

UNA DATA IMPORTANTE PER L'AI: 3 AGOSTO 2023

La possibilità di creare immagini e testi del tutto nuovi ha costituito un salto tecnologico che ha terremotato l'economia e i mercati finanziari, come abbiamo già visto. Ma le grammatiche generative per funzionare hanno bisogno di essere alimentate e cioè di imparare quel che si dice sul mondo fagocitando informazioni che sono catturabili in rete. Sono fonti disponibili a chiunque, ma in realtà alcune di queste informazioni depositate in rete sono coperte da diritti d'autore: articoli, libri, immagini, filmati, e così via. Dietro a molto di ciò che si trova su internet ci sono degli scienziati, ricercatori, scrittori, scenografi, fotografi, artisti e, spesso, i media a cui questi autori hanno venduto i diritti del loro ingegno. Fatto sta che, in questa prospettiva, il 3 agosto 2023 sarà probabilmente una data importante, uno spartiacque. Un tempo queste immagini venivano considerate libere, come quando Google nel 2012 aveva addestrato una grande rete neurale di 16mila processori a riconoscere le immagini di gatti distinguendole da tutte le altre immagini. Le cose sono cambiate quando un importante giornale come il New York Times si è accorto che i suoi articoli venivano sistematicamente usati per alimentare le grammatiche generative. Il 3 agosto 2023 che il New York Times ha cambiato i termini di servizio vietando l'uso dei suoi contenuti: testi, foto, video e quant'altro. Questo divieto vale per qualsiasi programma software che voglia sfruttare quanto pubblicato dal quotidiano. In particolare è espressamente fatto divieto di usare questo materiale per l'addestramento di un sistema di apprendimento automatico o di intelligenza naturale. Proprio quell'apprendimento diventato necessario a software come ChatGpt, creato dalla società OpenAI (oggi è il programma generativo più usato), che ha bisogno di essere alimentato con miliardi di dati. Ma se questi dati vengono presi indistintamente dalla rete si può incappare anche in informazioni protette dal diritto d'autore non solo dei singoli creatori ma anche del canale con cui sono stati depositati, come appunto è il caso del New York Times. Il giornale ha fatto e fa molta fatica per procurarsi e controllare tutto quel pubblica e, ora come ora, basterebbe abbonarsi a un prezzo risibile per saccheggiare questo ben di Dio. Di qui il cambiamento dei termini di servizio.

Questa strategia, a partire dal 3 agosto, si sta diffondendo gradualmente presso vari media durante la seconda metà del 2023. A questo punto, per le società sempre più numerose che offrono servizi di software generativi sono possibili sostanzialmente due vie. La prima è negoziare con le fonti dei dati, per esempio, con il New York Times, pagando i diritti così da poter utilizzare i loro contenuti: questa è la via classica, tradizionale, usata da sempre. La seconda via consiste nel creare da zero il materiale su cui addestrare le reti, come racconta il 21 luglio 2023 Madhumita Murgia sul Financial Times anticipando di due settimane proprio la decisione del New York Times che era nell'aria (<https://www.ft.com/madhumita-murgia>).

In certo qual modo i software generativi incominciano a incontrare problemi proprio per il loro affermarsi travolgente. Il lancio di ChatGPT della società OpenAI, controllata da Microsoft, ha avuto un tale successo che nel 2023 molte altre aziende hanno seguito il pioniere: ancor oggi è il programma generativo più seguito. La corsa finora è guidata da Google, Microsoft e Meta, ma altre case si stanno aggiungendo. Il materiale utilizzato per allenarsi e imparare è costituito dai dati che si trovano su Internet. Questo materiale è minacciato non solo dai regolamenti rigorosi che l'Europa sta introducendo, ma anche dalle richieste di diritti d'autore e dagli scioperi di organizzazioni, come gli scenografi statunitensi, che si sentono assediati e, magari, in parte o del tutto scalzati da questi nuovi sistemi "creativi".

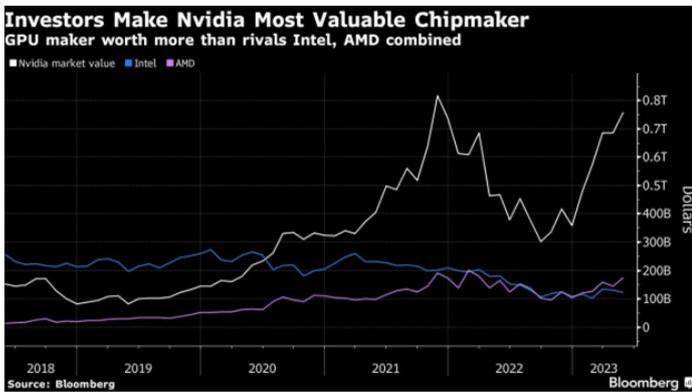
Aidan Gomez, addottoratosi a Oxford, fondatore e capo della società Cohere, commenta così questa seconda via che è spesso chiamata "sintetica" perché l'alimentazione è costruita e non racimolata in rete: "Se si potesse avere veramente tutti i dati dalla rete già ripuliti, questo in teoria sarebbe fantastico. Ma in realtà il web è pieno di rumore e di informazioni fuorvianti, se non infondate, vere e proprie frottole, e quindi non è una fonte ottimale per addestrare i sistemi generativi".



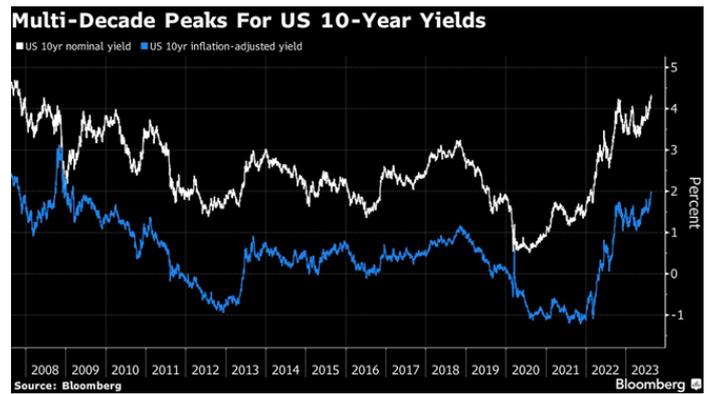
Prof. Paolo Legrenzi
Professore emerito di
psicologia cognitiva
presso l'Università Ca'
Foscari di Venezia

Per la verità già oggi i sistemi generativi, alimentati dalla rete, fanno meglio della media dei candidati “umani” nei test di ammissione all’università o negli esami per accedere alle professioni. Però, se vogliamo che costituiscano una vera e propria rottura innovativa, affidabile a lungo termine, in ambiti come la medicina, le scienze e le tecnologie, le basi-dati devono essere completamente pulite, affidabili, documentate e controllate. Di conseguenza queste possono essere create o arruolando scienziati, esperti e professionisti preparati in ogni campo dello scibile, oppure comprandole da università e centri di ricerca, banche, dalle grandi società finanziarie, farmaceutiche, e da ogni altro ambito di interesse. Fatto sta che in questo modo i software generativi diventano molto costosi e si tratta di trovare il punto di equilibrio più conveniente tra prezzo e precisione. Una via di mezzo possibile consiste nel far dialogare due sistemi su un dato argomento, poniamo la trigonometria, in modo che si migliorino l’un l’altro. Alla fine un esperto umano può limitarsi a fare un controllo della correttezza del risultato raggiunto. Sono tutte forme per ottenere banche dati “sintetiche”, evitando le imprecisioni e gli errori che si trovano in rete oppure la necessità di ottenere una qualità più elevata pagando i diritti di autore. Infine, un altro problema, discusso a lungo da Madhumita Murgia e da altri esperti, dipende dalla constatazione che, col passare del tempo: “il contenuto del web sarà sempre più generato dall’intelligenza artificiale e questa, mescolata a tutte le altre fonti, porterà a un intreccio inestricabile con conseguente degrado complessivo”.

Come vedremo nella prossima lezione tutte queste realtà e novità impatteranno anche sulla consulenza finanziaria.



Una sorta di paradosso: l’azienda che produce le componenti materiali che fanno funzionare i sistemi generativi è l’unica azienda manifatturiera che non solo ha quadruplicato il valore in borsa ma ha anche battuto i concorrenti tradizionalmente costruttori di chip come Intel. Fonte: Bloomberg modificata.



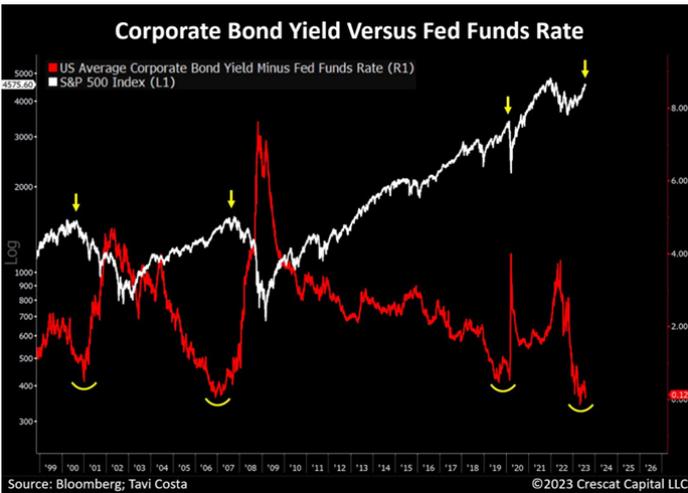
Il rendimento reale sul decennale USA è tornato ai livelli di inizio secolo e quindi si riafferma come una possibile componente di un portafoglio, soprattutto se mostrerà di non essere correlato con il rendimento azionario. In quest’ultimo caso è preferibile un portafoglio tutto composto da azioni statunitensi perché sui tempi medi e lunghi nulla riesce a batterlo, come ripeto da più di un decennio, soprattutto se composto con fondi che non costano quasi nulla. Fonte: Bloomberg modificata.



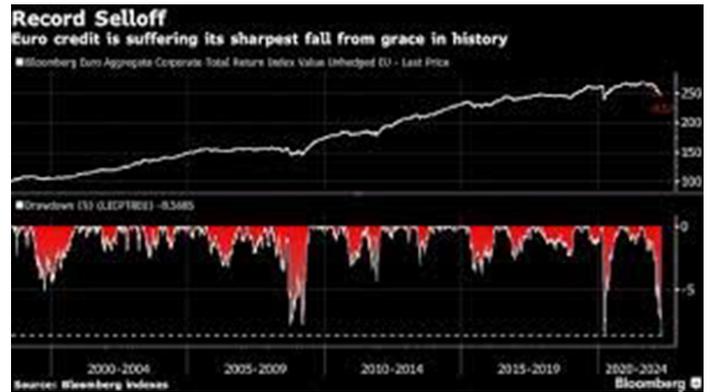
Pronti alla recessione? E’ dall’inizio del secolo che il differenziale dei rendimenti sui 2/10 anni non è stato così basso al punto da aver superato la temibile inversione. Fonte: Bloomberg modificata.



I prezzi al consumo hanno avuto nell’ultimo anno il più forte incremento dall’inizio degli anni ottanta. Fonte: Bloomberg modificata.



Nella figura si vede l'andamento dello S&P 500 in rapporto al differenziale tra i corporate bond e i treasury. Fonte: Bloomberg modificata.



Salgono i tassi e l'Euro credit è in forte calo. Fonte: Bloomberg modificata.



I rendimenti dei Treasury a 30 e a 10 anni è in forte salita dal 2020. Fonte: Bloomberg modificata.

Prof. Paolo Legrenzi

Laboratorio di Economia Sperimentale
 Paolo Legrenzi (Venezia, 1942) è professore emerito di psicologia cognitiva presso l'Università Ca' Foscari di Venezia, dove coordina il Laboratorio di Economia Sperimentale nato dalla partnership con GAM. È membro dell'Innovation Board della Fondazione Università Ca' Foscari.

Per maggiori informazioni visitate il sito GAM.com



Importanti avvertenze legali:

I dati esposti in questo documento hanno unicamente scopo informativo e non costituiscono una consulenza in materia di investimenti. Le opinioni e valutazioni contenute in questo documento possono cambiare e riflettono il punto di vista di GAM nell'attuale situazione congiunturale. Non si assume alcuna responsabilità in quanto all'esattezza e alla completezza dei dati. La performance passata non è un indicatore dell'andamento attuale o futuro.