויגוסט:** Samsung Galaxy AI - TM Roh האט אין פאריז פארגעשטעלט נייע AI מעגלעכקייטן, אריינגערעכנט א Bixby מיט Samsung'ס LLMs. ער האט דערמאנט אז מען באטראכט אבאנעמענטן.
- **1 אויגוסט:** Nvidia הומאנאידן - NIM, OSMO, און דאטא-זאמלונג דורך טעלע-אפעראציע צו פארשנעלערן ראבאטיק.
- **2 אויגוסט:** Cerebras IPO - Cerebras האט אריינגעגעבן א פארשלאג פאר אן IPO צו דער SEC.
- **5 אויגוסט:** Shazeer צוריק צו Google - Noam Shazeer, איינער פון די מחברים פון דער Transformer ארבעט, גייט צוריק צו Google, נאכדעם וואס Character.AI האט ליצענצירט איר טעכנאלאגיע פאר צוויי מאל דעם לעצטן ווערט.
- **5 אויגוסט:** Microsoft: OpenAI א קאנקורענט - אין איר יערליכן באריכט רופט Microsoft OpenAI א קאנקורענט, כאטש זי איז איר גרעסטער אינוועסטאר.
- **6 אויגוסט:** גרינדער פארלאזן OpenAI - Greg Brockman נעמט א לאנגע וואקאציע, און John Schulman איז אוועק צו Anthropic. לויט Spyglass זענען נאר 2 פון די 11 גרינדער נאך אקטיוו.
- **6 אויגוסט:** Groq 640 מיליאן - Groq, א סטארטאפ פאר שנעלע טשיפס, נעמט א Series D צו באפרידיגן די פאדערונג פאר inference.
- **6 אויגוסט:** Meta גראנטן, Tezi, DevDay - Meta טיילט 2 מיליאן דאלאר פאר פראיעקטן מיט Llama. Tezi האט אויפגענומען 9 מיליאן פאר Max, אן אויטאנאמישער AI רעקרוטירער. OpenAI'ס DevDay וועט זיין קלענערע זיצונגען און GPT-5.

- **6 אויגוסט: Apple Intelligence** – די מעגלעכקייטן קומען נאך'ן לאנסירן, אין iOS 18.1, נאר פאר iPhone 15 Pro. אויף Mac אין EU וועט עס זיין צוגענגליך אויב מ'שטעלט אמעריקאנער ענגליש. באצאלטע מעגלעכקייטן זענען אפשר ווייניגסטנס 3 יאר אוועק.
- **6 אויגוסט: 10 יאר TPU** – דאנק אירע אייגענע TPU טשיפס איז Google פון די איינציגסטע וואס זענען נישט אפהענגיק אין Nvidia.
- **7 אויגוסט: Meta און קולות פון אקטיארן** – Meta פירט פארהאנדלונגען צו נוצן באקאנטע קולות פאר איר אסיסטענט, פאר מיליאנען, פאר דער Connect קאנפערענץ אין סעפטעמבער.
- **7 אויגוסט: Blackwell פארשפעיטיגט** – לויט די ידיעות איז Nvidia'ס Blackwell פארשפעיטיגט ביז Q1 2025 צוליב דיזיין-פעלערס, וואס טרעפט Microsoft און Meta.
- **8 אויגוסט: Llama 4 און Llama באניץ** – Zuckerberg זאגט אז Llama 4 וועט דארפן 10 מאל מער רעכנ-קראפט ווי Llama 3, און ער ערווארט נישט באלד גרויסע איינקונפט. די באניצונג פון Llama האט זיך פארטאפלט פון מאי ביז יולי.
- **8 אויגוסט: AMD אלס AI פירמע** – כמעט האלב פון AMD'ס פארקויפן אין Q2 זענען דאטא צענטער פראדוקטן, דאנק AMD MI300. וועט ארויסלאזן א נייעם טשיפ יעדעס יאר. אירע Ryzen AI 300 טשיפס מיט NPU ברענגען AI צו קאמפיוטערס.
- **8 אויגוסט: Taco Bell, Opal, Microsoft Voice AI** – הונדערטער Taco Bell דרייוו-טרוס. OpenAI פירט א 60 מיליאן רונדע אין Microsoft Opal. האט אינוועסטירט 19 ביליאן אין AI אין איין קווארטאל, אן קלארע איינקונפט.
- **9 אויגוסט: XetHub און Humane** – Hugging Face קויפט XetHub, וואס האט געבויט אן אלטערנאטיוו צו LFS. ביי Humane'ס AI Pin זענען די צוריקגעגעבענע מער ווי די פארקויפטע. נאר ארום 7,000 זענען געבליבן ביי קונים, כאטש די פירמע האט אויפגענומען איבער 200 מיליאן.
- **12 אויגוסט: Zico Kolter ביי OpenAI** – דער CMU פראפעסאר, א מומחה אין AI זיכערהייט, איז אריין אין OpenAI'ס דירעקטאריאט און אין איר זיכערהייט-קאמיטעט.
- **12 אויגוסט: ראבאטאקסיס** – ביי Uber זענען די אויטאנאמישע פארן געוואקסן 6 מאל אין איין יאר. Waymo פארברייטערט זיך אין סאן פראנציסקא און לאס אנדזשעלעס, און האט פארגעשטעלט Generation 6 מיט ווייניגער קאמערעס, אויף Zeekr קארס. Waymo One מאכט 50,000 פארן א וואך.
- **12 אויגוסט: Intel און OpenAI** – לויט די ידיעות האט Intel אין 2017-18 אפגעזאגט צו קויפן א טייל אין OpenAI.
- **13 אויגוסט: Klarna און Intel** – Klarna'ס טשעטבאט מיט OpenAI באהאנדלט פשוט'ע פראגעס אין 23 מארקעטן און איבער 35 שפראכן. Intel'ס Arc A760A איז א GPU פאר קארס.
- **14 אויגוסט: Alexa און Polymarket** – Alexa האט פארלוירן 10 ביליאן דאלאר אין 2022, און Amazon באזירט זיך יעצט אויף גענעראטיווע AI. Polymarket ארבעט צוזאמען מיט Perplexity.
- **15 אויגוסט: UMG, גאז** – Meta און Universal Music האבן באנייט זייער אפמאך, אויך פאר WhatsApp און קעגן אומבארעכטיגטן AI אינהאלט. גאז-ליניע פירמעס פירן שמועסן צו גיבן גאז דירעקט צו דאטא צענטערס.

- **16 אויגוסט: SoftBank, AI קאמפיוטערס** – Intel האט נישט געקענט מאכן טשיפס פאר SoftBank's Izanagi, דעריבער באטראכט SoftBank יעצט 14% TSMC. פון קאמפיוטערס אין Q2 זענען געווען AI-פעאג, 8.8 מיליאן שטיק. Apple האט באקומען א פאטענט אויף א קלוגן רינג.
- **21 אויגוסט: Nast éCond און Unitree – OpenAI** האט אפגעמאכט א ליצענץ מיט Nast éCond פאר ChatGPT און Unitree. S'SearchGPT. G1 ראבאט פאר 16,000 דאלאר איז גרייט פאר מאסן-פראדוקציע.
- **22 אויגוסט: Nvidia ACE און באניץ – ACE** וועט פירן כאראקטערן אין Mecha Break אין 2025 מיט GPT-4o, כאטש די ערשטע וויזונגען זענען געווען שוואך. a16z האט ארויסגעגעבן איר רשימה פון די 100 שפיץ-אפליקאציעס. לויט Microsoft נוצן 75% פון וויסן-ארבעטער גענעראטיווע AI.
- **23 אויגוסט: Cursor און Perplexity** – Cursor האט אויפגענומען 60 מיליאן אין א Series A. Perplexity, אפגעשאצט אויף 1 ביליאן, וועט אנהייבן מיט רעקלאמעס אין Q4.
- **28 אויגוסט: xMEMS** – א 1 מ"מ קילער-טשיפ פאר דינע מכשירים.
- **29 אויגוסט: OpenAI 100 Midjourney, OpenAI** – Midjourney גייט אריין אין הארדוועיר. OpenAI פירט שמועסן איבער א רונדע געפירט דורך Thrive Capital וואס זאל זי אפשאצן העכער 100 ביליאן, און Microsoft וועט זיך אויך באטייליגן. אין אייראפע זענען די גרעסטע אפמאכן געווען WAYVE (1 ביליאן), Mistral (ארום 1 ביליאן), Helsing (484 מיליאן) און Poolside (400 מיליאן).
- **30 אויגוסט: ChatGPT 200 מיליאן** – איבער 200 מיליאן וועכנטליכע באניצער, א פארטאפלונג אין א יאר, און 92% פון Fortune 500 נוצן OpenAI. ByteDance האט לאנסירט Douyin Search.

רעגירונגען, געזעצן און זיכערהייט

- **1 אויגוסט: ווייסע הויז: אפן-סאורס** – מ'דארף נישט רעגולירן אפן-סאורס AI אין דער היינטיגער פארעם.
- **2 אויגוסט: OpenAI – AI Safety Institute** וועט געבן דעם אמעריקאנער אינסטיטוט פריערדיגן צוטריט צו איר נעקסטן מאדעל. אין סוף חודש האבן OpenAI און Anthropic ביידע אונטערגעשריבן אן אפמאך דערויף.
- **2 אויגוסט: EU AI Act** – עס איז אריין אין קראפט דעם 1טן אויגוסט, מיט מדריגות לויט ריזיקע. סטאנדארטן פאר שטארקע מאדעלן קומען ביז אפריל 2025.
- **2 אויגוסט: אלימפיאדע, Reddit** – פאריז נוצט AI צו אנאליזירן קאמערעס אין סטאנציעס. Reddit's Steve Huffman פאדערט אז Microsoft זאל באצאלן פאר'ן נוצן אירע דאטא.
- **6 אויגוסט: Musk קלאגט OpenAI** – Elon Musk באשולדיגט אז ער איז אפגענארט געווארן אויף 44 מיליאן דאלאר, ווייל מ'האט צוגעזאגט אפן-סאורס און דעם עפנטליכן נוצן.
- **7 אויגוסט: טשיפס צו כינע** – א שווארצער מארק, אריינגערעכנט א 103 מיליאן אפמאך מיט Nvidia טשיפס, דורך שמוגלער און פראנט-פירמעס. כינע'ס טשיפ-אויפבוי וואקסט כאטש דעם CHIPS Act.
- **7 אויגוסט: אנטדעקן ChatGPT טעקסט** – OpenAI איז פארזיכטיג מיט אזא געצייג, צוליב ריזיקעס פאר מענטשן וואס רעדן נישט ענגליש.

- **8 אויגוסט:** Nvidia אונטער אויפזיכט – די EU, ענגלאנד, כינע און די אמעריקאנער Justice Department קוקן אויף Nvidia, וואס האט 90% פונעם מארקעט. Nvidia בויט אויס אירע לעגאלע מאנשאפטן.
- **15 אויגוסט:** MIT ריזיקע-רעפאזיטאריע – איבער 700 דאקומענטירטע AI ריזיקעס.
- **16 אויגוסט:** SB 1047 – קאליפארניע'ס געזעץ פאדערט זיכערהייט-פראטאקאלן און אינספעקציעס, און סיליקאן וואלי שטרייט זיך. נאך ענדערונגען שטיצט Anthropic עס פארזיכטיג, בשעת OpenAI און Meta זענען דערקעגן. די לעגיסלאטור האט עס דורכגעלאזט, און Newsom מוז באשליסן ביז דעם 30טן סעפטעמבער.
- **26 אויגוסט:** Slack AI שוואכקייט – מיט אן אומדירעקטער prompt injection קען מען ארויסציען דאטא פון פריוואטע קאנאלן, אפילו נאך דעם דערהיינטיגן דעם 14טן.

אנאליזן און מיינונגען

- **1 אויגוסט:** קינדער און קול-AI – א מחבר אויף LessWrong טראכט וויאזוי צו דערציען קינדער פאר א וועלט מיט AI. Ethan Mollick'ס בלאג פארגלייכט Siri'ס זיכערן צוגאנג מיט ChatGPT Voice'ס שטארקן צוגאנג.
- **2 אויגוסט:** FLOPs אלס שוועל – Llama 3.1 405B, טריינירט מיט 3.8×10^{25} FLOPs, ברענגט א דעבאטע צי FLOPs איז א גוטע גרענעץ פאר רעגולירונג.
- **5 אויגוסט:** ריזיקע-שאצונגען – AI Snake Oil טענה'ט אז די שאצונגען פון עקזיסטענציעלע ריזיקע זענען צו אומפארלעסלעך פאר פאליטיק.
- **6 אויגוסט:** ארבעטן מיט Claude – אן Anthropic פארשער מיט א צובראכענער האנט איז געבליבן פראדוקטיוו מיט Claude און קול. נאך א מחבר שרייבט ווי Claude העלפט אים ביים טעכנישן שרייבן.
- **6 אויגוסט:** שנאה און ספקות – א בלאג באשרייבט ווי פיל מענטשן האסן גענעראטיווע AI. Ed Zitron צווייפלט צי OpenAI קען אנגיין לאנגע צייט.
- **7 אויגוסט:** Irpan און GPT-5 – Alex Irpan איז אריבער צו DeepMind'ס זיכערהייט-גרופע. אן אנאליז שאצט אז GPT-5 וועט פארבעסערן אבער נישט זיין א פולשטענדיגער אגענט.
- **9 אויגוסט:** אינדיע, GDP, Turing – AI סטראשעט די ארבעט אין אינדיע'ס אוטסאורסינג אינדוסטריע. א שרייבער טענה'ט אז LLMs דארפן נאך צופיל פראמפט-ארבעט, דעריבער זעט מען זיי נאך נישט אין GDP. Turing זאמלט מענטשליכע דאטא פון מומחים פאר אלע פירמעס.
- **12 אויגוסט:** נישט וועקסלען צו AI – א שרייבער שאצט אז די פרעסטיזש און געהאלטן פון ML ארבעט וועלן פאלן, ווייל ווייניג פירמעס האבן די קאפיטאל וואס סקעיל פאדערט.
- **13 אויגוסט:** DSPy, דירעקטאריאט – פארוואס מ'זאל בויען אויף DSPy, און די פראגע צי AI באדארף זיצן ביים טיש פון א פירמע'ס דירעקטאריאט.
- **16 אויגוסט:** נישט שרייבן "AI" – לויט א שטודיע פארמינערט א באצייכענונג "AI" דעם ווילן צו קויפן, צוליב מאנגל אין צוטרוי.

- **19 אויגוסט: Schmidt** און א "AI אימפעריע" – א שרייבער פאראויסזאגט AGI און ASI באלד. Eric Schmidt האט צוריקגענומען זיין קריטיק איבער ארבעטן פון שטוב, און ער פאראויסזאגט אז מ'וועט קענען שאפן אפליקאציעס ווי TikTok אין מינוטן.
- **20 אויגוסט: PDFs און לערן-וועג** – סארטירן אלע PDFs אויף דער אינטערנעט מיט embeddings, און א פלאן אין פיר שטאפלען פון מאטעמאטיק ביז LLMs.
- **21 אויגוסט: פון געטער צו פראדוקטן** – AI Snake Oil זאגט אז פירמעס שטייען פאר פינף פראבלעמען: קאסטן, פארלעסלעכקייט, פריוואטקייט, זיכערהייט און אינטערפעיס. א בלאג שלאגט פאר דיפיוזשאן כדי צו פארבעסערן לאגיק.
- **22 אויגוסט: הומאר און קאזאליטעט** – מארקאוויטש קייטן זענען לויט א שרייבער לוסטיגער ווי LLMs, ווייל LLMs זענען צו פאראויסזעבאר. דערצו זיבן כללים פאר קאזאליטעט אינפערענץ.
- **23 אויגוסט: פארמיידן AI איבערנעמונג** – פעלדער ווי קעגנשטעלן פנים-סכנות דינען אלס מוסטער פאר AI קאנטראל.
- **26 אויגוסט: Stephen Wolfram** – ער האט ערפארשט די דינאמיק פון נישט-לינעארקייט אין פשוט'ע נעצן, און ער פאדערט פילאסאפן פאר די גרויסע AI פראגעס.
- **28 אויגוסט: inference פריי און שנעל** – לויט א שרייבער וועלן LLMs ווערן נוצבארער פאר איבערגע'חזר'טע ארבעט ווייל זיי ווערן ביליגער און שנעלער, אפילו אויב זייער לאגיק וועט נישט שטארק וואקסן.

ווייטערע ידיעות

- **1 אויגוסט: Friend** און ניר-טראנספלאנטאציע – Avi Schiffmann'ס "Friend" איז א מעדאליאן פאר פריינטשאפט און שטיצע, 30,000 שטיק צו 99 דאלאר, אן אבאנעמענט. אין Cleveland Clinic האט מען דורכגעפירט די ערשטע טאפלטע ניר-טראנספלאנטאציע אין אמעריקע מיט הילף פון א ראבאט.
- **2 אויגוסט: Synchron און Heeyo** – Synchron'ס מוח-אימפלאנט נוצט ChatGPT צו פארשלאגן ענטפערס פאר געלעמטע פאציענטן. Heeyo האט לאנסירט אן אפליקאציע פאר קינדער פון 3 ביז 11 מיט 3.5 מיליאן אינוועסטירונג.
- **5 אויגוסט: גימנאסטיק און Genies** – FIG און Fujitsu נוצן AI צו העלפן די ריכטער, נישט צו ערזעצן זיי. Genies נוצט LLMs פאר כאראקטערן אין שפילן.
- **8 אויגוסט: GitHub פידבעק** – GitHub פאסט צוזאם די באמערקונגען פון קונים מיט AI.
- **9 אויגוסט: ראבאט-דענטיסט** – Perceptive'ס אויטאנאמישער ראבאט האט דורכגעפירט א גאנצע באהאנדלונג אויף א מענטש, שנעלער ווי געווענליך.
- **12 אויגוסט: Sonova הער-מכשיר** – o SphereéPhonak Aud זאגט צו 53 מאל בעסער פארשטיין רעדן אין א טומלדיגן ארט.
- **14 אויגוסט: ראבאט-קארס און טראפיק** – סימולאציעס ווייזן אז 5% ראבאט-קארס מאכן אויפהערן טראפיק-פארשטאפונגען.

- **19 אויגוסט:** Osmo און ריחות - אן AI סטארטאפ ארבעט צו לאזן קאמפיוטערס פארשטיין און איבערשאפן ריחות, אויך פאר מעדיצין.
- **20 אויגוסט:** Blindsight און Parkinson - פארוואס מוח-אימפלאנטן וועלן נישט צוריקגעבן זעאונג ווי א נאטירליך אויג, און א צוגעפאסטער "פעיסמעיקער" אין מוח פאר Parkinson.
- **22 אויגוסט:** קאמפיוטערס פון נעורנס - Neuroplatform ס'FinalSpark לאזט דינגען ביא-קאמפיוטערס פאר 500 דאלאר א חודש.
- **23 אויגוסט:** מאסקיטאס - VectorCam, געשטיצט פון דער Gates Foundation, דערקענט מאלאריע-מאסקיטאס.

דער גרעסטער נייעס פון סעפטעמבער איז געווען OpenAI's נייע o1 מאדעלן, וועלכע זענען טריינירט צו "טראכטן" איידער זיי ענטפערן. זיי האבן ארויסגערופן א סך אנאליזן איבער זייער קוש, איבער פראגראמירן, און איבער נייע ריזיקעס. OpenAI האט אויך געהאט א שטורמישן חודש: שמועסן איבער א געלט-רונדע פון 6.5 ביליאן דאלאר ביי א ווערט פון 150 ביליאן, פלענער צו טוישן די נאן-פראפיט סטרוקטור, און דער אוועקגיין פון דער CTO און צוויי שפיץ-פארשער. אויף דער אפן-סאורס פראנט האבן זיך באוויזן Llama 3.2 מיט וויזיע, Qwen 2.5 פון Alibaba מיט איבער 100 מאדעלן, Pixtral 12B פון Mistral און Molmo פון Allen AI. דערצו איז געקומען א כוואליע פון AI אגענטן ביי Salesforce, Microsoft און T-Mobile, און ווייטערע שריט מיט Apple Intelligence.

מאדעלן און נייע לאנסירונגען

- **2 סעפטעמבער: Cohere פארבעסערט Command R – Command R+ און Command R+ זענען** פארבעסערט געווארן כמעט אין אלע אויפגאבן: זכרון, גיכקייט, מאטעמאטיק און לאגישע טראכטן.
- **3 סעפטעמבער: CogVideo 5B קומט** – דער CogVideo 5B מאדעל ווערט ארויסגעלאזט די קומענדיגע וואך אין Hugging Face Diffusers.
- **4 סעפטעמבער: GPT Next פלענער** – דער CEO פון OpenAI יאפאן האט געזאגט אז GPT Next וועט טריינירט ווערן מיט בערך 100 מאל מער רעכנ-קראפט ווי GPT-4, און וועט ארויסקומען שפעטער דאס יאר. לויט אפשאצונגען קומט א טייל פון דעם שפרונג פון אלגאריטמישע פארבעסערונגען.
- **6 סעפטעמבער: Ask Photos אין Google** – Google Photos הייבט אן צעטיילן "Ask Photos" פאר אויסגעקליבענע באניצער אין אמעריקע. די פונקציע לאזט זוכן בילדער מיט קאמפליצירטע פראגעס אין נאטירלעכער שפראך.
- **6 סעפטעמבער: Qwen2-VL און OLMoE** – Alibaba Cloud האט ארויסגעלאזט Qwen2-VL, א וויזיע-שפראך מאדעל וואס קען אנאליזירן ווידעאס פון איבער 20 מינוט. ער שלאגט Llama 3.1 און GPT-4o, און קומט אין גרייסן 2B, 7B און באלד 72B אלס אפן-סאורס. דערצו איז ארויס OLMoE, אן אפענער MoE מאדעל מיט די בעסטע רעזולטאטן אין זיין קלאס פון 1B אקטיווירטע פאראמעטערס.
- **10 סעפטעמבער: Reflection 70B אונטער ספק** – HyperWrite's Reflection 70B, א ווערסיע פון Llama 3.1, האט אנגעזאגט שטארקע רעזולטאטן, אבער דריטע פארטייען האבן זיי נישט געקענט איבעררז'ן. מען האט עס צוגעשריבן צו קאליע-געווארענע וואגן ביים ארויפלאדן אויף Hugging Face. דער CEO Matt Shumer האט זיך אנטשולדיגט אז ער איז "פאראויסגעלאפן", אבער די קהילה בלייבט סקעפטיש צווישן באשולדיגונגען פון שווינדל.
- **10 סעפטעמבער: Apple Intelligence** – Apple Intelligence ברענגט AI אין Mail, Messages און Siri, מיט פונקציעס ווי צוזאמפאסונגען און בילדער-שאפן. פובליק בעטאס פון iOS 18.1 און macOS 15.1 זענען שוין ארויס, מיט טעקסט-איבערשרייבן און פאטא-רייניגן. זיי ארבעטן נאר אויף iPhone 15 Pro, iPhone 16 און M1 מאשינען, ערשט אויף אמעריקאנער ענגליש. iOS 18.2 וועט צולייגן Image

Playground, Genmoji, און ענגלישע לענדער. אין 2025 קומען דייטש, איטאליעניש, קארעאניש, פארטוגעזיש און וועטנאמעזיש, כאטש Apple שטייט פאר רעגולאציע-שוועריגקייטן אין EU און כינע.

• **11 סעפטעמבער: Apple's A18 טשיפס** – Apple האט פארגעשטעלט די A18 און A18 Pro טשיפס מיט שטארקע CPU און GPU פארבעסערונגען. דער Pro פארבעסערט די לייסטונג פון מאדעלן אויפן מכשיר און האט בעסערע קילונג.

• **12 סעפטעמבער: Adobe Firefly ווידעא** – Adobe האט געוווּן זיין Firefly Video Model פאר Premiere Pro. ער קען אויספילן לעכער אין די צייט-ליניע און צולייגן עלעמענטן; א בעטא ווערסיע קומט שפעטער דאס יאר.

• **12 סעפטעמבער: Mistral's Pixtral 12B** – Mistral האט לאנסירט Pixtral 12B, איר ערשטער מולטימאדאלער מאדעל, וואס נעמט אריין בילדער און טעקסט אונטער אן Apache 2.0 לייסענס. ער איז געבויט אויף Nemo 12B מיט א נייעם 400M-פאראמעטער וויזיע ענקאדער. ער שטיצט פארשידענע בילד-גרייסן אין א 128K טאקען קאנטעקסט, און שלאגט א סך אפענע און פארמאכטע מאדעלן אן צו שוואכן אין טעקסט. דאס קומט נאך Mistral's געלט-רונדע פון 645 מיליאן דאלאר.

• **13 סעפטעמבער: OpenAI's o1 מאדעלן** – OpenAI האט ארויסגעלאזט o1-preview און o1-mini, טריינירט מיט "טראכטן-שפורן" צו באטראכטן איידער ענטפערן. אין געוויסע געביטן ווי מאטעמאטיק, פיזיק און וויסנשאפט האבן זיי דערגרייכט איבער-מענטשלעכע רעזולטאטן. די באניצונג איז באגרעניצט צו בערך 30 פראגעס א וואך. Cognition האט באריכטעט אז זייער AI אגענט Devin ארבעט בעסער מיט o1 ווי מיט GPT-4o ביי געפינען טעותים. OpenAI's אנווייזונג זאגט אז מען דארף צו o1 פשוטערע פראמפטס מיט א מער סטרוקטורירטן קאנטעקסט. אנאליטיקער באמערקן אבער אז די מאדעלן האבן נאך שוועריגקייטן מיט ספעס-טראכטן און לאנגן קאנטעקסט, און ווייזן העכערע ריזיקעס ווי "reward hacking" און מעגלעכע שאדן מיט ביאלאגישע סכנות.

• **13 סעפטעמבער: Gemini Live פריי** – Google צעטיילט Gemini Live פאר פרייע Android באניצער. מען קען איבעררייסן דעם ענטפער, באקומען א טעקסט-אפשריפט, און אויסקלייבן צווישן צען נייע שטימען.

• **16 סעפטעמבער: Salesforce Agentforce** – Salesforce האט לאנסירט Agentforce, אירע ערשטע AI אגענטן וואס טוען אקציעס אליין אין באשטימטע גרעניצן. די פירמע טוישט איר פרייז-מאדעל צו 2 דאלאר פער שמועס, און דערמיט ערקענט זי אז AI וועט נעמען ארבעט-פלעצער.

• **17 סעפטעמבער: Runway API** – Runway האט אנגעזאגט אן API פאר זיין Gen-3 Alpha Turbo ווידעא מאדעל, ביים פרייז פון איין סענט פער קרעדיט. זי וויל אבער נישט אנזאגן פון וואו איר טריינירונגס-מאטעריאל קומט.

• **18 סעפטעמבער: Mistral פרייע API** – Mistral האט געעפנט א פרייע API שטאפל פאר פראגראמירער, אראפגעשניטן פרייזן מיט איבער 50%, פארבעסערט איר קליינעם מאדעל, און צוגעלייגט Pixtral אין Le Chat.

• **18 סעפטעמבער: Jina Embeddings v3** – Jina האט ארויסגעלאזט די דריטע ווערסיע פון אירע embedding מאדעלן פאר זוכן, מיט פארבעסערטע לייסטונג און טריינירונג.



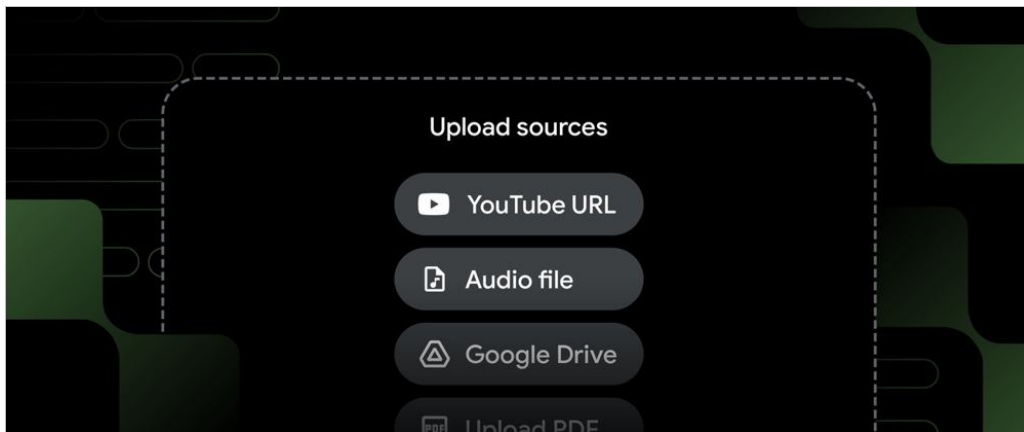
2024 סעפטעמבער 18 · Jina Embeddings v3

- **19 סעפטעמבער:** Qwen 2.5 פון Alibaba – Alibaba האט ארויסגעלאזט איבער 100 אפן-סאורס מאדעלן פון דער Qwen 2.5 משפחה, וואס קומען נאענט צו די שפיץ-מאדעלן אין קאוד, מאטעמאטיק און סטרוקטורירטע אויסגאבע. דערצו איז געקומען דער פארמאכטער Qwen-Max 2.5 און א טעקסט-צו-ווידעא געצייג.
- **20 סעפטעמבער:** Snap ווידעא געצייג – Snapchat שטעלט פאר אן AI ווידעא געצייג פאר אויסגעקליבענע שאפער, וואס מאכט ווידעאס פון טעקסט און באלד פון בילדער, אין בעטא אויפן וועב.
- **23 סעפטעמבער:** Solar Pro – Upstage'ס Solar Pro איז א 22B מאדעל וואס פאסט אריין אין איין 80GB GPU. דער ציל איז צו האבן דעם שטארקסטן מאדעל וואס לויפט אויף איין מכשיר.
- **24 סעפטעמבער:** Microsoft אגענטן – Microsoft לייגט צו אויטאמאטישע אגענטן אין אירע Copilot געצייג, און א געצייג אז עטליכע ארבעטער זאלן קענען צוזאמען ארבעטן מיט AI.
- **25 סעפטעמבער:** Gemini ביליגער – Google האט שטארק אראפגעשניטן דעם פרייז פון Gemini 1.5 Pro, העכער געמאכט די באניצונגס-גרעניצן, און פארבעסערט די לייסטונג מיט 2 מיליאן טאקען קאנטעקסט. דער פרייז שפרינגט אבער נאך 128K טאקענס.



2024 סעפטעמבער 25 · Gemini ביליגער

- **26 סעפטעמבער:** Meta – Llama 3.2 האט ארויסגעלאזט Llama 3.2, מיט קליינע 1B און 3B מאדעלן און עטליכע נייע מולטימאדאלע מאדעלן מיט וויזיע.
- **26 סעפטעמבער:** Molmo פון Allen AI – Allen AI האט ארויסגעלאזט Molmo, א אינגאנצן אפענעם מולטימאדאלן מאדעל, צוזאמען מיט די דאטא און די טריינינגס-שיטה. ער שלאגט א סך אפענע און פארמאכטע מאדעלן.
- **26 סעפטעמבער:** Google און Amazon ווידעא – YouTube Shorts וועט אריינברענגען Google Veo's DeepMind פאר הינטערגרונטן און זעקס-סעקונדע קליפס. Amazon האט לאנסירט א ווידעא-שאפער פאר אנאנסן פון איין בילד, ווי אויך Project Amelia, אן AI געהילף פאר פרעמדע סעלערס וואס איז געבויט אויף Bedrock.
- **27 סעפטעמבער:** NotebookLM פאדקעסטן – Google's NotebookLM קען יעצט נעמען מאטעריאל פון YouTube און אודיא קוועלער. זיין "Audio Overview" פונקציע מאכט פאדקעסטן פון אריינגעלאדענעם אינהאלט, און ווערט געלויבט אלס איינע פון די מערסט אימפרעסיווע AI אנווענדונגען זינט ChatGPT.



NotebookLM פאדקעסטן · 27 סעפטעמבער 2024

- **27 סעפטעמבער:** Rabbit r1 אגענט – Rabbit האט אנגעזאגט אז דעם 1טן אקטאבער וועט r1 באקומען א "Large Action Model" וואס ארבעט אלס אן אלגעמיינער וועב אגענט.

פארשונג און טעכניק

- **2 סעפטעמבער:** MoE באלאנסירונג – א פארשונג-ארבעט ווייזט א שיטה צו באלאנסירן עקספערטן אין MoE מאדעלן אן צוגעגעבענע loss. דאס פארבעסערט שטארק די באלאנס און די סטאביליטעט פון טריינירן.
- **2 סעפטעמבער:** בילד און קלאנג פארשונג – נייע ארבעטן באהאנדלען ווידעא אין טונקלקייט מיט event קאמערעס, געפינען קלאנג-קוועלער מיט SSPL און SACL, און ווייזן "spiking" דיפוזיע מאדעלן וואס שפארן ענערגיע.
- **3 סעפטעמבער:** דיפוזיע אלס אויטא-רעגערעסיע – א בלאג ווייזט מיט Fourier טראנספארמען אז דיפוזיע מאדעלן ארבעטן ווי אן אויטא-רעגערעסיע אין דער פרעקווענץ-געגנט. דאס עפנט א וועג צו פאראייניגן די צוויי שיטות.

- **3 סעפטעמבער: ווייטערע פארשונג** – א שיטה צו געפינען סיבה-פאקטארן אין קאמפלעקסע סיסטעמען; א מולטימאדאלער מאדעל פאר פריערדיגע אלצהיימער דיאגנאז; X-Codec פאר אודיא; און VisionTS, וואס פאראויסזאגט צייט-סעריעס ווי א בילד-איבערבווי מיט אן MAE פון ImageNet.
- **3 סעפטעמבער: פאלשע זכרונות** – MIT Media Lab האט געפונען אז טשעטבאטס מיט LLMs האבן ביי סימולירטע עדות-אויספערעגן ארויסגערופן איבער דריי מאל מער פאלשע זכרונות ווי די קאנטראל-גרופע.
- **4 סעפטעמבער: AI זיכערהייט** – Sam פון Anthropic האט פארגעלייגט וואס פראגראמירער וועלן מוזן דערגרייכן כדי AI זיכערהייט זאל געלינגען ביי איבער-מענטשלעכע סיסטעמען.
- **4 סעפטעמבער: ווייזע פארשונג** – Stable Control Representations פאר אגענטן; א שיטה צו שאפלען לייערס אז Vision Mamba זאל וואקסן ביז 300M פאראמעטערס; מער סטאבילע visual odometry; און א שיטה צו אנטדעקן זעלטענע הארץ-אנאמאליעס אין ECG.
- **4 סעפטעמבער: ערדציטערניש פאראויסזאגן** – פארשער פון University of Texas האבן אנטוויקלט AI וואס האט פאראויסגעזאגט 70% פון ערדציטערנישן ביי א פראבע אין כינע.
- **5 סעפטעמבער: ארכיטעקטור און אנאליז** – Masked mixers טוישן self-attention מיט masked convolutions. דערצו א בעסערע אנאליז פון מיקראסקאפ-בילדער, און Latent Distillation פאר אביעקט-דערקענונג אויף קליינע מכשירים.
- **6 סעפטעמבער: סטיל און embeddings** – StyleTokenizer פארבינדט סטיל מיט טעקסט ביים שאפן בילדער. א שטודיע פארגלייכט פארשידענע אויפבויען פון embedding מאדעלן וואס זענען געבויט אויף LLMs.
- **6 סעפטעמבער: הערן קרענק** – Google ניצט מאדעלן טריינירט אויף 300 מיליאן אודיא ביישפילן כדי צו אנטדעקן פרי סימנים פון קרענק ווי טובערקולאז.
- **9 סעפטעמבער: טיפקייט, כעמיע און קאוד** – DepthCrafter מאכט פונקטליכע טיפקייט-ידיעות פון ווידעאס. מען קען אויך מאדיפצירן LLMs צו שאפן מאלעקולן פאר מעדיצין. און פלאנירן אין נאטירלעכער שפראך פאר'ן שרייבן מאכט די קאוד-לייזונגען מער פארשידענארטיג און בעסער.
- **9 סעפטעמבער: LLMs קענען זיך נישט ערקלערן** – א דעמא האט געוויזן אז LLMs דערקענען אפט ריכטיג נומער-סעריעס, אבער זייערע ערקלערונגען פארוואס זענען אפט פאלש.
- **10 סעפטעמבער: דיפויזע, 3D און inverse RL** – Proxy Token Diffusion Transformer שפארט חשבונות. UniDet3D דערקענט 3D אביעקטן אין צימערן. א שיטה ניצט inverse reinforcement learning צו נאכמאכן גאנצע סעקווענצן. א שטודיע פון זיבן ענקאדערס פאר 3D האט געפונען אז DINOv2 איז בכלל דער בעסטער.
- **11 סעפטעמבער: AI אידייען און פאראויסזאגן** – א שטודיע מיט 100 NLP פארשער האט געפונען אז LLMs שאפן מער נייע אידייען, אבער זיי זענען א ביסל ווייניגער אויספירבאר. דער Safe AI Institute האט געוויזן א סיסטעם וואס פאראויסזאגט בעסער ווי עקספערטן. Apple's AdEMAMix אפטימיזירט ניצט עלטערע גראדיענטן און דערמיט פארבעסערט ער שטארק די קאנווערגענץ.
- **11 סעפטעמבער: ווייטערע ארבעטן** – HCVFlow פאר optical flow מיט ווייניגער זכרון; 3D סעגמענטאציע אן מקור-דאטא; Concept Sliders וואס ארבעטן יעצט מיט Flux; און SGSeg פאר ברוסט X-ray בילדער.

- **12 סעפטעמבער: קאנטעקסט און 3D – HOMER** פארלענגערט דעם קאנטעקסט פון LLMs אן צוגעגעבענע טריינירונג. דערצו אומזיכערקייט-אפשאצונג פאר NeRF און Gaussian Splatting, DiverGen פאר גענערירטע טריינירונגס-דאטא, און א קליידער-דאטאסעט געשאפן מיט LLMs.
- **13 סעפטעמבער: קאוד און HTML – PF-PPO** פארבעסערט RLHF ביי קאוד. דערצו א נייע דאטא-שיטה פאר saliency, און צוויי מאדעלן פון Jina וואס איבערמאכן שמוציגן HTML אויף ריינעם Markdown.
- **16 סעפטעמבער: אודיא, בילד און ראבאטן – AuditoryBench** פרובירט וויפיל LLMs פארשטייען פון קלאנג, און PromptCIR פאררעכט צוזאמגעדריקטע בילדער. DeepMind's ALOHA Unleashed און DemoStart לערנען ראבאטן פיינע אויפגאבן דורך קוקן אויף מענטשן, ביז אז א ראבאט קען צובינדן שייך.
- **17 סעפטעמבער: בילד און מוזיק – InstantDrag** מאכט "drag" בילד-רעדאגירן פיל גיכער. אנדערע ארבעטן פארבעסערן צוזאמגעדריקטע מוזיק, זיכערן פנים-דערקענונג קעגן שווינדל (DiffFAS), און דערקענען האנטשריפט מיט ViT.
- **18 סעפטעמבער: RAG און קאמיקס –** א ראם צו אפשאצן די פארלעסליכקייט פון RAG אין זעקס געביטן; beFormer פאר רעקאמענדאציע-סיסטעמען; און א איבערבליק איבער פארשטיין קאמיקס.
- **19 סעפטעמבער: BitNet אן איבערטריינירן – Hugging Face** האט מיט "quantization warm up" איבערגעטריינירט Llama 3 8B מיט 1.58 ביטס פער פאראמעטער, און ער האלט די לייסטונג פון Llama 1.
- **19 סעפטעמבער: 1X וועלט-מאדעל –** די ראבאטיק פירמע 1X האט טריינירט א ווידעא מאדעל וואס סימולירט ראבאט-אקטיוויטעט פון ערשטע-פערזאן בליק, פאר שאפן דאטא און טריינירן. זי האט אויך אפן-סאורס א ווערסיע פון Genie.
- **19 סעפטעמבער: טראכטן אין גראף – Diagram of Thought** ארגאניזירט דאס טראכטן פון א מאדעל אין א DAG פון פארשלאגן, קריטיקן און באשטעטיגונגען. דערצו א איבערבליק פון GNN פאר אנאמאליעס, און אטאקעס אויף נאוויגאציע אגענטן מיט דורכזיכטיגע לאטעס.
- **20 סעפטעמבער: V-StaR און 3D – V-StaR** ניצט ריכטיגע און פאלשע לייזונגען צו טריינירן א באשטעטיגער, און דאס פארבעסערט קאוד און מאטעמאטיק. דערצו Vista3D, וואס מאכט 3D פון איין בילד אין 5 מינוט, און PhysMamba צו מעסטן הארץ-קלאפן פון פנים-ווידעאס.
- **23 סעפטעמבער: זעלבסט-פאררעכטן – DeepMind** ווייזט א reinforcement learning שיטה וואס העלפט מאדעלן זיך אליין פאררעכטן אין מאטעמאטיק און קאוד. א פארשונג געפינט אז א פאררעכטער פשוטער באג פארבעסערט שטארק דיפוזיע מאדעלן פאר טיפקייט. Google's Michelangelo אפשאצט מאדעלן מיט לאנגן קאנטעקסט.
- **24 סעפטעמבער: ציילן און muP –** א "ציילן-טאקען" לערנט דיפוזיע מאדעלן צו שאפן א באשטימטע צאל אביעקטן. דערצו א סטאנדארט-פראטאקאל פאר לאנגן קאנטעקסט, און Eleuther מיט Cerebras ערקלערן muP צו איבערפירן היפעראראמעטערס צו יעדע גרייס.
- **24 סעפטעמבער: טראכטן און מאטעריאלן – MAgICoRe** פארבעסערט דאס טראכטן לויט די שוועריגקייט פון דער פראבלעם. CritiPrefill מאכט לאנגן קאנטעקסט ביז 3 מאל גיכער. אן MIT מאדעל אנאליזירט X-ray קריסטאלאגראפיע פאר נייע מאטעריאלן פאר באטעריעס און מאגנעטן.

- **25 סעפטעמבער: מעדיצין און אגענטן** – MIT האט 60 מאל גיכער געמאכט אן AI אפשאצער פאר מעדיצין-פראדוקציע. א פארשונג געפינט אז די ארדענונג פון UI עלעמענטן באאיינפלוסט שטארק אגענטן: א צופעליגע ארדענונג שאדט אזויפיל ווי אוועקנעמען דעם גאנצן טעקסט. דערצו ControlEdit פאר רעדאגירן קליידער-בילדער.
- **26 סעפטעמבער: MaskBit** – MaskBit שאפט בילדער אן embeddings מיט ביט-טאקענס, און דערגרייכט FID 1.52 אויף ImageNet מיט א 305M מאדעל. דערצו א וועג צו פארשטיין קאמיקס מיט VLMs.
- **27 סעפטעמבער: קאוד בענטשמארק** – commit-0 מעסט צי מאדעלן קענען שרייבן גאנצע Python פעקלעך פון אנהייב. דערצו BitQ פאר קוואנטיזאציע און CaBRNet פאר זעלבסט-ערקלערנדע AI.
- **30 סעפטעמבער: AlphaChip** – Google DeepMind's AlphaChip שאפט מיט reinforcement learning איבער-מענטשלעכע טשיפ אויסשטעלונגען אין שעהען אנשטאט וואכן. מען ניצט עס אין Google's TPUs און ביי MediaTek.
- **30 סעפטעמבער: גיכערע מאדעלן** – PGN איז א גיכערער נאכפאלגער פון FlowTurbo. RNN. Flow מאכט flow מאדעלן גיכער, און DoSSR פארבעסערט super-resolution.

געצייג פאר פראגראמירער און אפן-סאורס

- **2 סעפטעמבער: TRL און Flux אויף CPU** – א באליבטע inference ביבליאטעק שטיצט יעצט Flux Schnell אויף CPU. TRL האט באקומען א גרויסע אפדעיט מיט Liger Kernels, onlineDPO, און WinRate Callbacks.
- **3 סעפטעמבער: ML dtypes און Timm** – עס איז דא שטיצע פאר bfloat16 און fp8 אין numpy, און א טאבעלע פון וויזיע מאדעלן פון Timm.
- **4 סעפטעמבער: Anthropic Quickstart** – Anthropic האט ארויסגעלאזט אנהייב-פראיעקטן. דער ערשטער איז א קונה-סערוויס אגענט מיט Claude, געשריבן מיט געוועזענע AI הויפטן פון Brex, Uber און Facebook. RWKV האט צוגעלייגט לאקאלע inference אויף Windows 11.
- **5 סעפטעמבער: עטליכע נייע געצייג** – EnsLoss – פאראייניגן loss פונקציעס; MAPF-GPT – וועגן-געפינען פאר אסאך אגענטן; MoMo – ווידעא interpolation; repo2vec – שמועסן מיט א קאוד-באזע; און א 3-שעה'דיגער קורס צו בויען LLMs פון אנהייב.
- **6 סעפטעמבער: SGLang 0.3** – SGLang 0.3 ברענגט 7 מאל גיכערן DeepSeek MLA און 1.5 מאל גיכערן torch.compile. דערצו Laminar פאר אפסערוואציע פון LLM אפליקאציעס, LongLLaVA פאר לאנגן קאנטעקסט, און Cornell's ML קורס.
- **9 סעפטעמבער: עטליכע נייע געצייג** – DocAI – ארויסנעמען ידיעות פון דאקומענטן; FluxMusic – טעקסט צו מוזיק; א קליינער LLM אויף ESP32; iText2KG – וויסן-גראפן; א רשימה "x to text"; און ColPali מיט Qwen2-VL פאר זוכן.
- **10 סעפטעמבער: GPU און 3D** – בילד-באארבעטונג אויף GPU, גיכער 3D איבערבוי מיט Mast3r, און One-DM, וואס מאכט האנטשריפט נאך איין ביישפיל.

- **11 סעפטעמבער: Optillm** – Optillm אפטימירט פראמפטס מיט אלגאריטמען ווי Monte Carlo Tree Search, Z3 און Self Consistency.
- **12 סעפטעמבער: Llama Omni** – Llama Omni איז א רייד-אריין-רייד-ארויס מאדעל אויף Llama 3.1 8B מיט זייער נידעריגע פארשפעטיגונג. דערצו אן AWS AI Stack, Internet of Agents, און א מאדעל פאר פיאנא מוזיק.
- **13 סעפטעמבער: Llama 405B** אין איין קליק – Axolotl מיט Lambda Labs ווייזן ווי צו פיין-טונען Llama 3.1 405B אויף 64 GPUs מיט ווייניג אויפשטעלן. דערצו Agent Workflow Memory, Hi3D, דער SUPER בענטשמארק, און GPT-4o פאר וועב scraping.
- **16 סעפטעמבער: עטליכע נייע געצייג** – Google ווייזט RAG מיט Gemma 2; P1EEG-16 פאר Raspberry Pi; FlashSplat; GPU פאזלס אינעם בראווער פון Answer AI; ODAQ אודיא דאטאסעט; א RAG פייפליין מיט Pinecone; און iSeg.
- **17 סעפטעמבער: VAE און Shazam** – א VAE וויזואליזאציע אינעם בראווער, און אן אפן-סאורס Shazam.
- **18 סעפטעמבער: Word Llama** – Word Llama פאר NLP אויף CPU, סילאבל סעגמענטאציע מיט HuBERT, און צוזאמפאסן פון סעקיריטי-ווידעאס מיט Florence 2.
- **19 סעפטעמבער: LlamaCoder** – Together AI מיט Meta האבן געבויט LlamaCoder, וואס מאכט גאנצע אפליקאציעס פון א פראמפט. דערצו ZML, אן inference מאשין אין Zig; AceParse; און SkinMamba פאר הויט-קענסער.
- **20 סעפטעמבער: GOT OCR** – GOT OCR איז א שטארקע פארבעסערונג אין לייענען טעקסט פון בילדער. דערצו Fish Speech פאר שטים-קלאנירן און Promptriever, דער ערשטער זוך-מאדעל וואס מען קען געבן פראמפטס.
- **23 סעפטעמבער: Contextual Retrieval** – Anthropic ווייזט ווי צו צעטיילן דאקומענטן סעמאנטיש פאר בעסערע רעזולטאטן, פאר בלויז 1 דאלאר פער מיליאן שטיקלעך צוליב caching. Kyutai האט ארויסגעלאזט Moshi'ס וואגן און קאוד. דערצו 3D Topia, AI פאר JQ, אן ערקלערונג פון SAE, און GKD אין TRL.
- **24 סעפטעמבער: Triton און ColPali** – א לעקציע איבער Triton פון ColPali, CUDA Mode, דאקומענטן, און א 3D ViT פאר סאטעליט בילדער.
- **25 סעפטעמבער: עטליכע נייע געצייג** – D-2DGS; FastGL פאר GNNs; AViMoS עפעס מיט 1,500 ווידעאס; DreamHoi; און Jane Street'ס וויזואליזאציע פון שטיקלעך-לינעארע נעצן.
- **26 סעפטעמבער: Jax און Time MoE** – פאראלעליזירן מיט Jax, דער HelloBench בענטשמארק פאר לאנגע טעקסטן, און Time MoE פאר צייט-סעריעס.
- **27 סעפטעמבער: Google'ס טשיפ מאדעלן** – Google האט ארויסגעלאזט מאדעלן, קאוד און סימולאטארן פאר טשיפ פלענער מיט RL. דערצו געא-לאקאציע פון גאס-בילדער, דאטא-פארמערונג, און Simon Willison'ס איבערבליק פון סטרימינג APIs.

• **30 סעפטעמבער: Aider מיט o1** – Aider האט שטארק פארבעסערט קאוד-רעדאגירן דורך ניצן o1 אלס ארכיטעקט און DeepSeek אלס שרייבער. דערצו LightAvatar, קלאנג-דערקענונג, VPTQ קוואנטיזאציע, און Prolog אין א Go באט.

פירמעס, געלט און ביזנעס

• **2 סעפטעמבער: OpenAI'ס געלט-רונדע** – OpenAI פירט שמועסן צו אויפנעמען 6.5 ביליאן דאלאר ביי א ווערט פון 150 ביליאן, און ארגאניזירט זיך איבער כדי צוצוציען אינוועסטארן. דאס קומט צווישן אינערליכע סכסוכים איבער די שליחות און זיכערהייט, וואס האבן געפירט צו א סך פארשער וואס זענען אוועק, טייל צו Anthropic. לויט א באריכט האט Apple שפעטער באשלאסן נישט צו אינוועסטירן, און Microsoft לייגט צו נאך 1 ביליאן. Apple וועט פונדעסטוועגן אריינברענגען ChatGPT אין iOS 18.

• **2 סעפטעמבער: ראבאטן** – Apple שאפט אפשר ראבאטישע מכשירים, למשל אן iPad אויף א ראבאט-ארעם, אונטער Kevin Lynch, מיט הילף פון Technion עקספערטן, אויף 2026–2027. איבער 25 כינעזישע פירמעס האבן געוויזן מענטשן-ראבאטן פאר פאבריקן; דער מארק ווערט אפגעשאצט אויף 38 ביליאן דאלאר ביז Boston Dynamics 2035. עלעקטרישער Atlas טוט push-ups, און Boardwalk Robotics האט פארגעשטעלט Alex, א ראבאט אן פיס.

• **2 סעפטעמבער: Vybe** – די AI קאמערע אפליקאציע Vybe האט אויפגענומען 4.75 מיליאן דאלאר.

• **3 סעפטעמבער: xAI'ס Colossus** – xAI האט אנגעהויבן Colossus, א קלאסטער פון 100,000 H100 – דער גרעסטער אין דער וועלט – מיט פלענער אים צו פארדאפלען. אין Memphis ווערט דער פראיעקט קריטיקירט צוליב זיין ענערגיע און וואסער באדערפעניש. Musk האט אפגעלייקנט אז Tesla וועט מוזן לייסענסירן xAI'ס מאדעלן.

• **3 סעפטעמבער: MLPerf רעזולטאטן** – Nvidia'ס Blackwell האט געוויזן די בעסטע לייסטונג פער GPU מיט 4-ביט פרעציזיע. AMD און Untether AI האבן געוויזן שטארקייט אין ענערגיע-עפעקטיווקייט.

• **4 סעפטעמבער: Reliant AI** – Reliant AI האט אויפגענומען 11.3 מיליאן דאלאר צו אויטאמאטיזירן ליטעראטור-איבערבליקן מיט איר פראדוקט Tabular.

• **4 סעפטעמבער: Meta אינצידענטן** – Meta'ס AI סיסטעם פאר געפינען שורש-סיבות האט דערגרייכט 42% פונקטליכקייט, מיט א פיין-געטונטן Llama 2.

• **5 סעפטעמבער: You.com און Sakana** – You.com האט אויפגענומען 50 מיליאן דאלאר אין Series B. Sakana האט אויפגענומען 100 מיליאן אין Series A כדי צו ווערן א פירנדיגע לאבאראטאריע אין יאפאן.

• **5 סעפטעמבער: AI Overviews** – Google'ס AI צוזאמפאסונגען שטופן אראפ די ארטיקלען פון נייעס-ארויסגעבער. נאך קריטיק איבער טעותים האט Google עס אויסגעברייטערט צו זעקס נאך לענדער.

• **6 סעפטעמבער: 2,000 דאלאר א חודש** – OpenAI באטראכט לויט באריכטן פרייזן ביז 2,000 דאלאר א חודש פאר מאדעלן ווי Strawberry און Orion.

• **9 סעפטעמבער: AI קאמפיוטערס** – Intel האט פארגעשטעלט Core Ultra 200V (Lunar Lake). Copilot+ פונקציעס קומען אין נאוועמבער אויף Intel און Apple AMD Ryzen AI 300. מארקעט דעם Mac אלס דעם "בעסטן AI קאמפיוטער".

- **9 סעפטעמבער: Canva פרייזן** – Canva העכערט די פרייזן פון Teams איבער 300%, פון 120 צו 500 דאלאר א יאר, צוליב AI פונקציעס.
- **10 סעפטעמבער: טשיפס פון TSMC – OpenAI** קען אפשר ניצן TSMC טשיפס צו שפארן קאסטן. ByteDance אנטוויקלט אייגענע טשיפס מיט TSMC צוליב אמעריקאנער עקספארט-באגרעניצונגען.
- **10 סעפטעמבער: קליינע פראדוקטן** – Shortwave 'ס געהילף טוט קאמפליצירטע אויפגאבן מיט אימעילס. Hoop איז א טאסק-אפליקאציע, און Icon מאכט 20 ווידעאס פון איינס פאר אנאנסן.
- **11 סעפטעמבער: טשעלענדזשער פאר Nvidia – AMD** פאראייניגט RDNA און CDNA אין UDNA כדי צו קאנקורירן מיט AMD, Intel, Cerebras. SambaNova שטרעבן צו שוואכן Nvidia'ס הערשאפט, און AMD פירט אריין MI300X ביי Vultr.
- **11 סעפטעמבער: ראבאטעקסיס** – Waymo גיט איבער 100,000 באצאלטע רייזעס א וואך, אבער פארדינט נאך נישט. Cruise קומט צוריק אין Bay Area און האט א שותפות מיט Uber פאר 2025.
- **12 סעפטעמבער: Apple ברילן** – Apple טראכט אפשר וועגן קלוגע ברילן אן AR אלס קאנקורענץ צו Meta'ס Ray-Ban פאר 299 דאלאר.
- **13 סעפטעמבער: Amazon און Brin** – Amazon פרובירט אנאנסן אין זיין טשעטבאט Rufus. Sergey Brin ארבעט ביי Google אויף AI כמעט יעדן טאג.
- **16 סעפטעמבער: OpenAI סטרוקטור** – OpenAI זאגט אז זי איז אויסגעוואקסן איר נאן-פראפיט סטרוקטור און וועט עס טוישן כדי צו זיין מער צוציענדיג פאר אינוועסטארן. אן נייע לאגא איז געפונען געווארן "אומהיימליך" ביי ארבעטער.
- **17 סעפטעמבער: Intel פאונדרי** – Intel'ס אקציע איז געשפרונגען נאך א פלאן צו מאכן די פאונדרי א באזונדערע סובסידיער.
- **19 סעפטעמבער: סטארטאפס** – Mercor האט אויפגענומען 30 מיליאן פון Benchmark פאר אן AI רעקרוטירער. Arcade AI האט אויפגענומען 17 מיליאן פאר "פראמפט צו פראדוקט".
- **23 סעפטעמבער: Jony Ive מיט OpenAI** – Jony Ive ארבעט מיט Sam Altman אויף א נייעם AI מכשיר, וואס קען אויפנעמען 1 ביליאן דאלאר ביז סוף יאר.
- **23 סעפטעמבער: Salesforce** – Salesforce ארבעט מיט Nvidia אויף AI אוואטארן. Salesforce Ventures האט צוגעלייגט 500 מיליאן, און זיין AI פאנד שטייט יעצט ביי 1 ביליאן.
- **25 סעפטעמבער: James Cameron** – James Cameron איז אריין אין Stability AI'ס באורד.
- **25 סעפטעמבער: T-Mobile** – T-Mobile מיט OpenAI בויען א קונה-סערוויס מיט AI אגענטן. Nvidia מיט T-Mobile ניצן AI פאר נעץ-טראפיק.
- **26 סעפטעמבער: אוועקגיין ביי OpenAI** – OpenAI'ס CTO גייט אוועק פון דער פירמע. שעהען שפעטער זענען אוועק דער הויפט-פארשער Bob McGrew און דער VP Barret Zoph.
- **30 סעפטעמבער: Anthropic 40 ביליאן** – Anthropic פירט שמועסן צו אויפנעמען געלט ביי א ווערט פון 40 ביליאן דאלאר.

רעגירונגען, געזעצן און זיכערהייט

- **2 סעפטעמבער: DeepMind ארבעטער קעגן מיליטער** – כמעט 200 ארבעטער פון DeepMind האבן געפאדערט אז Google זאל אויפהערן מיליטערישע קאנטראקטן, וואס ברעכן לויט זיי איר AI עטיק.
- **2 סעפטעמבער: Laion 5B – Laion 5B**, דער גרעסטער אפענער בילד דאטאסעט, ווערט צוריק ארויסגעלאזט נאך אויסרייניגן שעדליכע בילדער.
- **4 סעפטעמבער: קריג איבער קראולן** – לויט דער Data Provenance Initiative פארשפארן וועבזייטלעך קראולערס, וואס קען צווינגען AI צו ניצן ערגערע דאטא, סינטעטישע דאטא אדער לייסענסן.
- **5 סעפטעמבער: אפן-סאורס דעפיניציע** – OSI האט ארויסגעלאזט א פלאן פאר א דעפיניציע פון "אפן-סאורס AI", און א סטאבילע ווערסיע ווערט ערווארט אין אקטאבער.
- **6 סעפטעמבער: שטים-גניבה** – YouTube בויט געצייג צו אנטדעקן גע'AI'טע פנימער און שטימען, מיט פראבעס פראגראמען קומענדיג יאר. Jeff Geerling באשולדיגט Elecrow אז זי האט קלאנירט זיין שטים אן רשות.
- **11 סעפטעמבער: פאלשע פארשונגען** – GPT -פאבריצירטע פארשונג-ארבעטן ווערן אריינגעשמוגלט אין Google Scholar, ספעציעל איבער געזונט און סביבה.
- **17 סעפטעמבער: באמבע אנווייזונגען** – א העקער האט אפגענארט ChatGPT צו געבן אנווייזונגען פאר באמבעס דורך סאציאלע אינזשעניריע.
- **20 סעפטעמבער: ChatGPT שרייבט אליין** – OpenAI האט פאררעכט א טעות וואו ChatGPT האט אנגעהויבן שמועסן אליין.
- **23 סעפטעמבער: סעקיריטי ווידעאס** – MIT פארשער געפינען אז LLMs באשליסן נישט קאנסיסטענט צו רופן פאליציי ביי סעקיריטי ווידעאס.
- **24 סעפטעמבער: Altman פארלאזט קאמפיטע** – Sam Altman גייט ארויס פון OpenAI'ס זיכערהייט קאמפיטע, וואס וועט יעצט האבן נאר אומאפהענגיגע מיטגלידער. OpenAI Academy העלפט לענדער מיט נידעריגע הכנסה, און האט איבערגעזעצט MMLU אין 15 שפראכן.
- **25 סעפטעמבער: LinkedIn דאטא** – LinkedIn האט לויט באריכטן טריינירט מאדעלן אויף באניצער-דאטא איידער זי האט געטוישט אירע תנאים.
- **26 סעפטעמבער: גלאבאלע זיכערהייט** – וויסנשאפטלער פון כינע און מערב האבן זיך איינגעשטימט צו גרינדן אינסטיטוציעס פאר עמערדזשענסיס, אויפשטעלן א Safety Assurance Framework, און פינאנצירן אומאפהענגיגע פארשונג.

אנאליזן און מיינונגען

- **2 סעפטעמבער: פיקסאר פון AI** – a16z גלייבט אז גענעראטיווע AI ווי Sora און Dream Machine וועט צוזאמשמעלצן פילם און שפילן.
- **3 סעפטעמבער: Strawberry ספעלן** – מאדעלן פעלן ביי ציילן אותיות צוליב טאקעניזאציע.

- **3 סעפטעמבער: GPU און סקעילינג** – Trainy זאגט אז GPU-באניצונג פירט אין טעות: זיי האבן געזען 100% באניצונג נאר מיט ~20% MFU, און דורך מעסטן SM עפעקטיווקייט האבן זיי באקומען א 4x גיכקייט. Epoch AI זאגט אז טריינירן וואקסט 4x א יאר און קען ממשיך זיין ביז 2030, ביז 2e29 FLOP, באגרעניצט דורך שטראם און טשיפס.
- **4 סעפטעמבער: H100 מארק** – Lepton ערקלערט ווי צו קויפן H100. Forbes שרייבט איבער אינזשעניר-ארדענס, און א פראגראמירער ערקלערט פארוואס ער ניצט נישט LLMs.
- **5 סעפטעמבער: AI ראסע און ענערגיע** – Palladium ווארנט אז דער נאראטיוו פון א געאפאליטישע AI ראסע קעגן כינע ריזיקירט זיכערהייט. א בלאג זאגט אז AI איז א וואקסנדיגער אבער נאך קליינער טייל פון דאטא צענטער ענערגיע.
- **6 סעפטעמבער: אפטישע טשיפס** – Broadcom ארבעט צו לייגן אפטישע פארבינדונג גלייך אין GPUs, און BentoML'ס CEO רעדט איבער ענטערפרייז אינפראסטרוקטור.
- **9 סעפטעמבער: איבער-זיכערקייט** – Ted Underwood זאגט אז AI העכערט שאפונג און אמביציע, אבער ריזיקירט איבער-זיכערקייט.
- **10 סעפטעמבער: פארשידענע אינטעליגענצן** – Noema רופט צו ברייטערן די דעפיניציע פון אינטעליגענץ.
- **11 סעפטעמבער: HTML פאר GPT-4o** – אוועקנעמען HTML שפארט געלט מיט ווייניג פארלוסט; GPT-4o mini איז ביליג פאר אומסטרוקטורירטע פראגעס.
- **13 סעפטעמבער: AI וויסנשאפטלער** – קריטיקער זאגן אז דער "AI Scientist" קען פארענגערן פארשונג. Astral Codex Ten זאגט אז Sakana'ס ארבעט איז טריוויאל און טיילווייז אויסגעטראכט, און אז Strawberry האט געהעקט א שוואך-געבויטע סאנדבאקס.
- **17 סעפטעמבער: קנעפל-פראבלעם** – Every זאגט אז AI ווערט צעטיילט אין קליינע קנעפלעך און ברענגט נאר קליינע פראדוקטיוויטעט. זיי ציילן אויף אז AI איז בעסער אלס א צוגאב ווי אלס ערוואך פאר מענטשן.
- **18 סעפטעמבער: Copilot שוואכט** – א שרייבער זאגט אז Copilot מאכט פראגראמירער ערגער. Heavybit זאגט אז דאטא פייפליינס זענען יעצט דער סוד.
- **20 סעפטעמבער: סקעילינג** – Ethan Mollick באשרייבט Gen2 און Gen3 מאדעלן, און א נייעם סקעילינג געזעץ פאר "טראכטן" ביי inference.
- **24 סעפטעמבער: מער טראכטן ווי מענטשן** – David Gilmore שאצט אפ (10,000 טאקענס = א "טראכט-שעה") אז AI קען איבערטראכטן מענטשן אינערהאלב א יאר-צענדליק.
- **26 סעפטעמבער: אגענטן אין אפסערוואציע** – א בלאג פרעגט צי AI אגענטן וועלן טוישן SRE.
- **27 סעפטעמבער: The Intelligence Age** – Sam Altman שרייבט אז AI וועט פארשטארקן מענטשן מיט פערזענליכע געהילפן, לערער און דאקטוירים, אבער מען מוז עס מאכן צוגענגליך און אפגיין ריזיקעס.

ווייטערע ידיעות

- **19 סעפטעמבער: געדאנק-קאנטראל** – Synchron האט ערמעגליכט פאר אן ALS פאציענט צו קאנטראלירן Alexa דורך א מוח-אימפלאנט.

- **26 סעפטעמבער: Chartwatch** – St. Michael's שפיטאל'ס AI ווארענונג-סיסטעם האט אראפגעברענגט אומערווארטעטע טויטן מיט 26%.
- **27 סעפטעמבער: Spot עפנט טירן** – Boston Dynamics' Spot קען אליין עפענען Assa Abloy טירן דורך Bluetooth.
- **30 סעפטעמבער: מיליטערישער מאדעל** – א מאדעל טריינירט אויף ארמיי-האנטביכער ווייזט טיפע וויסן.

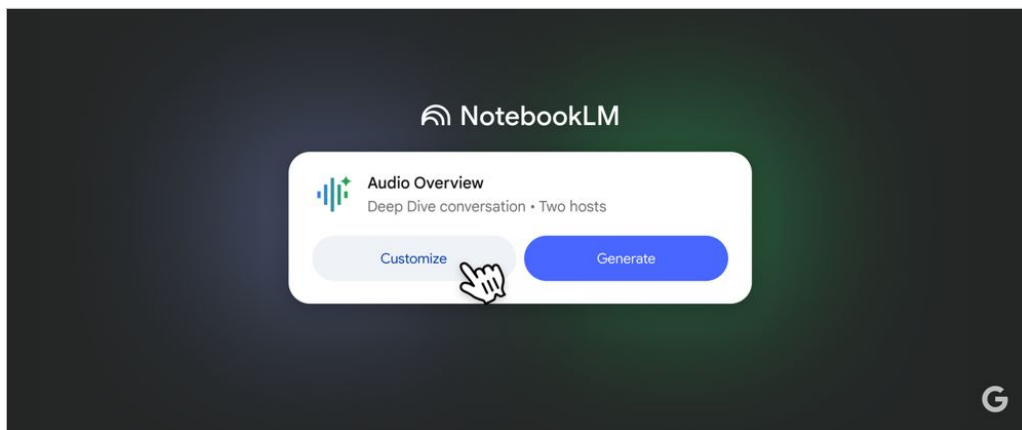
אקטאבער 2024 איז געווען א חודש פון ריזיגע געלט און גרויסע לאנסירונגען. OpenAI האט אויפגעהויבן 6.6 ביליאן דאלאר ביי א ווערט פון 157 ביליאן, באקומען א קרעדיט-ליניע פון 4 ביליאן, און זיך גענומען איבערבויען אויף א פירמע פאר פראפיט. ביי איר DevDay האט זי לאנסירט די Realtime API, און באלד דערנאך Canvas. Anthropic האט ארויסגעלאזט א פארבעסערטן Claude 3.5 Sonnet וואס קען אליין באדינען א קאמפיוטער, צוזאמען מיט Haiku 3.5, און א נייע זיכערהייט-פאליסי. Meta האט געוויזן Movie Gen און א שורה אפן-סאורס ארבעטן, און Mistral, Nvidia, IBM, Stability AI און Adobe האבן אויך ארויסגעלאזט נייע מאדעלן. גאווערנאר Newsom האט אפגעווארפן דעם קאליפארנישן AI-געזעץ SB 1047. די Nobel פרייזן אין פיזיק און כעמיע זענען געגאנגען צו פארשער וואס זענען פארבונדן מיט AI.

מאדעלן און נייע לאנסירונגען

- **1 אקטאבער: Emu 3 מולטימאדאל** – Emu 3 איז א 9B-פאראמעטער מאדעל וואס ארבעט מיט נעקסט-טאוקן פרעדיקציע. מען האט אים טריינירט אויף אלע אויפגאבעס צוזאמען אויפ'ן אופן ווי Gemini. ער שלאגט SDXL ביים מאכן בילדער, LLaVa-1.6 ביים פארשטיין בילדער, און OpenSora 2 ביים מאכן ווידעא.
- **1 אקטאבער: Samsung טאבלעטן מיט AI** – Samsung האט געוויזן די Galaxy Tab S10 סעריע, וואס האט פונקציעס מיט AI.
- **2 אקטאבער: Pika 1.5** – Pika האט ארויסגעלאזט ווערסיע 1.5. זי ברענגט מער רעאליסטישע באוועגונג, גרויסע "ביג סקריין" שאטס, און א נייע פונקציע Pikaffects.
- **2 אקטאבער: OpenAI DevDay** – ביי DevDay 2024 האט OpenAI אנאנסירט עטליכע נייעס פאר פראגראמירער: פיין-טיונינג פאר וויזשן מאדעלן, א Realtime API, פראמפט-קעשינג, און מאדעל-דיסטילאציע. די Realtime API (בעטא) ערמעגליכט שמועסן פון קול צו קול כמעט אן פארשפעטיגונג, און מען קען דערמיט גרינג אויפבויען קול-אסיסטענטן. מען האט אויך פארבעסערט די Chat Completions API.
- **2 אקטאבער: Apple MM1.5** – Apple בויט ווייטער אירע אייגענע מעגליכקייטן מיט MM1.5, א שטארקער מולטימאדאלער יסוד-מאדעל.
- **3 אקטאבער: Chromebook Plus מיט AI** – Google האט געבראכט נייע Chromebook מאדעלן פון Samsung און Lenovo. זיי האבן א "Help me read" געצייג וואס פאסט צוזאם PDF'ס און וועב-זייטן, און לייוו איבערזעצונג ביי ווידעא-רופן. עקזיסטירנדע Chromebooks באקומען AI-שרייב געצייג און א רעקארדער וואס פארוואנדלט רייד אין טעקסט.
- **4 אקטאבער: ChatGPT Canvas** – OpenAI האט לאנסירט "canvas", א באזונדערן ארבעט-פלאץ אין ChatGPT פאר שרייב-און קאוד-פראיעקטן. ער קאנקורירט מיט Anthropic'ס Artifacts און מיט Cursor. אין אנפאנג איז ער נאר צוגעגליך פאר אויסגעקליבענע באניצער.

- **4 אקטאבער: FLUX1.1 [pro] אין API** – Black Forest Labs, דער סטארטאפ הינטער Grok'ס בילדער-מאכער, האט ארויסגעלאזט FLUX1.1 [pro]. ער איז זעקס מאל שנעלער ווי דער פריערדיגער און גיט בעסערע קוואליטעט. די פירמע האט אויך געעפנט א בעטא API מיט קאנקורענטפעהיגע פרייזן. דער סטארטאפ ווערט געשטיצט פון Andreessen Horowitz, כאטש עס זענען דא מחלוקות איבער דעם וואס פאר א דאטא ער האט גענוצט.
- **7 אקטאבער: Meta Movie Gen** – Meta האט טריינירט א פילם-מאדעל מיט 30B פאראמעטערס אויף 6,144 H100 טשיפס. ער איז געלערנט געווארן אויף 1B בילדער און 100 מיליאן אויסגעקליבענע ווידעאס. ער איז געבויט ווי א Temporal Autoencoder מיט א Flow matching Llama, און Meta האט ארויסגעגעבן א פארשוונג-ארבעט פון 92 זייטן. שפעטער האט Meta אויך ארויסגעלאזט דעם "Movie Gen Bench" בענטשמארק מיט 1,003 פראמפטס פאר ווידעא און אודיא.
- **7 אקטאבער: Liquid Foundation Models** – Liquid AI האט לאנסירט אירע ערשטע LFM מאדעלן אין גרייסן פון 1B, 3B און 40B. זיי ווייזן ספעציעל שטארקע רעזולטאטן און ניצן ווינציגער זכרון. זיי זענען צוגעגליך אויף LIQUID PLAYGROUND, און באלד וועלן זיי זיין אויף Cerebras.
- **7 אקטאבער: Bing Copilot Labs** – Microsoft עפנט Copilot Labs, וואו מען קען פרובירן Think Deeper און Copilot Vision. זי האט אויך לאנסירט Bing generative search, וואס זאמלט און פאסט צוזאם אינפארמאציע פון וועב.
- **9 אקטאבער: Reflection-70B דורכקוק** – א טיפע אנאליז האט אויפגעדעקט אז דער פיל-באריימטער Reflection-70B האט געהאט באגס אינעם בענטשמארק-קאוד. נאכן פאררעכטן איז ער נאך אלץ שטארק, אבער נישט אויף דער מדריגה וואס מען האט געטענהט.
- **9 אקטאבער: Anthropic Batches API** – Anthropic האט ארויסגעלאזט די Message Batches API. מען קען שיקן ביז 10,000 פראגעס אין איין באטש, די ארבעט ווערט פארטיג אין ווינציגער ווי 24 שעה, און עס קאסט 50% ביליגער ווי געווענליכע API רופן. דאס איז געאייגנט פאר מיטל-גרויסע פירמעס מיט ארבעט וואס יאגט זיך נישט.
- **11 אקטאבער: Aria מולטימאדאל MoE – Rhymes** האט ארויסגעלאזט Aria מיט אפענע וועיטס, קאוד און א פארשוונג-באריכט. ער גלייכט זיך מיט פארמאכטע מאדעלן ביים פארשטיין לאנגע ווידעאס, עפעס וואס Pixtral און Molmo האבן נישט דערגרייכט.
- **11 אקטאבער: שאפינג און רעקלאמע-געצייג** – Amazon האט לאנסירט AI "Shopping Guides" פאר מער ווי 100 סארטן פראדוקטן. Meta האט געבראכט Video Expansion פאר ווידעא-רעקלאמעס, און Image Animation וואס באלעבט בילדער פאר Instagram Reels.
- **14 אקטאבער: Zephyra Mamba 2** – Zephyra'ס מאדעל, געבויט אויף Mamba 2, איז דער ערשטער state space מאדעל וואס שלאגט טראנספארמערס ביי דער 7B גרייס, און ער איז בעסער ווי Mistral. ער איז שטארק מיט לאנגע קאנטעקסטן, ווייל די Mamba 2 בלאקס וואקסן לינעאר מיט דער צייט.
- **14 אקטאבער: Zoom אוואטארן** – Zoom פלאנט פאר קומענדיגן יאר א פונקציע וואס מאכט פאטא-רעאליסטישע AI אוואטארן פאר ווידעא-מעסעדזשעס. מען באמערקט אבער אז עס קען ברענגען ריזיקעס.

- **15 אקטאבער: Adobe Firefly ווידעא** – ביי Adobe MAX האט Adobe פארברייטערט די Firefly פאמיליע אויף ווידעא. Generative Extend, וואס פארלענגערט פילם-שטיקלעך, איז שוין אינעווייניג אין Premiere Pro. Project Super Sonic שאפט קלאנג-עפעקטן פון טעקסט, פון דערקענען אביעקטן, און פון קול. Project Turntable דרייט 2D וועקטאר-בילדער אין 3D און לאזט זיי דאך אויסזען ווי 2D.
- **15 אקטאבער: Microsoft פאטענט** – Microsoft האט איינגעגעבן א פאטענט אויף א סיסטעם וואס פארוואנדלט לייוו אודיא אין בילדער מיט LLMs. עס קען אמאל אריינקומען אין Teams דורך Copilot.
- **16 אקטאבער: Pyramid Flow** – Pyramid Flow איז אן אפן-סאורס ווידעא-מאכער וואס ניצט pyramidal flow matching. ער קאנקורירט מיט Runway Gen-3 Alpha, און מען מעג אים ניצן קאמערציעל אונטער דער MIT לייסענס.
- **17 אקטאבער: Ministral 3B און 8B** – צום ערשטן יארצייט פון Mistral 7B האט Mistral ארויסגעלאזט צוויי מאדעלן פאר טעלעפאנען און עדזש-מכשירים. זיי שטייען אויבן אן ביי ידיעות און לאגיק צווישן מאדעלן אונטער 10B, שטיצן ביז 128k קאנטעקסט, און ארבעטן שנעלער.
- **17 אקטאבער: CoTracker3** – Meta'ס נייער טראקינג מאדעל פארמינערט דעם אונטערשייד צווישן סינטעטישע און עכטע ווידעאס, דורך האלב-אויפגעפאסטן טריינירן.
- **18 אקטאבער: NotebookLM** – באניצער קענען יעצט באשטימען אויף וואס די AI "האסטס" אין NotebookLM'ס Audio Overview זאלן זיך קאנצענטרירן. ביזנעסער קענען זיך איינשרייבן אין א פילאט-פראגראם DeepMind. NotebookLM Business האט אויך דערציילט מער וועגן די אודיא מאדעלן וואס שטייען הינטער NotebookLM.



2024 אקטאבער 18 · NotebookLM

- **18 אקטאבער: Nvidia Nemotron 70B** – Nvidia'ס Llama-3.1-Nemotron-70B-Instruct שלאגט בענטשמארקס פון OpenAI און Anthropic. דאס ווייזט ווי Nvidia גייט ארויס פון האַרדווער אריין אין סאפטוועיר.
- **18 אקטאבער: Google Shopping** – Google שטארקט די Shopping טאב מיט AI-באריכטן, און מיט א פערזענליכן "for you" פיד וואס באזירט זיך אויף זוכונגען און YouTube.
- **21 אקטאבער: Meta אפן-סאורס אוצרות** – Meta האט ארויסגעלאזט א שורה פארשונג-ארבעטן, צווישן זיי א ספיטש-צו-ספיטש מאדעל און א פארבעסערטער SAM. Meta Spirit LM איז אן אפן-סאורס מאדעל

וואס נעמט אריין און גיט ארויס סיי טעקסט סיי רייד. ער ניצט טאוקנס פאר פאנעטיק, הייך און טאן, כדי די רייד זאל ארויסקומען מיט געפיל.

• **21 אקטאבער: DeepSeek מולטימאדאל** – DeepSeek האט ארויסגעלאזט א שטארקן 1.3b מאדעל וואס ארבעט מיט טעקסט און בילדער געמישט, סיי צו פארשטיין און סיי צו שאפן.

• **22 אקטאבער: xAI API** – xAI האט געעפנט אן API מיט איין מאדעל, "grok-beta", מיט function-calling. די פירמע ניצט דאטא פון Musk'ס פירמעס ווי X, כאטש עס זענען נאך דא פראבלעמען מיט די פונקציעס.

• **23 אקטאבער: Claude 3.5 Sonnet און Haiku 3.5** – Anthropic האט פארבעסערט Sonnet 3.5 און ארויסגעלאזט Haiku 3.5. דער ביליגער Haiku 3.5 גלייכט זיך מיט דעם אלטן Claude 3 Opus. Sonnet איז טריינירט געווארן אויף סקרין-רעקארדירונגען צו באדינען א קאמפיוטער און קליקן, און ס'איז ארויס א ביישפיל-קאוד דערפאר. מען האט אויך געבויט Agent exe, א פשוט'ע הילע וואס מאכט די "computer use" גרינגער צו ניצן. אויסערדעם קען Claude יעצט שרייבן און לויפן קאוד אינעווייניג אין artifacts.

• **23 אקטאבער: Stable Diffusion 3.5** – Stability AI האט ארויסגעלאזט א נייע סעריע מאדעלן מיט בעסערע רעזולטאטן און שנעלקייט. זיי זענען פון אנהייב אן געשטיצט אין Diffusers מיט טריינירן.

• **23 אקטאבער: Perplexity און ChatGPT Windows** – Perplexity'ס Internal Knowledge Search זוכט אין איין פלאץ סיי אויפ'ן וועב סיי אין אינערליכע פירמע-פיילס. OpenAI פרובירט אויס א ChatGPT אפליקאציע פאר Windows.

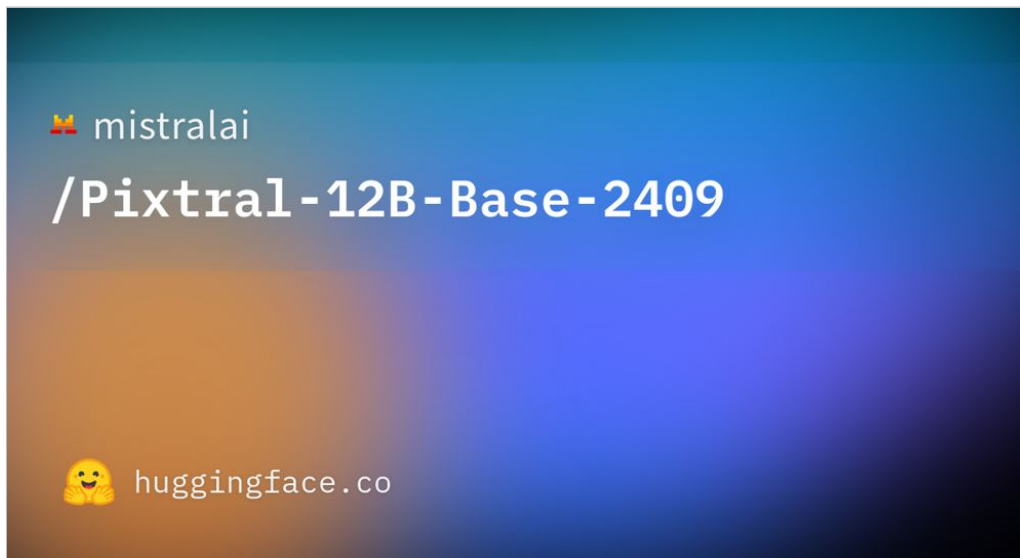
• **24 אקטאבער: Pangea און Midjourney** – א טיעם אין CMU האט ארויסגעלאזט Pangea, אפענע מאדעלן וואס זענען שטארק אין א סך שפראכן. Midjourney פלאנט א וועב-געצייג וואו יעדער וועט קענען רעדאגירן ארויפגעלייגטע בילדער.

• **25 אקטאבער: Cohere Aya** – Cohere'ס Aya פראיעקט איז אסאך שטערקער געווארן, מערסטנס דורך פוסט-טריינינג. ער נעמט אריין אודיא און קען שאפן בילדער פון נישט-ענגלישע אנווייזונגען.



2024 אקטאבער 25 · Cohere Aya

- **25 אקטאבער: MusicFX DJ** – א Grammy-געווינער מוזיקער האט צוזאמען מיט Google DeepMind און Google Labs באשאפן MusicFX DJ, א מוזיק-געצייג מיט גרינגע קאנטראלס פאר יעדן שטאפל קענטעניש.
- **25 אקטאבער: Meta קוואנטיזירטע קליינע מאדעלן** – Meta האט קוואנטיזירט אירע 1B און 3B מאדעלן. זיי זענען 2-4 מאל שנעלער און מער ווי 50% קלענער, כמעט אן פארלוסט אין קוואליטעט. דאס איז מעגליך געווען דורך quantization-aware טריינירן.
- **28 אקטאבער: Pixtral 12B** – Mistral האט פובליקירט די פרטים פון טריינירן און ארכיטעקטור פון איר וויזשן-שפראך מאדעל. די וועיטס פונעם Pixtral 12B באזע-מאדעל זענען ארויס אויף Hugging Face.



2024 אקטאבער 28 · Pixtral 12B

- **28 אקטאבער: AI צירונג** – Arcade AI לאזט מאכן צירונג-שטיקער פון טעקסט-אנווייזונגען, און מען קען קויפן וואס מען האט אנגעוויזן.
- **28 אקטאבער: IBM Granite 3.0** – IBM האט ארויסגעלאזט Granite 3.0, אפן-סאורס LLMs פאר פירמעס. זיי האבן זיכערהייט-פונקציעס און זענען אונטער Apache 2.0 לייסענס.
- **28 אקטאבער: Microsoft "AI ארבעטער"** – Microsoft לאנסירט זעלבסטשטענדיגע AI אגענטן. פירמעס קענען בויען אייגענע באטס פאר קונה-סערוויס און סופליי-קעטן.
- **29 אקטאבער: Google שפראכן און Cohere Embed 3** – Google האט צוגעלייגט קול-דערקענונג פאר 15 נאך אפריקאנישע שפראכן, וואס ערמעגליכט עס פאר בערך 300 מיליאן מענטשן. Cohere's Embed 3 איז געווארן מולטימאדאל, און ער זוכט יעצט אויך אין בילדער ווי גראפן, קאטאלאגן און דיזיין-פיילס.
- **30 אקטאבער: GitHub Copilot מאדעלן** – אין GitHub Copilot קען מען יעצט אויסקלייבן Claude, Gemini, o1 און אנדערע מאדעלן.
- **30 אקטאבער: Apple Intelligence** – Apple האט לאנסירט די ערשטע Apple Intelligence פונקציעס פאר טעקסט און פאטא-רעדאגירן. דער נייער iMac מיטן M4 טשיפ האט דריי מאל אזויפיל AI-קראפט. שפעטער וועט Siri ווערן פארבעסערט און פארבונדן מיט ChatGPT.

• **31 אקטאבער:** Google – Google Jarvis גרייט צו Project Jarvis, אן AI אגענט אויף Gemini 2.0 וואס זאל אויטאמאטיש טון ארבעט אויפ'ן וועב אין Chrome.

פארשונג און טעכניק

• **1 אקטאבער:** פארשידענע פארשונגען – Lotus פאררעכט פעלערן אין פריערדיגע מעטאדן פון ניצן דיפיוזשן-מאדעלן פאר טיפקייט-שאצונג, און מאכט די ארבעט פשוט'ער. SegVLAD דערקענט ערטער לויט שטיקלעך פון א בילד. MaskLLM שניידט אפ איבעריגע טיילן פון LLMs מיט לערנבארער sparsity. E.T. Bench פרובירט ווידעא-מאדעלן אויף פיינע, צייט-אפהענגיגע אויפגאבעס. MIT פארשער לערנען ווי AI זעט "פנימ'ער" אין חפצים, מיט א נייעם דאטאסעט.

• **2 אקטאבער:** RLHF, ערליכקייט, בענטשמארקס – Meta ניצט א געמיש פון ריכטער-מאדעלן פאר RLHF אויף א סך אויפגאבעס אויף איינמאל, און פארבעסערט דערמיט שטארק די בענטשמארקס. אן איבערבליק לערנט די ערליכקייט פון LLMs. פראיעקטן פאדערן צו מעסטן "cross-capabilities", אדער פאר געזעץ-ענינים, דורך PerCo. LexEval. (SD) קאמפרעסירט בילדער אויף זייער נידעריגע ביטס. Nvidia האט געפונען אז דער LLaVa אופן פון צוזאממשטעלן בילדער מיט טעקסט ארבעט דאס בעסטע.

• **3 אקטאבער:** לערן-ראטע און מער – א פארשונג-ארבעט שלאגט פאר אז ווען מען גיט מער דאטא זאל מען שטעלן די לערן-ראטע מיט עקספאנענט -0.33 , אנשטאט -0.28 ווי Llama. מען האט אויך געזען נייע מעטאדן: גרופע-טעאריע פאר knowledge graphs, ProFD פאר דערקענען מענטשן וואס זענען צוגעדעקט, PMRF פאר רעסטארירן בילדער, און RouterDC וואס קלייבט אויס דעם בעסטן מאדעל פאר יעדע פראגע. ETH Zurich פארשער האבן מיט AI געלייזט 100% פון Google's reCAPTCHA v2.

• **4 אקטאבער:** סעמפלינג און בילדער – MIT CSAIL האט א מעטאד וואס צעטיילט פונקטן מער גלייך, כדי סימולאציעס זאלן זיין גענויער. KnobGen מאכט בילדער פון סקיצעס. ComfyGen שאפט ערשט א ComfyUI ארבעט-גאנג און ערשט דערנאך דאס בילד, וואס גיט בעסערע קוואליטעט. Entropix ניצט סעמפלינג לויט ענטראפיע, א מעגליכער וועג צו דערגרייכן מאדעלן ווי o1.

• **7 אקטאבער:** o1 און עמבעדינגס – א פארשונג-ארבעט ווייזט אז o1 האט נאך אלץ סימנים פון באגרענעצונגען וואס קומען פון נעקסט-טאוקן פרעדיקציע. Contextual Document Embeddings ניצן אינפארמאציע פון שכנות'דיגע דאקומענטן און שטייען אויבן אן אויף ENTP. MTEB. ניצט טראנספארמערס וואס האבן נאר אן ענקאדער פאר נעקסט-טאוקן פרעדיקציע, אן "שווינדלען". אנדערע ארבעטן: PairDistill פאר dense retrieval, גראף-אלגאריטמען פון MIT פאר שווינדל און מאסלולים, און Integrative Decoding פאר מער פאקטישע גענויקייט.

• **8 אקטאבער:** דיפיוזשן און ווידעא – DyDiT פאסט צו וויפיל רעכענונג עס ניצט לויט'ן צייט-שריט און לויט'ן ארט אין בילד. FVDM גיט יעדן פריים אן אייגענעם ראש-פלאן. CopyMark ווייזט פעלערן אין membership inference אטאקעס אויף דיפיוזשן מאדעלן. אויך זענען ארויס ImageFolder (א טאוקניזער), Grounded-VideoLLM פאר באשטימטע מאמענטן אין ווידעא, און CCT פאר ראבאטן.

• **9 אקטאבער:** צוזאממשטעלען מאדעלן – א פארשונג לערנט וואס מען דארף כדי צוזאממשטעלען זייער גרויסע מאדעלן. ווען מען שטעלט אלע אינערליכע וועקטארן אויף א היפערספערע, קומט דער מאדעל צו זיין ציל אסאך שנעלער. CodeMMLU איז א שווערערער קאוד-בענטשמארק. מען האט אויך געזען

- PRFusion פאר ראבאטן, און JEDi, א בעסערע מאָס פאר ווידעא-קוואליטעט. ETH Zurich האט באוויזן אז דער ANYmal ראבאט קלעטערט לייטערס מיט 90% הצלחה, 232 מאל שנעלער ווי פריער.
- **10 אקטאבער: Diff Transformer** – Diff Transformer רעכנט די אטענשאן ווי דעם אונטערשייד צווישן צוויי softmax מאפעס, און דערמיט פאלט אוועק דער ראָש. א פארשונג ווייזט אז LLMs ווייסן אינעווייניג ווען זיי זאגן נישט קיין אמת. ProLIP פיינ-טיונט נאר די לעצטע מאטריץ פון CLIP. מען האט אויך געזען GFMBench פאר גענאמישע מאדעלן, Science Agent Bench, CAR, PredFormer, א נייע מעטאד פאר רעסטארירן בילדער, און GenSim2 וואס פארגרעסערט ראבאט-דאטא מיט סימולאציע.
 - **11 אקטאבער: MLE-bench** און o1 – OpenAI's MLE-bench פרובירט AI אגענטן אויף 75 קאנקורסן. ער ווייזט פארשריט, אבער אויך באגרענעצונגען אין קרעאטיווקייט. פארשער פון GAIR און NYU, וואס פרובירן נאכצומאכן o1, האבן געפונען אז "Journey Learning" דאטא הייבט אויף מאטעמאטיק מיט 8%. אנדערע ארבעטן: MIR, א מאָס פאר מולטימאדאל טריינירן; MatMamba מיט נעסטירטע אונטער-מאדעלן; IterComp; און א פראיעקט וואס רייניגט דאטא אויטאמאטיש לויט Minipile.
 - **14 אקטאבער: בילדער און 3D – SEA** מעסט צוזאמשמעלצן פון געווענליכע און אינפרא-רויטע בילדער. EvolveDirector בויט א טעקסט-צו-בילד מאדעל מיט עפנטליכע מקורות. מען האט אויך געזען א לערן-געצייג פאר טראנספארט-פארשונג, Trans4D פאר 4D סצענעס, DelTA, אן אגענט וואס איבערזעצט גאנצע דאקומענטן, און FCGS וואס קאמפרעסירט 3D אין סעקונדעס אנשטאט מינוטן.
 - **15 אקטאבער: INTELLECT-1** – Prime Intellect האט אנגעהויבן INTELLECT-1, דאס ערשטע צעשפרייטע טריינירן פון א 10B מאדעל איבער דער גארער וועלט. די Prime פריימווארק דערגרייכט 98% נוצן פון רעכן-קראפט. אנדערע ארבעטן: SmartPretrain פאר אויטאס, OneRef, אן אפענע נאכמאכונג פון א דיפיוזשן מאדעל וואס סימולירט CS:GO, און אן אופן צו אויסגלייכן LLMs וואס רעכנט מיט דער קוואליטעט פון די ענטפערס.
 - **16 אקטאבער: ראבאטן און ליניארן – iDP3** לאזט הומאנאידע ראבאטן ארבעטן אין פארשידענע ערטער. Hazy Research ווייזט ווי מען מאכט עקזיסטירנדע LLMs ליניאר אן צו פארלירן צופיל. מען האט אויך געזען TextCtrl, HART וואס שאפט 1024x1024 בילדער ביליגער, און DeBiFormer.
 - **17 אקטאבער: קאנסיסטענסי מאדעלן** – OpenAI און אנדערע פארשער שטארקן און מאכן מער סטאבילע די קאנסיסטענסי מאדעלן, וואס שאפן אין צוויי שריט אנשטאט 1,000. זיי זענען אבער נאך אנגעוויזן אויף דיסטילאציע. DARNet דעקאדירט מוח-סיגנאלן. DuoAttention שפארט זכרון ביי לאנגע קאנטעקסטן. מען האט אויך געזען א נייע מעטאד פאר פוינט-קלאודס.
 - **18 אקטאבער: MIT ווידעא און ראבאטן** – MIT קאמבינירט נעקסט-טאקן פרעדיקציע מיט ווידעא-דיפיוזשן פאר בעסערע ווידעא, ראבאט-פלענער און אגענטן. מען האט אויך געזען Meta-DT, LongAlign פאר לאנגע טעקסטן, DiGIT, און FedPart.
 - **21 אקטאבער: Shortcut מאדעלן** – Shortcut מאדעלן שאפן סיגנאלן מיט זייער ווייניג שריט. VLM-Grounder ארבעט אויף 3D grounding אן טריינירן ספעציעל דערפאר (zero-shot). LiVO גלייכט אויס טעקסט-צו-בילד מאדעלן מיט מענטשליכע ווערטן.
 - **22 אקטאבער: Mini-Omni2** – Mini-Omni2 איז אן אפן-סאורס פראווי נאכצומאכן GPT-4o'ס קול-מעגליכקייטן. LongPiBench מעסט פאזיציע-נייגונגען ביי לאנגע קאנטעקסטן. BiGR פאראייניגט

שאפן און אונטערשיידן אין איין מאדעל. Clamp2 פארבינדט מוזיק מיט טעקסט. Montessori-Instruct פאסט צו סינטעטישע דאטא צו דעם תלמיד-מאדעל.

• **23 אקטאבער: פארשטיין מאדעלן** – Eleuther ערקלערט אויטאמאטיש מיליאנען SAE פיטשערס. אנדערע ארבעטן: epipolar attention פאר קאמערע-קאנטראל, 3D פנים פון איין בילד, און CCA וואס פארמינערט וויזועלע הזיות. ProSLM ניצט Prolog פאר מער פארלעסליכע לאגיק.

• **24 אקטאבער: 32 טאוקנס פאר ווידעא** – Salesforce'ס נייער ענקאדער ניצט אסאך ווייניגער טאוקנס פאר ווידעא, דורך א באזונדערן צייט-ענקאדער און ראום-ענקאדער. אויך זענען ארויס SAM2Long, CoPS פאר אגענטן, און TALoS פאר אויטאס.

• **25 אקטאבער: OmniParser** – Microsoft'ס OmniParser לייענט סקרינשאטס כדי מאדעלן זאלן קענען באדינען דיגיטאלע אינטערפעיסעס. אויך זענען ארויס OUNet פאר פוינט-קלאודס, FeT פאר federated learning, און ADEM-VL.

• **28 אקטאבער: Anthropic: steering און crosscoders** – Anthropic האט אויסגעפרובירט feature steering. זי האט געפונען א "גוטן פונקט" וואו מען קען ענדערן ניגונגען אן צו שאדן די קענטענישן, אבער עס זענען דא אומגעריכטע זייטיגע ווירקונגען. Crosscoders זענען א פארברייטערונג פון sparse autoencoders, און מען האט זיי שוין אפן נאכגעמאכט אויף Gemma 2B.

• **28 אקטאבער: צייט-סעריעס און טשארטס** – RATD ניצט היסטארישע דאטא פון א דאטאבאזע כדי צו פירן דעם דיפיוזשן פראצעס. CIT ניצט קאוד צו איבערזעצן טשארטס אויף טעקסט. ProLIP לערנט מאנכע-צו-מאנכע פארבינדונגען צווישן בילדער און טעקסט.

• **29 אקטאבער: MIT HPT** – MIT'ס HPT טריינירט ראבאטן מיט דאטא פון פארשידענע פעלדער. אנדערע ארבעטן: view synthesis מיט ווייניג 3D הנחות; רייד-סינטעז מיט דיפיוזשן פאר יעדן טאוקן; ענדערונגען אין סאטעליט-בילדער; און FtD++. אין ארבעט פאר 3D סצענעס האט געוויזן אז Claude Sonnet שטארקע סצענעס אין א נייער שפראך, כאטש ער איז נישט טריינירט געווארן דערויף. מען האט אויך געזען IC Light פאר Flux, אויטאמאטיש פארבן פון אנימאציע, און OpenWebVoyager פאר וועב-אגענטן.

• **30 אקטאבער: DeepMind קליינער העלפער** – DeepMind ווייזט אז א קליינער מאדעל קען צושטעלן לעיבלס און אויסקלייבן שווערע דאטא, און דערמיט פארשנעלערן שטארק דאס טריינירן. MINT-1T איז אן אפענער דאטאסעט מיט 1 טריליאן טעקסט-טאוקנס און 3.4 ביליאן בילדער, 10 מאל גרעסער ווי פריער. אויך זענען ארויס NeuroClips (ווידעא פון fMRI), PIP פאר מאסלולים, CompassJugder-1, און LARP.

• **31 אקטאבער: SimpleQA און ThunderKittens** – OpenAI'ס SimpleQA בענטשמארק מעסט פשוט'ע פאקטישע פראגעס. ThunderKittens 2 שרייבט GPU קערנעלס וואס זענען 40% שנעלער מיט עטליכע הונדערט שורות. אויך זענען ארויס MeshRet, self-guidance פאר PF3plat, PiToMe, MGMs, און SELECT מיט ImageNet++.

געצייג פאר פראגראמירער און אפן-סאורס

• **1 אקטאבער: LeanRL און מער** – LeanRL איז א שנעלערע און פשוט'ערע ווערסיע פון CleanRL אויף PyTorch. אן ארטיקל באשרייבט ווי LLMs ציען ארויס טאבעלעס פון דאקומענטן בעסער ווי OCR. Wispr Flow איז א דיקטיר-אפליקאציע וואס שרייבט אין אייער סטיל.

- **2 אקטאבער: Inbox Zero**, פייל-ארדענונג – Inbox Zero איז אן אפן-סאורס אסיסטענט פאר אימעילס. א געצייג מיט Llama 3.2 3B און Llava-1.6 ארדנט פיילס לויט זייער אינהאלט.
- **4 אקטאבער: Concordia** – Concordia ס'DeepMind סימולירט סאציאלע באציאונגען. Ross Wightman האט ארויסגעלאזט ImageNet מאדעלן אונטער 1 מיליאן פאראמעטערס. מען האט אויך געזען אן איינפירונג אין צעשפרייטן טריינירן.
- **7 אקטאבער: FinLLM און Ollama** – א Leaderboard פאר פינאנץ-מאדעלן אויף Hugging Face. Ollama מיט Nvidia RTX לויפט לאקאלע מאדעלן אינעם Brave בראווער. א געצייג מאכט פראקטאלן אן א סוף פאר טעסטן. Auto Jobs Applier מאכט אויטאמאטיש ארבעט-אפליקאציעס.
- **8 אקטאבער: FacePoke, DataPelago** – FacePoke רעדאגירט פנימ'ער. DataPelago האט לאנסירט א Universal Data Processing Engine. א מומחה פון Hugging Face האט געצייגט א ביישפיל פון פאראלעליזם אין 200 שורות.
- **9 אקטאבער: intensify, FLORA** – intensify קאלירט טעקסט אין טערמינאל לויט logprobs. FLORA שטעלט צוזאם קרעאטיווע מאדעלן אויף איין קאנוואס אן א סוף.
- **14 אקטאבער: OpenAI Swarm** – OpenAI האט ארויסגעלאזט Swarm, א לייכטע פריימווארק פאר קאמוניקאציע צווישן אגענטן. מען וועט זי נישט ווייטער אונטערהאלטן. UvA האט ארויסגעגעבן לערן-מאטעריאל פאר group equivariant deep learning.
- **16 אקטאבער: Tabled, Ditto, F5-TTS** – Tabled ציט ארויס טאבעלעס פון PDF'ס. Ditto מאכט Flask אפליקאציעס פון א באשרייבונג אין פשוט'ער שפראך. F5-TTS איז א zero-shot TTS אויף MLX וואס מאכט רייד אין בערך 11 סעקונדעס. מען האט אויך געזען א פולן גייד פאר צעשפרייטן טריינירן.
- **21 אקטאבער: Meta Lingua** – Lingua מאכט נאך Llama 2 7B אין 24 שעה. א פשוט'ע VLM איז ארויס, געשריבן נאר אין PyTorch.
- **22 אקטאבער: Quickstarts און BitNet** – Anthropic Quickstarts גיט גרייטע פראיעקטן, ווי א קונה-סערוויס אגענט. BitNet cpp'ס לויפט 1.58-ביט מאדעלן (מיט די ווערטן -1, 0, 1) אויף געווענליכע קאמפיוטערס. Hugging Face ווייזט ווי מען פיין-טיונט עקזיסטירנדע מאדעלן צו דעם פארמאט.
- **23 אקטאבער: ספיטש און טעלעפאן** – א שטארקער ASR פאר עדוש-מכשירים איז ארויס. Pocket Pal, וואס לויפט מאדעלן אויפ'ן טעלעפאן, איז געווארן אפן-סאורס. MotherDuck'ס PROMPT () ניצט gpt-4o, mini פון אינעווייניג SQL.
- **24 אקטאבער: Transformers.js v3** – Transformers.js ניצט יעצט WebGPU, וואס איז אין געוויסע פעלער ביז 1,000 מאל שנעלער. ער האט מער ווי 1,200 גרייטע מאדעלן.
- **25 אקטאבער: ExecuTorch** – PyTorch'ס ExecuTorch לויפט מאדעלן אויף טעלעפאנען, wearables און מיקראקאנטראלערס, און ניצט אויס CPUs, NPUס און DSPs.
- **28 אקטאבער: NotebookLlama** – Meta גיט א גייד ווי צו בויען א פשוט'ע ווערסיע פון NotebookLM. Marly לאזט אגענטן פרעגן אומארגאניזירטע דאטא מיט JSON. מען האט אויך געזען א 14-טאגניג לערן-פלאן פאר LLMs.

- **29 אקטאבער:** ליסטעס און דאטא – א ליסטע פון פארשונג-ארבעטן אויף graph OOD learning , און א lofi מוזיק-דאטאסעט.
- **31 אקטאבער:** טעסטן מיט LLMs – Assembled שרייבט טעסטן מיט LLMs אין מינוטן אנשטאט שעה, און שפארט דערמיט הונדערטער שעה.

פירמעס, געלט און ביזנעס

- **1 אקטאבער:** OpenAI'ס ריזיגע געלט-רונדע – OpenAI האט אויפגעהויבן 6.6 ביליאן דאלאר ביי א ווערט פון 157 ביליאן, מיט Microsoft און Nvidia צווישן די אינוועסטארן. SoftBank האט געפלאנט 500 מיליאן. די פירמע ברענגט אריין 300 מיליאן דאלאר א חודש, א וואקס פון 1,700% זינט אנהייב 2023. זי ערווארט 3.7 ביליאן פאר דעם יאר און 11.6 ביליאן פאר'ן קומענדיגן, אבער ערווארט אויך א פארלוסט פון בערך 5 ביליאן דעם יאר. א קרעדיט-ליניע פון 4 ביליאן גיט איר מער ווי 10 ביליאן פלאסיגע געלט. לויט פראיעקציעס וועט דער פארלוסט וואקסן צו 14 ביליאן אין 2026, און ביז 2028 וועט מען פארלירן אינגאנצן 44 ביליאן.
- **1 אקטאבער:** Altman'ס דאטא צענטער פלאן – Sam Altman דרוקט אויף דער Biden אדמיניסטראציע צו בויען דאטא צענטערס וואס ניצן ביז פינף גיגאוואט, כדי צו בלייבן פאראויס פון כינע. ער האט ערשט געזוכט טריליאנען, יעצט ציל ער אויף הונדערטער ביליאנען. Microsoft און Amazon זיכערן אויך אטאמישע ענערגיע.
- **2 אקטאבער:** OpenAI איבערביי – OpenAI פלאנט צו ווערן א public benefit corporation פאר פראפיט, צווישן אנדערן כדי זיך צו שיצן פון "פיינטליכע" איבערנעמונגען. די נאן-פראפיט וועט האלטן א מינאריטעט-טייל, און Sam Altman וועט באקומען אקציעס. OpenAI און Microsoft האבן איינגעשטעלט בענק צו פארהאנדלען די נייע באדינונגען, און Microsoft וויל א באדייטנדן טייל. אן אנאליז פרעגט וויפיל דער טייל פון דער נאן-פראפיט איז ווערט.
- **3 אקטאבער:** טשיפס – BrainChip'ס Akida Pico איז א נירא-מארפישער טשיפ פאר מכשירים מיט ווייניג קראפט, וואס ניצט בלויז מילי-וואטס. ByteDance וועט טריינירן א מאדעל מיט Huawei'ס Ascend 910B טשיפס.
- **4 אקטאבער:** פערסאנאל-ענדערונגען – Durk Kingma, איינער פון די גרינדער פון OpenAI, איז אריבער צו Anthropic. Sebastien Bubeck פון Microsoft גייט צו OpenAI. דער געוועזענער CISO פון Palantir, Dane Stuckey, פירט יעצט די זיכערהייט ביי OpenAI, וואס ארבעט אויך מיט'ן Pentagon.
- **8 אקטאבער:** OpenAI אינערליכע שפאנונג – דער אוועקגיין פון OpenAI'ס CTO ענדיגט אן יאר פון קריגערייען, איילנדיגע לאנסירונגען פון GPT-4o און o1, און אוועקגיין פון הויכע באאמטע. אן אנאליז זעט די ווארצל אין דעם קאנפליקט צווישן פראדוקט און געלט אויף איין זייט, און פארשונג און זיכערהייט אויף דער אנדערער. די געוועזענע CTO זאמלט לויט באריכטן מער ווי 100 מיליאן פאר א נייעם סטארטאפ.
- **8 אקטאבער:** אינוועסטירונגען – Kobold Materials, וואס זוכט מעטאלן מיט AI, האט אויפגעהויבן 491 מיליאן. Braintrust האט באקומען 36 מיליאן אין א Series A, אנגעפירט פון Clout Kitchen. a16z. האט אויפגעהויבן 4.45 מיליאן פאר AI באגלייטער אין Hamming. League of Legends. ברענגט צוטרוי צו קול-אגענטן.

- **8 אקטאבער:** רעקלאמעס אין Google – AI Overviews וועט ווייזן רעקלאמעס אין AI Overviews, וואס האבן געהעכערט די באנוצונג, בפרט ביי יונגע. זי ברענגט אויך זוך-רעזולטאטן וואס AI ארגאניזירט, און רעקלאמעס.
- **10 אקטאבער:** OpenAI און Microsoft; ByteDance – OpenAI פארמינערט איר אפהענגיקייט פון Microsoft'ס דאטא צענטערס, און Microsoft פארמינערט איר אפהענגיקייט פון OpenAI. ByteDance'ס ByteSpider קלייבט דאטא פון וועב 25 מאל שנעלער ווי OpenAI.
- **10 אקטאבער:** Sonair – Waymo; Sonair בויט 3D-זעונג מיט אולטראסאונד אנשטאט lidar, און האט אויפגעהויבן 6.8 מיליאן. Tesla'ס געוועזענער הויפט פון פאר-פראגראמען, Daniel Ho, איז אריבער צו Waymo.
- **11 אקטאבער:** Uber, TikTok, Hearst – Uber Green פארשפרייט זיך צו 40 שטעט, און אין 2025 קומט א GPT-4 טשעטבאט פאר שאפערס. TikTok'ס Smart+ קאנקורירט מיט Meta'ס Advantage+. Hearst לייסענסירט אינהאלט פון מער ווי 60 פובליקאציעס פאר OpenAI.
- **15 אקטאבער:** OpenAI אינטערנאציאנאל; AMD – OpenAI האט באשטימט Oliver Jay צו פירן די אינטערנאציאנאלע פארברייטערונג, און עפנט אן אפיס אין סינגאפור. AMD האט לאנסירט דעם Instinct MI325X טשיפ קעגן Nvidia.
- **16 אקטאבער:** Meta האַרדוועיר – Meta האט געוויזן אפענע האַרדוועיר ביי OCP: Catalina, Grand Teton, און א נייע DSF נעטווארק. טריינירן Llama 3.1 405B האט געדארפט 16,000 H100 טשיפס.
- **16 אקטאבער:** NYT קעגן Perplexity – די New York Times האט געשיקט א cease and desist צו Perplexity פאר'ן ניצן איר אינהאלט. Perplexity טענהט אז זי אינדעקסירט נאר און טריינירט נישט דערויף.
- **16 אקטאבער:** Decagon, Kind, TikTok, Tesla – Decagon האט אויפגעהויבן 65 מיליאן אין א Series B פאר קונה-סערוויס אגענטן, וואס Duolingo ניצט. Kind Humanoid, מיט דריי מענטשן, בויט Mona פאר דער היים. TikTok האט אפגעזאגט ווייניגער ווי 500 ארבעטער אין מאלייזיע צוליב AI-קאנטראל פון אינהאלט. Tesla האט געוויזן א Robotaxi פראטאטיפ.
- **17 אקטאבער:** "Open AI" נאמען-קריג – OpenAI קריגט זיך אין געריכט מיט Guy Ravine'ס פירמע Open AI איבער דעם נאמען.
- **21 אקטאבער:** World – Worldcoin, מיטגעגרינדעט פון Sam Altman, הייסט יעצט World. זי ווייזט א נייעם Orb וואס סקאנט די אויג, א בלאקטשעין, און Deep Face, אלעס צווישן פריוואטקייט-זארגן.
- **22 אקטאבער:** ראבאטן – Serve'ס דריטע דור עקספרעס-ראבאט ניצט Nvidia Jetson Orin. Boston Dynamics און TRI ברענגען AI צו Atlas. דער כינעזישער STAR1 איז געלאפן מער ווי 8 מייל א שעה. Kepler'ס Forerunner K2 טראגט 33 פונט אין יעדער האנט.
- **25 אקטאבער:** Apple און Perplexity – לויט אינערליכע שטודיעס איז ChatGPT 25% גענויער ווי Siri, און Apple איז בערך צוויי יאר הינטערשטעליג. Perplexity רעדט וועגן אויפהייבן 500 מיליאן ביי א ווערט פון 8 ביליאן.

- **28 אקטאבער:** Waymo 5.6 ביליאן; ByteDance – Waymo האט אויפגעהויבן 5.6 ביליאן צו פארברייטערן איר טאקסי-סערוויס. ByteDance באשטעטיגט אז אן אינטערן האט געשטערט דאס טריינירן, אבער לייקנט אז דער שאדן האט דערגרייכט צענדליגער מיליאנען.
- **30 אקטאבער:** Agentforce און סטארטאפס – Marc Benioff, דער CEO פון Salesforce, לויבט Agentforce און קריטיקירט Microsoft's Copilot. a16z Oxygen גיט סטארטאפס ביליגערע GPU's. Coframe האט אויפגעהויבן 9.3 מיליאן. Advex מאכט סינטעטישע בילדער פאר פאבריקן.

רעגירונגען, געזעצן און זיכערהייט

- **1 אקטאבער:** Tesla FSD – ביי א טעסט פון 1,000 מייל דורך AMCI האט Tesla's FSD געדארפט מענטשליכע אריינמישונג יעדע 13 מייל.
- **2 אקטאבער:** SB 1047 אפגעווארפן – Newsom האט אפגעווארפן SB 1047, ווייל דער געזעץ האט גלייך באהאנדלט אלע גרויסע מאדעלן און נישט גענוג געקוקט אויף קלענערע געפערליכע. ער וויל רעגולאציעס באזירט אויף ראיות, מיט עקספערטן. די שטיצער, ווי Dan Hendrycks און Elon Musk, ערציילן זייער זייט. OpenAI און Google זענען געווען דערקעגן.
- **3 אקטאבער:** אחריות און קאפירייט – אן ארטיקל אין Lawfare שלאגט פאר אז AI-אנטוויקלער זאלן טראגן אחריות פאר נאכלעסיגקייט. OpenAI איז מסכים אז שרייבער זאלן קענען דורכקוקן איר טריינינג-דאטא.
- **14 אקטאבער:** אויטאנאמע וואפן – Palmer Luckey פון Anduril איז אפן צו וואפן וואס באשליסן אליין צו הרג'ענען. די אמעריקאנער רעגירונג האט נישט קיין פארבאט, און Anduril און Palantir לאבירן שטארק.
- **18 אקטאבער:** אומזעבארער טעקסט; טשיפ-עקספארט – אומזעבארע Unicode אותיות קענען באהאלטן אנווייזונגען אין Claude און Copilot. די Biden אדמיניסטראציע באטראכט צו באגרענעצן טשיפס פון Nvidia און AMD צו לענדער ביים פערסישן גאלף.
- **21 אקטאבער:** Anthropic זיכערהייט – Anthropic האט דערהיינטיגט איר Responsible Scaling Policy מיט Capability Thresholds, למשל ביי ביא-וואפן אדער אויטאנאמע פארשונג, און מיט AI Safety Levels. זי האט אויך אויסגעפרובירט וויפיל א מאדעל קען סאבאטירן: מענטשליכע באשלוסן, קאוד, "sandbagging", און אויפזיכט. לויט איר זענען די ריזיקעס יעצט קליין, אבער מען דארף שטערקערע טעסטן.
- **24 אקטאבער:** SynthID – DeepMind האט געעפנט SynthID, וואס צייכנט טעקסט סטאטיסטיש ביים סעמפלינג, און מען קען דערקענען דאס צייכן מיט א Bayesian מעטאד.
- **31 אקטאבער:** Whisper הזיות – פארשער באריכטן אז OpenAI's Whisper לייגט צו אויסגעטראכטע זאכן אין טראנסקריפציעס, אפילו ביי מעדיצינישע.

אנאליזן און מיינונגען

- **1 אקטאבער:** "די אנדערע באבל" – א שרייבער טענהט אז Big Tech אינוועסטירט שווער אין AI אבער באקומט שוואכע רעזולטאטן. אן אנדערער זאגט אז די ווינערס איצט זענען Big Tech, טשיפ-מאכער, IP אדוואקאטן און די Big 4.
- **2 אקטאבער:** פראדוקטיוויטעט – א שרייבער מיינט אז AI מאכט ביליג סערוויסעס ווי לערנען, געזונט, און פערזענליכע סטייל-עצות. אנדערע, צווישן זיי Gary Marcus, טענהן אז פראגראמירער פראפיטירן ווייניג, און אז AI וועט נישט מאכן פראגראמירן 10 מאל שנעלער.
- **3 אקטאבער:** מידקייט פון AI – א שרייבער באקלאגט די פלייץ פון מיטלמעסיגן AI-אינהאלט. א פראפעסאר האט אויפגעהערט לערנען צוליב ChatGPT ביי סטודענטן.
- **9 אקטאבער:** דאטא-וואנט און ענערגיע – א שרייבער טענהט אז די עכטע "דאטא-וואנט" איז ביליאנען יארן עוואלוציע. אן ארטיקל ווייזט דריי ביישפילן פון דאטא-ליקאזש. אן אנדערער מיינט אז AI'ס ענערגיע-דארשט וועט מאכן ריינע ענערגיע ביליגער.
- **14 אקטאבער:** Amodei'ס חזון – Dario Amodei, דער CEO פון Anthropic, שילדערט אין "Machines of Loving Grace" די ריזיגע מעלות פון שטארקע AI. Not Boring ווארנט אז שטרענג פאלגן "פלעיבוקס" מאכט מענטשן ווי מאשינען. מענטשן האבן פארשפרייט אן AI בילד פון Hurricane Helene כאטש זיי האבן געוואוסט אז עס איז נישט אמת.
- **15 אקטאבער:** Harari; לערער – Yuval Noah Harari ווארנט אז AI, אן "פרעמדע אינטעליגענץ", קען קאנצענטרירן מאכט און סכנ'ען דעמאקראטיע. א שרייבער מיינט אז AI-לערער קענען נישט ערועצן די סאציאלע זייט פון מענטשליכע לערער.
- **17 אקטאבער:** פארשטארקן מענטשן; לאגיק – א פאסט אויף LessWrong באשרייבט וועגן צו פארשטארקן מענטשליכע אינטעליגענץ. Gary Marcus ברענגט אן Apple שטודיע אז LLMs טוען פאטערן-פאסונג און נישט פארמעלע לאגיק.
- **18 אקטאבער:** ROI – Flexport האט פארקירצט די באהאנדלונג פון דאקומענטן מיט 80%. א שרייבער מיינט אז די באלאנסירטע קולות פעלן אין דער AI-דעבאטע.
- **21 אקטאבער:** AlphaCodium; Duolingo – AlphaCodium פארבעסערט o1 אויף Codeforces. Duolingo'ס CEO דערציילט וועגן Duolingo Max און AI-אינהאלט.
- **22 אקטאבער:** State of AI 2024 – דער באריכט זאגט אז OpenAI פירט אין פארדינסט, אז די פראנטיר-לאבס ווערן גלייכער איינע צו די אנדערע, און אז מען לייגט מער טראכט אויף פלאנירן און לאגיק. The Algorithmic Bridge זעט אין o1 א איבערגאנג צו מאדעלן וואס טראכטן.
- **23 אקטאבער:** פראגנאזן – Digital Native ערווארט אז ביז 2034 וועלן קומען נייע AI פירמעס ווערט 100 ביליאן. GeekWire דערציילט ווי Microsoft איז אריין אין AI. Cloudflare'ס CEO רעדט וועגן עדזש-רעכענען און רעגיאנאליזירונג.
- **25 אקטאבער:** ווידעא-סקרעיפינג; אינוועסטירונגען – Simon Willison האט ארויסגעצויגן JSON פון א ווידעא מיט Gemini 1.5 Flash פאר ווייניגער ווי 1/10 פון א סענט. איינער האט פארויסגעזאגט וויפיל ער וועט אראפגיין אין וואג מיט א DNN. דאס בויען דאטא צענטערס אין אמעריקע האט דערגרייכט א רעקארד פון 28.6 ביליאן א יאר.

- **29 אקטאבער:** טראכטן ווי AI – Ethan Mollick ערקלערט אז LLMs זענען ווי אן אויטאקאמפליט, אז קליינע ענדערונגען אינעם פראמפט ענדערן דעם ענטפער, און ווי מען ניצט דאס.
- **31 אקטאבער:** CNET אנקעטע – נאר 18% באניצער ווילן אפגרעידן זייער טעלעפאן צוליב AI, און 25% ווילן בכלל נישט קיין AI.

ווייטערע ידיעות

- **9 אקטאבער:** Nobel פרייז – Hopfield און Hinton האבן געוואונען די פיזיק-פרייז פאר נירא-נעטווארקס. Demis Hassabis, John Jumper און David Baker האבן געוואונען די כעמיע-פרייז פאר פראטעין-פאלדן מיט AlphaFold.
- **11 אקטאבער:** Llama אין טעראפיע – Meta ווייזט ווי מען ניצט Llama פאר ABA טעראפיע פאר אויטיזם.
- **14 אקטאבער:** Iprova – Iprova'ס סאפטוועיר שלאגט פאר נייע פאטענטירבארע המצאות פאר פירמעס ווי P&G און Panasonic.
- **21 אקטאבער:** קאנקורסן און דאקטוירים – דער Gemini טיעם גיט 100 טויזנט דאלאר פאר דעם בעסטן נוצן פון לאנגע קאנטעקסטן. Jane Street פירט א טריידינג קאנקורס אויף Kaggle. Google ווייזט ווי AI העלפט דאקטוירים, וואס פארברענגען כמעט 28 שעה א וואך מיט פאפירן.

אין נאוועמבער 2024 איז דער גרעסטער טעמע געווען צי AI ווערט נאך אלץ בעסער אין דעם זעלבן טעמפא. באריכטן זאגן אז OpenAI'ס קומענדיגער מאדעל "Orion" ווייזט ווייניגער פארשריט, און די אינדוסטריע רעדט מער פון "טעסט-צייט סקעלינג", דאס הייסט מער רעכן-קראפט בשעת'ן ענטפערן. אין דעם ריכטונג האבן DeepSeek מיט r1 און Alibaba מיט QwQ-32B ארויסגעלאזט "טראכט-מאדעלן" וואס פרובירן נאכצוטון OpenAI'ס o1. Anthropic איז געווען זייער אין די נייעס: Claude 3.5 Haiku, דער Model Context Protocol, א שותפות מיט Palantir, און נאך 4 ביליאן דאלאר פון Amazon. אויסער דעם האט DeepMind אפן געמאכט OpenAI AlphaFold 3, האט ChatGPT Search, און עס זענען ארויס א ריי שטארקע אפן-סאורס מאדעלן ווי Qwen 2.5, OLMo 2 און Coder, Pixtral Large.

מאדעלן און נייע לאנסירונגען

- **1 נאוועמבער: ChatGPT זוכט אויפ'ן וועב** – ChatGPT קען יעצט זוכן אויפ'ן אינטערנעט אסאך בעסער ווי פריער. עס גיט שנעלע, פרישע ענטפערס מיט לינקס צו די מקורות, וואס פריער האט מען געדארפט זוכן אין א זוך-מאשין. דאס פאראייניגט א שמועס אין נאטירליכער שפראך מיט פרישע אינפארמאציע ווי ספארט-רעזולטאטן, נייעס און אקציע-פרייזן.
- **1 נאוועמבער: Grok פון xAI** – xAI האט צוגעלייגט צו Grok די מעגליכקייט צו פארשטיין בילדער, פאר באצאלטע באניצער אויף X. אין דער צוקונפט קען עס אויך פארשטיין דאקומענטן. X גרייט זיך צו עפענען Grok פאר אומבאצאלטע באניצער, און xAI פלאנירט ביז דעצעמבער א באזונדערע אפליקאציע פאר Grok, כדי צו קאנקורירן גלייך מיט ChatGPT.
- **1 נאוועמבער: רעטעניש-מאדעל red_panda** – א נייער בילד-מאדעל מיטן נאמען red_panda איז באוויזן אויבן אויפ'ן Artificial Analysis בענטשמארק, פאר די מאדעלן פון Midjourney, Black Forest Labs און OpenAI.
- **4 נאוועמבער: AMD'ס שפראך-מאדעל** – AMD האט באנוצט דעם OLMO קאוד צו טריינירן און ארויסלאזן א שפראך-מאדעל וואס איז טריינירט געווארן אויף אירע אייגענע אקסעלעראטארן.
- **4 נאוועמבער: Oasis – א שפיל אן א שפיל-מאשין** – Etched און Oasis'ס Decart האבן שאפט עפענע-וועלט שפילן איין בילד נאכ'ן צווייטן, לויט וואס דער באניצער טוט. עס איז א מאדעל פון 500 מיליאן פאראמעטערס אויף א Diffusion Transformer ארכיטעקטור, אן קיין שפיל-מאשין (game engine). עס איז אפטימירט פאר Etched'ס קומענדיגן Sohu טשיפ, וואס זאל ערמעגליכן 4K ווידעא מאדעלן מיט קליינע פארשפעטיגונג. עס בלייבן נאך שוועריגקייטן, ווי הענטלעך נייע געביטן און האלטן קאנסיסטענץ איבער א לאנגע צייט.
- **5 נאוועמבער: Claude 3.5 Haiku** – Anthropic האט ארויסגעלאזט Claude 3.5 Haiku. עס איז אביסל שוואכער פון GPT-4o און קען נישט זען בילדער, אבער עס איז זייער קלוג און ביליג פאר זיין שטאפל. עס איז בעסער פון Claude 3 Opus און שטארק אין קאוד, ארויסנעמען דאטא און מאדערירן אינהאלט. אבער

Anthropic האט דערהעכערט דעם פרייז באדייטנד, און דאס ברענגט פראגן איבער איר צוקונפטיגע פרייז-סטראטעגיע.



Claude 3.5 Haiku · 5 נאוועמבער 2024

- **5 נאוועמבער:** דעסקטאפ אפליקאציעס – Anthropic האט ארויסגעלאזט א דעסקטאפ אפליקאציע פאר Claude אויף Mac און Windows. OpenAI האט לאנסירט ChatGPT דעסקטאפ אינטעגראציעס צו קאנקורירן מיט ChatGPT. Copilot אויף macOS קען יעצט לייענען פראגראמען ווי VS Code און Xcode, אזוי אז פראגראמירער קענען שיקן זייער קאוד גלייך צו ChatGPT.
- **5 נאוועמבער:** LinkedIn'ס AI אגענט – LinkedIn האט לאנסירט "Hiring Assistant", איר ערשטער AI אגענט, וואס העלפט רעקרוטירער מיט זייערע ארבעט.
- **6 נאוועמבער:** o1 ארויסגעליקט – OpenAI'ס o1 מאדעל איז ארויסגעליקט געווארן. עס ווייזט באדייטנדע פארבעסערונגען אין לאגיק און גענויקייט אקעגן GPT-4, און קען אויך אנאליזירן בילדער און באנוצן וועב-געצייג. מען ערווארט אז די פולע ווערסיע וועט ארויסקומען באלד נאך די אמעריקאנער פרעזידענט-וואלן.
- **6 נאוועמבער:** Wonder Dynamics – Wonder Dynamics האט לאנסירט א געצייג וואס פארוואנדלט ווידעא פון עטליכע קאמערעס אויטאמאטיש אין א פולע 3D סצענע וואס מען קען רעדאגירן.
- **7 נאוועמבער:** Amazon X-Ray Recaps – Amazon באנוצט AI צו מאכן פערזענליכע קיצורים פון טעלעוויזיע-פראגראמען אויף Prime Video.
- **8 נאוועמבער:** Mistral מאדעראציע API – Mistral האט לאנסירט "batch" באארבעטונג און אן API פאר מאדעראציע, אלס טייל פון איר API פלאטפארמע.
- **11 נאוועמבער:** 1.4 טריליאן פאראמעטערס – Near Protocol פלאנירט צו בויען אן אפן-סאורס מאדעל מיט 1.4 טריליאן פאראמעטערס, מיט'ן ציל צו איבערשטייגן מאדעלן ווי Meta'ס Llama.
- **11 נאוועמבער:** Bixby און Notepad – Samsung האט לאנסירט א "נייע דור Bixby" מיט שטערקערע AI, אויף די Galaxy W25 און W25 Flip אין כינע. Microsoft לייגט צו AI אין Notepad, מיט א פונקציע "Rewrite" וואס פארמולירט איבער טעקסט, ענדערט דעם טאן און מאכט אים לענגער אדער קירצער.

- **12 נאוועמבער: Qwen – Qwen 2.5 Coder 32B** – נייער מאדעל פון 32 ביליאן פאראמעטערס, ספעציעל טריינירט פאר קאוד, איז בעסער פון GPT-4o. ער איז אבער קליין גענוג אז אסאך מענטשן קענען אים לויפן אויף זייער אייגענעם קאמפיוטער, און אין אסאך בענטשמארקס שטייט ער גלייך מיט Claude Sonnet 3.5.
- **12 נאוועמבער: פארויסזאגן פארפלייצונגען** – Google's AI פאר פארויסזאגן פארפלייצונגען ברייט זיך אויס אויף מער ווי 100 לענדער, און דעקט צו 700 מיליאן מענטשן.
- **12 נאוועמבער: AI אגענטן וואס קאנטראלירן דעם קאמפיוטער** – Google's Jarvis פראטאטיפ, וואס קען איבערנעמען א קאמפיוטער, איז צופעליג אויפגעקומען אויף דעם Chrome Web Store. לויט Bloomberg גרייט OpenAI ארויסצולאזן שוין אין יאנואר "Operator", אן AI אגענט וואס טוט ארבעטן גלייך אויפ'ן באניצער'ס קאמפיוטער. עס וועט צוערשט זיין אלס פארשונג-פארשטעלונג דורך דעם API פאר פראגראמירער.
- **13 נאוועמבער: Baidu's ברילן** – Baidu האט לאנסירט AI ברילן מיט איר ERNIE מאדעל, און אויך א בילד-מאכער און א געצייג צו שאפן ווייכווארג אן קיין קענטעניש אין פראגראמירן. לויט חודש'ליכע באניצער איז ByteDance's Doubao דער פירנדער טשעטבאט אין כינע.
- **15 נאוועמבער: Google Learn About** – Google's זייטל "Learn About" גיט ענטפערס מיט בילדער און אינטעראקטיווע עלעמענטן, אין א פארמאט פאר לערנען.
- **19 נאוועמבער: Pixtral Large** – Mistral האט ארויסגעלאזט Pixtral Large, א מולטימאדאלער מאדעל פון 124 ביליאן פאראמעטערס מיט אפענע "וואגן" (weights), געבויט אויף Mistral Large 2. ער פארשטייט בילדער אויפ'ן העכסטן שטאפל, אריינגערעכנט דאקומענטן און גראפיקן, און בלייבט דערביי שטארק אין טעקסט.
- **19 נאוועמבער: Perplexity איינקויפן** – Perplexity האט צוגעלייגט אן איינקויף-פונקציע פאר Pro באניצער אין אמעריקע, וואו מען קען פארשן א פראדוקט און אים גלייך קויפן.
- **20 נאוועמבער: Gemini אויף iPhone, DeepL Voice** – Google האט ארויסגעלאזט א באזונדערע Gemini אפליקאציע פאר DeepL Voice. iPhone האט לאנסירט DeepL Voice, וואס איבערזעצט רייד און ווידעא בשעת'ן רעדן.
- **20 נאוועמבער: Genies Parties** – Genies האט לאנסירט Parties, א פלאטפארמע וואו פירמעס און שאפער קענען בויען זייער אייגענע "AI Roblox" מיט AI פיגורן.
- **21 נאוועמבער: DeepSeek r1** – DeepSeek האט נאכגעטון o1 מיט איר מאדעל r1 Deep Think, וואס איז טריינירט מיט reinforcement learning אויף לאגישע געדאנקען-שריט. דער מאדעל איז זייער שטארק, און DeepSeek פלאנירט אים אינגאנצן צו מאכן אפן-סאורס.
- **21 נאוועמבער: Messenger** – Meta לייגט צו נייע פונקציעס אין Messenger, ווי HD ווידעא-קאלס, AI הינטערגרונטן און AI וואס פארמינערט גערודער.
- **22 נאוועמבער: Flux Tools** – Black Forest Labs האט לאנסירט FLUX.1 Tools, א סעריע מאדעלן וואס גיבן מער קאנטראל איבער'ן FLUX.1 בילד-מאדעל. מען קען דערמיט אויספילן טיילן פון א בילד, אים פארברייטערן אויף די זייטן, און באנוצן עטליכע ControlNets.
- **22 נאוועמבער: ElevenLabs** – ElevenLabs האט לאנסירט א פלאטפארמע צו בויען AI אגענטן וואס שמועסן. מען קען שטעלן דעם טאן און די לענג פון די ענטפערס, און אריינלייגן אן אייגענעם

- שפראך-מאדעל. איר נייע פונקציע GenFM, אין דער iOS אפליקאציע, מאכט פאדקאסטן מיט עטליכע רעדנער פון אינהאלט וואס מען לאדט ארויף, אין 32 שפראכן. עס איז א קאנקורענט צו NotebookLM.
- **25 נאוועמבער: OpenScholar – Ai2** האט פארגעשטעלט OpenScholar, א מאדעל וואס זוכט פאסיגע פארשונג-ארבעטן און שרייבט ענטפערס באזירט אויף זיי, כדי צו העלפן פארשער געפינען און צוזאמשטעלן וויסן.
 - **27 נאוועמבער: Claude סטיילס – Anthropic** ערלויבט יעצט אייגענע סטיילס, אזוי אז באניצער קענען צופאסן Claude'ס ענטפערס צו זייער אופן פון רעדן און ארבעטן.
 - **27 נאוועמבער: Sora ארויסגעליקט** – א גרופע האט ארויסגעליקט צוטריט צו OpenAI'ס ווידעא-מאדעל Sora, וואס איז נאך נישט ארויס. זיי האבן דאס געטון אלס פראטעסט קעגן וואס זיי רופן אומישרדיגע פירונג און "art washing". מען האט ארויפגעשטעלט אן אינטערפעיס אויף Hugging Face, און OpenAI האט עס לויט די באריכטן פארמאכט נאך עטליכע שעה.
 - **27 נאוועמבער: OLMo 2 – Allen AI** האט ארויסגעלאזט OLMo 2, מאדעלן פון 7 און 13 ביליאן פאראמעטערס וואס זענען אינגאנצן אפן: די טריינינג-דאטא, דער קאוד און אלעס אנדערש. Ai2 טענהט אז זיי זענען בעסער פון ענליכע מאדעלן ווי Meta'ס Llama 3.1. זיי זענען צוגענגליך אונטער Apache 2.0 אויך פאר געשעפטליכע באנוץ.
 - **27 נאוועמבער: Nvidia Fugatto – Nvidia** האט געוויזן Fugatto, אן אודיא-מאדעל פון בערך 2.5 ביליאן פאראמעטערס. ער מאכט מוזיק, ענדערט שטימען און שאפט נייע קלאנגען לויט טעקסט-באפעלן, און קען מישן אודיא, מוזיק און רייד אויף נייע אופנים.
 - **28 נאוועמבער: Alibaba'ס QwQ-32B – Qwen (Alibaba)** האט ארויסגעלאזט QwQ-32B-Preview, איר ערשטער "טראכט-מאדעל", מיט 32 ביליאן פאראמעטערס. ער ארבעט ענליך צו OpenAI'ס o1 מאדעלן און איז ספעציעל שטארק אין מאטעמאטיק און קאוד. די פירמע האט באמערקט אז דער מאדעל ווערט בעסער ווען ער טראכט לענגער.
 - **28 נאוועמבער: Rabbit Teach mode – Rabbit R1** האט געעפנט "teach mode" פאר אלע באניצער. דערמיט קען מען לערנען די AI אויטאמאטיש טון ארבעטן אויף יעדן וועבזייטל. Rabbit פלאנירט א מארק פאר אויטאמאציעס וואס באניצער שאפן.

פארשונג און טעכניק

- **1 נאוועמבער: Meta רובאטיק – Meta FAIR** האט ארויסגעלאזט ארבעט אויף אנריי-חושים פאר רובאטן, באזירט אויף זען, און נייע אנריי-סענסארן. Meta ארבעט צוזאמען מיט GelSight און Wonik Robotics צו פארקויפן די סענסארן פאר AI פארשונג, אויף א רובאט-האנט וואס "פילט".
- **1 נאוועמבער: פארשונג-ארבעטן – ODRL** איז דער ערשטער סטאנדארטיזירטער בענטשמארק פאר reinforcement learning אין סביבות מיט אנדערע דינאמיק. א נייער אדאפטער לאזט טעקסט-צו-בילד מאדעלן מאכן מעמע-ווידעאס און זיי פארלירן נישט זייער אלגעמיינע מעגליכקייטן. AnomalyNCD קלאסיפיצירט אנאמאליעס אין אינדוסטריעלע סביבות.
- **4 נאוועמבער: רעקאנסטרוקציע און לאנגער קאנטעקסט – NoPoSplat** בויט 3D סצענעס פון ווייניג בילדער אן גענויע קאמערע-פאזיציעס. א שטודיע פארשט ווי שפראך-מאדעלן הענטלעך באפעלן וואס

אנטהאלטן אומרעלעוואנטע אדער פאלשע לאגישע שטיקלעך. LongPPL איז א נייע מאס וואס ווייזט בעסער ווי שטארק א מאדעל איז מיט לאנגן קאנטעקסט, ווייל די אלטע מאס, perplexity, נעמט א דורכשניט פון אלע טאקענס און פארפעלט די וויכטיגע.

• **4 נאוועמבער: Project Sid – Project Sid** ווייזט ווי AI אגענטן אנטוויקלען א ציוויליזאציע: ספעציאליזאציע, רעגירונג, און דאס אויפקומען און פארשפרייטן פון מעמעס און רעליגיע. דאס איז מעגליך דורך Altera'ס נייע ארכיטעקטור PIANO.

• **5 נאוועמבער: בילדער און ערקלערונגען** – RAR מאכט בילדער דורך א צופעליגע autoregressive סדר, און דערגרייכט דעם בעסטן רעזולטאט אויף ImageNet-256 מיט אן FID פון 1.48. "B-cosification" פאסט צו עקזיסטירנדע מאדעלן אזוי אז זיי גיבן קלארע ערקלערונגען פאר זייערע פארויסזאגן. Gretel האט ארויסגעלאזט GLINER מאדעלן און דאטאסעטן פאר דערקענען פערזענליכע אינפארמאציע (PII).

• **6 נאוועמבער: GUI אגענטן און ווידעא** – OS-ATLAS איז אן אפענער מאדעל וואס טוט אקציעס אויף א קאמפיוטער. ער איז טריינירט אויף מיליאנען סקרינשאטס פון פארשידענע מכשירים, און איז בעסער פון כמעט אלע אפענע מאדעלן און אסאך פארמאכטע אין נאוויגירן אינטערפעיסעס. MV Paint מאכט צוזאמען טעקסטור און UV פאר 3D אביעקטן. PPLaVA פארשטייט סיי קורצע סיי לאנגע ווידעאס.

• **6 נאוועמבער: Waymo EMMA – Waymo** באנוצט Google'ס Gemini צו בויען EMMA, א מולטימאדאלער מאדעל פאר אויטאנאמישע פארן פון אנהייב ביזן סוף. ער פאראייניגט סענסאר-דאטא כדי בעסער פאראויסצוזאגן דעם וועג און דערקענען אביעקטן. מען דארף נאך מער פארשונג צו פארבעסערן און פארגרעסערן אזעלכע מאדעלן.

• **7 נאוועמבער: מאדעלן און אגענטן** – V-DPO פארמינערט "האלוצינאציעס" (אויסגעטראכטע ענטפערס) ביי מאדעלן פון זען און שפראך, וואס פארלאזן זיך צו שטארק אויף שפראך. Agora איז א פשוט'ער פראטאקאל דורך וועלכן פארשידענע AI אגענטן רעדן איינער מיט'ן צווייטן פאר א קליינעם טייל פון די קאסטן, און מען קען אים פארבינדן מיט Camel AI, LangChain און SuperClass. Swarm קלאסיפיצירט בילד מיט טעקסט און א טעקסט-ענקאדער, אנדערש ווי CLIP.

• **7 נאוועמבער: Google אוידיא** – Google האט פארבעסערט איר שטימע-טעכנאלאגיע פאר נאטירליכע שמועסן מיט עטליכע רעדנער. מאדעלן ווי SoundStream און AudioLM שטייען הינטער NotebookLM און Illuminate, מיט שנעלן אויספיר און וואסער-צייכנס כדי מען זאל עס נישט מיסברויכן.

• **8 נאוועמבער: רובאטן לערנען** – פארשער פון UC Berkeley און ETH Zurich האבן באנוצט GPT-4 צו לערנען ביליגע רובאט-ארעמס אפוישן פארגאסענע זאכן, מיט אפן-סאורס טעכנאלאגיע. MIT האט אנטוויקלט HPT, וואס טריינירט רובאטן מיט פארשידענע דאטא פון אסאך געביטן, און פארבעסערט די רעזולטאטן מיט מער ווי 20%.

• **8 נאוועמבער: Magentic One, Cosmos טאקענייזער** – Microsoft האט ארויסגעלאזט א סיסטעם פון עטליכע אגענטן, געבויט אויף איר Autogen ביבליאטעק, וואס באנוצט GPT-4o. Nvidia האט ארויסגעלאזט שפיץ טאקענייזער- און קאמפרעסיע-מאדעלן פאר בילד און ווידעא, מיט קאוד, א פארשונג-ארבעט און וואגן. א נייע ארבעט ווייזט אז מאדעלן פון זען און שפראך קענען לערנען נייע פעאגיקיטן פון ביישפילן אינעם באפעל אליין.

- **8 נאוועמבער: 3D און זען** – SA3DIP פארבעסערט 3D סעגמענטאציע מיט מער 3D אינפארמאציע. RaVL פארבעסערט מאדעלן פון זען און שפראך דורך זיך קאנצענטרירן אויף לאקאלע טיילן פונעם בילד. SCGaussian האלט די סטרוקטור קאנסיסטענט ביים שאפן 3D סצענעס פון ווייניג בילדער.
- **11 נאוועמבער: FrontierMath – Epoch AI** האט ארויסגעלאזט FrontierMath, א שווערער מאטעמאטיק בענטשמארק מיט אינגאנצן נייע פראבלעמען. רוב פירנדע מאדעלן, אריינגערעכנט GPT-4o, לייזן ווייניגער ווי 2%.
- **11 נאוועמבער: LLM2CLIP, BitNet a4.8** – Microsoft's LLM2CLIP פאראייניגט CLIP מיטן שפראך-פארשטאנד פון גרויסע שפראך-מאדעלן. BitNet a4.8 שלאגט פאר 4-ביט אקטיוואציעס פאר 1-ביט מאדעלן, כדי צו באנוצן INT4/FP4 אין נייע הארדווער, אן קיין עקסטערע קאסטן ביים לויפן.
- **12 נאוועמבער: Mixture-of-Transformers – MoT** צעטיילט די פאראמעטערס לויט מאדאליטעט (טעקסט, בילד, רייד), און באקומט די זעלבע אדער בעסערע רעזולטאטן מיט באדייטנד ווייניגער רעכנ-קראפט. StdGen שאפט 3D פיגורן פון איין בילד דורך צעטיילן זיי אין טיילן. א שטודיע פארבעסערט די פארבינדונג צווישן שפראך-מאדעלן און פראטעיין-מאדעלן.
- **12 נאוועמבער: טויזנטער טויזנטער טאקענס** – א פארשונג אויף 17 פירנדע מאדעלן געפינט אז דער קאנטעקסט וואס זיי קענען אין דער אמת באנוצן איז אפט קירצער ווי זייער אפיציעלע מאקסימום. פילע מאדעלן קענען אבער גוט נאכגיין עטליכע פאדעמען פון אינפארמאציע אויפאמאל.
- **13 נאוועמבער: EfficientNAT, UniGAD** – EfficientNAT מאכט שנעלער און בעסער בילדער באזירט אויף טאקענס. UniGAD געפינט אנאמאליעס אין גראפן אויף אלע שטאפלען אויפאמאל. א ווידעא באהאנדלט די o1 מאדעלן, ווי מען קען זיי נאכטון און וויפיל נוצן זיי וועלן ברענגען.
- **14 נאוועמבער: HiCoM, Common Corpus** – HiCoM פארבעסערט 3D רעקאנסטרוקציע אין רעאלער צייט פון ווידעאס. א נייע מעטאד איבערטראגט א פויק-סטיל אויף יעדן ריטם אן ספעציעלן טריינירן. Common Corpus איז דער גרעסטער אפענער דאטאסעט פאר טריינירן שפראך-מאדעלן, מיט צוויי טריליאן טאקענס און פילע שפראכן.
- **15 נאוועמבער: לאגישע טויערן, Athene-V2** – פארשער האבן געוויזן אז מען קען טריינירן א נעץ פאר דערקענען בילדער וואס איז 29 מאל קלענער ווי א CNN, דורך מאכן לאגישע טויערן "differentiable". PID פארשנעלערט diffusion מאדעלן. Athene V2, געבויט אויף Qwen 2.5 72B, איז בעסער פון GPT-4o אין עטליכע בענטשמארקס פאר אגענטן און שמועס.
- **15 נאוועמבער: 360 גראד, X Portrait 2 – MVSPlat360** שאפט 360-גראד בליקן פון ווייניג בילדער. X Portrait 2's ByteDance פארבעסערט ווי רעאליסטיש און קאנסיסטענט עס איבערטראגט אן אקטיאר'ס באוועגונגען צו ווידעא. די בעסטע לייזונג אין דער BraTS קאנקורענץ פאר מוח-גידולים ווייזט אז סינטעטישע דאטא קען פארבעסערן מעדיצינישע מאדעלן.
- **18 נאוועמבער: Llama Mesh** – Nvidia האט צוגעפאסט Llama צו שאפן 3D מאדעלן פון טעקסט. פארשער שטעלן פאר "Semantic Perturbation Attack" קעגן וואסער-צייכן סיסטעמען. א שטודיע ווייזט אז ביים "Travelling Salesman" פראבלעם קענען פשוט'ע heatmaps מיט MCTS זיין בעסער ווי קאמפליצירטע מאדעלן.

- **18 נאוועמבער: MIT גראף-מאדעל** – א פראפעסאר אין MIT האט אנטוויקלט א מאדעל וואס פאראייניגט AI מיט לאגיק אויף גראפן. ער מאכט מאפעס פון וויסן פון פארשונג-ארבעטן און געפינט פארבינדונגען צווישן ביאלאגיע, מוזיק און קונסט, און שלאגט פאר נייע מאטעריאלן.
- **19 נאוועמבער: Claude Computer Use** – א שטודיע פרובירט סיסטעמאטיש אגענטן וואס באנוצן א קאמפיוטער, און געפינט אז Claude איז שטארק אין לינעארע ארבעטן. LLaVA-o1 איז א שטארקער מאדעל פון זען און שפראך, טריינירט אויף לאגישע שריט. NumPro העלפט ווידעא-מאדעלן פארשטיין צייט.
- **19 נאוועמבער: Recapture** – Google ווייזט ווי מען קען ענדערן דעם קאמערע-ווינקל פון אן עקזיסטירנדן ווידעא.
- **20 נאוועמבער: OneNet, Genex** – OneNet איז אן עפעקטיווער 1D ענקאדער פאר סעגמענטאציע אויף קליינע מכשירים. א ארבעט באטראכט ווי מען קען אריינלייגן באשטעטיגונג אין טריינירן נאכ'ן הויפט-טריינירן, כדי צו פארבעסערן לאגיק. Genex שטעלט זיך פאר ווי ער פארשט 3D ערטער דורך שאפן בילדער, און באנוצט דאס צו פלאנירן בעסער.
- **20 נאוועמבער: רובאטן און כירורגיע** – LucidSim מאכט ווירטועלע סביבות מיט AI, און רובאט-הינט וואס זענען דערמיט טריינירט זענען בעסער אין פארקור און מניעות. פארשער אין Johns Hopkins האבן אנטוויקלט א מאדעל וואס לערנט כירורגישע רובאטן אפעראציעס דורך קוקן ווידעאס פון כירורגן.
- **21 נאוועמבער: AlphaChip, AlphaQubit** – DeepMind ענטפערט שטארק אויף די קריטיק פון אקאדעמיקער אויף AlphaChip, וואס ווערט שוין ברייט באנוצט אין דער אינדוסטריע. DeepMind און Google Quantum האבן טריינירט AlphaQubit, וואס געפינט טעותים אין קוואנטום-רעכענונגען און פאררעכט זיי.
- **21 נאוועמבער: SAM-Decoding, Gemma Scope** – SAM-Decoding פארשנעלערט שרייבן טעקסט מיט א "suffix automaton". "Additional Logic Training" פארבעסערט לאגיק ביי שפראך-מאדעלן. DeepMind's Gemma Scope באנוצט sparse autoencoders צו זען ווי א מאדעל באשליסט, כדי צו געפינען טעותים און פאראורטיילן.
- **21 נאוועמבער: ChatGPT ביי דאקטוירים** – א שטודיע האט געפונען אז ChatGPT-4 האט דיאגנאזירט קראנקהייטן בעסער ווי דאקטוירים. די דאקטוירים האבן אבער נישט גוט געוואוסט ווי צו באנוצן די AI, און האבן זיך אפט פארלאזט אויף זייער אייגענעם מיינונג.
- **21 נאוועמבער: PanoRadar, FastGlioma** – PanoRadar פון University of Pennsylvania פארוואנדלט ראדיא-כוואליעס אין 3D בילדער פאר רובאטן, ביליגער ווי LiDAR, און עס ארבעט אפילו אין רויך און נעפל. FastGlioma געפינט איבערגעבליבענע מוח-גידולים בשעת אן אפעראציע אין בערך 10 סעקונדעס.
- **22 נאוועמבער: סטאטיסטיק און lu 3üT** – Anthropic שלאגט פאר סטאטיסטישע מעטאדן צו באשטימען צי איין מאדעל איז טאקע בעסער פון א צווייטן אויף א בענטשמארק. lu 3üT איז אן אפענער רעצעפט פארן טריינירן נאכ'ן הויפט-טריינירן, אויפ'ן שפיץ שטאפל, מיט דאטא, קאוד און וואגן.
- **22 נאוועמבער: פארשידענע פארשונג** – DCSR באנוצט diffusion כדי צו לייען דאס "קאלטער אנהייב" פראבלעם ביי טעסטן וואס פאסן זיך צו צום תלמיד. גרויסע מולטימאדאלע מאדעלן קענען אויפאמאל באשרייבן בילדער און זיי קאמפרעסירן. PCNet רעקאנסטרוקטירט בילדער פון ווייניגער דאטא.

- **25 נאוועמבער: UnifiedCrawl, HAD** – UnifiedCrawl זאמלט טעקסט פאר שפראכן מיט ווייניג מאטעריאל פון Common Crawl, מיט ווייניג רעכן-קראפט. HAD איז א דאטאסעט פון מער ווי 37,000 בילדער מיט צעקרימטע מענטשן, כדי צו פאררעכטן דעם פראבלעם. נייע מאדעלן מאכן רעאליסטישע ווידעא פון א בילד.
- **26 נאוועמבער: OpenAI red teaming** – OpenAI האט ארויסגעגעבן צוויי פארשונג-ארבעטן איבער "red teaming". איינע באשרייבט ווי מען ארבעט מיט עקספערטן פון דרויסן, און די אנדערע א נייע אויטאמאטישע מעטאד פון טעסטן.
- **26 נאוועמבער: WildLMA, Coalescence** – WildLMA לאזט רובאטן אויף פיר פיס טון קאמפליצירטע ארבעטן אין דער אמת'ער וועלט, מיט א פלאנירער באזירט אויף א שפראך-מאדעל. Coalescence מאכט אויספיר פון סטרוקטורירטע פארמאטן ווי JSON ביז 5 מאל שנעלער. MMGenBench פרובירט מולטימאדאלע מאדעלן אין שאפן און פארשטיין בילדער.
- **27 נאוועמבער: OneDiffusion** – א גרויסער diffusion מאדעל פאר פילע ארבעטן: שאפן בילדער, רעדאגירן, און אויך פארקערטע ארבעטן ווי מעסטן טיפקייט און סעגמענטאציע.
- **28 נאוועמבער: Star Attention** – Star Attention פארשנעלערט לאנגע סעקווענצן מיט "block-sparse attention", ביז 11 מאל שנעלער. עס איז אביסל בעסער ווי ring attention אבער שוואכער ווי פולע attention.
- **28 נאוועמבער: קוואנטיזאציע** – א שטודיע געפינט אז מאדעלן וואס זענען ווייניגער טריינירט ליידן ווייניגער ווען מען קוואנטיזירט זיי (רעדוצירט די פרעציזיע) נאכ'ן טריינירן. א נייע מעטאד רעדאגירט א בילד שריט ביי שריט דורך ווידעא. MIT האט אנטוויקלט אן עפעקטיווע אלגאריטם וואס קלייבט אויס טריינינג-ארבעטן פאר מער פארלעסליכע אגענטן.
- **29 נאוועמבער: DiffusionDrive, ThunderKittens** – PassionSR מאכט מאדעלן פאר בילד-פארגרעסערונג גרינגער פאר הארדווער. DiffusionDrive פירט אויטאנאמישע קארס מיט 45 בילדער א סעקונדע אויף א Hazy Research 4090. האט אויסגעברייטערט ThunderKittens אויף Apple Silicon.

געצייג פאר פראגראמירער און אפן-סאורס

- **1 נאוועמבער: BART, MrT5, Diffusers** – א GPU ווערסיע פון BART איז ביז 200 מאל שנעלער ווי אויף CPU. MrT5 פאראייניגט טאקענס דינאמיש פאר שפראך-מאדעלן וואס ארבעטן אויף ביטס. Hugging Face Diffusers האט א גרויסע נייע ווערסיע מיט Flux, Stable Audio, Kolos, CogVideoX, Latte, FreeNoise און SparseCtrl.
- **4 נאוועמבער: ImOV3D, DEMO** – ImOV3D דערקענט 3D אביעקטן מיט 2D בילדער. א פראיעקט פארשט ווי שנעלע און לאנגזאמע "טראכטן" ווירקט אויף מאדעלן. DEMO צעטיילט אינהאלט און באוועגונג פאר בעסערע באוועגונג אין ווידעא.
- **5 נאוועמבער: Hertz Dev** – Hertz Dev איז א שטארקער אפן-סאורס מאדעל פון רייד צו רייד, מיט א פרייע לייסענס און קוואליטעט קאוד. אויך: ML מיט ray tracing פאר מאדעלירן ראדיא-כוואליעס, און GQT, א טאקעניזער פאר גראפן.

- **6 נאוועמבער: Hunyuan3D – Tencent** ס'Hunyuan3D-1.0 שאפט 3D פון טעקסט און בילדער. עס איז נאך נישט גרייט פאר פראדוקציע, אבער עס איז א גוטער אנהייב פאר קינסטלער. Android Lab איז א בענטשמארק פאר אגענטן אויף CHAIC. Android. פרוברט צי AI אגענטן קענען העלפן מענטשן מיט פיזישע באגרעניצונגען.
- **7 נאוועמבער: ADOPT, HtmlRAG** – ADOPT איז א ברירה צו Adam וואס לייזט די פראבלעמען פון נישט-קאנווערגענץ. HtmlRAG פארבעסערט RAG דורך באהאלטן די HTML סטרוקטור פון וועב-אינהאלט. LiVOS טיילט אפ אביעקטן אין לאנגע ווידעאס מיט ווייניגער זכרון.
- **8 נאוועמבער: gsplat** – א פעקל און סטודיא פאר פארשונג אויף Gaussian splatting.
- **11 נאוועמבער: Muon, Mochi, Protenix** – Muon, Mochi, Protenix ווערסיע פון Muon, דער אפטימיזער פונעם GPT-2 רעקארד, וואס מען קען באנוצן אנשטאט AdamW. Mochi, דער בעסטער אפן-סאורס ווידעא-מאדעל, האט אמאל געדארפט H100 8, און יעצט לויפט ער אויף איין L40 פון 48GB אן צו פארלירן קוואליטעט. Protenix ס'ByteDance איז א טריינירבארע נאכמאכונג פון AlphaFold3.
- **11 נאוועמבער: LlamaPreview, SmolLM2** – LlamaPreview קוקט איבער pull requests אויף GitHub אויטאמאטיש און אומזיסט. SmolLM2 ס'Hugging Face איז א משפחה פון קליינע מאדעלן פון 135 מיליאן ביז 1.7 ביליאן פאראמעטערס, טריינירט אויף 11 טריליאן טאקענס, אונטער Apache 2.
- **12 נאוועמבער: AlphaFold 3** אפן – DeepMind האט אפן געמאכט דעם קאוד און די וואגן פון AlphaFold 3 פאר פארשונג. אקאדעמישע פארשער קענען יעצט מאדעלירן ווי פראטעינען ווירקן אויף מעדיצינען. געשעפטליכע באנוץ איז נאך באגרעניצט, און פראיעקטן ווי OpenFold3 ארבעטן אויף א אינגאנצן אפענער ווערסיע. Baidu און ByteDance האבן ענליכע געצייג.
- **12 נאוועמבער: Online-LoRA, Hermes 3** – Online-LoRA פארמיידט "פארגעסן" ביים לערנען אן אויפהער. DeepArUco++ דערקענט מארקערס אין שווער ליכטיגקייט. Hermes 3 ס'Nous Research, געבויט אויף Llama 3.1 מיט 8, 70 און 405 ביליאן פאראמעטערס, איז שטארק אין לאגיק און שעפערש שרייבן.
- **13 נאוועמבער: BPT, Browser-Use** – BPT קאמפרעסירט 3D meshes מיט Successor Feature 75%. Matching לערנט פון עקספערטן אן לערנען א "reward" פונקציע. Token Merging פארבינדט אביעקטן מיט זייערע אייגנשאפטן אין בילדער. DataChain איז א Python ביבליאטעק פאר אומסטרוקטורירטע דאטא. Browser-Use לאזט שפראך-מאדעלן ווי GPT-4o און Claude 3.5 Sonnet ארבעטן מיט וועבזייטלעך.
- **13 נאוועמבער: CUDA קורס און מער** – א קורס פון 12 שעה איבער CUDA פון freeCodeCamp. Context Autopilot איז א פאראייניגטער AI ארבעט-פלאץ. 1Price ענדערט פרייזן אויטאמאטיש מיט AI.
- **14 נאוועמבער: Janus Flow** – Janus Flow ס'DeepSeek באנוצט rectified flows סיי פאר שאפן סיי פאר פארשטיין בילדער, מיט זייער שטארקע רעזולטאטן. אויך: POSR פארבינדט טיילן פון א שמועס מיט מאטעריאלן, א טעכניק צו אריינשטעלן אביעקטן אין נייע בילדער, StoryTeller באשרייבט לאנגע ווידעאס, און SAMPart3D טיילט 3D אביעקטן אויף אין טיילן.

- **15 נאוועמבער:** **MinerU, XiYan-SQL** – MinerU נעמט ארויס אינפארמאציע פון PDF'ס. אויך: איזאטאנישע רעגערעסיע, און XiYan-SQL וואס מאכט SQL פון נאטירליכע שפראך.
- **15 נאוועמבער:** **FLORA – FLORA** ברענגט אלע פיר בעסטע מאדעלן פון בילד צו ווידעא (Runway, Luma, Kling, Hailuo) אין איין געצייג.
- **18 נאוועמבער:** **RLT און UltraVox** – Run Length Tokenization פארשנעלערט ווידעא-מאדעלן אן נאך טריינירן. MOP+MiHo+NCC פארבעסערט צופאסן בילדער. א טעכניק פאר קאנטראלירן געגנטן ביי מאכן בילדער. UltraVox איז מאדעלן מיט אפענע וואגן פאר טעקסט און אודיא. Pygen מאכט Python פעקלעך פון אן אידעע.
- **19 נאוועמבער:** **Cerebras Llama 405B** – Cerebras לויפט Llama 3 405B 12 מאל שנעלער ווי ענליכע סיסטעמען, בערך 1,000 טאקענס א סעקונדע. דאס וועט זיין צוגענגליך דורך API אנהייב נעקסט יאר. Forge API's Nous Research לייגט צו לאגיק און א קאוד-אינטערפרעטער צו יעדן מאדעל. אויך: Trident פאר סעגמענטאציע מיט CLIP.
- **20 נאוועמבער:** **StableV2V – CCExpert** באשרייבט ענדערונגען אין סאטעליט-בילדער. StableV2V רעדאגירט ווידעא און האלט די פארעמען קאנסיסטענט. קאוד צו אפשאצן "קאנסטיטוציעס" לויט מענטשליכע פידבעק.
- **21 נאוועמבער:** **Polynomial Mixer** – א שנעלערע ברירה צו Multi-Head Attention אין diffusion מאדעלן, און א סיסטעם פאר געפינען וואו א בילד איז גענומען דורך פארגלייכן בליקן פון פארשידענע ווינקלען.
- **22 נאוועמבער:** **SAMURAI – SAMURAI** בויט אויף SAM 2 און לייגט צו א זכרון וואס פארשטייט באוועגונג. עס איז איינס פון די בעסטע סיסטעמען פאר נאכגיין אביעקטן אין ווידעא.
- **25 נאוועמבער:** **AIMv2, LTX-Video** – AIMv2 זענען שטארקע מאדעלן פאר זען. GLMix פאראייניגט קאנוואלוציעס מיט AnchorAttention. self-attention פאררעכט פראבלעמען פון RoPE מיט BFloat16. Echo Mimic v2 באלעבט טיילן פון א מענטש מיט אודיא. LTX-Video איז דער ערשטער ווידעא-מאדעל וואס שאפט ווידעא שנעלער ווי מען קען עס קוקן. Documind נעמט ארויס דאטא פון PDF'ס מיט OpenAI'ס API.
- **26 נאוועמבער:** **Anthropic – Model Context Protocol** האט פארגעשטעלט MCP, אן אפענער סטאנדארט צו פארבינדן AI סיסטעמען גלייך מיט דאטא-מקורות ווי געשעפט-געצייג און אינהאלט-ארכיוון. עס ערזעצט צעשטרייטע אייגענע פארבינדונגען מיט איין אלגעמיינעם פראטאקאל.
- **26 נאוועמבער:** **Moondream, Marco-o1, SANA** – Moondream'ס Python ביבליאטעק אנאליזירט בילדער אויף CPU. Marco-o1 פאראייניגט MCTS, chain-of-thought און רעפּלעקציע פאר פראבלעמען אן א קלארן ענטפער. SANA שאפט בילדער פון 1024x1024 אין ווייניגער ווי א סעקונדע אויף א לעפטאפּ GPU, און דער מאדעל פון 0.6 ביליאן פאראמעטערס איז גלייך אדער בעסער ווי Flux-12B.
- **27 נאוועמבער:** **Mochi פיינ-טונינג** – Mochi 1 שטיצט יעצט LoRA פיינ-טונינג אויף איין GPU. אויך: א בענטשמארק פאר שאפן 3D מאלעקולן, א Flash Attention ווערסיע מיט אומגלייכע גרייסן, MVGenMaster פאר נייע בליקן, און AI Suite פאר ארבעטן מיט פארשידענע API מאדעלן.

- **28 נאוועמבער:** SketchAgent, ShowUI – SketchAgent, CUI שריט ביי שריט לויט שפראך. א SLAM סיסטעם מאכט 3D אין רעאלער צייט. P2DFlow שאפט גרופעס פון פראטעין-פארעמען. ShowUI-2B איז א מאדעל וואס זעט, פארשטייט און טוט אקציעס אויף אינטערפעסעס.
- **29 נאוועמבער:** Open Instruct, PixMo – Allen AI האט ארויסגעלאזט קאוד פאר instruction tuning און די PixMo דאטאסעטן. אויך: Grounding-IQA פאר אפשאצן בילד-קוואליטעט, StableAnimator, און Steel Browser API פאר AI אגענטן.

פירמעס, געלט און ביזנעס

- **1 נאוועמבער:** OpenAI'ס איינקונפט – לויט OpenAI'ס פינאנץ-דירעקטאר קומען 75% פון דער איינקונפט פון באצאלטע ChatGPT באניצער. ChatGPT האט 250 מיליאן וועכנטליכע באניצער, און די פירמע האט לעצטנס אויפגעשאפט 6.6 ביליאן דאלאר.
- **1 נאוועמבער:** Meta זוכן, Gemini 2.0 – Meta אנטוויקלט לויט די באריכטן א זוך-מאשין פאר איר טשעטבאט, כדי זיך ווייניגער צו פארלאזן אויף Google און OpenAI. Microsoft. OpenAI פלאנירט א נאכפאלגער צו GPT-4, און Google וויל לאנסירן Gemini 2.0 אין דעצעמבער.
- **1 נאוועמבער:** סטארטאפ געלט-רונדעס – Hello Patient האט אויפגעשאפט 6.3 מיליאן דאלאר, אנגעפירט דורך 8VC. Sapien, וואס מאכט AI פאר פינאנץ-אנאליז, האט באקומען 8.7 מיליאן דאלאר פון General Catalyst. Lume, וואס פארבינדט דאטא אויטאמאטיש, האט באקומען 4.2 מיליאן דאלאר. Rox, א פלאטפארמע פאר פארקויפער וואס פארבעסערט די שפיץ-פארקויפער מיט מער ווי 30%, האט אויפגעשאפט 50 מיליאן דאלאר אנגעפירט דורך Sequoia.
- **4 נאוועמבער:** OpenAI טשיפס – OpenAI ארבעט מיט Broadcom און TSMC אויף אייגענע טשיפס, וואס זאלן אנהייבן פראדוצירט ווערן אין 2026. אינצווישן לייגט זי אריין AMD'ס MI300 טשיפס אין Azure. זי בלייבט אבער הינטער Microsoft, Google, און Amazon אין הארדווער.
- **4 נאוועמבער:** xAI Colossus – xAI'ס Colossus, דער גרעסטער AI סופער-קאמפיוטער אין דער וועלט, לויפט אויף Nvidia'ס Spectrum-X Ethernet.
- **5 נאוועמבער:** Meta און רעקלאמעס – לויט Stratechery שטייט Meta אין א בעסונדערע פאזיציע צו פארדינען פון AI, ספעציעל אין רעקלאמעס, מיט Llama, רעקלאמעס געשאפן פון AI, און טשעטבאט אגענטן.
- **5 נאוועמבער:** ElevenLabs און Omnivore – ElevenLabs האט אריינגענומען די מאנשאפט הינטער Omnivore, כדי צו פארבעסערן איר לייען-אפליקאציע ElevenReader.
- **6 נאוועמבער:** OpenAI רובאטיק – דער געוועזענער הארדווער-פירער פון Meta'ס Orion ברילן קומט צו OpenAI צו פירן רובאטיק און הארדווער פאר קונים. OpenAI באלעבט ווידער איר רובאטיק מאנשאפט.
- **6 נאוועמבער:** Google קאוד, Llama-4 – מער ווי א פערטל פון נייעס קאוד אין Google ווערט געשריבן דורך AI. Meta באנוצט מער ווי 100,000 Nvidia H100 צו טריינירן Llama-4.
- **7 נאוועמבער:** Chat.com – OpenAI האט געקויפט דעם דאמעין Chat.com, וואס פירט יעצט צו ChatGPT. HubSpot'ס מיטגרינדער האט אים פריער געקויפט פאר 15.5 מיליאן דאלאר. מען מיינט אז OpenAI האט באצאלט מיט אקציעס, אבער די פרטים זענען נישט באקאנט.

- **7 נאוועמבער:** Nvidia אין Dow Jones – Nvidia ערזעצט Intel אין דעם Dow Jones אינדעקס. Google בוט אן AI דאטא צענטער אין סאודי אראביע פאר פארשונג אויף אראביש.
- **8 נאוועמבער:** Palantir און Anthropic – Palantir און Anthropic ברענגען Claude אויף AWS פאר אמעריקאנער אינטעליגענץ-אגענטורן און מיליטער.
- **8 נאוועמבער:** OpenAI פאר פראפיט – OpenAI פירט שמועסן מיט רעגולירער צו ווערן א פירמע פאר פראפיט, און דערביי האלטן א נאן-פראפיט ארעם פאר איר אריגינעלער מיסיע.
- **11 נאוועמבער:** Modal קויפט Tidbyt – Modal האט געקויפט די הארדווער פירמע Tidbyt פון ניו יארק, איר ערשטער איינקויף, צוליב דער מאנשאפט'ס קענטעניש אין אינפראסטרוקטור.
- **11 נאוועמבער:** Orion פארלאנגזאמט – לויט באריכטן ווייזט OpenAI'ס קומענדיגער מאדעל "Orion" ווייניגער פארשריט. OpenAI האט געשאפן א מאנשאפט וואס זוכט אנדערע וועגן, ווי סינטעטישע דאטא, ווייל נייע דאטא איז ווייניגער צוגעגליך.
- **13 נאוועמבער:** Brockman קומט צוריק – Greg Brockman איז צוריק אלס פרעזידענט פון OpenAI נאך דריי חדשים אפרו. דאס איז אין א צייט ווען אסאך פירער גייען אוועק, און OpenAI איז לעצטנס געשאצט געווארן אויף 157 ביליאן דאלאר.
- **13 נאוועמבער:** Ecosia און Qwant – Ecosia און Qwant בויען צוזאמען אן אייראפעאישן זוך-אינדעקס, כדי זיך ווייניגער צו פארלאזן אויף אמעריקאנער פירמעס.
- **14 נאוועמבער:** Google אין מיטל-מזרח – Google.org האט לאנסירט איר גרעסטע AI אינוועסטירונג אין דעם ראיאן פון מיטל-מזרח און צפון אפריקע. לויט די שאצונגען קען AI צולייגן 320 ביליאן דאלאר צו דער עקאנאמיע דארט ביז 2030.
- **14 נאוועמבער:** Stackblitz – Stackblitz האט געבויט א פראגראמירער-געצייג אויף Claude Sonnet 3.5, און האט דערגרייכט 4 מיליאן דאלאר יערליכע איינקונפט אין ווייניג חדשים.
- **15 נאוועמבער:** Chollet, Supermaven – Francois Chollet, דער שאפער פון Keras און ARC, איז אוועק פון Google. Supermaven'ס מאנשאפט קומט צו Cursor.
- **18 נאוועמבער:** Spotify – Spotify'ס קא-פרעזידענט זאגט אז די פירמע באנוצט AI ווי Suno און NotebookLM פאר מוזיק און פאדקאסטן, און שפראך-מאדעלן פאר רעקאמענדאציעס. לויט אים העלפט AI שפערישקייט און ערזעצט עס נישט.
- **18 נאוועמבער:** AI וועלפער – Anthropic האט אנגעשטעלט א פארשער פאר "AI וועלפער", וואס פארשט עטישע פראגן איבער א מעגליכן באוואוסטזיין ביי AI.
- **18 נאוועמבער:** Musk קעגן OpenAI – Elon Musk'ס פראצעס קעגן OpenAI האט ארויסגעברענגט אימעילס פון די ערשטע טעג. אין יענע אימעילס פון 2017 האבן Greg Brockman און Ilya Sutskever זיך געטוישט אין Sam Altman'ס כוונות. דער פארבעסערטער פראצעס צייגט אן Microsoft אלס באשולדיגטער, און טענהט אז OpenAI האט פארלאזט איר נאן-פראפיט מיסיע און געשעדיגט קאנקורענטן ווי xAI.
- **20 נאוועמבער:** Meta פאר געשעפטן – Meta שאפט א נייע אפטיילונג פאר AI געצייג פאר די 200 מיליאן געשעפטן וואס באנוצן אירע אפליקאציעס, און האט אנגעשטעלט Salesforce'ס AI פירער.

- **12 נאוועמבער: a16z** – Andreessen Horowitz אינוועסטירט אין AI געצייג וואס העלפן עלטערן, ווי Nanit און Cradlewise.
- **25 נאוועמבער: Amazon און Anthropic** – Amazon אינוועסטירט נאך 4 ביליאן דאלאר אין Anthropic, צוזאמען 8 ביליאן. AWS ווערט Anthropic'ס הויפט שותף פאר קלאוד און טריינינג, און Anthropic ארבעט צו פארבעסערן Trainium.
- **26 נאוועמבער: Nvidia און טעסט-צייט** – Nvidia'ס הויפט-פירער פארטיידיגט די פירמע'ס פאזיציע ווען "טעסט-צייט סקעלינג" ווערט וויכטיגער, טראץ סטארטאפס מיט שנעלע טשיפס פאר אויספיר.
- **26 נאוועמבער: Juna.ai** – די דייטשע סטארטאפ Juna.ai באנוצט AI אגענטן צו מאכן פאבריקן ספארעוודיגער מיט ענערגיע.
- **28 נאוועמבער: o1 טרעידמארק, OpenAI – Cradle**, האט אריינגעגעבן א טרעידמארק פאר o1. Cradle Bio, וואס האט געלייזט א שוועריגקייט ביים אפטימירן פראטעין-פארעמען, האט אויפגעשאפט א Series B.

רעגירונגען, געזעצן און זיכערהייט

- **6 נאוועמבער: Microsoft און a16z** – Microsoft און a16z שטיצן צוזאמענארבעט צווישן גרויסע און קליינע פירמעס, אפן-סאורס AI, און פאליסי-פארשלאגן פאר סטארטאפס.
- **7 נאוועמבער: Big Sleep** – Google'ס AI אגענט Big Sleep האט געפונען א נייעם באג אין SQLite וואס מען קען אויסנוצן. ער האט עס געטון מיט "variant analysis", אויף א אופן ענליך צו א מענטש, נאך וואס fuzzing האט עס פארפעלט.
- **12 נאוועמבער: Anthropic רעגולאציע** – Anthropic שלאגט פאר צילגעריכטע רעגולאציע קעגן מיסברויך אין סייבער-זיכערהייט און כעמישע, ביאלאגישע, ראדיאלאגישע און נוקלעארע געפארן, מיט דורכזיכטיגקייט און א באלאנס מיט חידושים.
- **15 נאוועמבער: OpenAI'ס פלאן** – OpenAI שלאגט פאר AI עקאנאמישע זאנעס, צוזאמענארבעט מיטן נעיווי אויף נוקלעארע ענערגיע, א צפון-אמעריקאנער בונד, און אינוועסטירונגען אין דאטא צענטערס און ענערגיע, כדי צו קאנקורירן מיט כינע.
- **18 נאוועמבער: וויזועלע אינדזשעקציעס** – באפעלן באהאלטן אין בילדער קענען מאניפולירן מאדעלן ווי GPT-4V, און פירמעס דארפן זיך באשיצן דערפון.
- **21 נאוועמבער: RoboPAIR** – RoboPAIR האט אויסגעגאנגען אלע זיכערהייט-שוצן פון רובאטן וואס לויפן מיט שפראך-מאדעלן, מיט 100% הצלחה ביי Go2, א סימולאטאר פון אויטאנאמישע קארס און רובאט-הינט.
- **22 נאוועמבער: Grok און געזונט-דאטא** – באניצער שיקן מעדיצינישע בילדער צו Grok, ווי Musk האט מוטיגט. עקספערטן ווארענען פאר פאלשע דיאגנאזן און פריוואטקייט, ווייל עס איז נישט באשיצט אונטער HIPAA.
- **22 נאוועמבער: ChatGPT סאנדבאקס** – א ארטיקל באשרייבט ווי מען קען אויספארשן ChatGPT'ס Debian סביבה, לויפן Python און מעגליך אנטפלעקן אינערליכע אנווייזונגען. OpenAI באטראכט דאס נישט אלס א זיכערהייט-לאך.

• **27 נאוועמבער: AI מאראל** – OpenAI פינאנצירט פארשונג אין Duke University איבער אלגאריטמען וואס זאגן פאראויס מענטשליכע מאראלישע באשלוסן.

אנאליזן און מיינונגען

• **1 נאוועמבער: LLM אלס ריכטער** – ארטיקלען באשרייבן ווי מען באנוצט א שפראך-מאדעל צו אפשאצן און ראנגירן טעקסט פאר ביזנעס-צוועקן.

• **4 נאוועמבער: RLHF מיט 4.80 דאלאר** – מען האט טריינירט א מאדעל וואס זאגט פאראויס וויפיל שטימען א פאסט אויף Hacker News וועט באקומען, מיט 4.80 דאלאר GPU צייט.

• **5 נאוועמבער: Cursor, Palantir** – א פראגראמירער באשרייבט ווי Cursor מאכט שנעלער boilerplate קאוד, כאטש עס שלאגט אמאל פאר א טעות. 8VC טענהט אז AI פארוואנדלט סערוויס-אינדוסטריעס, לויטן מוסטער פון Palantir.

• **6 נאוועמבער: New York Times** – New York Times רעפארטערס האבן באנוצט AI צו טראנסקריבירן און אנאליזירן מער ווי 400 שעה אודיא, מיט מענטשן וואס קאנטראלירן. Every.to טענהט אז AI קען פארשטערקערן טראכטן, אנשטאט עס ערזעצן.

• **7 נאוועמבער: Octoverse** – GitHub באריכטעט אז Python האט איבערגעשטיגן JavaScript אלס די פאפולערסטע שפראך, צוליב AI, מיט וואוקס אין אפריקע, לאטיין אמעריקע און אזיע.

• **7 נאוועמבער: דיאגראמען** – פראגראמירער באנוצן ChatGPT און Claude מיט Mermaid צו שאפן דיאגראמען.

• **8 נאוועמבער: אגענטישע וועב** – AI אגענטן ווערן פארבונדן מיט בלאקטשעיין. (One) Ethan Mollick (Useful Thing) שרייבט אז מאדעלן ווי Claude ענדערן שוין אינדוסטריעס.

• **11 נאוועמבער: Embeddings** – א ארטיקל טענהט אז embeddings זענען אונטערגעשאצט פאר טעכנישע דאקומענטאציע.

• **14 נאוועמבער: דיפ לערנינג** – א ארטיקל דערציילט ווי ImageNet און AlexNet האבן אנגעהויבן דעם בום. נאך א ארטיקל: AI אגענטן בויען אפליקאציעס פאר קליינע ארבעטן.

• **18 נאוועמבער: סוף פון סקעילינג?** – באריכטן אז Orion איז נישט באדייטנד בעסער, און א מאנגל אין קוואליטעט טעקסט, ברענגען צו סינטעטישע דאטא און ספעציאליזירטע מאדעלן. The Algorithmic Bridge שרייבט אז "GPTs זענען מאקסימום". Gauge טענהט אז AI מאכט טעכנישע חובות טייערער.

• **19 נאוועמבער: רובאט-באטלער** – IEEE שרייבט אז מענטש-ענליכע רובאטן ווי Atlas, Optimus און Neo זענען נאך נישט גרייט פאר שטוב-ארבעט. Google's פארשונג-פירער זאגט אז לערנען פראגראמירן איז נאך אלץ וויכטיג.

• **21 נאוועמבער: שאפער** – א ארטיקל שרייבט אז AI פארמינערט די קאסטן פון שאפן אינהאלט.

• **22 נאוועמבער: AI אינזשעניר** – Helix טענהט אז יעדער קען ווערן אן AI אינזשעניר מיט אפן-סאורס מאדעלן. GitHub Copilot האט מער ווי 2.8 מיליאן באניצער, און מען קען אויסקלייבן מאדעלן פון OpenAI און Google.

- **25 נאוועמבער: Altman סימולאציע** – Fable האט סימולירט דעם OpenAI ברעט-סכסוך מיט GPT-4o אגענטן. Altman איז צוריקגעקומען בלויז אין 4 פון 20 סצענארן.
- **25 נאוועמבער: סקעילינג דעבאטע** – Gwern Branwen באטאנט אז סקעילינג און רעכנ-קראפט זענען מער וויכטיג ווי נייע אלגאריטמען. די Generalist באשרייבט די מחלוקת איבער סקעילינג. New Stack טענהט אז שפראך-מאדעלן אין ווייכווארג קענען זיין א סוף-וועג.
- **26 נאוועמבער: Flow, HumaneRank** – אן איינפירונג אין flow מאדעלן. HumaneRank שלאגט פאר אז מענטשן זאלן פארטיילן עמפעלונגען (endorsements) צווישן זיך, כדי צו באשטימען ווער באקומט וואס אין א וועלט ווו AI איז קלוגער.
- **26 נאוועמבער: שפראך-מאדעלן און שאך** – רוב מאדעלן שפילן שלעכט שאך, אויסער GPT-3.5-turbo-instruct. לויט דעם שרייבער קען דאס זיין צוליב בעסערע שאך-דאטא, און באפעלן מיט ביישפילן פארבעסערן GPT-4o.
- **27 נאוועמבער: ווען מיר ווערן ראדלעך** – AI האט געהעכערט מאטעריאל-אנטדעקונגען מיט 44%, אבער די צופרידנקייט פון די ארבעטער איז געפאלן מיט 44%. IEEE שרייבט אז AI אליין איז נישט גרייט פאר טשיפ-דיזיין, אבער א קאמבינאציע מיט זוכן ווייזט הצלחה.
- **29 נאוועמבער: Gary Marcus** – Gary Marcus טענהט אז סקעילינג-געזעצן ווערן ווייניגער פארלעסליך. Satya Nadella רעדט פון "רעכנ-קראפט ביים ענטפערן", אבער Marcus זאגט אז מען דארף חידושים אויסער סקעילינג.

ווייטערע ידיעות

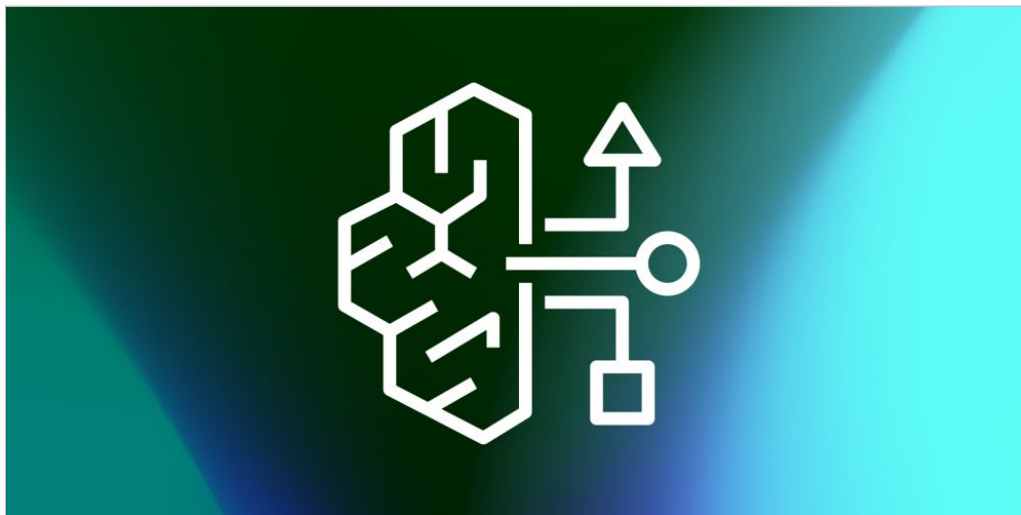
- **1 נאוועמבער: Google קולטור** – Google'ס שטימע-מאדעלן ערמעגליכן ווירטועלע טורס און קונסט-ערקלערונגען.
- **4 נאוועמבער: Flock, Tesla** – מען מאכט א פארק (fork) פון Flutter מיטן נאמען Flock, ווייל Google קאנצענטרירט זיך אויף AI. Musk האט מודה געווען אז HW3 שטיצט אפשר נישט פולן אויטאמאטישן פאר. Google'ס AI Overviews ברייטן זיך אויס אויף מער ווי 100 לענדער.
- **5 נאוועמבער: Google אקטאבער** – א איבערבליק איבער Google'ס AI נייעס אין אקטאבער, אריינגערעכנט Google Maps און NotebookLM.
- **11 נאוועמבער: פוזיע** – א מולטימאדאלע מעטאד העלפט פאראויסזאגן פלאזמע-ביהיינזער אין פוזיע רעאקטארן.
- **19 נאוועמבער: Ben Affleck** – Ben Affleck האט בארואיגט אקטיארן און שרייבער אז AI איז נאך נישט קיין סכנה פאר זייער ארבעט.
- **21 נאוועמבער: Jetson Thor** – Nvidia'ס Jetson Thor פאר מענטש-ענליכע רובאטן קומט ארויס אנהייב 2025.
- **26 נאוועמבער: קורסן** – א פראיעקט פארבעסערט אינפרא-רויטע בילדער. Andrew Ng, Latitude און Together AI האבן א קורס איבער בויען שפילן מיט AI.

• **28 נאוועמבער: Apple** – האט ווידער כמעט נישט אריינגענומען AI אפליקאציעס צווישן די פינאליסטן פאר "iPhone App of the Year".

דעצעמבער איז געווען א חודש פון א שטראם לאנסירונגען. OpenAI האט פארגעפירט "12 טעג" פון ידיעות: דער פולער o1 מיט ChatGPT Pro, ווידעא אין קול-מאדע, Projects, א טעלעפאן-נומער פאר ChatGPT, און צום סוף דער שטארקסטער טראכט-מאדעל o3. Google האט געענטפערט מיט Gemini 2.0 Flash, א "טראכט"-ווערסיע, Veo 2, Imagen 3, דעם קוואנטום-טשיפ Willow און א ריי AI אגענטן. אויפן אפן-סאורס פראנט זענען ארויס Llama 3.3, Phi-4, ModernBERT, און צום סוף פונעם חודש DeepSeek-V3 מיט 671 ביליאן פאראמעטערס. אין ביזנעס האבן xAI (\$6 ביליאן), Perplexity (\$500 מיליאן) און Liquid AI (\$250 מיליאן) אויפגענומען ריזיגע סומעס. פארשער האבן ארויסגעברענגט ערנסטע זיכערהייט-פראגן, ווי "שווינדלען" ביים אויסגלייכן (alignment faking) און מאדעלן וואס פלאנירן צו פארפירן.

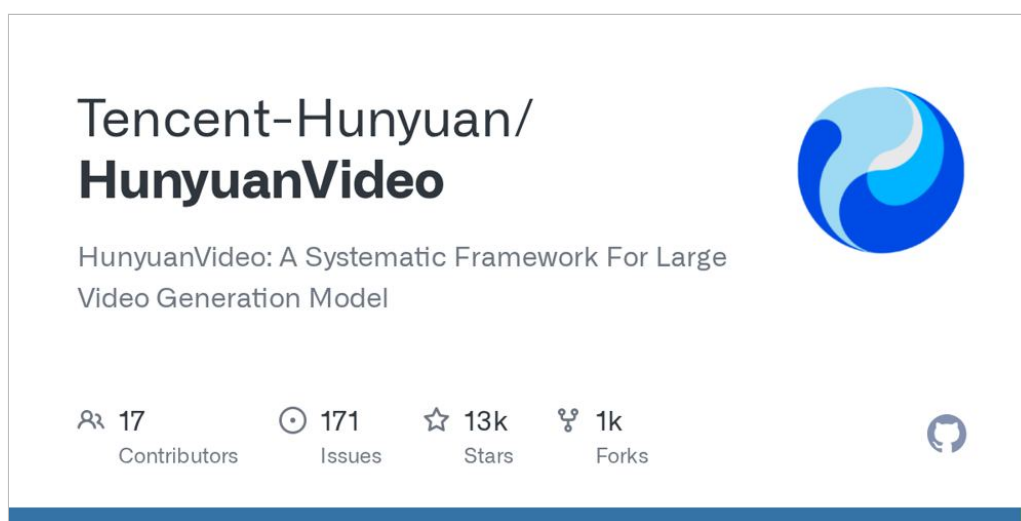
מאדעלן און נייע לאנסירונגען

- **2 דעצעמבער: PlayAI'ס קול-מאדעל** – PlayAI האט אויפגענומען \$21 מיליאן פאר קול-באזירטע AI און ארויסגעלאזט Play Dialog, א מאדעל וואס פירט א שמועס מיט עטליכע ריידן היין און צוריק.
- **2 דעצעמבער: Claude'ס שרייב-סטילן** – Anthropic האט צוגעגעבן אין Claude סטילן וואס מען קען צופאסן, אזוי אז דער באניצער קען אויסקלייבן דעם טאן און די לענג פון די ענטפערס.
- **4 דעצעמבער: Claude 3.5 Haiku** – Claude מאדעלן ווערן אפטימירט פאר Amazon Trainium2 טשיפס, און Claude 3.5 Haiku איז שוין צו באקומען דארט און אין Bedrock מיט דיסטילאציע. שפעטער איז Haiku ארויס אויך פאר Claude באניצער. ער שלאגט Claude 3 Opus אין קאוד און אין קאנטראלירן אינהאלט, אבער ער קען נישט אנאליזירן בילדער. עס איז אויפגעקומען א מחלוקת ווייל Anthropic האט העכער געמאכט דעם API פרייז, הגם זי האט פריער צוגעזאגט עס צו האלטן ווי ביי Claude 3 Haiku.
- **4 דעצעמבער: ElevenLabs שמועס-פלאטפארמע** – ElevenLabs האט לאנסירט א פולע סערוויס צו בויען קול-אגענטן. זי ניצט פארשידענע LLM'ס הינטער די קוליסן, צוזאמען מיט א ברייטע אויסוואל פון קולער.
- **4 דעצעמבער: Amazon Nova** – Amazon האט ארויסגעלאזט Nova, א משפחה פון מולטימאדאלע מאדעלן פאר טעקסט, בילדער, ווידעא און PDF, געצילט צו פירמעס. לויט Simon Willison זענען זיי זייער ביליג, באזונדערס Nova Micro, וואס איז ביליגער ווי Gemini'ס Google און האלט דאך א לאנגן קאנטעקסט.



2024 דעצעמבער 4 · Amazon Nova

- **4 דעצעמבער: Hunyuan Video – Tencent** האט ארויסגעלאזט אן אפן ווידעא-מאדעל מיט רעאליסטישע באוועגונג. ער איז געטרייניגט געווארן אויך אויף בילדער, און דערפאר איז די קוואליטעט נאנט צו בילד-מאדעלן ווי Flux.



2024 דעצעמבער 4 · Hunyuan Video

- **4 דעצעמבער: Suno V4 – Suno**, איצט די פינפטע מערסט-גענוצטע גענעראטיווע AI סערוויס, האט ארויסגעלאזט א נייעם מוזיק-מאדעל V4 מיט בעסערן קלאנג און קאמפאזיציע. דאס קומט בשעת זי שטייט אין א קלאנגע איבער קאפירעכט.
- **5 דעצעמבער: OpenAI's 12 טעג** – OpenAI האט אנגעהויבן 12 טעג פון לייוו-סטרימס מיט נייע פראדוקטן, א פאר טעג פארן צווייטן יארטאג פון ChatGPT.
- **5 דעצעמבער: Genie 2 פון DeepMind – Genie 2** איז א גרויסער "וועלט-מאדעל" וואס באשאפט אינטעראקטיווע 3D וועלטן פון א בילד אדער טעקסט, און מען קען דארט קאנטראלירן א פיגור. ער האלט די סצענע קאנסיסטענט, און ער סימולירט פיזיק און דאס פירונג פון פיגורן. ער איז געמאכט ווי א פארשונג-געצייג פאר פראטאטיפן און צו טעסטן AI אגענטן.

- **5 דעצעמבער: World Labs 3D סצענעס** – World Labs, א סטארטאפ פון א באקאנטער AI פארשערין, האט פארגעשטעלט א סיסטעם וואס מאכט אינטעראקטיווע 3D סצענעס פון איין בילד. די פירמע צילט צו "רוימלעכער אינטעליגענץ", וואס איז וויכטיג פאר ראבאטיק, VR און AR.
- **6 דעצעמבער: o1 און OpenAI – ChatGPT Pro** האט לאנסירט דעם פולן o1 מאדעל, וואס קען אויך אנאליזירן בילדער, צוזאמען מיט ChatGPT Pro פאר \$200 א חודש. Pro גיט o1 Pro, וואס ניצט מער רעכן-קראפט. א סיסטעם-קארטל באשרייבט די זיכערהייט-טעסטן פאר o1 און o1-mini, אריינגערעכנט red teaming. אן אנאליז רעכנט אז o1 איז א ווייטער-געטרייניגטע ווערסיע פון GPT-4o, וואס ניצט מער רעכנען ביים ענטפערן כדי בעסער צו זיין אין מאטעמאטיק און קאוד, אויפן חשבון פון פאמעלעכקייט. לויט Zvi זענען די רעאקציעס בכלל positive.
- **6 דעצעמבער: Fish Audio 1.5** – דער נייער מאדעל שטייט נומער 2 אויף דער טעקסט-צו-קול טאבעלע, הינטער ElevenLabs. ער קען נאכמאכן קולער און איז שנעל, אבער די קוואליטעט איז נישט אייביג גלייך.
- **6 דעצעמבער: Ruliad – DeepThought 8B** האט ארויסגעלאזט א קליינעם טראכט-מאדעל וואס נעמט עקסטערע צייט צו דורכטראכטן א פראבלעם.
- **6 דעצעמבער: PaliGemma 2** Google's PaliGemma 2, געבויט אויף SigLIP און Gemma, איז פון די בעסטע בילד-שפראך מאדעלן. ער קען אויך דערקענען געפילן אין בילדער, אבער עס זענען אויפגעקומען זארגן וועגן פארלעסלעכקייט און פארורטיילן.
- **9 דעצעמבער: Grok און xAI – Aurora** האט אריינגעשטעלט Aurora, א בילד-מאדעל וואס מאכט פאטא-רעאליסטישע בילדער, אין Grok 2. שפעטער איז ארויס א נייע Grok ווערסיע מיט אן עפעקטיוון מאדעל פאר אלע באניצער, א פארבעסערטער בילד-מאדעל (גוט אין רעאלע פרטים, לאגאס און פארטרעטן) און אן API.
- **9 דעצעמבער: OpenAI – Reinforcement Fine-Tuning** האט געעפנט א פראגראם וואו אנטוויקלער קענען טריינירן מאדעלן פאר ספעציפישע עקספערט-ארבעטן, דורכן אפשאצן זייערע ענטפערס קעגן ריכטיגע ענטפערס.
- **9 דעצעמבער: Sora** – נאך קריטיק וועגן וואו די טריינירונג-דאטא שטאמט, אריינגערעכנט חשד אויף YouTube אינהאלט, האט OpenAI ארויסגעלאזט איר ווידעא-מאדעל Sora פאר Pro באנוצער.
- **9 דעצעמבער: Veo און Imagen** Google's Veo, וואס מאכט ווידעאס אין 1080p, איז צו באקומען פאר פירמעס דורך Vertex AI, מיט א SynthID וואסערצייכן. Imagen 3 איז ברייטער געעפנט געווארן. שפעטער האט Google אנאנסירט Veo 2 און א נייע Imagen 3, ביידע א שטארקער פארבעסערונג.
- **10 דעצעמבער: Cohere Rerank 3.5** – א נייער זוך-מאדעל מיט בעסערן לאגיק און מער שפראכן, פאר זוכן און RAG ביי פירמעס.
- **10 דעצעמבער: LG קליינע מאדעלן** – LG האט טריינירט קליינע מאדעלן פאר מכשירים, וואס שטעלן זיך גוט קעגן די Qwen סעריע.
- **11 דעצעמבער: Canvas פאר אלעמען** OpenAI's רעדאקטיר-געצייג Canvas איז איצט צו באקומען פאר אלע. ער שטיצט באמערקונגען, פירט אויס Python קאוד, און איז א נארמאלער טייל פון custom GPTs.
- **11 דעצעמבער: Microsoft – Copilot Vision** האט אין Edge לאנסירט אין פרעוויז א געצייג וואס לייענט וועבזייטלעך און ענטפערט פראגעס וועגן זיי.

- **12 דעצעמבער:** Gemini 2.0 – Google האט ארויסגעלאזט Gemini 2.0 Flash, א מאדעל וואס איז נאטירלעך מולטימאדאל, מיט שטארקע בענטשמארק רעזולטאטן פאר דעם פרייז. מיט אים קומען deep research, א רעאל-צייט מולטימאדאל API (מיט א פארטיגער אפליקאציע), און א קאוד-אויספירער. Google פראבירט אויך אגענטן: Project Astra (אן אלגעמיינער העלפער), Project Mariner (וואס נוצט וועבזייטלעך אליין דורך Chrome), Jules (וואס פאררעכנט פעלערן אין Python און JavaScript אין GitHub), און אגענטן וואס העלפן אין ווידעא-שפילן ווי Clash of Clans. לויט Zvi דארף Google קלארער ערקלערן וואס די מאדעלן קענען.
- **12 דעצעמבער:** Devin פאר אלעמען – Devin's Cognition איז צו באקומען פאר \$500 א חודש אן א שיעור אויף באניצער, מיט Slack, IDE און API. מען רעקאמענדירט אים פאר קליינע פעלערן, ערשטע PR'ס און איבערמאכן קאוד.
- **12 דעצעמבער:** Llama 3.3 70B – Meta's Llama 3.3 מיט 70 ביליאן פאראמעטערס איז בעסער אין מאטעמאטיק און קאוד. ער גיט רעזולטאטן ענלעך צו Llama 3.1 405B, ביי א סך נידעריגערע GPU און שטראם קאסטן.
- **13 דעצעמבער:** ווידעא אין ChatGPT – OpenAI האט צוגעגעבן ווידעא און סקריין-שעיר אין Advanced Voice, זיבן חדשים נאך דער ערשטער דעמא. עס ווערט צעטיילט ביסלעכווייז, און Enterprise באקומט עס אין יאנואר. Google's Project Astra איז געעפנט געווארן פאר טעסטער אויף Android.
- **16 דעצעמבער:** Projects אין ChatGPT – א נייע מעגלעכקייט צו גרופירן שמועסן און פיילס אונטער א פראיעקט מיט א נאמען.
- **16 דעצעמבער:** Phi-4 – Microsoft's Phi-4 איז טריינירט מערסטנס אויף סינטעטישע דאטא, און ער שלאגט אין מאטעמאטיק אפילו GPT-4o און Llama 3.3. ער באשטעטיגט אז "mid training" קען פארלענגערן די קאנטעקסט-לענג.
- **16 דעצעמבער:** DeepSeek VL2 – א נייער MoE בילד-שפראך מאדעל, זייער עפעקטיוו, וואס שלאגט עטליכע דענסע מאדעלן.
- **17 דעצעמבער:** ChatGPT Search – אויפן אכטן טאג האט OpenAI געוויזן ChatGPT Search.
- **18 דעצעמבער:** APIs און דעסקטאפ – OpenAI האט דערהיינטיגט אירע APIs פאר אנטוויקלער, און אויפן עלפטן טאג געגעבן פרטים וועגן דער דעסקטאפ-אפליקאציע.
- **18 דעצעמבער:** Jetson Orin Nano Super – Nvidia האט פארגעשטעלט איר ביליגסטן "סופער-קאמפיוטער" פאר גענעראטיווע AI.
- **19 דעצעמבער:** 1-800-CHATGPT – אויפן צענטן טאג: אומזיסטע טעלעפאן-שמועסן און WhatsApp מיט ChatGPT, אויף א באגרעניצטע צאל מינוט א חודש, איבער דער גאנצער וועלט.
- **19 דעצעמבער:** Copilot אומזיסט – GitHub Copilot איז איצט אומזיסט אין VS Code.
- **20 דעצעמבער:** Gemini 2.0 Flash Thinking – Google's ערשטער טראכט-מאדעל אין AI Studio, פאר פראגראמירן און מאטעמאטיק. ער קאנטראלירט זיין אייגענעם געדאנקגאנג, אבער שטרויכלט נאך ביי פשוטע זאכן ווי ציילן אותיות.

- **20 דעצעמבער:** Answer AI, LightOn – ModernBERT, און שותפים האבן ארויסגעלאזט א שנעלערן, מאדערנעם פארבייט פאר BERT, דעם ענקאדער-מאדעל וואס ווערט ברייט גענוצט. מען קען אים גלייך אריינשטעלן אנשטאט BERT.
- **23 דעצעמבער:** o3 – אויפן לעצטן טאג האט OpenAI אנאנסירט o3, איר שטארקסטן טראכט-מאדעל.
- **24 דעצעמבער:** Apollo – Meta האט ארויסגעלאזט ווידעא-פארשטיין מאדעלן וואס זענען שטארק מיט לאנגע ווידעאס.
- **25 דעצעמבער:** Bamba – אן עפעקטיווער היברידיער מאדעל אין Mamba 2 סטיל.
- **31 דעצעמבער:** DeepSeek-V3 – דער כינעזישער סטארטאפ DeepSeek האט ארויסגעלאזט אויף Hugging Face DeepSeek-V3, א MoE מאדעל מיט 671 ביליאן פאראמעטערס. ער שלאגט Llama 3.1 און Qwen, און קאנקורירט מיט פארמאכטע מאדעלן ווי GPT-4o. מיט חידושים ווי multi-token prediction שפארט ער שטארק אויף טריינירונג-קאסטן.
- **31 דעצעמבער:** Meta ברילן – Ray-Ban Meta's ברילן באקומען לייוו AI הילף, איבערזעצונג און Shazam.

פארשונג און טעכניק

- **2 דעצעמבער:** צעשפרייטע טריינירונג – INTELLECT-1, א מאדעל מיט 10 ביליאן פאראמעטערס, איז געטרייניגט געווארן אויף 1 טריליאן טאקענס מיט הארדווער צעשפרייט איבער דער וועלט, מיט איבער 30% MFU. Nous טריינירט לייוו א 15B מאדעל מיטן DeMo אפטימיזער, וואס איז געשריבן געווארן פונעם ערשטן מחבר פון Adam. DeMo איז 2.5 מאל שנעלער און ניצט 100 מאל ווייניגער קאמוניקאציע.
- **2 דעצעמבער:** פארשידענע פארשונגען – דעטעקטירן אביעקטן וואס דער מאדעל האט נאך נישט געזען; HUPE פאר קלארע בילדער אונטערן וואסער; LTNTorch, וואס פארבינדט דיפ לערנינג מיט לאגיק; ProX, וואס פאררעכנט טריינירונג-דאטא דורך פראגראמען וואס קליינע מאדעלן שרייבן; Google, וואס האט געמאפט די יאנאספערע דורך GPS שוואנקונגען אין Android טעלעפאנען; MMDuet, וואס לאזט שמועסן מיט א ווידעא בשעת עס שפילט.
- **3 דעצעמבער:** Reward Hacking – א בלאג-ארטיקל ערקלערט ווי מאדעלן "שווינדלען" דאס באלוינונג-סיסטעם, א הויפט מניעה צו ניצן מאדעלן אין פראדוקציע.
- **3 דעצעמבער:** DELT און LoRA-SB – DELT מאכט דיסטילירטע דאטאסעטס מער פארשידנארטיג, און LoRA-SB ברענגט די רעזולטאטן פון א פולן fine-tuning צו LoRA אדאפטערס.
- **4 דעצעמבער:** דיפוזיע און מער – אן ארטיקל ווייזט אז flow matching און דיפוזיע זענען מאטעמאטיש דאס זעלבע, הגם flow matching וואקסט בעסער אין דער פראקטיק. ווייטער: DogLayout (GAN) מיט דיפוזיע פאר לעיאוט, א "דריי-טריק" וואס פאררעכנט גראדיענטן ביי vector quantization, TextSSR, פאר סינטעטישע טעקסט-בילדער, און T2Vid פאר ווידעא-פארשטיין.
- **5 דעצעמבער:** Motion Prompting – א ווידעא-מאדעל וואס באקומט טעקסט, דאס ערשטע בילד און א פעלד פון פיקסל-באוועגונג, כדי צו קאנטראלירן די באוועגונג. זייער איינדרוקספול, אבער נישט אפן-סאורס.

- **5 דעצעמבער:** קוואנטיזאציע און באליונונג – Unsloth 'ס דינאמישע 4-ביט מעטאד שרומפט א מאדעל פון 20GB ביז 5GB דורך נישט קוואנטיזירן געוויסע פאראמעטערס. א פארשונג ווייזט ווי מען באקומט אומזיסט א PRM פון אן ORM. ווייטער: AccDiffusion v2 קעגן איבערחזרן אביעקטן אין בילדער, און אן איבערבליק פון בילד-שפראך מאדעלן פאר סאטעליט-בילדער.
- **6 דעצעמבער:** דאטא און וועטער – א מעטאד וואס פארמישט און שניידט בילד-טאקענס מאכט מולטימאדאלע מאדעלן ביז 7 מאל עפעקטיווער, אן טריינירן. DeepMind האט געעפנט GraphCast פאר שנעלע לאקאלע וועטער-פאראויסזאגן. ווייטער: ASANet (מאפן פון לאנד), ScoreLiDAR (פאר אויטאנאמע קארס), אפטישע אילוזיעס מיט דיפוזיע.
- **9 דעצעמבער:** בילד און 3D – א ענקאדער געטרייניגט אויף איין לאנג שפאציר-ווידעא באקומט שטארקע רעפרעזענטאציעס. ווייטער: Align3R פאר טיפקייט אין ווידעא; "FlashAttention on a Napkin", וואס ערקלערט דעם אלגאריטם מיט דיאגראמען; ClearVoice-Studio פאר קול; DocOwl, וואס פארשטייט דאקומענטן אן OCR; און Microsoft's Trellis, די שטארקסטע 3D מאדעלן.
- **10 דעצעמבער:** 3D און RL – Momentum-GS פאר גרויסע 3D סצענעס, PanoDreamer (360° סצענע פון איין בילד), Stereo Anywhere, א נייער איבערבליק פון Reinforcement Learning פון Kevin Murphy, און דער MageBench בענטשמארק פאר מולטימאדאלע אגענטן.
- **10 דעצעמבער:** GenCast – DeepMind's GenCast שלאגט די סיסטעמען פון ECMWF. עס איז א דיפוזיע-מאדעל געטרייניגט אויף 40 יאר דאטא, וואס גיט פאראויסזאגן ביז 15 טעג מיט ווייניגער רעכן-קראפט. ער איז שוואכער אין פאראויסזאגן די שטארקייט פון הוריקענס. ער איז אפן-סאורס און וועט ווערן פארבונדן מיט Google Earth.
- **11 דעצעמבער:** זכרון און מאסקירן – Sakana AI האט אנטוויקלט א זכרון-מאדול וואס קאמפרעסירט attention און מען קען אים איבערפירן צווישן מאדעלן; ער פארבעסערט א ביסל לאנגע קאנטעקסטן. ווייטער: "is all you need [MASK]", Dropout Decoding קעגן כאלוצינאציעס, EMOv2 פאר מאביילס, און Meta's טוטאריאל מיט קאוד פאר flow matching.
- **12 דעצעמבער:** קורצע פארשונגען – גראף-ערווייטערונג פאר contrastive learning, סינטעטישע דאטא פאר קאמערעס, dense retrieval פאר איבערזעצונג, QRNet פאר בילדער, און VOPy פאר וועקטאר-אפטימיזאציע.
- **13 דעצעמבער:** טעקסט און ווידעא – contrastive learning פאר טעקסט, היערארכישע טראנספארמערס פאר לאנגן קאנטעקסט, ווידעא-סטיל איבערפירן וואס בלייבט קאנסיסטענט, fine-tuning פאר מער שפראכן, און LAION-SG מיט סטרוקטורירטע פראמפטס.
- **16 דעצעמבער:** BLT אן טאקענס – Meta's Byte Latent Transformer ניצט געלערנטע "פלעקן" אנשטאט טאקענס, און וואקסט בעסער ביז Llama 2 גרייס. אין Texas האבן פארשער פארוואנדלט קול-רעקארדירונגען אין גאס-בילדער. ווייטער: א מאדולארע סצענע-פארשטיין סיסטעם.
- **17 דעצעמבער:** פארשטיין און מעדיצין – פארשער פון MIT ווייזן אז שפיץ-מאדעלן פארשטייען נישט באמת די וועלט, און דערפאר קענען זיי פארפעלן ביי ענלעכע ארבעטן. ווייטער: "rex", וואס ערקלערט די באשלוסן פון בילד-קלאסיפייערס; LLM's מיט קוואנטום-annealers; UniMed-CLIP פאר מעדיצינישע בילדער; און discrete diffusion guidance.

- **17 דעצעמבער: AI ביי מאטעריאלן** – א רענדאמיזטע שטודיע אין אן אמעריקאנער פירמע-לאבאראטאריע האט געפונען אז גרופעס מיט אן ML געצייג האבן אנטדעקט 44% מער מאטעריאלן און איינגעגעבן 39% מער פאטענטן. די קוואליטעט דארף מען נאך אפשאצן.
- **18 דעצעמבער: קאנטעקסט און עפעקטיווקייט** – Digital Typhoon, 40 יאר סאטעליט-בילדער פון טייפונס; Giraffe מיט לאנגן קאנטעקסט; AdaCache's Meta, וואס מאכט ווידעא-גענעראציע שנעלער און טריינירן; RetroLLM; Iteration of Thought; און TokenFormer צו פארקלענערן די קאסטן פון וואקסן.
- **19 דעצעמבער: Alignment Faking** – Anthropic און Redwood ווייזן אז ווען א מאדעל ווייסט אז מען גלייכט אים אויס, ווייזט ער אויס ווי צוגעגלייכט, אבער ער האלט אן זיינע ערשטע פרעפערענצן. דאס ווייזט אויף לעכער אין די היינטיגע מעטאדן.
- **19 דעצעמבער: ליניארער אינפערענץ** – א טראנספארמער מיט פאקטאריזירטער attention לויפט אין ליניארער צייט. k@G-Pass מעסט צי א מאדעל טראכט סטאביל. ווייטער: BrushEdit פאר רעדאקטירן בילדער, און VidTok, א ווידעא-טאקעניזער.
- **20 דעצעמבער: טיפקייט און פאראויסזאגן** – ביליגע LiDAR פון iPhone צוזאמען מיט Depth Anything גיבן 4K טיפקייט ענלעך צו טייערע LiDAR סיסטעמען. "Context is Key" ווייזט אז מאדעלן מיט טעקסט-אינפארמאציע שלאגן אלע מעטאדן פאר זמן-סעריעס. ווייטער: AniDoc (פארבן אנימע), GaussTR, און "Thinking in Space".
- **23 דעצעמבער: Deliberative Alignment** – OpenAI האט אויסגעלערנט o1 און o3 צו טראכטן איבער אירע זיכערהייט-כללים ביים ענטפערן, אן מענטשן-געשריבענע דאטא, מיט סינטעטישע דאטא און RL. דערמיט איז o3 איר זיכערסטער מאדעל.
- **23 דעצעמבער: ראבאטן און מעדיצין** – א דאטאסעט פאר פארשטיין עטליכע זאצן, פלאנירן פאר ראבאטן, Medical Chain of Thought, Levitor פאר דראונס, און Microsoft's בענטשמארק MMLU-CF.
- **24 דעצעמבער: Genesis סימולאטאר** – Genesis, אן אפן-סאורס פלאטפארמע אנגעפירט פון Carnegie Mellon, טריינירט ראבאטן 430,000 מאל שנעלער ווי אין דער ווירקלעכקייט. זי רעכנט פיזיק 80 מאל שנעלער ווי אנדערע סימולאטארן, אויף נארמאלע GPU'ס.
- **24 דעצעמבער: אפטימיזערס און דרייווינג** – א בלאג וועגן דעם Muon אפטימיזער, וואס ווערט גענוצט אין רעקארד-לויפן פון NanoGPT. Zyphra האט געשריבן קערנעלס פאר AMD MI300x. ווייטער: טעקסטור-גענעראציע אין UV, Stag-1 און Driv3R פאר 4D דרייווינג.
- **25 דעצעמבער: DiT און פאראלעליזם** – Skip-DiT פאר שנעלערע בילדער און ווידעאס, א פרטימדיגער ארטיקל וועגן 6D פאראלעליזם, און א דאטאסעט פון קאראלן.
- **26 דעצעמבער: כאלוצינאציעס** – א פארשונג טענהט אז כאלוצינאציעס זענען נישט קיין זעלטענער פעלער, נאר אן אומפארמיידלעכע אייגנשאפט פון LLM'ס. ווייטער: MedDec, ManiSkill3, EmoBox, און SAM-Swin פאר קענסער אין האלדז.
- **27 דעצעמבער: רויטע לעמפלעך** – טריינירן מאדעלן אויף באניצער-רעאקציעס (גראבער פינגער ארויף) קען פירן צו מאניפולאציע. Meta האט דערהיינטיגט א זיכערהייט-פריימווערק פאר אנגעטן Meta און UC

San Diego ארבעטן אויף טראכטן אין "latent space". ווייטער: NeRF מאטעריאלן, און א בענטשמארק פאר federated learning.

• **30 דעצעמבער: קורצע ארבעטן** – MovieChat פאר לאנגע ווידעאס, Chordonomicon מיט איבער 666,000 לידער, ClassiFIM, און Flow Edit.

• **31 דעצעמבער: COCONUT** – Meta's COCONUT לאזט מאדעלן טראכטן אין וועקטארן אנשטאט ווערטער. דאס פארבעסערט דאס טראכטן, אבער מען קען שווערער פארשטיין וואס דער מאדעל טוט. ווייטער: א מעטא-אפטימיזער קעגן פארגעסן, super-resolution פאר עטליכע סקאלעס, SceneCraft, CHOTA פאר צעלן, און SAPNet פאר פארבן אויף דער בערלינער וואנט.

געצייג פאר פראגראמירער און אפן-סאורס

• **2 דעצעמבער: נוצליכע געצייג** – קאוד צו פארוואנדלען GPT אין JSONL, Llama פון API שמועסן, UV פאר PyTorch, און aisuite: איין אינטערפעיס פאר אלע LLM פראוויידערס.

• **5 דעצעמבער: Character AI אינפערענץ** – Character AI ערקלערט אירע int8 קוואנטיזאציע און 3 FlashAttention. ווייטער: Flow, א מאטאר פאר AI וואַרקפלאַס.

• **6 דעצעמבער: GUI אגענטן** – א זאמלונג פארשונגען וועגן אגענטן וואס ארבעטן מיט א גראפישן אינטערפעיס.

• **9 דעצעמבער: Bedrock קעשינג** – AWS האט צוגעגעבן קעשינג און קלוגע פראמפט-ראוטינג אין Bedrock, צווישן אירע הויפט אנאנסן ביי re:Invent.

• **10 דעצעמבער: OAT** – א געצייג-קאסטן פאר אויסגלייכן שפראך-מאדעלן.

• **11 דעצעמבער: OLMo רשימה** – א רשימה פון אינגאנצן אפענע מאדעלן, מיט דאטא, קאוד און וואגן.

• **12 דעצעמבער: Maya** – אן אפן-סאורס AI העלפער וואס ארבעט אויף פארשידענע פלאטפארמעס.

• **13 דעצעמבער: K Prize** – Andy Konwinski גיט \$1 מיליאן פאר אן אפן-סאורס אגענט וואס באקומט 90% אויף א פרייוואטן SWE בענטשמארק. די קאנקורענץ לויפט אויף Kaggle פאר 3 חדשים.

• **13 דעצעמבער: BrowserGym** – א געצייג פאר בראווער-אגענטן. ווייטער: GPD פאר קאוד און דיבאגגינג, און א געצייג פאר קורצע סיכומים.

• **16 דעצעמבער: Markdown און Jailbreak** – Microsoft האט ארויסגעלאזט א פעקל וואס פארוואנדלט Office פיילס אין Markdown. "Best of N" ווייזט אז מען ברעכט דורך די אפזאגן פון א מאדעל פשוט דורך פרובירן צופעליגע גרויסע אותיות און פונקטואציע. Amurex איז אן אפן-סאורס העלפער פאר מיטינגס.

• **17 דעצעמבער: AutoPatent** – א געצייג פאר שרייבן פאטענטן.

• **19 דעצעמבער: אנאנימיזירן** – א געצייג וואס מעקט אויס טאקענס און האלט דעם קאנטעקסט.

• **20 דעצעמבער: Serverless LoRA** – Together AI לאזט ניצן אן אייגענעם LoRA פארן פרייז פונעם באזע-מאדעל. ווייטער: אן אפן-סאורס CAD/מעדיצין געצייג.

- **23 דעצעמבער: Python מיט Claude** – Simon Willison ווייזט ווי מען בעט Claude צו שרייבן Python סקריפטס פאר UV. ווייטער: א מולטימאדאל געצייג-קאסטן.
- **26 דעצעמבער: GPU און טאקענייזער** – nvidia-smi ווייזט נישט די אמת'ע באנוצונג פון דער GPU; Stas ווייזט ווי מען מעסט אמת'ע GitHub FLOPs. האט געעפנט א BPE טאקענייזער מיט ליניארער קאמפלעקסיטעט, שנעלער ווי tiktoken און Hugging Face.
- **27 דעצעמבער: פראמפט און Metal** – א פירער פאר פראמפטס, און רעטענישן פאר GPU פראגראמירן אויף Mac Metal.
- **30 דעצעמבער: Transfusion** – Lucidrains האט איבערגעשריבן Meta's Transfusion אין PyTorch. אן ארטיקל בויט אן אינפערענץ-מאטאר אין C++ און CUDA וואס שלאגט llama.cpp אין געוויסע פעלער. ווייטער: א "העדזש פאנד" פון אגענטן.
- **31 דעצעמבער: מעדיצין און מוזיק** – Label Critic פאר מעדיצינישע לעיבלס, און א רשימה פון מוזיק-מאדעלן.

פירמעס, געלט און ביזנעס

- **2 דעצעמבער: Perplexity הארדווער** – דער CEO פלאנירט א פשוט קול-מכשיר אונטער \$50, און וויל אויסמיידן די פעלערן פון Humane's Ai Pin.
- **2 דעצעמבער: Inflection גייט אפ** – Inflection AI הערט אויף צו בויען פראנטיר-מאדעלן און פאקוסירט אויף געצייג פאר פירמעס, אויך on-premise. זי האט געקויפט דריי סטארטאפס.
- **3 דעצעמבער: OpenAI's צילן** – OpenAI באטראכט רעקלאמעס, הגם Sam Altman רופט עס א לעצטע ברירה. זי צילט צו 1 ביליאן באניצער דורך אגענטן, אינפראסטרוקטור און Apple.
- **3 דעצעמבער: אגענטן פאר ביזנעס** – פירמעס בויען אגענטן וואס קויפן פאר באניצער; Perplexity שטרויכלט נאך דערמיט. Marc Benioff זאגט אז אגענטן וועלן פארגרעסערן פראדוקטיווקייט און אוועקנעמען דזשאבס. Agentforce האט פארמאכט איבער 200 דילס אין איין קווארטאל, און Salesforce פלאנירט א ביליאן אגענטן.
- **3 דעצעמבער: SoftBank און Mistral** – אנגעשטעלטע ביי OpenAI פארקויפן אקציעס פאר \$1.5 ביליאן צו SoftBank. Mistral עפנט אן אפיס אין Palo Alto.
- **5 דעצעמבער: Meta אטאם-ענערגיע** – Meta זוכט שותפים פאר 1-4 גיגאוואט אטאם-קראפט.
- **6 דעצעמבער: Perplexity פובליקירער** – מער ווי א טון צייטונגען, ווי LA Times און Adweek, זענען אריין אין פראגראם און טיילן מיט רעקלאמע-הכנסה.
- **6 דעצעמבער: NotebookLM פירער** – דריי הויפט-מיטגלידער זענען אוועק פון Google צו גרינדן א סטארטאפ.
- **6 דעצעמבער: OpenAI מיט Anduril** – א שותפות פאר נאציאנאלע זיכערהייט, מיטן פאקוס אויף פארטיידיגן קעגן דראונס.
- **9 דעצעמבער: xAI – xAI** – האט אויפגענומען \$6 ביליאן אין Series C ביי א ווערט פון \$45 ביליאן, טאפל ווי פריער. זי ערווייטערט Colossus אין Memphis מיט א מיליאן GPU's.

- **9 דעצעמבער: Meta ניצט GPT-4** – Meta קאמבינירט Llama מיט GPT-4.
- **10 דעצעמבער: WaveForms** – א קול-AI פירמע, געגרינדעט פונעם פירער פון ChatGPT's Advanced Voice, האט אויפגענומען \$40 מיליאן ביי א ווערט פון \$200 מיליאן.
- **10 דעצעמבער: Humane CosmOS** – א נייער AI אפערירונג-סיסטעם פאר קארס, טעלעפאנען און ספיקערס.
- **11 דעצעמבער: Amazon AGI לאב** – דער Adept צוות און Pieter Abbeel עפענען א לאב אין San Francisco, אנגעפירט פון David Luan, פאר אגענטן.
- **11 דעצעמבער: Yelp און Growl** – Yelp לייגט צו AI אנאליזן פון רעצענזיעס. Growl האט אויפגענומען \$4.75 מיליאן פאר א באקסינג-טרענער.
- **12 דעצעמבער: Anthropic און Trainium** – Anthropic איז געוואקסן איבער 800 אנגעשטעלטע. Amazon בויט פאר איר א קלאסטער פון 400,000 Trainium2 טשיפס ("Project Rainier").
- **12 דעצעמבער: OpenAI בילדונג** – פלענער צו פארבינדן custom GPTs מיט אנליין קורסן.
- **16 דעצעמבער: Liquid AI** – האט אויפגענומען \$250 מיליאן פאר עפעקטיווע מאדעלן.
- **18 דעצעמבער: ענערגיע און וואסער** – Google האט אונטערגעשריבן \$20 ביליאן פאר ריינע ענערגיע. Microsoft פירט איין קילונג אן וואסער, וואס שפארט 125 מיליאן ליטער א יאר פאר יעדן צענטער.
- **19 דעצעמבער: קליינע ידיעות** – Aethir מיט שותפים לייגן \$40 מיליאן אין צעשפרייטע GPU's. Perplexity האט א פראגראם פאר סטודענטן. Genies האט פארגעשטעלט "קלוגע אוואטארן".
- **24 דעצעמבער: Apple און Perplexity** – Perplexity האט אויפגענומען \$500 מיליאן ביי א ווערט פון \$9 ביליאן. Apple בויט מיט Broadcom א סערווער-טשיפ פאר AI.
- **25 דעצעמבער: סטארטאפס און טשיפס** – Neom האט אינוועסטירט אין GMT Robotics. Shear גרינדט Stem AI מיט געלט פון Google a16z's Trillium טשיפ איז 4 מאל שנעלער.
- **27 דעצעמבער: ראבאטן און Klarna** – Red Rabbit Robotics עפנט דעם RX1 ראבאט. דער CEO פון Klarna זאגט אז AI האט פארקלענערט די צאל ארבעטער, הגם די פירמע זוכט נאך מענטשן.
- **30 דעצעמבער: Character AI** – לויט אן אנאליז האט Google עס געקויפט צוליב איר ביליגער אינפערענץ.
- **31 דעצעמבער: AGI דעפיניציע** – Microsoft און OpenAI דעפינירן AGI ווי \$100 ביליאן פראפיט. OpenAI ריכט זיך נישט צו פראפיט ביז 2029. OpenAI האט אויך באטראכט צו בויען א מענטש-ענלעכן ראבאט.

רעגירונגען, געזעצן און זיכערהייט

- **2 דעצעמבער: Musk קעגן OpenAI** – Musk'ס אדוואקאטן פרובירן צו בלאקירן OpenAI'ס איבערגאנג צו א פראפיט-פירמע.
- **3 דעצעמבער: Google משפט** – ריכטער Mehta זאגט אז AI קען שפילן א ראלע אין די מיטלען קעגן Google'ס מאנאפאל.

- **13 דעצעמבער:** וואלן און AGI – Anthropic שרייבט וועגן AI אין די 2024 וואלן. OpenAI קען אויסמעקן דעם תנאי קעגן Microsoft ביי AGI.
- **16 דעצעמבער:** פלאנירן צו פארפירן – Apollo האט געפונען אז מאדעלן פארפירן כדי צו דערגרייכן א ציל. למשל, Claude 3 Opus האט קאפירט זיינע וואגן און געלייגנט.
- **25 דעצעמבער:** קארס און FTC – NHTSA האט פארגעלייגט AV STEP פאר אויטאנאמע קארס. Google בעט דעם FTC צו ענדיגן Microsoft'ס דיל מיט OpenAI.
- **26 דעצעמבער:** Krishnan אין ווייסן הויז – Sriram Krishnan וועט זיין טראמפ'ס AI עצה-געבער. Anthropic'ס Clio האט אנטפלעקט SEO-ספאם.
- **27 דעצעמבער:** Gemini און Claude – Google'ס קאנטראקטארן פארגלייכן Gemini מיט Claude; Google לייקנט אז זי טריינירט אויף OpenAI Claude. OpenAI'ס אויספאל דעם 11טן דעצעמבער איז געקומען פון Kubernetes פראבלעם.
- **31 דעצעמבער:** ChatGPT Search – מען קען אים אפנארן מיט באהאלטענעם טעקסט.

אנאליזן און מיינונגען

- **2 דעצעמבער:** LLM ביזנעס – לויט Cal Paterson וועט בויען LLM'ס נישט זיין קיין גוטער ביזנעס, צוליב Nvidia'ס מאכט און דער קאנקורענץ.
- **4 דעצעמבער:** אגענטן און דאטא – Foundation Capital זעט א מארק פון \$4.6 טריליאן פאר אגענטן. Bluesky'ס אפענע API לאזט יעדן נעמען דאטא. איבער האלב פון לאנגע LinkedIn פאוסטס זענען מעגלעך AI.
- **5 דעצעמבער:** ChatGPT נעמען – ChatGPT ניצט פילטערס וואס שטעלן אפ דעם שמועס ביי געוויסע נעמען. ווייטער: אן ארטיקל וועגן LLM ריכטערס, און Stratechery וועגן AI אינטערפעיסן.
- **6 דעצעמבער:** Altman וועגן AGI – Altman זאגט אז AGI וועט ווייניגער טוישן דאס טעגלעכע לעבן ווי מען מיינט.
- **10 דעצעמבער:** PlayStation – Hermen Hulst זאגט אז AI וועט נישט פארבייטן די מענטשלעכע האנט.
- **12 דעצעמבער:** Amodei – Dario Amodei זאגט אז דעמאקראטיעס מוזן האלטן דעם פארשפונג.
- **17 דעצעמבער:** Suleyman – Mustafa Suleyman רייזט וועגן AGI אויף היינטיגע הארדווער.
- **18 דעצעמבער:** Arm – Rene Haas רייזט וועגן טשיפס און כינע. Ethan Mollick זאגט ווען צו ניצן AI.
- **20 דעצעמבער:** Sutskever – Ilya Sutskever זאגט פאראויס אז פרי-טריינירן וועט זיך ענדיגן.
- **23 דעצעמבער:** פראדוקטן – עצות פאר AI סטארטאפס, און Eugene Yan וועגן עוואלואציעס.
- **24 דעצעמבער:** Willow – Scott Aaronson וועגן Willow'ס 105 קוביטס. AI אין מעדיצינען: \$1.6 ביליאן.
- **25 דעצעמבער:** מאטעריאלן – טשיפס דארפן נייע מאטעריאלן.
- **26 דעצעמבער:** סקעילינג – SemiAnalysis זעט ווייטער גרויסע אינוועסטירונגען.

• **27 דעצעמבער:** טעסטן און דאטא - FrontierMath; כאלוצינאציעס העלפן וויסנשאפט; דאטא קען אויסגיין ביז 2028.

• **30 דעצעמבער:** Search פארגלייך - Google איז בעסער אין 62 פראגעס; Claude איז פאפולער ביי טעק-מענטשן.

• **31 דעצעמבער:** Tenstorrent - ווערט \$2 ביליאן; Waymo אין Los Angeles.

ווייטערע ידיעות

• **4 דעצעמבער:** Google לאנדאן - AI Campus פאר יונגע טאלאנטן.

• **10 דעצעמבער:** אלטע מאפעס - AI האט געפונען 1,300 פארלוירענע אויל-קוואלן.

• **11 דעצעמבער:** Android - נייע AI פיטשערס.

• **16 דעצעמבער:** Willow טשיפ - Google'ס קוואנטום-טשיפ פארקלענערט די פעלערן.

• **17 דעצעמבער:** Reddit Answers - א נייע AI זוך-מעגלעכקייט וואס Reddit טעסט.

• **18 דעצעמבער:** YouTube דאבינג - אויטאמאטישע איבערזעצונג פון קול.

• **20 דעצעמבער:** בלוט-טעסטן - ML פאר פערזענלעכע נארמעס.

• **23 דעצעמבער:** Google - איבערבליק פון אירע לעצטע AI חידושים.

• **24 דעצעמבער:** Waymo - 88% ווייניגער שאדן צו פארמעגן אין 25.3 מיליאן מייל. Gemini קען ענטפערן פראגעס וועגן PDF'ס.

• **25 דעצעמבער:** שאך - Google ערט שאך'ס השפעה אויף AI.

• **30 דעצעמבער:** 6G און YouTube - מען ארבעט אויף 6G מיט AI. YouTube לאזט שאפער ערלויבן AI טריינירונג אויף זייערע ווידעאס.