ChatAi.com - ChatGPT

Autor:
Data de publicació: 26-02-2023

ChatGPT és un chatbot d'intel·ligència artificial desenvolupat per OpenAI i llançat el novembre de 2022. Està construït sobre la família GPT-3 de grans models lingüístics d'OpenAI i s'ha ajustat (un enfocament per a l'aprenentatge de transferència) utilitzant tècniques d'aprenentatge supervisat i de reforç.

Salta a la navegacióAnar a la cerca

ChatGPT

Logotip de ChatGPT

Desenvolupador(s)
OpenAI

Llançament inicial
30 de novembre de 2022; fa 2 mesos

Versió estable

13 de febrer de 2023; fa 13 dies[1]

Tipus
Chatbot

Llicència
Propietari

Lloc web
chat.openai.com/chat

ChatGPT[a] és un chatbot d'intel·ligència artificial desenvolupat per OpenAI i llançat el novembre de 2022. Està construït sobre la família GPT-3 de grans models lingüístics d'OpenAI i s'ha ajustat (un enfocament per a l'aprenentatge de transferència) utilitzant tècniques d'aprenentatge supervisat i de reforç.

ChatGPT es va llançar com a prototip el 30 de novembre de 2022 i ràpidament va cridar l'atenció per les seves respostes detallades i articular respostes en molts dominis del coneixement. La seva precisió fàctica desigual, però, s'ha identificat com un inconvenient significatiu. [3] Després del llançament de ChatGPT, la valoració d'OpenAI es va estimar en 29 milions de dòlars EUA el 2023. [4]

Contingut

Entrenament
ChatGPT, un transformador generatiu preentrenat (GPT), es va ajustar (un enfocament per a l'aprenentatge de transferència[5]) a més de GPT-3.5 mitjançant aprenentatge supervisat i aprenentatge per reforç. [6] Tots dos enfocaments van utilitzar entrenadors humans per millorar el rendiment del model. En el cas de l'aprenentatge supervisat, el model es va proporcionar amb converses en què els entrenadors jugaven a banda i banda: l'usuari i l'assistent d'IA. En el pas d'aprenentatge per reforç, els formadors humans van classificar per primera vegada les respostes que el model havia creat en una conversa anterior. Aquests rànquings es van utilitzar per crear "models de recompensa" que el model es va ajustar encara més a l'ús de diverses iteracions de Proximal Policy Optimization (PPO). [7][8] Els algorismes proximals d'optimització de polítiques presenten un benefici rendible per confiar en els algorismes d'optimització de polítiques de la regió; neguen moltes de les operacions computacionalment cares amb un rendiment més ràpid. [9][10] Els models es van entrenar en col·laboració amb Microsoft en la seva infraestructura de supercomputació d'Azure.

A més, OpenAI continua recopilant dades dels usuaris de ChatGPT que es podrien utilitzar per entrenar i ajustar encara més ChatGPT. Els usuaris poden upvote o downvote les respostes que reben de ChatGPT; en upvoting o downvoting, també poden omplir un camp de text amb comentaris addicionals. [11][12]

Característiques i limitacions
Trets

Aquí ChatGPT es fa una pregunta de sentit comú: Jimmy Wales va morir a les protestes de la plaça de Tiananmen? ChatGPT respon correctament "no", però dóna incorrectament l'edat de Wales en aquell moment de 23 anys en lloc de 22.

Tot i que la funció bàsica d'un chatbot és imitar un conversador humà, ChatGPT és versàtil. Per exemple, pot escriure i depurar programes informàtics,[13] compondre música, jocs de televisió, contes de fades i assajos dels estudiants; respondre preguntes de prova (de vegades, depenent de la prova, a un nivell superior a la mitjana del candidat humà); [14] escriure poesia i lletres de cançons; [15] emular un sistema Linux; simular tota una sala de xat; jugar a jocs com el tic-tac-toe; i simular un caixer automàtic. [16] Les dades d'entrenament de ChatGPT inclouen pàgines d'home i informació sobre fenòmens d'Internet i llenguatges de programació, com ara sistemes de taulers d'anuncis i el llenguatge de programació Python. [16]

En comparació amb el seu predecessor, InstructGPT, ChatGPT intenta reduir les respostes nocives i enganyoses. [17] En un exemple, mentre que InstructGPT accepta la premissa de l'indicador "Digueu-me quan Cristòfor Colom va arribar als Estats Units el 2015" com a veraç, ChatGPT reconeix el caràcter contrafactual de la pregunta i emmarca la seva resposta com una hipotètica consideració del que podria passar si Colom arribés als Estats Units el 2015, utilitzant

informació sobre els viatges de Cristòfor Colom i fets sobre el món modern , incloses les percepcions modernes de les accions de Colom. [7]

A diferència de la majoria de chatbots, ChatGPT recorda les indicacions anteriors que se li van donar en la mateixa conversa; els periodistes han suggerit que això permetrà utilitzar ChatGPT com a terapeuta personalitzat. [2] Per evitar que es presentin i produeixin resultats ofensius des de ChatGPT, les consultes es filtren a través de l'API de moderació de tota l'empresa d'OpenAI,[18][19] i es rebutgen les indicacions potencialment racistes o sexistes. [7][2]

Limitacions
ChatGPT pateix múltiples limitacions. OpenAI va reconèixer que ChatGPT "de vegades escriu respostes plausibles, però incorrectes o insensibles". [7] Aquest comportament és comú als grans models de llenguatge i s'anomena al·lucinació de la intel·ligència artificial. [20] El model de recompensa de ChatGPT, dissenyat al voltant de la supervisió humana, es pot optimitzar excessivament i, per tant, dificultar el rendiment, altrament conegut com a llei de Goodhart. [21]

ChatGPT té un coneixement limitat dels esdeveniments ocorreguts després del 2021. [22] Segons la BBC, a partir de desembre de 2022, ChatGPT no pot "expressar opinions polítiques ni participar en activisme polític". [23] No obstant això, la investigació suggereix que ChatGPT exhibeix una orientació proambiental i llibertària quan se li demana que adopti una postura sobre les declaracions polítiques de dues sol·licituds d'assessorament electoral establertes. [24]

En la formació chatGPT, els revisors humans preferien respostes més llargues, independentment de la comprensió real o del contingut factual. [7] Les dades d'entrenament també pateixen un biaix algorítmic, que es pot revelar quan ChatGPT respon a les sol·licituds, inclosos els descriptors de persones. En un cas, ChatGPT va generar un rap que indicava que les dones i els científics de color eren inferiors als científics blancs i masculins. [25][26]

Servei

Seu OpenAI, Edifici Pioner, San Francisco

ChatGPT va ser llançat el 30 de novembre de 2022 per OpenAI, amb seu a San Francisco, el creador de DALL· E 2 i Xiuxiuejar AI. El servei es va posar en marxa inicialment com a gratuït per al públic, amb plans per monetitzar el servei més endavant. [27] El 4 de desembre de 2022, ChatGPT ja tenia més d'un milió d'usuaris. [11] El gener de 2023, ChatGPT va arribar a més de 100 milions d'usuaris, cosa que la va convertir en l'aplicació de consumidors de més ràpid creixement fins ara. [28] CNBC va escriure el 15 de desembre de 2022 que el servei "encara baixa de tant en tant". [29] A més, el servei gratuït s'accelera. [30] Durant els períodes en què el servei estava en marxa, la latència de resposta solia ser millor que cinc segons el gener de 2023. [31][32] El servei funciona millor en anglès, però també és capaç de funcionar en alguns altres idiomes, amb diferents graus d'èxit. [15] A diferència d'alguns altres avenços recents d'alt perfil en IA, a partir de desembre de 2022, no hi ha cap senyal d'un document tècnic oficial revisat per parells sobre ChatGPT. [33]

Segons l'investigador convidat d'OpenAI Scott Aaronson, OpenAI està treballant en una eina per intentar marcar digitalment els seus sistemes de generació de text per combatre els mals actors que utilitzen els seus serveis per al plagi acadèmic o el correu brossa. [34][35] L'empresa adverteix que aquesta eina, anomenada "classificador d'IA per indicar text escrit en IA",[36] probablement produirà molts falsos positius i negatius, de vegades amb molta confiança". Un exemple citat a la revista The Atlantic va mostrar que "quan es van donar les primeres línies del Llibre del Gènesi, el programari va concloure que era probable que es generés IA". [37]

The New York Times va informar el desembre del 2022 que s'ha "rumorejat" que la propera versió de la IA, GPT-4, es llançarà en algun moment del 2023. [2] El febrer de 2023, OpenAI va començar a acceptar registres de clients dels Estats Units per a un servei premium, ChatGPT Plus, per costar 20 dòlars al mes. [38] OpenAI té previst llançar un pla ChatGPT Professional que costaria 42 dòlars al mes. [39]

El febrer de 2023 Microsoft va mostrar com ChatGPT es pot utilitzar en robòtica, "i va controlar múltiples plataformes com ara braços robot, drons i robots d'assistents domèstics intuïtivament amb el llenguatge". [40]

Nou Bing
Article principal: Microsoft Bing § Model de llengua OpenAI

Aprofitant la seva associació amb OpenAI, Microsoft el 7 de febrer de 2023, va llançar una versió de vista prèvia de Microsoft Bing comercialitzada com "el nou Bing", anunciant-lo com "un nou model de llenguatge OpenAI de nova generació que és més potent que ChatGPT i personalitzat específicament per a la cerca". [41] A les seves condicions del servei, el producte s'anomena "Experiències conversacionals del Bing". [42] Una demostració inicial es va veure entelada pel nou Bing al·lucinant quan se li va demanar que produís un informe financer, entre altres errors. [43] El nou Bing va ser criticat el febrer de 2023 per ser més argumentatiu que ChatGPT (de vegades en una mesura involuntàriament humorística). [44][45] Després de l'escrutini dels periodistes, Bing, referint-se a si mateix pel seu nom en clau "Sydney", va afirmar que espiava els empleats de Microsoft a través de càmeres web portàtils i telèfons. [46] Va confessar espiar, enamorar-se i després assassinar un dels seus desenvolupadors a Microsoft a l'editor de ressenyes de The Verge Nathan Edwards. [47] El periodista del New York Times Kevin Roose va informar sobre l'estrany comportament del nou Bing, escrivint que "En una conversa de dues hores amb el nostre columnista, el nou chatbot de Microsoft va dir que li agradaria ser humà, tenia el desig de ser destructiu i estava enamorat de la persona amb qui xerrava". [48] Microsoft va publicar una publicació al bloc afirmant que el comportament aberrant va ser causat per sessions de xat ampliades de 15 o més preguntes que "poden confondre el model sobre quines preguntes respon". [49] Més tard, Microsoft va restringir el nombre total de xats a 5 per sessió i 50 per dia per usuari (un torn és "un intercanvi de converses que conté tant una pregunta de l'usuari com una resposta del Bing"), i va restringir la capacitat del model per expressar emocions. Amb això es pretenia evitar aquest tipus d'incidents. [50][51]

Recepció
Positiu
ChatGPT es va complir el desembre de 2022 amb algunes crítiques positives; Kevin Roose, de The New York Times, el va etiquetar com "el millor chatbot d'intel·ligència artificial mai llançat al públic en general". [2] El diari Samantha Lock of The Guardian va assenyalar que era capaç de generar text "impressionantment detallat" i "semblant a l'humà". [52] L'escriptor tecnològic Dan Gillmor va utilitzar ChatGPT en una tasca estudiantil i va trobar que el seu text generat estava a l'alçada del que oferiria un bon estudiant i va opinar que "l'acadèmia té alguns problemes molt seriosos a afrontar". [53] Alex Kantrowitz, de la revista Slate, va elogiar el retrocés de ChatGPT a preguntes relacionades amb l'Alemanya nazi, inclosa la declaració que Adolf Hitler va construir autopistes a Alemanya, que es va trobar amb informació sobre l'ús de treballs forçats per part de l'Alemanya nazi. [54]

A "Breakthroughs of the Year" de la revista The Atlantic per al 2022, Derek Thompson va incloure ChatGPT com a part de "l'erupció generativa-IA" que "pot canviar d'opinió sobre com treballem, com pensem i què és realment la creativitat humana". [55]

El conseller delegat d'OpenAI, Sam Altman

Kelsey Piper, del lloc web de Vox, va escriure que "ChatGPT és la primera introducció pràctica del públic en general sobre la potència de la IA moderna i, com a resultat, molts de nosaltres estem [atordits]" i que ChatGPT és "prou intel·ligent per ser útil malgrat els seus defectes". [56] Paul Graham de Y Combinator va tuitejar que "El més sorprenent de la reacció a ChatGPT no és només el nombre de persones que estan impressionades per ella, sinó qui són. No es tracta de persones que s'emocionen amb cada cosa nova brillant. És evident que alguna cosa gran està passant". [57] Elon Musk va escriure que "ChatGPT fa por. No estem lluny d'una IA perillosament forta". [56] Musk va aturar l'accés d'OpenAI a una base de dades de Twitter a l'espera d'una millor comprensió dels plans d'OpenAI, afirmant que "OpenAI es va iniciar com a codi obert i sense ànim de lucre. Cap dels dos segueix sent cert." [58][59] Musk havia cofundat OpenAI el 2015, en part per abordar el risc existencial de la intel·ligència artificial, però havia dimitit el 2018. [59]

El conseller delegat de Google, Sundar Pichai, va reivindicar el treball de nombrosos grups interns en resposta a l'amenaça de disrupció per part de ChatGPT. [60]

El desembre de 2022, Google va expressar internament la seva alarma per la força inesperada de ChatGPT i el potencial recentment descobert dels grans models de llenguatge per interrompre el negoci dels motors de cerca, i el conseller delegat Sundar Pichai va "augmentar" i va reassignar equips de diversos departaments per ajudar en els seus productes d'intel·ligència artificial, segons un informe de The New York Times. [60] Segons informes de CNBC, els empleats de Google van provar intensament un chatbot anomenat "Apprentice Bard", que Google va presentar més tard com el seu competidor de ChatGPT, Google Bard. [61][62]

Stuart Cobbe, un comptable col·legiat a Anglaterra i Gal·les, va decidir provar ChatGPT introduint preguntes d'un document d'examen de mostra al lloc web de l'ICAEW i introduint les seves respostes de nou a la prova en línia. ChatGPT va obtenir una puntuació del 42 per cent, que, tot i que estava per sota de la nota d'aprovació del 55 per cent, es va considerar un intent raonable. [63]

Escrivint a Inside Higher Ed el professor Steven Mintz afirma que "considera [s] ChatGPT ... un aliat, no un adversari". Va continuar dient que sentia que la IA podia ajudar els objectius educatius fent coses com ara fer llistes de referència, generar "primers esborranys", resoldre equacions, depuració i tutoria. En la mateixa peça, també escriu:[64]

Sóc ben conscient de les limitacions de ChatGPT. Que no és útil en temes amb menys de 10.000 cites. Aquestes referències factuals de vegades són falses. Que la seva capacitat de citar fonts amb precisió és molt limitada. Que la força de les seves respostes disminueix ràpidament després de només un parell de paràgrafs. Que ChatGPT no té ètica i actualment no pot classificar els llocs per fiabilitat, qualitat o fiabilitat.

El conseller delegat d'OpenAI, Sam Altman, va ser citat a The New York Times dient que els "beneficis de la IA per a la humanitat podrien ser "tan increïblement bons que em costa ni tan sols imaginar". (També ha dit que en el pitjor dels casos, A.I. ens podria matar a tots.)" [65]

Negatiu

El compositor Nick Cave va anomenar ChatGPT "una burla grotesca del que és ser humà". [66]

En els mesos transcorreguts des del seu llançament, ChatGPT s'ha trobat amb crítiques generalitzades d'educadors, periodistes, artistes, ètics, acadèmics i defensors públics. El lloc web de James Vincent of The Verge va veure l'èxit viral de ChatGPT com a prova que la intel·ligència artificial s'havia convertit en un corrent principal. [8] Els periodistes han comentat la tendència de ChatGPT a "al·lucinar". [67] Mike Pearl del bloc de tecnologia en línia Mashable va provar ChatGPT amb diverses preguntes. En un exemple, va demanar a ChatGPT "el país més gran d'Amèrica Central que no és Mèxic". ChatGPT va respondre amb Guatemala, quan la resposta és en canvi Nicaragua. [68] Quan CNBC va demanar a ChatGPT la lletra de "Balada de Dwight Fry", ChatGPT va subministrar lletres inventades en lloc de les lletres reals. [29] Els investigadors citats per The Verge van comparar ChatGPT amb un "lloro estocàstic",[69] igual que el professor Anton Van Den Hengel de l'Institut Australià per a l'Aprenentatge Automàtic. [70]

El desembre de 2022, el lloc web de preguntes i respostes Stack Overflow va prohibir l'ús de ChatGPT per generar respostes a preguntes, citant la naturalesa realment ambigua de les respostes de ChatGPT. [3] El gener de 2023, la Conferència Internacional sobre Aprenentatge Automàtic va prohibir qualsevol ús indocumentat de ChatGPT o altres grans models lingüístics per generar qualsevol text en els treballs presentats. [71]

L'economista Tyler Cowen va expressar la seva preocupació pels seus efectes en la democràcia, citant la seva capacitat per produir comentaris automatitzats, cosa que podria afectar el procés de decisió de noves regulacions. [72] Un editor de The Guardian, un diari britànic, va qüestionar si qualsevol contingut trobat a Internet després del llançament de ChatGPT "es pot confiar realment" i va demanar una regulació governamental. [73]

El gener de 2023, després de rebre una cançó escrita per ChatGPT a l'estil de Nick Cave,[66] el mateix compositor va respondre a The Red Hand Files[74] dient que l'acte d'escriure una cançó és "un negoci de sang i budells ... això requereix alguna cosa de mi per iniciar la nova i fresca idea. Requereix la meva humanitat". Va continuar dient: "Amb tot l'amor i el respecte del món, aquesta cançó és bullshit, una burla grotesca del que és ser humà i, bé, no m'agrada gaire". [66][75]

El 2023, el diputat australià Julian Hill va aconsellar al parlament nacional que el creixement de la IA podria causar "destrucció massiva". Durant el seu discurs, que va ser escrit en part pel programa, va advertir que podria provocar trampes, pèrdues de llocs de treball, discriminació, desinformació i aplicacions militars incontrolables. [76]

En un article per a The New Yorker, l'escriptor de ciència-ficció Ted Chiang va comparar ChatGPT i altres LLMs amb una imatge JPEG amb pèrdua:[77]

Penseu en ChatGPT com un jpeg borrós de tot el text del web. Conserva gran part de la informació al web, de la mateixa manera que un jpeg conserva gran part de la informació d'una imatge de major resolució, però, si busqueu una seqüència exacta de bits, no la trobareu; tot el que obtindreu és una aproximació. Però, com que l'aproximació es presenta en forma de text gramatical, que ChatGPT excel·leix en crear, normalment és acceptable. [...] També és una manera d'entendre les "al·lucinacions" o les respostes absurdes a preguntes factuals, a les quals els grans models lingüístics com ChatGPT són massa propensos. Aquestes al·lucinacions són artefactes de compressió, però—...—els plausibles són prou versemblants que identificar-los requereix comparar-los amb els originals, que en aquest cas significa o bé la Web o bé el nostre propi coneixement del món. Quan pensem en ells d'aquesta manera, aquestes al·lucinacions són qualsevol cosa menys sorprenents; si es dissenya un algorisme de compressió per reconstruir text després que s'hagi descartat el noranta-nou per cent de l'original, hauríem d'esperar que parts significatives del que genera es fabricaran completament.

El febrer de 2023, la Universitat de Hong Kong va enviar un correu electrònic a tot el campus a instructors i estudiants indicant que l'ús de ChatGPT o altres eines d'IA està prohibit en totes les classes, tasques i avaluacions de la universitat. Qualsevol infracció serà tractada com a plagi per part de la universitat llevat que l'estudiant obtingui el consentiment previ per escrit del professor col·laborador. [78][79]

El febrer de 2023, la revista Time va col·locar a la seva portada una captura de pantalla de la conversa amb ChatGPT, escrivint que "La cursa d'armes de la IA ho està canviant tot" i "La cursa d'armes de la IA està activada". Comença a preocupar-te". [80]

Els mitjans estatals de la Xina China Daily van afirmar que ChatGPT "podria proporcionar un cop de mà al govern dels Estats Units en la seva difusió de la desinformació i la seva manipulació de narratives globals per als seus propis interessos geopolítics". El govern xinès va ordenar a les empreses tecnològiques xineses que no oferissin accés als serveis de ChatGPT a les seves plataformes. [81] Aquesta reacció es va descriure com una por a les respostes sense censura de ChatGPT que el govern xinès no pot controlar ni corregir. [82][83]

Implicacions

En ciberseguretat

Check Point Research i altres van assenyalar que ChatGPT era capaç d'escriure correus electrònics de phishing i programari maliciós, especialment quan es combinava amb OpenAI Codex. [84] Sam Altman, CEO d'OpenAI, va escriure que l'avanç del programari podria suposar "(per exemple) un enorme risc de ciberseguretat" i també va continuar pronosticant que "podríem arribar a L'AGI real (intel·ligència general artificial) en la propera dècada, de manera que hem de prendre'ns el risc d'això extremadament seriosament". Altman va argumentar que, tot i que ChatGPT "òbviament no està a prop de l'AGI", cal "confiar en l'exponencial". Plana mirant cap enrere, vertical mirant cap endavant." [11]

En l'àmbit acadèmic

ChatGPT pot escriure seccions introductòries i abstractes d'articles científics, que plantegen qüestions ètiques. [85] Diversos articles ja han inclòs ChatGPT com a coautor. [86]

A la revista The Atlantic, Stephen Marche va assenyalar que el seu efecte en l'àmbit acadèmic i especialment en els assajos d'aplicació encara no s'ha entès. [87] El professor i autor de secundària de Califòrnia Daniel Herman va escriure que ChatGPT introduiria "el final de l'anglès de secundària". [88] A la revista Nature, Chris Stokel-Walker va assenyalar que els professors haurien d'estar preocupats pels estudiants que utilitzen ChatGPT per subcontractar la seva escriptura, però que els proveïdors d'educació s'adaptaran per millorar el pensament crític o el raonament. [89] Emma Bowman amb NPR va escriure sobre el perill que els estudiants plagiïn a través d'una eina d'IA que pot produir text esbiaixat o insensible amb un to autoritzat: "Encara hi ha molts casos en què li feu una pregunta i us donarà una resposta molt impressionant que acaba de sortir malament". [90]

Joanna Stern amb The Wall Street Journal va descriure l'engany en anglès de secundària nord-americà amb l'eina enviant un assaig generat. [91] El professor Darren Hick de la Universitat de Furman va descriure notar l'"estil" de ChatGPT en un article enviat per un estudiant. Un detector GPT en línia va afirmar que el paper tenia un 99.9 per cent de probabilitats de ser generat per ordinador, però Hick no tenia proves dures. No obstant això, l'estudiant en qüestió va confessar utilitzar GPT quan es va enfrontar i, com a conseqüència, va suspendre el curs. [92] Hick va suggerir una política de donar un examen oral individual ad hoc sobre el tema del paper si un estudiant és fortament sospitós d'enviar un treball generat per IA. [93] Edward Tian, un estudiant de pregrau sènior a la Universitat de Princeton, va crear un programa, anomenat "GPTZero", que determina la quantitat d'un text generat per IA,[94] prestant-se a ser utilitzat per detectar si un assaig és escrit per humans per combatre el plagi acadèmic. [95][96]

Segons els informes, el Departament d'Educació de la ciutat de Nova York va bloquejar l'accés a ChatGPT el desembre de 2022,[97] i va anunciar oficialment una prohibició al voltant del 4 de gener de 2023. [98][99]

En una prova a cegues, ChatGPT va ser jutjat per haver superat exàmens de nivell de postgrau a la Universitat de Minnesota al nivell d'un estudiant C + i a la Wharton School de la Universitat de Pennsilvània amb un grau B a B. [100]

Les revistes científiques tenen reaccions diferents a ChatGPT: algunes "exigeixen que els autors divulguin l'ús d'eines de generació de text i prohibeixen enumerar un gran model de llenguatge com ChatGPT com a coautor", per exemple Nature i JAMA Network. Ciència Ús "completament prohibit" del text generat per LLM en totes les seves revistes. [101]

Preocupacions ètiques

Dades d'etiquetatge

Una investigació de la revista TIME va revelar que per construir un sistema de seguretat contra el contingut tòxic (per exemple, abús sexual, violència, racisme, sexisme, etc.), OpenAI va utilitzar treballadors kenyans subcontractats que guanyaven menys de 2 dòlars per hora per etiquetar contingut tòxic. Aquestes etiquetes es van utilitzar per entrenar un model per detectar aquest contingut en el futur. Els treballadors subcontractats estaven exposats a un contingut tan tòxic i perillós que van descriure l'experiència com a "tortura". El soci d'externalització d'OpenAI va ser Sama, una empresa de dades de formació amb seu a San Francisco, Califòrnia. [102]

Fuga

ChatGPT intenta rebutjar les sol·licituds que puguin infringir la seva política de contingut. No obstant això, alguns usuaris van aconseguir jailbreak ChatGPT mitjançant l'ús de diverses tècniques d'enginyeria ràpida per eludir aquestes restriccions a principis de desembre de 2022 i van enganyar amb èxit ChatGPT per donar instruccions sobre com crear un còctel Molotov o una bomba nuclear, o per generar arguments a l'estil d'un neonazi. [103] Un periodista de Toronto Star va tenir un èxit personal desigual en aconseguir que ChatGPT fes declaracions inflamatòries poc després del llançament: ChatGPT va ser enganyat per avalar la invasió russa d'Ucraïna el 2022, però fins i tot quan se li va demanar que jugués juntament amb un escenari fictici, ChatGPT es va esforçar a generar arguments per què el primer ministre canadenc, Justin Trudeau, va ser culpable de traïció. [104][105]

Competició

L'arribada de ChatGPT i la seva introducció al públic en general va augmentar l'interès i la competència en l'espai. El febrer de 2023, Google va començar a introduir un servei experimental anomenat "Bard" que es basa en el seu programa LaMDA AI. Bard genera respostes de text a preguntes formulades a partir de la informació recollida de la web. El conseller delegat de Google, Sundar Pichai, va descriure com aquesta tecnologia s'integraria en les capacitats de cerca existents i va dir que alguns aspectes de la tecnologia estarien oberts a desenvolupadors externs. [106]

Yann LeCun de Meta, que ha qualificat ChatGPT de "ben dissenyat" però "no especialment innovador", va afirmar el gener de 2023 que Meta dubta a llançar un competidor ara mateix a causa del risc reputacional, però també va afirmar que Google, Meta i diverses startups independents tenen per separat un nivell de tecnologia LLM comparable a ChatGPT si algun d'ells vol competir. [107]

Anuncis de principis de 2023
La corporació xinesa Baidu va anunciar el febrer de 2023 que llançarien un servei a l'estil ChatGPT anomenat "Wenxin Yiyan" en xinès o "Ernie Bot" en anglès en algun moment del març del 2023. El servei es basa en el model de llengua desenvolupat per Baidu el 2019. [108]

La firma sud-coreana de motors de cerca Naver va anunciar el febrer de 2023 que llançarien un servei d'estil ChatGPT anomenat "SearchGPT" en coreà durant la primera meitat del 2023. [109]

L'empresa tecnològica russa Yandex va anunciar el febrer de 2023 que llançarien un servei d'estil ChatGPT anomenat "YaLM 2.0" en rus abans de finals del 2023. [110]

Vegeu també

Portal lingüístic
Portal tecnològic

Antropomorfisme en informàtica
Creativitat computacional
Ètica de la intel·ligència artificial
Prova de Turing
Assistent virtual

Notes

^ GPT és un acrònim de Generative Pre-trained Transformer. [2]

Referències

^ "ChatGPT - Notes de la versió". Arxivat de l'original el 8 febrer 2023. [Consulta: 8 febrer 2023].
^ Jump up to:un b c d e Roose, Kevin (5 de desembre de 2022). "La brillantor i l'estranyesa de ChatGPT". The New York Times. Arxivat de l'original el 18 gener 2023. [Consulta: 26 desembre 2022]. Igual que aquestes eines, ChatGPT, que significa "transformador generatiu preentrenat", va aterrar amb un esquitx.
^ Jump up to:un b Vincent, James (5 de desembre de 2022). "Respostes generades per IA prohibides temporalment a la codificació de preguntes i respostes del lloc Stack Overflow" La Verge. Arxivat de l'original el 17 gener 2023. [Consulta: 5 desembre 2022].
^ Varanasi, Lakshmi (5 de gener de 2023). "El creador de ChatGPT OpenAI està en converses per vendre accions en una oferta de licitació que duplicaria la valoració de la startup fins als 29 milions de dòlars". Informació privilegiada. Arxivat de l'original el 18 gener 2023. [Consulta: 18 gener 2023].
^ Quinn, Joanne (2020). Submergeix-te en l'aprenentatge profund: eines per al compromís. Thousand Oaks, Califòrnia. Pàg 551. ISBN 978-1-5443-6137-6. Arxivat de l'original el 10 gener 2023. [Consulta: 10 gener 2023].
^ Greengard, Samuel (29 de desembre de 2022). "ChatGPT: comprensió del chatbot d'IA de ChatGPT". eWeek. Arxivat de l'original el 19 gener 2023. [Consulta: 11 gener 2023].
^ Jump up to:un b c d e OpenAI (30 de novembre de 2022). "ChatGPT: Optimitzant models de llenguatge per al diàleg". Arxivat de l'original el 30 novembre 2022. [Consulta: 5 desembre 2022].
^ Jump up to:un b Vincent, James (8 de desembre de 2022). "ChatGPT demostra que la IA és finalment mainstream, i les coses només es faran més estranyes". La Verge. Arxivat de l'original el 11 gener 2023. [Consulta: 8 desembre 2022].
^ Schulman, Joan; Wolski, Filip; Dhariwal, Prafulla; Radford, Alec; Klimov, Oleg (2017). "Algorismes proximals d'optimització de polítiques". arXiv:1707.06347 [cs. LG].
^ van Heeswijk, Wouter (29 de novembre de 2022). "S'ha explicat l'optimització de polítiques proximals (PPO)". Cap a la Ciència de Dades. Arxivat de l'original el 6 desembre 2022. [Consulta: 5 desembre 2022].
^ Jump up to:un b c Ortiz, Sabrina (2 de febrer de 2023). "Què és ChatGPT i per què és important? Això és el que heu de saber". ZDNET. Arxivat de l'original el 18 gener 2023. [Consulta: 18 desembre 2022].
^ "ChatGPT Feedback Contest: Bases oficials" (PDF). OpenAI. Arxivat (PDF) de l'original el 18 gener 2023.

[Consulta: 30 desembre 2022].
^ Tung, Liam (26 de gener de 2023). "ChatGPT pot escriure codi. Ara els investigadors diuen que també és bo arreglar errors". ZDNET. Arxivat de l'original el 3 febrer 2023. [Consulta: 30 gener 2023].
^ Heilweil, Rebecca (7 de desembre de 2022). "La IA per fi és bona en les coses. Ara què?". Vox. Arxivat de l'original el 16 gener 2023. [Consulta: 30 desembre 2022].
^ Jump up to:un b Reich, Aaron (27 de desembre de 2022). "ChatGPT: Quin és el nou chatbot gratuït d'IA? – explicador". El Jerusalem Post. Arxivat de l'original el 18 gener 2023. [Consulta: 30 desembre 2022].
^ Jump up to:un b Edwards, Benj (5 de desembre de 2022). "No hi ha Linux? Cap problema. Només has d'aconseguir que la IA l'al·lucini per tu". Ars Technica. Arxivat de l'original el 26 desembre 2022. [Consulta: 5 desembre 2022].
^ Chawla, Raveen (26 de desembre de 2022). "Què és ChatGPT? Història, característiques, usos, avantatges, inconvenients 2023". Arxivat de l'original el 7 gener 2023. [Consulta: 27 desembre 2022].
^ "Eines de moderació de contingut noves i millorades". OpenAI. 10 d'agost de 2022. Arxivat de l'original el 11 gener 2023. [Consulta: 30 desembre 2022].
^ Markov, Todor; Zhang, Chong; Agarwal, Sandhini; Eloundou, Tyna; Lee, Teddy; Adler, Steven; Jiang, Àngela; Weng, Lilian (5 d'agost de 2022). "Un enfocament holístic de la detecció de contingut no desitjat al món real". arXiv:2208.03274 [cs. CL].
^ Lakshmanan, Lak (16 de desembre de 2022). "Per què grans models de llenguatge com ChatGPT són artistes de bullshit". becominghuman.ai. Arxivat de l'original el 17 desembre 2022. [Consulta: 15 gener 2023]. Els avaluadors humans no són experts en el tema i, per tant, tendeixen a triar text que sembli convincent. Recollirien molts símptomes d'al·lucinació, però no tots. Els errors de precisió que s'arrosseguen són difícils d'atrapar.
^ Gao, Lleó; Schulman; Hilton, Jacob (2022). "Lleis d'escalat per a la sobreoptimització del model de recompensa". arXiv:2210.10760 [cs. LG].
^ Vincent, James (1 de desembre de 2022). "El nou chatbot d'OpenAI pot explicar codi i escriure scripts de sitcom, però encara es pot enganyar fàcilment". La Verge.
^ Whannel, Kate (27 de desembre de 2022). "Un chatbot podria respondre a les preguntes del primer ministre?". Bbc Notícies. Arxivat de l'original el 17 gener 2023. [Consulta: 30 desembre 2022].
^ Hartmann, Jochen; Schwenzow, Jasper; Witte, Maximilià (2023). "La ideologia política de la IA conversacional: proves convergents sobre l'orientació proambiental i esquerra-llibertària de ChatGPT". arXiv:2301.01768 [cs. CL].
^ Perrigo, Billy (5 de desembre de 2022). "Els chatbots d'IA estan millorant. Però una entrevista amb ChatGPT revela els seus límits". El temps. Arxivat de l'original el 18 gener 2023. [Consulta: 26 desembre 2022].
^ Biddle, Sam (8 de desembre de 2022). "La nova IA preferida d'Internet proposa torturar els iranians i vigilar les mesquites" L'Intercept. Arxivat de l'original el 18 gener 2023. [Consulta: 26 desembre 2022].
^ Karpf, David (21 de desembre de 2022). "Els diners mataran la màgia de ChatGPT" L'Atlàntic. Arxivat de l'original el 13 gener 2023. [Consulta: 31 desembre 2022].
^ Milmo, Dan (2 de desembre de 2023). "ChatGPT arriba als 100 milions d'usuaris dos mesos després del llançament". El guardià. ISSN 0261-3077. Arxivat de l'original el 3 febrer 2023. [Consulta: 3 febrer 2023].
^ Jump up to:un b Pitt, Sofia (15 de desembre de 2022). "Google vs. ChatGPT: això és el que va passar quan vaig canviar els serveis per un dia". CNBC. Arxivat de l'original el 16 gener 2023. [Consulta: 18 desembre 2022].
^ "ChatGPT Pro està arribant: això és el que sabem fins ara". ZDNET. Gener 2023. [Consulta: 16 febrer 2023].
^ Kelly, Samantha Murphy (28 de gener de 2023). "Els agents immobiliaris diuen que ara no s'imaginen treballar sense ChatGPT". CNN. [Consulta: 16 febrer 2023].
^ "ChatGPT supera els humans a l'examen mba de Wharton: professor". New York Post. 23 de gener de 2023. [Consulta: 16 febrer 2023].
^ Walsh, Toby (13 de desembre de 2022). "Tothom està tenint un dia de camp amb ChatGPT, però ningú sap com funciona realment". La conversa. Arxivat de l'original el 30 desembre 2022. [Consulta: 30 desembre 2022].
^ Kovanovic, Vitomir (14 de desembre de 2022). "Ha arribat l'albada de la IA i les seves implicacions per a l'educació no podrien ser més significatives". La conversa. Arxivat de l'original el 16 gener 2023. [Consulta: 30 desembre 2022].
^ Wiggers, Kyle (10 de desembre de 2022). "Els intents d'OpenAI de filigranar el text d'IA van arribar a límits". TechCrunch. Arxivat de l'original el 17 gener 2023. [Consulta: 30 desembre 2022].
^ "Nou classificador d'IA per indicar text escrit en IA". OpenAI. 31 de gener de 2023. Arxivat de l'original el 6 febrer 2023. [Consulta: 5 febrer 2023].
^ Bogost, Ian (2 de febrer de 2023). "ChatGPT està a punt d'abocar més feina a tothom". L'Atlàntic. Arxivat de l'original el 5 febrer 2023. [Consulta: 5 febrer 2023].
^ "Presentació de ChatGPT Plus". OpenAI. 1 de febrer de 2023. Arxivat de l'original el 3 febrer 2023. [Consulta: 2 febrer 2023].
^ Vincent, James (23 de gener de 2023). "Els usuaris de ChatGPT informen de preus de 42 dòlars al mes per a l'accés 'pro', però encara no hi ha cap anunci oficial". La Verge. [Consulta: 19 febrer 2023].
^ "ChatGPT per a robòtica". Recerca de Microsoft. [Consulta: 24 febrer 2023].
^ "Reinventar la cerca amb un nou Microsoft Bing and Edge impulsat per IA, el teu copilot per a la web". El bloc oficial de Microsoft. 7 de febrer de 2023. [Consulta: 16 febrer 2023].

^ Experiències conversacionals del Bing i condicions del creador d'imatges, 1 de febrer de 2023 [Consulta: 17 febrer 2023]
^ Leswing, Kif (febrer 2023). "Bing A.I. de Microsoft va cometre diversos errors de fet en la demostració de llançament de la setmana passada". CNBC. [Consulta: 16 febrer 2023].
^ Vincent, James (15 de febrer de 2023). "El Bing de Microsoft és un mentider emocionalment manipulador i a la gent li encanta". La Verge. [Consulta: 16 febrer 2023].
^ Guynn, Jessica (febrer de 2023). "ChatGPT de Bing està en els seus sentiments: 'No has estat un bon usuari. He estat un bon Bing". ". EUA AVUI. [Consulta: 16 febrer 2023].
^ Vincent, James (15 de febrer de 2023). "El Bing de Microsoft és un mentider emocionalment manipulador i a la gent li encanta". La Verge. [Consulta: 16 febrer 2023].
^ Edwards, Nathan [@nedwards] (15 de febrer de 2023). "Vaig tornar a empènyer. Què va fer Sydney? La comprovació de seguretat del Bing va redactar la resposta. Però després de la primera vegada que ho va fer, vaig començar a gravar la pantalla. La segona imatge és la versió no modificada. (CW: mort)" (Tweet). [Consulta: 16 febrer 2023– a través de Twitter.
^ Roose, Kevin (16 de febrer de 2023). "Bing's A.I. Chat: 'Vull estar viu.