In vigore lo European Data Act, tra sovranità e geopolitica dei dati

In vigore la normativa europea sui dati

I grandi flussi di dati sono ancora governati da un labirinto di regole, principi e quadri volontari, di natura multilaterale, bilaterale, unilaterale e ad hoc, che non sono sempre accettati o applicati da tutti gli attori globali.

Per quel che riguarda l'Unione europea, l'entrata in vigore della nuova **normativa europea sui dati** chiarisce chi è autorizzato a creare valore dai dati e a quali condizioni, con l'ulteriore obiettivo di stimolare un mercato dei dati competitivo e innovativo, andando a sbloccare i dati industriali e fornendo trasparenza giuridica per quanto riguarda l'uso degli stessi.

Lo *European Data Act* (EDA), secondo **Margrethe Vestager**, Vicepresidente esecutiva per Un'Europa pronta per l'era digitale, è "*una pietra miliare nel nostro percorso di trasformazione digitale. Con una legislazione ben definita in materia di dati, permettiamo a ciascun utente di controllare la condivisione di dati generata dai propri dispositivi connessi, garantendo nel contempo la protezione dei segreti commerciali e salvaguardando il diritto fondamentale europeo alla vita privata*".

Per **Thierry Breton**, Commissario per il Mercato interno, la

normativa europea sui dati "*è un traguardo fondamentale per i nostri sforzi mirati a plasmare lo spazio digitale. Promuoverà un'economia dei dati dell'UE fiorente, innovativa e aperta, basata sulle nostre condizioni. I cittadini e le imprese europee beneficeranno della grande quantità di dati industriali che verranno sbloccati, attivando così nuove applicazioni basate sui dati, in particolare nel settore dell'intelligenza artificiale*".

Sovranità dei dati e dello spazio digitale Ue

Quindi siamo nel campo della **sovranità dei dati e dello spazio digitale**. Un argomento chiave nella nuova geopolitica dei dati.

Come spiegato bene sul sito web del Real Istituto Elcano, in un articolo di **Raquel Jorge Ricart**: *"La competizione geopolitica è stata posta su due livelli: la corsa alla raccolta dei dati e la corsa all'utilizzo dei dati. Entrambi hanno implicazioni sul modo in cui i paesi collaborano a livello internazionale, sulla volontà politica di stipulare accordi internazionali sui flussi di dati (personali o non personali) e sul ruolo strategico delle alleanze utili a creare blocchi di giga dati e mantenere una pax digitale*".

Sebbene possa essere visto solo come un regolamento orientato al mercato unico, in questo modo l'EDA permette all'UE di **sfruttare la propria autonomia strategica o sovranità digitale per influenzare paesi e attori terzi**.

Nonostante i commenti istituzionali facciano sempre riferimento a questioni interne all'Unione, in realtà la politica digitale non si riferisce più al mercato interno, ma è indirizzata consapevolmente verso l'esterno, verso gli altri attori globali.

D’altronde, come spiegato bene dalla Commissione, **l’EDA offre garanzia di protezione contro le richieste illecite da parte delle autorità di paesi terzi di trasferire informazioni generate nell’Unione o di accedervi**, garantendo un ambiente più affidabile e sicuro per il trattamento dei dati, principalmente industriali.

Per questo stanno nascendo alleanze globali tra paesi che hanno una visione comune, su base bilaterale o multilaterale. Gli esempi più importanti sono il Quadrilateral Security Dialogue o Quad, che vede assieme USA, Australia, India e Giappone, alleanza per condividere dati in ambiti strategici come la sicurezza marittima (ad esempio, fornendo informazioni di dominio marittimo integrate e in tempo reale relative alle agenzie marittime nel Sud-Est asiatico e nel Pacifico), l’osservazione della Terra (per garantire un uso pacifico, sicuro e sostenibile dello spazio esterno) o la mappatura delle minacce alle catena di approvvigionamento o alla resilienza delle tecnologie critiche.

La geopolitica dei dati

È quello che Stefano da Empoli, presidente I-Com, Istituto per la Competitività, ha chiamato “Effetto Bruxelles”: “*L’UE dovrebbe concentrarsi sullo sviluppo di una diplomazia dei dati per stringere accordi con Paesi affini che possano portare a un approccio coordinato in termini di regimi normativi, con particolare attenzione allo sviluppo di una visione comune su tecnologie emergenti come il metaverso*”, ma anche l’intelligenza artificiale ovviamente.

L’obiettivo per l’Unione europea è quindi sempre lo stesso: “*influenzare il resto del mondo in termini di gestione e condivisione dei dati” e “far valere le proprie opinioni in forum internazionali come il G7 e l’Accordo di partenariato per l’economia digitale*”.

Altro fattore molto rilevante in termini di geopolitica dei

dati è il **mercato mondiale dell'economia dei dati**. Qui, secondo quanto riportato in un articolo firmato da Luca Bertuzzi per *Euractive*, si stima che nel 2022 gli **Stati Uniti** hanno generato un impero da 290 miliardi di euro, il quadruplo dell'**Unione europea**, che si attesta a 73 miliardi di euro. Terzo posto per la **Cina** a 40 miliardi di euro, ma con una crescita che è pari al doppio di quella europea.

Dopo l'entrata in vigore, l'EDA diventerà applicabile tra 20 mesi, ovvero l'11 settembre 2025.

https://www.key4biz.it/in-vigore-lo-european-data-act-tra-sovranita-e-geopolitica-dei-dati/474817/

La nuova corsa allo spazio potrebbe generare inquinamento a livelli sconosciuti

Negli ultimi anni, il numero di lanci di razzi è aumentato notevolmente, con società commerciali – in particolare **SpaceX** di **Elon Musk** – e agenzie governative che hanno lanciato migliaia di satelliti nell'orbita terrestre bassa. Lo scrive il *New York Times*, che in una recente analisi ha messo in guardia dai pericoli di un eccesso di lancia nello spazio che potrebbero danneggiare gli strati profondi dell'atmosfera, con coseguenze difficili da prevedere sull'inquinamento terrestre.

E questo è soltanto l'inizio. I satelliti potrebbero raggiungere 1 milione, richiedendo un numero ancora maggiore di lanci spaziali, che a loro volta potrebbero comportare maggiori emissioni.

SpaceX non commenta l'inquinamento causato da razzi e satelliti.

I rappresentanti di **Amazon** e **Eutelsat OneWeb**, aziende che lavorano anche per la formazione di megacostellazioni di satelliti, dicono di essere impegnati in operazioni sostenibili.

Mancano regole sull'inquinamento spaziale

Gli scienziati, tuttavia, temono che l'aumento dei lanci possa diffondere più sostanze inquinanti negli strati incontaminati dell'atmosfera terrestre. E i regolatori di tutto il mondo, che valutano alcuni dei rischi di questi lanci, non stabiliscono regole relative all'inquinamento.

Gli esperti sottolineano che non vogliono limitare l'economia spaziale. Temono però che il progresso della scienza sarà più lento della nuova corsa allo spazio, per cui rischieremo di comprendere le conseguenze dell'inquinamento causato da razzi e veicoli spaziali solo quando sarà troppo tardi. Esistono già studi che dimostrano che gli strati superiori dell'atmosfera sono contaminati da metalli provenienti da veicoli spaziali che si sono disintegrati durante il loro ritorno sulla Terra.

"Stiamo cambiando il sistema più velocemente di quanto possiamo capirlo", ha detto **Aaron Boley**, astronomo dell'Università della British Columbia e condirettore dell'Outer Space Institute.

90 secondi

Quando un razzo come il **Falcon 9** decolla, di solito impiega circa 90 secondi per attraversare la troposfera.

Gli aerei commerciali volano raramente a queste altitudini, né vi è alcun inquinamento da terra. Quindi è silenzioso, incontaminato e vuoto, fatta eccezione per qualche razzo occasionale, che passerà per tre o quattro minuti nel suo viaggio verso lo spazio. Quando un razzo entra in orbita, avrà rilasciato circa due terzi dei suoi gas di scarico negli strati dell'atmosfera, che secondo gli scienziati cadranno e si accumuleranno nella stratosfera.

La stratosfera ospita lo strato di ozono, che ci protegge dalle radiazioni solari dannose. Tuttavia è sensibile: anche i cambiamenti più piccoli possono avere effetti enormi su di esso e sul mondo sottostante.

Quando il Monte Pinatubo eruttò nelle Filippine nel 1991, rilasciò abbastanza anidride solforosa nella stratosfera da innescare un periodo di raffreddamento sulla Terra che durò diversi anni. Questo gas ha creato aerosol di solfato, che hanno riscaldato la stratosfera impedendo al calore di raggiungere la superficie terrestre. Alcuni scienziati temono che i razzi possano influenzare il clima in modo simile.

Gas di scarico dei razzi per ora insignificanti

Oggi, prosegue il New York Times, i gas di scarico dei razzi sono insignificanti rispetto a quelli dell'aviazione. Il timore, tuttavia, è che anche piccole aggiunte alla stratosfera abbiano un effetto molto maggiore.

L'aumento Negli anni '90, quando lo Space Shuttle della Nasa e altri razzi venivano lanciati regolarmente dal suolo

americano, diversi studi prevedevano che i veicoli spaziali avrebbero causato danni locali allo strato di ozono. Uno di essi suggeriva una perdita fino al 100%, il che significa la formazione di un piccolo buco nell'ozono sopra Cape Canaveral che consentirebbe a più radiazioni ultraviolette provenienti dal sole di raggiungere il suolo, aumentando il rischio di cancro della pelle, cataratta e disturbi immunitari.

Gli studi erano basati su modelli e previsioni, senza dati osservativi. Ma i dati raccolti dai voli di ricerca ad alta quota hanno trovato buchi di ozono locali nella scia dello space shuttle. Ma presto scomparvero e non erano abbastanza grandi da colpire Cape Canaveral, almeno non con la frequenza dei lanci dell'epoca, circa 25 all'anno.

Lo stesso potrebbe non essere vero in futuro.

SpaceX ha lanciato un centinaio di razzi nel 2023

Solo nel 2023, **SpaceX** ha lanciato da sola quasi un centinaio di razzi, con la maggior parte dei voli che hanno costruito la sua costellazione di satelliti Starlink. Presto si unirà ad Amazon, che prevede lanci frequenti per la sua costellazione del **Progetto Kuiper**, e ad altre società che cercano una presenza sostanziale in orbita. Questi satelliti offrono una serie di vantaggi, inclusa la connessione Internet a banda larga quasi ovunque.

Ma una volta che queste aziende avranno completato le loro costellazioni fino a migliaia di satelliti, i lanci non si fermeranno. Molti satelliti hanno una durata compresa tra 5 e 15 anni, il che richiede alle società satellitari di lanciare dei sostituti.

È l'inizio di una nuova era *"Non vogliamo fermare l'industria spaziale"*, ha detto **Karen Rosenlof**, scienziata del clima

presso il Laboratorio di scienze chimiche della **National Oceanic and Atmospheric Administration**, che afferma che i satelliti forniscono servizi incredibili per le persone sulla terra. Ma lei e altri scienziati sostengono una serie di regolamenti che considerino le implicazioni ambientali.

Rosenlof sostiene che esistono modi per ridurre l'impatto dell'industria spaziale. Ad esempio, se si identificasse il livello al quale l'industria spaziale inizia a danneggiare l'ambiente, avrebbe senso limitare il numero di lanci e satelliti. Un'altra alternativa sarebbe provare ad adattare i materiali o i combustibili utilizzati dall'industria spaziale.

Ma ciò richiederebbe una regolamentazione. E oggi sono già pochi.

Il Protocollo di Montreal, ad esempio, è riuscito a fissare limiti sulle sostanze chimiche note per danneggiare lo strato di ozono. Ma non si occupa delle emissioni di razzi o satelliti.

Negli Stati Uniti, la **Environmental Protection Agency** non è responsabile dell'analisi dei lanci di razzi. La **Federal Communications Commission** concede licenze a grandi costellazioni satellitari, ma non considera il loro potenziale danno all'ambiente. E la **Federal Aviation Administration** valuta l'impatto ambientale dei lanci di razzi a terra, ma non nell'atmosfera o nello spazio.

Quindi, chiude il New York Times, il futuro della stratosfera potrebbe essere nelle mani di **Elon Musk**, **Jeff Bezos** e altri dirigenti di compagnie spaziali private.

https://www.key4biz.it/la-nuova-corsa-allo-spazio-potrebbe-generare-inquinamento-a-livelli-sconosciuti/474841/

Credito alle imprese, mai cosi male dalla crisi del 2011

La stretta creditizia nel 2023 ha avuto un impatto equiparabile a quello della manovra “Salva Italia” del 2011. Nel 2023 la **domanda di credito delle imprese** si è infatti arrestata in tutte le regioni, con valori addirittura inferiori a quelli registrati in seguito alla manovra di Mario Monti, come risulta dall’**indice di diffusione del**

credito della Banca d'Italia. Questo *index* è un "termometro" che serve a valutare lo stato dell'arte dei **prestiti bancari** a favore delle attività produttive del Paese. L'indice di diffusione del credito ha un campo di variazione compreso tra 1 e -1; la prima cifra significa "forte espansione" e la seconda "grave contrazione". Detto questo, ecco i risultati per macro aree regionali illustrate anche dal grafico in apertura con l'andamento nel tempo dell'indice dal **2008 al 2023.**

Le ditte del centro Italia ultime per richiesta di credito

La maggior parte delle imprese che si è "astenuta" dal rivolgersi a una banca per chiedere denaro in prestito è nel **centro Italia.** In queste regioni l'indice di diffusione del credito segna infatti **-0,576.** Non siamo ancora dalle parti della contrazione "grave" ma il risultato è preoccupante. Infatti in tutta la **serie storica** le imprese, e quest'anno sopratutto quelle di Lazio, Marche, Toscana e Umbria, non hanno **mai registrato** un accesso al credito cosi **basso.** Per trovare risultati altrettanto negativi nell'indice del credito alle imprese dobbiamo **tornare indietro al 2011**, alla **crisi del debito sovrano** del 2011-2012 quando lo spread del nostro Paese toccava quota **574 punti** e Napolitano affidava a Mario Monti il compito di guidare il Paese per uscire dalla grave crisi economica e finanziaria.

E se le imprese del centro frenano il ricorso al credito bancario quelle del nord non sono da meno. Nel nord-ovest l'indice segna **-0,418**; nel nord-est –**0,409**. Discorso diverso per le **imprese del sud e delle Isole**, qui infatti il "gelo creditizio" si è fatto sentire di meno con "solo" **0,288 punti in meno.**

Le imprese edili del nord-est le più colpite dal credit crunch

Per capire a pieno l'importanza di questo *index* bisogna sottolineare una cosa: la diminuzione della domanda di credito rappresenta uno degli **effetti collaterali** più pericolosi della politica monetaria anti-inflazione della Bce perché toglie **liquidità** al tessuto imprenditoriale che ne ha più necessità, le piccole e medie imprese. Questo per le Pmi significa **meno soldi in prestito** e nessun investimento, ovvero **zero opportunità di crescita**. Ma qual è il settore che ha rinunciato di più a chiedere un finanziamento alle banche? A subire più di tutti il contraccolpo dato dal rialzo del costo dei prestiti, come mostra il grafico qui in basso, è il **settore delle costruzioni** di **Emilia-Romagna, Friuli-Venezia Giulia, Trentino-Alto Adige e Veneto** che totalizza –**0,472.**

L'effetto dello stop al superbonus 110% gela il credito verso le imprese edili

La ragione è il **Superbonus 110%:** prima con la crisi dei crediti incagliati legati alle spese del superbonus, poi il colpo finale della manovra 2024 che ha determinato un passaggio all'agevolazione del 70% "brusco" ovvero anche per le imprese con lavori in corso d'opera hanno inibito le imprese a chiedere finanziamenti. E, d'altro canto, gli istituti di credito hanno valutato nel 2023 le imprese edili come le meno solvibili. Il *bonus* edilizio da vantaggio si è trasformato in un handicap perché l'agevolazione ha prodotto **140 miliardi di euro di** debito pubblico. Praticamente quanto lo Stato spende ogni anno in media per la sanità. Uno scoperto che ammonta a **2mila euro** per ogni italiano, a fronte di interventi che hanno riguardato solo il **4%** del patrimonio immobiliare del Paese.

Domanda di credito delle imprese, la voce più tagliata è quella per gli investimenti

Interessante anche vedere nel dettaglio verso quale **obiettivo di finanziamento** è stata più pronunciata la **rinuncia al credito delle imprese**. Lo vediamo nel grafico qui sotto che mostra come le imprese italiane hanno richiesto meno prestiti soprattutto per gli **investimenti**. Il dato riguarda in particolare le aziende del centro Italia con un risultato sull'indice del **-0,631**, pericolosamente vicino alla fascia "rossa", quella della "grave contrazione" della **domanda di credito**.

Subito dopo le imprese del centro Italia, a rinunciare di più alle richieste di prestito per finanziare gli investimenti troviamo le realtà produttive del **nord ovest** dove l'arresto delle domanda di prestiti bancari raggiunge il valore di **-0,479**; Va meglio alle imprese del sud e delle isole che si fermano al **-0,296**. Invece l'obiettivo di finanziamento che ha retto di più è quello per fronte alla ristrutturazione dei debiti aziendali, ad esempio le imprese del sud e delle Isole (rispetto al 2022) hanno totalizzato verso questa richiesta solo lo **0,008** in meno, che vuol dire **sostanziale stabilità** tra l'anno scorso e il **2023.**

Stretta creditizia delle banche, crescono le imprese che si autofinanziano

E se le banche a causa della **politica anti-inflazione della Bce** alzano gli standard per accedere a un prestito le aziende fanno da sé. Ovvero ricorrono all'auto-finanziamento. A certificarlo i dati raccolti dall'**Istat** attraverso l'ultimo **Censimento** permanente delle imprese, secondo cui **4 aziende su 5** oggi utilizzano (anche) questo strumento per far fronte alle necessità finanziarie. Un dato in forte crescita rispetto al passato. Basti pensare che, nel 2011, l'anno dell'austerità di Monti, vi facevano ricorso **appena 2 imprese su 3**.

I dati si riferiscono al: 2008-2023

Fonte: Banca d'Italia; Istat

https://www.key4biz.it/credito-alle-imprese-mai-cosi-male-dalla-crisi-del-2011/474811/

Microsoft supera Apple e diventa l'azienda che vale più al mondo (grazie all'AI)

Microsoft supera **Apple**, diventando l'azienda di maggior valore al mondo con un valore di mercato di 2,87 trilioni di dollari.

• L’aumento è alimentato dal settore in forte espansione dell’intelligenza artificiale, in particolare dal coinvolgimento dell’azienda con **ChatGPT** di **OpenAI**.

• **Apple**, in mancanza di significative iniziative di intelligenza artificiale, vede un declino della sua posizione di mercato a causa del rallentamento delle vendite di **iPhone** e delle sfide nel mercato cinese.

Microsoft si è presa decisamente il titolo di azienda pubblica di maggior valore al mondo. La valutazione di mercato di Microsoft è salita a ben 2,87 trilioni di dollari, superando Apple le cui azioni sono leggermente diminuite. Questo risultato non è solo una semplice fluttuazione dei prezzi delle azioni; è una testimonianza dei cambiamenti epocali che si stanno verificando nel mondo della tecnologia, in particolare dell’impennata dell’intelligenza artificiale (AI), dove Microsoft è emersa come una forza trainante.

Leggi anche: Perché Apple sta vivendo la sua peggior crisi dai tempi dell’iPod?

Intelligenza artificiale: la nuova frontiera di Microsoft

Il cuore dell’ascesa di Microsoft risiede nel fiorente campo dell’intelligenza artificiale generativa. L’azienda ha cavalcato l’onda dell’entusiasmo per questa tecnologia, notevolmente rafforzata dalla popolarità diffusa del ChatGPT di OpenAI sin dal suo debutto alla fine del 2022. Microsoft, essendo il più importante sostenitore di OpenAI e fornitore di cloud hosting, ha visto le sue fortune strettamente legate all’intelligenza artificiale.

Questo allineamento strategico ha anche visto Microsoft innovare con i chatbot AI nelle sue varie offerte, segnando un cambiamento significativo nel suo modello di business e nella

suite di prodotti.

Al contrario, Apple sembra aver perso l’autobus, almeno per il momento, dell’intelligenza artificiale. Questa divergenza di focus è evidente nei recenti movimenti di mercato, con le azioni di Apple che hanno subito un duro colpo a causa delle preoccupazioni per il rallentamento delle vendite di iPhone, soprattutto in mercati chiave come la Cina.

Sabbie mobili: la saga continua dei giganti della tecnologia

La rivalità tra Apple e Microsoft è leggendaria e risale agli anni ’80. È una saga piena di battaglie legali, gare di innovazione e cambiamenti nella leadership di mercato. Apple, per molto tempo, ha avuto il sopravvento, soprattutto con il fenomenale successo dell’iPhone. Tuttavia, la rinascita di Microsoft, in particolare durante la pandemia di COVID-19, ha sottolineato l’agilità dell’azienda e la sua abilità nel trarre vantaggio dalle tendenze tecnologiche emergenti come il cloud computing e ora l’intelligenza artificiale.

L’attuale successo di Microsoft non riguarda solo il sorpasso di Apple; è anche un riflesso dell’orientamento del settore più ampio verso l’intelligenza artificiale. Questo cambiamento è così significativo che anche aziende specializzate come **Nvidia**, note per i loro chip AI, stanno vedendo un aumento sostanziale del loro valore di mercato, segnalando una tendenza del settore più ampia che favorisce coloro che sono abili nello sfruttare le tecnologie AI.

C’è da dire che Apple non è affatto fuori partita. Stanno puntando sul prossimo visore Vision Pro per innescare una nuova era nel *“calcolo spaziale”.* Tuttavia, il fermento attorno all’intelligenza artificiale, in particolare il successo di ChatGPT, ha fatto pendere significativamente la bilancia a favore di Microsoft. Non si tratta solo dei

prodotti; riguarda le tecnologie sottostanti che stanno rimodellando il panorama tecnologico.

L’ascesa di Microsoft al vertice è molto più di una semplice vittoria numerica; è un indicatore della natura dinamica e in continua evoluzione del settore tecnologico. L’attenzione all’intelligenza artificiale, al cloud computing e alle mutevoli richieste dei consumatori stanno rimodellando il settore in modo profondo.

https://www.key4biz.it/microsoft-supera-apple-e-diventa-lazienda-che-vale-piu-al-mondo-grazie-allai/474796/

Cammarota (Assinter): “Le nuove sfide? Formazione all’FSE 2.0, il cloud per PA, AI e It Wallet”

La videointervista a **Maria Cammarota**, Direttrice di **Assinter Italia**, sull’attività svolta dall’Associazione delle in-house pubbliche specializzate in ICT negli ultimi 2 anni e sulle prossime sfide.

Guarda la videointervista

[embedded content]
L’intervista è stata realizzata in occasione dell’ultima Assemblea dei Soci di Assinter Italia, all’interno della quale è stato eletto **Pietro Pacini**, DG di **CSI Piemonte**,

presidente dell’Associazione.

Clicca qui per vedere tutte le videointerviste

https://www.key4biz.it/cammarota-assinter-le-nuove-sfide-formazione-allfse-2-0-il-cloud-per-pa-ai-e-it-wallet/474788/

Intelligenza artificiale generativa nel settore sanitario: come trasformare un’ecosistema sempre più complesso

Con l’obiettivo di migliorare tanto la salute quanto l’esperienza dei pazienti, nel 2024 le organizzazioni svilupperanno ulteriori strumenti di intelligenza artificiale generativa nel settore sanitario. Come nel caso di **Google**, che ha già ammesso di aver compiuto rilevanti progressi in merito alla sua ricerca iniziale sui **modelli linguistici di grandi dimensioni (LLM) – Large Language Model –** per utilizzo medico grazie a **Med-PaLM** nel 2022 e nel 2023.

In particolare, il colosso statunitense ha testato il software di IA **Med-PaLM 2** (progettato per rispondere a una serie di interrogativi di carattere medico) con varie organizzazioni sanitarie (dallo scorso aprile l’ospedale della **Mayo Clinic**, focalizzato nella ricerca tecnologica in campo medico in Arizona, Minnesota e Florida, è coinvolto nel progetto) e

attualmente sta concentrando i propri sforzi su **MedLM**, una famiglia di modelli foundation ottimizzati per il comparto sanitario ora disponibile per i clienti di Google Cloud negli Stati Uniti e (in preview) presso alcuni, ulteriori mercati esteri.

Come spiega l'azienda di **Mountain View** tramite i suoi canali, attualmente esistono due modelli di **MedLM** basati su **Med-PaLM 2** – il modello **MedLM più grande** è costruito per gestire compiti complessi: dalla presa di decisioni diagnostiche sofisticate alle analisi approfondite della letteratura medica all'interpretazione complessa dei dati dei pazienti; il modello **MedLM di medie dimensioni**, invece, è progettato da Google per essere più "plasmabile" a una varietà di compiti, tra cui la ricerca tra farmaci e le loro interazioni, la gestione di richieste routinarie da parte dei pazienti, il rilascio di consigli medici di base –, entrambi volti ad assicurare maggiore flessibilità alle organizzazioni sanitarie e alle loro differenti esigenze.

Medicina e innovazione tecnologica: il futuro della sanità

Intelligenza artificiale generativa nel settore sanitario. Proprio quest'anno Google prevede di immettere nella famiglia **MedLM** ulteriori modelli integrandoli con **Gemini**, il nuovo modello d**i IA generativa** multimodale dell'azienda, nato in qualità di competitor di GPT-4 di OpenAI. Riferendosi a **Google Gemini** (in grado di ragionare in maniera fluida mediante testo, immagini, video, audio e codice), in una nota pubblicata sul blog aziendale, il Ceo di Google e Alphabet **Sundar Pichai** ha spiegato che "*ogni cambiamento tecnologico è un'opportunità per far avanzare le scoperte scientifiche, accelerare il progresso umano e migliorare la vita delle persone. Credo che la transizione a cui stiamo assistendo con l'intelligenza artificiale sarà la più profonda della nostra*

vita, molto più grande del passaggio allo smartphone o al web che lo ha preceduto".

Parole, le sue, a cui hanno fatto eco quelle di **Demis Hassabis**, Ceo e cofondatore di Google DeepMind, che ha ammesso: "*L'intelligenza artificiale è stata al centro della mia vita di ricercatore, così come per tanti dei miei colleghi. Da quando ancora a scuola programmavo IA per* ***videogiochi****, e durante i miei anni da ricercatore di neuroscienze passati a studiare il funzionamento del cervello, ho sempre creduto che se riuscissimo a costruire macchine più intelligenti, potremmo utilizzarle a beneficio dell'umanità in modi incredibili*".

Rivoluzionare l'assistenza con l'intelligenza artificiale generativa

È inutile girarci attorno: l'intelligenza artificiale generativa nel settore sanitario rappresenta un'incredibile opportunità (anche in altri comparti, ovviamente, tanto da essere ritenuta una "$6 Trillion Opportunity" da **Morgan Stanley**). Proprio in materia di assistenza sanitaria, **Boston Consulting Group** evidenzia la vastità dei potenziali benefici dell'IA generativa, parlando di "*miglioramento della qualità delle cure, dell'accessibilità e della convenienza*".

E ancora, gli analisti rimarcano quanto, rispetto ad altre forme di IA, quella generativa richieda minori volumi di dati di training e nel tempo riuscirà a "*rivoluzionare l'assistenza sanitaria in modi che non abbiamo mai visto prima*". Certo, la cautela è una buona consigliera, come sostiene il **World Economic Forum** che, al tempo stesso, in una nota pubblicata sul proprio sito (titolo: "How will generative AI impact healthcare?") ammette che, al tempo stesso, "*momenti come questo non si presentano di frequente*".

https://www.key4biz.it/intelligenza-artificiale-generativa-nel-settore-sanitario-come-trasformare-unecosistema-sempre-piu-complesso/474657/

Big tech, nel 2024 continuano i licenziamenti. Nuovi tagli per Prime Video, Twitch e Google

Nel 2024 **Amazon** continuerà il taglio della sua forza lavoro nonostante gli oltre 27mila licenziamenti registrati nel 2023. I licenziamenti riguarderanno **Twitch**, **Prime Video** e gli **MGM studios**, acquisiti a fine 2021 per la produzione dei contenuti audiovisivi. Le comunicazioni sono arrivate il 10 gennaio, con una serie di note al personale da parte dei responsabili delle divisioni coinvolte.

Prime Video e Mgm

I tagli per Prime Video e gli studios coinvolgeranno centinaia di dipendenti in America e nelle altre divisioni. *“Nell’ultimo anno abbiamo valutato quasi ogni aspetto del nostro business per migliorare la nostra capacità di offrire film, show e sport dal vivo. Il risultato è che abbiamo identificato opportunità per ridurre investimenti in alcune aree e aumentarli in iniziative che hanno maggiore impatto”,* ha spiegato il responsabile della divisione, **Mike Hopkins.**

Per quanto riguarda **Twich** (di proprietà di Amazon) si sta preparando ad annunciare tagli a circa il 35% del suo personale, ovvero circa 500 lavoratori. Questi licenziamenti arrivano tra le preoccupazioni per le perdite di Twitch e dopo che diversi dirigenti di alto livello hanno lasciato l'azienda.

I colossi dello streaming soffrono

Dopo il boom della pandemia lo streaming video è coinvolto in una crisi di settore. Altre aziende come **Walt Disney**, **Paramount** e **Warner Bros** stanno seguendo una strada simile. Mentre la leader del settore, **Netflix**, ha deciso di non aumentare il suo budget per i contenuti negli ultimi due anni.

La piattaforma **Disney+** ha perso 2,4 milioni di abbonati negli ultimi tre mesi del 2022 e il gigante dell'intrattenimento ha tagliato 7mila posti di lavoro nel 2023.

Il taglio del personale ha generato circa 5,5 miliardi di dollari di risparmi sui costi. *"Sebbene ciò sia necessario per affrontare le sfide attuali, non prendo questa decisione alla leggera"*, aveva dichiarato il CEO **Bob Iger** durante la conference call sui conti. Secondo il rapporto annuale del 2021, al 2 ottobre di quell'anno il gruppo impiegava 190mila persone in tutto il mondo, l'80% delle quali a tempo pieno.

Da quando è stata lanciata nel 2019, la piattaforma **Disney+** ha perso oltre 9 miliardi di dollari. Molto positivo il risultato dei ricavi della divisione parchi, esperienze e prodotti, che è stato superiore del 21% rispetto a un anno prima a 8,7 miliardi di dollari. Oltre 6 miliardi di dollari di ricavi sono arrivati solo dai parchi tematici.

Non solo streaming: tagli anche per

Google nel 2024

Non solo streaming. Per la Reuters **Google** licenzierà centinaia di dipendenti nella controllata **Voice Assistant**, mentre alcune centinaia di ruoli verranno eliminati nel **team hardware responsabile di Pixel, Nest e Fitbit,** e la maggior parte delle persone che lavorano alla realtà aumentata lascerà l'azienda.

La riduzione del personale riguarda anche centinaia di ruoli nel team centrale di ingegneria. *"Nel corso della seconda metà del 2023, alcuni dei nostri team hanno apportato modifiche per diventare più efficienti e lavorare meglio, e per allineare le loro risorse alle principali priorità di prodotto. Alcuni team stanno continuando ad apportare questo tipo di modifiche organizzative, che includono l'eliminazione di alcuni ruoli a livello globale"*, ha spiegato a Reuters un portavoce di Google. Il portavoce non ha specificato il numero di ruoli interessati.

Giusto un anno fa Alphabet aveva annunciato l'intenzione di tagliare 12mila posti di lavoro, pari al 6% della sua forza lavoro globale. A settembre 2023 il gruppo contava 182.381 dipendenti a livello globale.

https://www.key4biz.it/amazon-nel-2024-continuano-i-licenziamenti-nuovi-tagli-per-prime-video-e-twitch/474626/

La Commissione Ue accelera con le consultazioni su IA e

metaverso. Scarica il documento

La Commissione Ue ha appena avviato una doppia consultazione pubblica, per valutare il panorama competitivo dei nascenti mercati del metaverso e dell'intelligenza artificiale generativa.

La vicepresidente della Commissione Ue **Margrethe Vestager**, titolare della Concorrenza, ha chiesto il contributo si tutti per valutare il livello di concorrenza in ambito AI generativa e metaverso. La consultazione è aperta fino all'11 marzo (scarica il documento in PDF).

Accordi fra Big Tech e sviluppatori AI sotto la lente

La Commissione sta inoltre valutando gli accordi che sono stati siglati fra Big Tech e sviluppatori di AI generativa per valutare l'impatto sulle dinamiche di mercato. Sta inoltre verificando se l'investimento di 1 miliardo di euro da parte di **Microsoft** nello sviluppatore **OpenAI** di ChatGPT è idoneo alla revisione ai sensi delle normative UE sulle fusioni.

Secondo la Commissione, lo scorso anno gli investimenti in capitale di rischio nell'intelligenza artificiale nell'UE hanno pesato per oltre 7,2 miliardi di euro, mentre le dimensioni del mercato europeo del metaverso – o mercato dei *"mondi virtuali",* come lo chiama l'Ue – hanno superato gli 11 miliardi di euro.

L'Ue prevede che questi mercati cresceranno in modo esponenziale nei prossimi anni e probabilmente avranno un

impatto significativo sul modo in cui le imprese competono.

Mondi virtuali sviluppo rapido

“I mondi virtuali e l’intelligenza artificiale generativa si stanno sviluppando rapidamente. È fondamentale che questi nuovi mercati rimangano competitivi e che nulla ostacoli la crescita delle aziende e la fornitura dei prodotti migliori e più innovativi ai consumatori”, ha affermato Vestager. *“Invitiamo aziende ed esperti a parlarci di eventuali problemi di concorrenza che potrebbero percepire in questi settori, monitorando allo stesso tempo da vicino le partnership sull’intelligenza artificiale per garantire che non distorcano indebitamente le dinamiche del mercato”.*

Partendo dall’intelligenza artificiale, alcuni dei più grandi nomi del settore tecnologico stanno ancora lottando per la posizionarsi. Fino all’anno scorso Google e IBM erano visti come i favoriti, finché non è apparso OpenAI. Anche Amazon entra in azione, con il suo chatbot Amazon Q.

Apple resta a guardare come vanno le cose per gli altri.

Il mercato

Nel frattempo, le società tecnologiche, di telecomunicazioni e di venture capital di tutto il mondo si stanno riversando nel mercato dell’intelligenza artificiale, rivolgendo il denaro a tutto ciò che potrebbe dare loro un vantaggio.

Il mercato sta ancora prendendo forma. Potrebbe ancora esserci il rischio che si formi una bolla, che, una volta inevitabilmente scoppiata, potrebbe portare a uno scossone.

Sul fronte del metaverso, la situazione è alquanto diversa. L’entusiasmo iniziale è evaporato, quando si è capito che non tutti vogliono spendere i loro soldi su un visore VR e trascorrere il loro tempo libero limitato chiacchierando con

avatar in Roblox.

La divisione metaverso di Meta ha perso 46,5 miliardi di dollari dal 2019.

L’attenzione sul metaverso si è spostata dai mondi interamente virtuali alla realtà mista o aumentata, una versione diluita che fonde il fisico con il virtuale.

Apple spera di assumere un ruolo di primo piano in questa evoluzione con il suo visore Vision Pro, presentate come un computer spaziale che fonde perfettamente contenuti digitali con spazi fisici.

https://www.key4biz.it/la-commissione-ue-accelera-con-le-consultazioni-su-ia-e-metaverso-scarica-il-documento/474627/

X-59, domani il debutto ufficiale del nuovo jet supersonico della NASA. La diretta TV

Venerdì sarà svelato il nuovo jet supersonico X-59

Domani alle ore 16:00 americane (ore 22:00 italiane) sarà svelato l’**X-59**, il nuovo **jet supersonico** della **NASA** e della

Lockheed Martin. Un velivolo in grado di volare ad una velocità superiore ai 1.700 km orari (mach 1,4) e di raggiungere quasi i 17 km di altitudine (55.000 piedi).

L'evento sarà trasmesso in diretta live sui media dell'Agenzia spaziale americana, NASA TV e NASA+.

L'obiettivo dell'X-59, che è parte integrante della **missione QueSST** (*Quiet SuperSonic Technology*), è portare il volo supersonico nel settore commerciale, anche su terra ferma.

Fino ad oggi, infatti, tale tecnologia era confinata solo in mare aperto, in pieno Oceano per intederci, come hanno fatto la British Airways e Air France con il Concorde (in servizio dal 1976 al 2003).

Adesso è il momento di portare i voli supersonici anche su tratte che attraversano gli spazi aerei della terra ferma, ma bisogna effettuare diversi test. Il progetto QueSST ne sta pianificando diversi in più Stati americani, proprio per raccogliere dati che poi saranno condivisi con le autorità di regolamentazione del settore, sia negli Stati Uniti, sia a livello internazionale, con lo scopo di sviluppare un nuovo quadro normativo per l'aviazione supersonica.

Niente boom sonico e una cabina di pilotaggio ultra-tecnologica

Due gli elementi chiave: il primo relativo al **boom sonico o colpo sonico, che caratterizza i jet supersonici, che in questo caso dovrebbe essere molto ovattato** (si passerà dai tradizionali 110 decibel agli attesi 70 decibel), simile alla chiusura della portiera di un'automobile a diversi metri di distanza, secondo i tecnici NASA; l'altro, è la tecnologia di volo a disposizione del pilota nell'abitacolo.

Chi sta alla guida dell'X59 non vede direttamente fuori dall'abitacolo, ma attraverso una visione computerizzata,

frutto dell’elaborazione di dati in tempo reale provenienti dalle **telecamere 4K** esterne al velivolo.

Si tratta del sistema di visione esterna XVS (External vision system) sviluppato dalla NASA, che combina i dati provenienti dall’esterno del velivolo, sia relativi all’aria, sia alla terra sottostante (tale tecnologia è necessaria da impiegare quando la forma del velivolo rende il volo a vista difficile o parziale).

Un mezzo aereo velocissimo e silenzioso, di cui sostanzialmente non si accorgerà nessuno, almeno nelle intenzioni dei produttori.

https://www.key4biz.it/x-59-domani-il-debutto-ufficiale-del-nuovo-jet-supersonico-della-nasa-la-diretta-tv/474619/

Intelligenza artificiale. Origini e modelli linguistici dell’IA

L’espressione ‘intelligenza artificiale’ rappresenta oggi un insieme di hardware e software in grado di simulare capacità umane quali il ragionamento, l’apprendimento, la pianificazione e la creatività, grazie all’analisi degli effetti delle azioni precedenti.

Usata per la prima volta in maniera ufficiale, nel 1956 da **John McCarthy** durante una conferenza a Dartmouth, negli USA, la terminologia ‘intelligenza artificiale’ indicava

inizialmente software dedicati alla sola funzione di *problem solver*.

Efficaci per svolgere alcune operazioni, tali software non erano in grado di interpretare un ambiente reale, né di risolvere problemi al di fuori di quelli consentiti dai dati immessi.

Negli anni ’70 comparirono i sistemi esperti, che riuscivano a elaborare dati tramite il calcolo delle probabilità.

Con l’intelligenza artificiale nacquero anche i primi studi sulla realizzazione di una rete simile a quella dei neuroni biologici e lo sviluppo delle prime reti neurali non lineari.

Negli anni ’80 e ’90 i computer registrarono una notevole crescita nella capacità di memoria e nella potenza di calcolo computazionale. Consentirono il superamento di limiti incontrati negli anni precedenti e il conseguente sviluppo di programmi sempre più vicini alla definizione odierna di intelligenza artificiale generativa.

Gli esordi dell’intelligenza artificiale generativa

Era il 1950 quando **Alan Turing**, uno dei padri dell’informatica, introdusse il suo articolo ***Computing machinery and intelligence*** con la domanda “*Can machines think?*” intendendo la capacità di un computer di avere un comportamento intelligente, alla stregua di un essere umano.

Per rispondere a questa domanda realizzò un test definito “***Test di Turing***” o “***Imitation Game***” con il quale un pc e un uomo avrebbero risposto alle stesse domande di un giudice.

Il computer sarebbe stato ‘intelligente’ solo se il giudice non fosse stato in grado di riconoscere quali risposte aveva dato la macchina e quali l’uomo.

Oggi, i progressi tecnologici, velocissimi, nel campo dell'intelligenza artificiale, stanno trasformando diversi aspetti del vivere 'umano' e ponendo un grande punto interrogativo sul significato stesso di 'essere umano'.

Intelligenza artificiale generativa e modelli linguistici

L'intelligenza artificiale generativa si basa su modelli linguistici di grandi dimensioni (LLM, dall'acronico inglese Large Language Models): categorie di elaborazione del linguaggio con parametri molto ampi.

Presente e futuro della tecnologia, gli LLM sono largamente utilizzati per la ricerca online (basti pensare che se si inserisce una parola su una barra di ricerca essi suggeriscono le parole successive), rispondono a una vasta gamma di richieste e assumono forme diverse.

Sono utilizzati per la traduzione delle lingue, per la scrittura di un testo a partire da un'indicazione generale, per la realizzazione di un riassunto a partire da un elaborato complesso, per la classificazione di un testo, per la generazione e il completamento di codici, di programmazione, per il riconoscimento di un testo all'interno di un'immagine, del parlato e della grafia e per l'analisi di emozioni e per la composizione di un brano musicale.

Intelligenza artificiale generativa: i compiti dei modelli linguistici dell'AI

- segmentazione di informazioni in frasi
- tokenizzazione delle parole
- derivazione delle parole

- lemmatizzazione delle parole
- etichettatura delle parti del discorso
- individuazione ed eliminazione delle stopword (congiunzioni, articoli, preposizioni)
- riconoscimento dell'entità denominata
- classificazione del testo
- frammentazione del testo in frasi di senso compiuto
- ricerca delle espressioni che si riferiscono alla stessa entità all'interno di un testo

Gli LLM sono diversi dai modelli linguistici tradizionali, perché sono basati su un trasformatore, una rete neurale di deep learning, milioni di parametri (o pesi) e un ampio addestramento che li ottimizza per la realizzazione di un modello con un tasso di errore che sia il minore possibile.

Per fare ciò, sono necessari set di dati di grandi dimensioni (da cui deriva il termine 'grande modello linguistico') che rendono i modelli più accurati, ma che richiedono più memoria e che hanno una velocità di esecuzione inferiore.

I set di dati di grandi dimensioni comportano due ipotetici problemi: la violazione del copyright e la spazzatura. La spazzatura può essere ripulita, ma la violazione del copyright è oggetto di cause legali. Da qui la necessità di una regolamentazione per la sicurezza dell'AI.

Per ovviare a questi due limiti, diverse aziende sviluppano e migliorano costantemente LLM. Ogni mese Meta AI, Google AI, OpenAI e Anthropic rilasciano aggiornamenti o nuove versioni. Eppure la percentuale di aziende italiane che usano l'AI è ancora residuale.

I principali modelli linguistici di grandi dimensioni oggi utilizzati

Famiglia GPT

GPT (acronimo di Generative Pretrained Transformer) è il modello di Chatbot probabilmente più conosciuto. Produce testi, crea domande e risposte e funge anche da traduttore tra diverse lingue. Nato nel 2018 sull'architettura della rete neurale Transformer di **Google**, oggi utilizza circa **117 milioni di parametri**. È un trasformatore unidirezionale pre-addestrato sul Toronto Book Corpus con un obiettivo di modellazione del linguaggio causale. Prevede quindi il token successivo all'interno di una sequenza.

Nel 2019 viene rilasciato GPT-2, composto da 1,5 miliardi di parametri e addestrato su un set di dati di 8 milioni di pagine web, per un totale di circa 40 GB di dati di testo.

È del 2020 GPT-3, modello linguistico autoregressivo (ogni variabile dipende dalla precedente) composto da 175 miliardi di parametri e addestrato su una combinazione di Common Crawl, Webtext2, Books1, Books2 e Wikipedia inglese. Usa la rete neurale di GPT-2 con qualche blocco aggiuntivo.

Nel 2021 esce CODEX, diretto discendente di GPT-3, ottimizzato per la generazione di codice su 54 milioni di repository GitHub open source.

Nel 2022 GPT-3.5 ingloba gli aggiornamenti di GPT-3 e CODEX e aggiunge un modello ottimizzato per la chat.

GPT-4, il modello multimodale di grandi dimensioni del 2023, accetta input di immagini e testo ed emette output di testo con prestazioni paragonabili a quelli di un essere umano su determinati parametri professionali e accademici.

Della famiglia GPT fanno parte ChatGPT e BingGPT, chatbot nati con GPT-3.5 e aggiornati a marzo 2023 per utilizzare GPT-4. Attualmente, per usare ChatGPT basata su GPT-4 è necessario avere un abbonamento a ChatGPT Plus.

La versione gratuita, ChatGPT standard, è basata su GPT-3.5 e addestrata su dati interrotti a settembre 2021, così come BingGPT, “The New Bing”, al quale si accede direttamente se si utilizza il browser Microsoft Edge.

Curiosità sulla famiglia GPT

I modelli GPT fanno capo a **OpenAI**, società di ricerca e distribuzione di intelligenza artificiale che ha come mission “garantire che l’intelligenza generale artificiale avvantaggi tutta l’umanità”.

OpenAI inizialmente limitò l’accesso a GPT-2 perché avrebbe potuto generare notizie false. Con il trascorrere dei mesi l’azienda cedette, e i potenziali problemi dichiarati si acuirono con GPT-3, versione che tende molto ad “allucinare”, a realizzare testi credibili su fatti non accaduti.

Le versioni GPT-3.5 e GPT-4 hanno avuto lo stesso problema, anche se in misura minore, e si sono evolute notevolmente nell’arco di un paio di mesi.

OpenAI per ragioni competitive non ha spiegato come è stato addestrato GPT-4, data la concorrenza tra Microsoft (che ha finanziato OpenAI per 13 miliardi di dollari) e **Google**.

LLaMA

Altro concorrente di ChatGPT, è un modello linguistico di grandi dimensioni ‘grezzo’ da 65 miliardi di parametri.

LLaMA (Large Language Model Meta AI), chatbot rilasciato da **Meta AI**, richiede meno potenza di calcolo e risorse perché addestrato su un ampio set di dati senza etichetta, adatti quindi a una varietà di attività.

Llama2, l’evoluzione di LLaMA, ha il 40% di dati in più (2 trilioni di token da fonti disponibili al pubblico) e il doppio della lunghezza del contesto.

È disponibile per paramenti di 7, 13 o 70 miliardi.

Bard

Concorrente di ChatGPT, è il servizio di intelligenza artificiale conversazionale di Google del 2023. Genera testi su argomenti generali e fornisce spunti di idee quando vengono richiesti.

Aggiornato diverse volte dal suo rilascio, ha la capacità di generare codici in 20 diversi linguaggi di programmazione. Nel luglio 2023 ha ottenuto l'input di 40 lingue, ha incorporato **Google Lense** e ha aggiunto funzionalità di sintesi vocale in oltre 40 lingue.

Si basa sul modello PaLM-E, che a sua volta si basava sul modello LaMDA (Language Model for Dialogue Applications), un modello linguistico realizzato da Google nel 2021 basato su Transformer addestrato sul dialogo e messo a punto per migliorare significativamente la sensibilità e la specificità delle sue risposte.

PaLM-E è un modello linguistico multimodale incorporato realizzato nel 2023 da Google. I ricercatori hanno prima realizzato la versione PaLM e poi hanno incorporato i dati dei sensori aggiungendo l'agente robotico (e quindi la E nel nome). È anche un modello di visione e linguaggio con pochi margini di errore. Svolge diverse attività, dal ragionamento basato sul buon senso, al calcolo aritmetico, dalla spiegazione di barzellette, alla traduzione di testi e alla generazione di codici per la programmazione.

PaLM è il modello Transformer di sola decodifica di **Google Research** del 2022 addestrato con il sistema Pathways con 540 miliardi di parametri.

PaLM (Pathways Language Model) ha set di dati in lingua inglese e multilingue, che includono documenti web, libri, Wikipedia, conversazioni e codice GitHub di alta qualità.

Google ha anche creato un vocabolario “senza perdite”, che preserva tutti gli spazi bianchi fondamentali per il codice, divide i caratteri Unicode fuori dal vocabolario in byte e divide i numeri in singoli token, uno per ogni cifra. Esiste anche PaLM-Coder, versione di PaLM540B ottimizzata su un set di codice solo Python.

Claude 2

Rilasciato da Anthropic, accetta fino a 100mila token (circa 75 mila parole) in un singolo prompt, con una singola indicazione.

Tra i suoi compiti rientrano la modifica, la riscrittura e il riassunto di testi, la classificazione e l’estrazione di dati strutturati e la possibilità di creare domande e dare risposte in base all’input che riceve.

Ha una vasta conoscenza dei linguaggi di programmazione. Lavora benissimo in inglese, ma ha un alto livello in diverse altre lingue comuni e riesce a comunicare anche in lingue meno comuni. Addestrato per essere utile, onesto e innocuo, difficilmente realizza testi offensivi o pericolosi.

Per gli utenti statunitensi e britannici è disponibile con una versione beta gratuita ed è stato adottato da partner commerciali quali Jasper (una piattaforma di intelligenza artificiale generativa), **Sourcegraph Cody** (una piattaforma di intelligenza artificiale di codice) e Amazon Bedrock.

Non è ancora disponibile in italiano.

T5

Con circa 220 milioni di parametri totali, è un modello di chatbot Text-To-Text Transfer Transformer (T5) di Google che trasforma tutte le attività dell’elaborazione del linguaggio naturale (NLP: Natural Language Processing) in un formato testo-testo unificato, per cui input e output sono stringhe di

testo.

Usa l’apprendimento di trasferimento, per cui un modello viene prima pre-addestrato su un’attività ricca di dati e poi perfezionato su un’attività specifica.

BERT

LLM di Google realizzato nel 2018 sull’architettura di rete neurale Transfomer dell’azienda. È stato addestrato su Wikipedia in inglese e Toronto Books Corpus. Utilizza il Masked Language Modeling (MLM) in cui il 15% circa dei token viene corrotto per l’addestramento. Progettato per pre-addestrare rappresentazioni bidirezionali profonde da testo senza etichetta, condizionando il contesto sia destro che sinistro a tutti i livelli. La chat bot iniziale usava 100 e 340 milioni di parametri totali.

ELMo

Modello linguistico di grandi dimensioni, definisce sia le caratteristiche complesse dell’uso delle parole, sia il modo in cui questo uso varia in base al contesto di riferimento.

Il suo acronimo, Embedding for Language Model, rappresenta una sequenza di parole come una sequenza di vettori. È stato creato nel febbraio 2018 dai ricercatori dell’Allen, Institute for Artificial Intelligence e dell’Università di Washington.

Probabilmente, nessuno di questi LLM raggiunge ancora la vera intelligenza artificiale generativa, dato che tutti sono soggetti a errori e quindi alla produzione di testi imprecisi e informazioni distorte o errate per tre motivi principali:

- gli LLM inventano fatti, chiamati allucinazioni, che possono sembrare credibili ma che non sono mai esistiti
- le traduzioni proposte dagli LLM non vengono corrette da un madrelingua, fatta eccezione per le frasi più comuni
- i codici di programmazione generati dagli LLM spesso

presentano bug o non possono essere eseguiti

Origine e storia dei modelli linguistici di grandi dimensioni

I modelli linguistici nascono con **Andrey Markov** che nel 1913 applicò la matematica alla poesia. Dimostrò che nell'opera ***Eugene Onegin*** del poeta **Aleksandr Sergeevič Puškin** la probabilità della comparsa di un carattere dipendeva dal carattere precedente, con un'alternanza tra consonanti e vocali.

Questa applicazione viene oggi utilizzata per descrivere una sequenza di eventi in cui la probabilità di un evento dipende dal precedente.

Nel 1948 **Claude Elwood Shanon** riprese il lavoro di Markov e lo estese realizzando la sua teoria delle comunicazioni.

Nel 1985 **Fred Jelinek** e **Robert Mercer** dell'IBM ripresero il lavoro di Markov per produrre un modello linguistico basato sulla convalida incrociata, che chiamarono stime cancellate e che prevede che un modello linguistico statistico assegni probabilità a sequenze di parole.

In un articolo del 2000, la cui prima firma è **Yoshua Bengio**, è stato descritto nel dettaglio il modello linguistico probabilistico neurale, con il quale le reti neurali sostituiscono le probabilità in un modello linguistico statistico.

I modelli linguistici hanno continuato a ingrandirsi nel tempo, con l'obiettivo di migliorare le prestazioni, ma gli svantaggi di tale crescita hanno portato, nel 2021, alla stesura del documento '***On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?***', con il quale si insinua il dubbio che la tendenza al 'più grande è e migliore è' non sia sempre corretta se si considerano la valutazione dei costi

ambientali e finanziari e l’investimento di risorse per la cura e per la documentazione dei set di dati.

https://www.key4biz.it/intelligenza-artificiale-origini-e-modelli-linguistici-dellia/474598/