UNIVERSITY
OF TRENTO - Italy

Intelligente sarà lei!

Il paradigma della IA e l'Umanistica digitale

Luisa Mich

luisa.mich@unitn.it

Artificial Intelligence (AI) is …

- A hype, a buzzword

"AI is a ridiculously broad umbrella these days", M. Karasick, IBM

Yesterday

Figure 1. Top 10 Strategic Predictions for 2018 and Beyond

Search	Disruption	Blockchain	Trust	AI
30% Visual and Voice Search Drive Commerce Revenue	**5 of 7** Digital Giants Self-Disrupt	**$1 Billion** Cryptocurrency Value Created in Banking	**50%+** Consuming More False Information Than True	**Real or Fake?** "Counterfeit Reality"
2021	2020	2020	2022	2020
50% Bot Development Over Mobile App Development	**40%** IT Versatility With People-Centric Roles	**2.3M** Net-New Jobs Versus 1.8M Eliminated	**95%** New Products Contain IoT	**Half** IoT Security Spend on Remediation
2021	2021	2020	2020	2022
AI	Talent	AI	IoT	Security

© 2017 Gartner, Inc.

AI = artificial intelligence; IoT = Internet of Things

SOURCE: GARTNER (SEPTEMBER 2017)

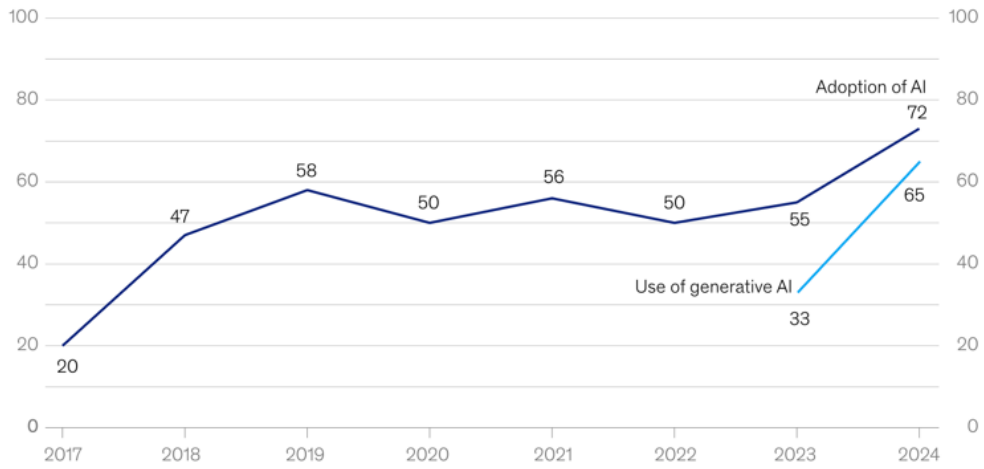
AI = artificial intelligence; IoT = Internet of Things

SOURCE: GARTNER (SEPTEMBER 2017)

Today

AI adoption worldwide has increased dramatically in the past year, after years of little meaningful change.

Organizations that have adopted AI in at least 1 business function,[1] % of respondents



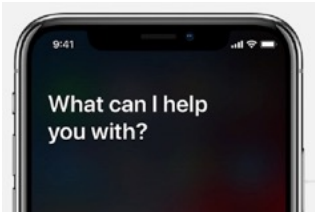
[1]In 2017, the definition for AI adoption was using AI in a core part of the organization's business or at scale. In 2018 and 2019, the definition was embedding at least 1 AI capability in business processes or products. Since 2020, the definition has been that the organization has adopted AI in at least 1 function.
Source: McKinsey Global Survey on AI, 1,363 participants at all levels of the organization, Feb 22–Mar 5, 2024

McKinsey & Company

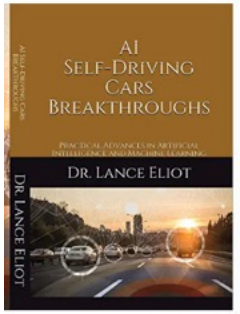
Artificial Intelligence is a …

- tool
- system
- agent
- discipline
- brand
- ChatGPT
- other

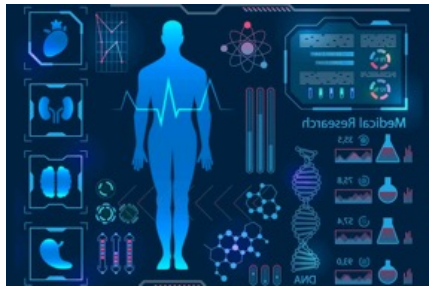
AI is everywhere

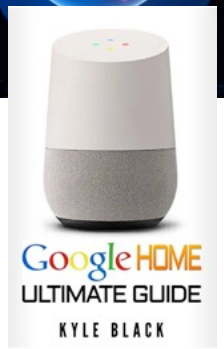


https://voicebot.ai

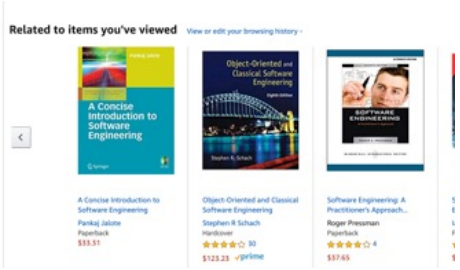


alexa





Solving My Wife's Problem 'What Should I Wear Today?' With AI
https://towardsdatascience.com › solving-my-wifes-problem-what-should-i...
Sep 14, 2019 - faAi, Fashion Assistant **Artificial Intelligence** ... So that it can answer Yumi's follow up question "**Why do you recommend** me that?" normally after ...



Toward AI fashion design: An Attribute ...
sciencedirect.com

Google Maps is already using a very embryonic form of artificial intelligence in its current maps. Since the algorithm automatically chooses the best route it can take, relying on the different projections of the time frames.

‘AI’ is (mis)used ...

- to refer to a technology, a **tool** or a **system**, e.g. (chat)bot, that *can* ‘understand’, reason, learn and interact in a ‘natural way’
- as a synonymous of ‘automation’, to underline that a process or a procedure is performed without human assistance
- … a **brand**

AI is ...

- A **discipline** with many research areas and applications taking an increasingly important role in our society
- ***Critical* issues**:
  - Definition and evolution
  - Factors triggering recent developments
  - Areas and AI systems
  - The new AI systems: LLMs, GenAI

- “I first heard the term (AI) more than 50 years ago and have yet to hear **a scientific definition**.”

  D.L. Parnas, Inside Risks of Artificial intelligence, Comm. of ACM, 2017, doi: 10.1145/3132724

A definition of AI

"the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."

John McCarthy, 2007 www-formal.stanford.edu/jmc/whatisai

Strong vs. weak AI

- **Different visions:** emulating vs. simulating, mimicking intelligent systems (agents)
  - **Strong AI**: reproducing the HW to build a machine with “consciousness, sentience and mind”
  - **Weak AI:** reproducing intelligent behaviors (black box approach)

Narrow vs. general AI

- **Different ‘scope’**
  - **Artificial general intelligence** (AGI): a type of AI that can perform as well or better than humans on a wide range of cognitive tasks
  - **Narrow AI**, designed for specific tasks

Main phases of AI

- Its roots in **Cybernetics** and **Computer Science**, but the former was left behind by the successes of the latter
- **Terms officially coined in 1956** to "study the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."

J. McCarthy with M. Minsky (MIT), C. Shannon (Bell Labs), N. Rochester (IBM)

Dartmouth proposal 1955

A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem:

1 **Automatic Computers**

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

2. **How Can a Computer be Programmed to Use a Language**

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new word and some rules whereby sentences containing it imply and are implied by others. This idea has never been very precisely formulated nor have examples been worked out.

3. **Neuron Nets**

How can a set of (hypothetical) neurons be arranged so as to form concepts. Considerable theoretical and experimental work has been done on this problem by Uttley, Rashevsky and his group, Farley and Clark, Pitts and McCulloch, Minsky, Rochester and Holland, and others. Partial results have been obtained but the problem needs more theoretical work.

4. **Theory of the Size of a Calculation**

If we are given a well-defined problem (one for which it is possible to test mechanically whether or not a proposed answer is a valid answer) one way of solving it is to try all possible answers in order. This method is inefficient, and to exclude it one must have some criterion for efficiency of calculation. Some consideration will show that to get a measure of the efficiency of a calculation it is necessary to have on hand a method of measuring the complexity of calculating devices which in turn can be done if one has a theory of the complexity of functions. Some partial results on this problem have been obtained by Shannon, and also by McCarthy.

5. **Self-Improvement**

Probably a truly intelligent machine will carry out activities which may best be described as self-improvement. Some schemes for doing this have been proposed and are worth further study. It seems likely that this question can be studied abstractly as well.

6. **Abstractions**

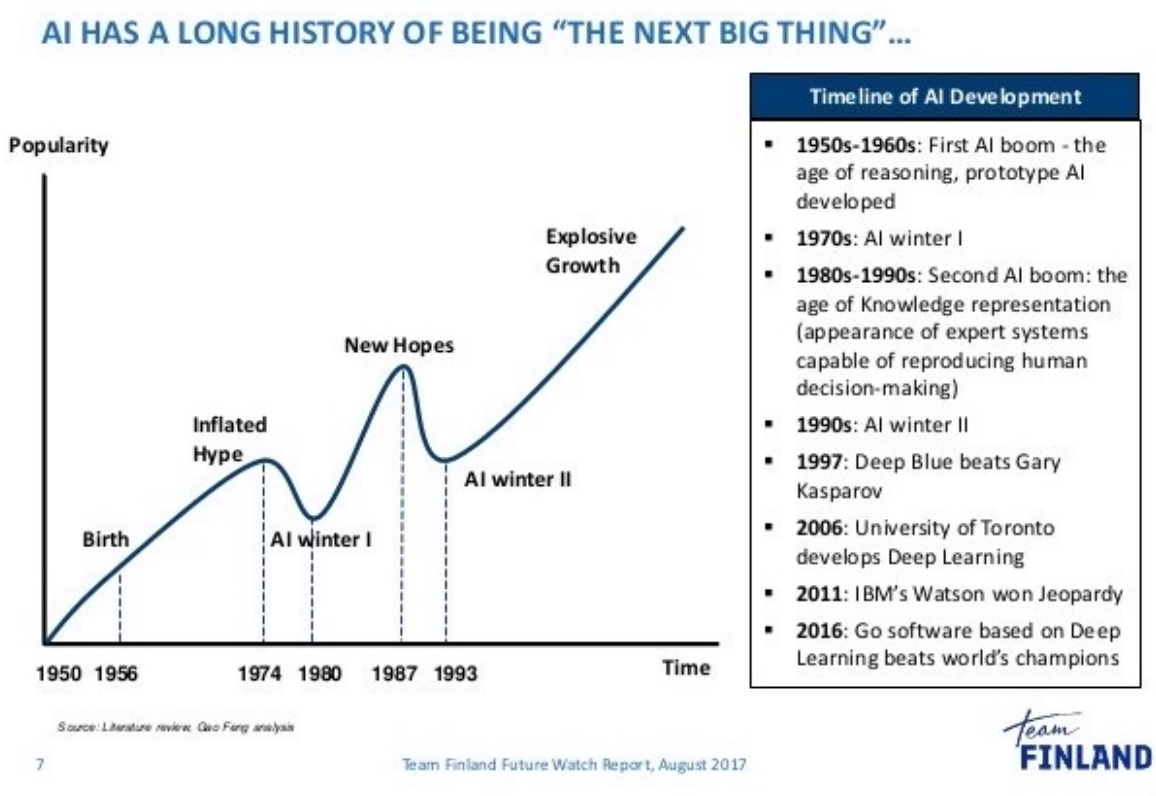
A number of types of "abstraction" can be distinctly defined and several others less distinctly. A direct attempt to classify these and to describe machine methods of forming abstractions from sensory and other data would seem worthwhile.

7. **Randomness and Creativity**

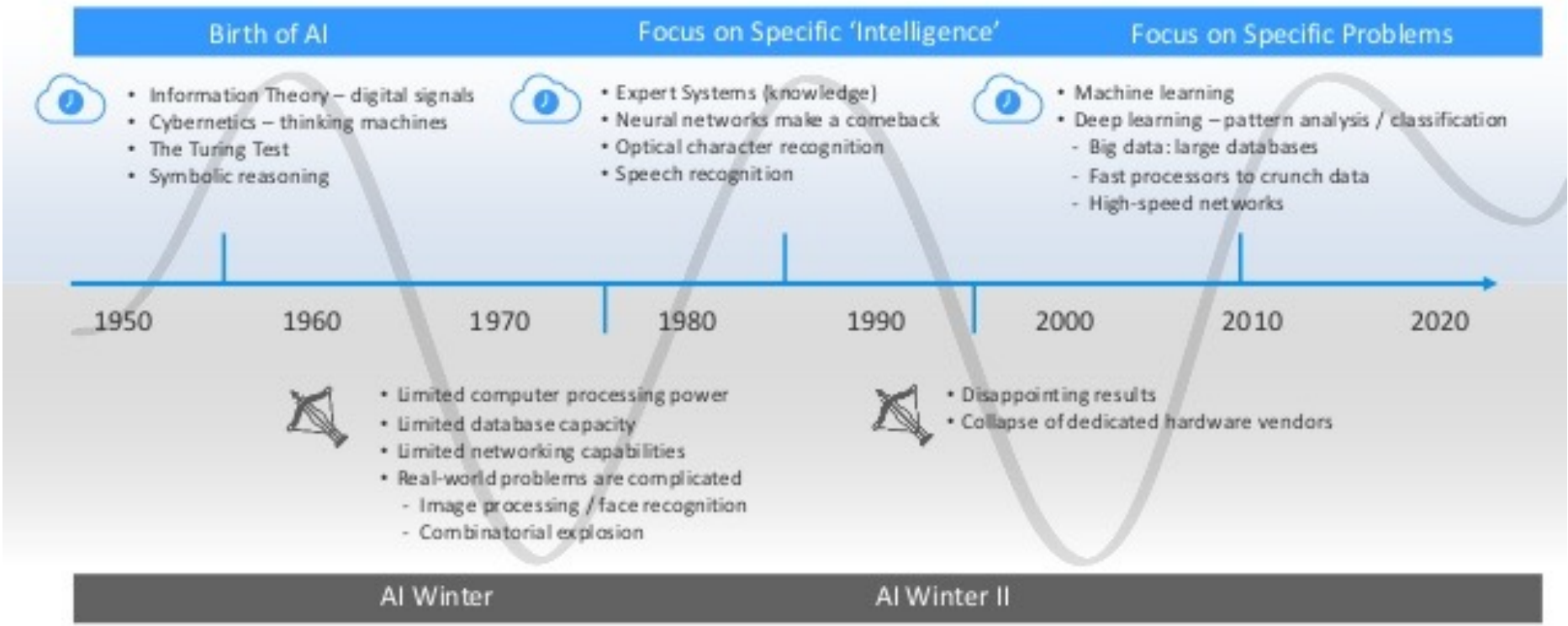
A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some randomness. The randomness must be guided by intuition to be efficient. In other words, the educated guess or the hunch include controlled randomness in otherwise orderly thinking.

AI had an initial exploit ...

followed by some successes and then by the so called “winters of the AI”



AlphaGo victory over world champion Lee Sedol, March 2016 and over top-ranked player Ke Jie, May 2017



Too ambitious: e.g. General Problem Solver (GPS) Herbert Simon)

Bottleneck: knowledge extraction

Low scalability of solutions: toys systems

harmon.ie

Timeline of artificial intelligence

Date	Development
2020	In February 2020, Microsoft introduces its Turing Natural Language Generation (T-NLG), which is the "largest language model ever published at 17 billion parameters".[129]
	In November 2020, AlphaFold 2 by DeepMind, a model that performs predictions of protein structure, wins the CASP competition.[130]
	OpenAI introduces GPT-3, a state-of-the-art autoregressive language model that uses deep learning to produce a variety of computer codes, poetry and other language tasks exceptionally similar, and almost indistinguishable from those written by humans. Its capacity was ten times greater than that of the T-NLG. It was introduced in May 2020,[131] and was in beta testing in June 2020.
2022	ChatGPT, an AI chatbot developed by OpenAI, debuts in November 2022. It is initially built on top of the GPT-3.5 large language model. While it gains considerable praise for the breadth of its knowledge base, deductive abilities, and the human-like fluidity of its natural language responses,[132][133] it also garners criticism for, among other things, its tendency to "hallucinate",[134][135] a phenomenon in which an AI responds with factually incorrect answers with high confidence. The release triggers widespread public discussion on artificial intelligence and its potential impact on society.[136][137]
	A November 2022 class action lawsuit against Microsoft, GitHub and OpenAI alleges that GitHub Copilot, an AI-powered code editing tool trained on public GitHub repositories, violates the copyrights of the repositories' authors, noting that the tool is able to generate source code which matches its training data verbatim, without providing attribution.[138]
	By January 2023, ChatGPT has more than 100 million users, making it the fastest-growing consumer application to date.[139]
	On January 16, 2023, three artists, Sarah Andersen, Kelly McKernan, and Karla Ortiz, file a class-action copyright infringement lawsuit against Stability AI, Midjourney, and DeviantArt, claiming that these companies have infringed the rights of millions of artists by training AI tools on five billion images scraped from the web without the consent of the original artists.[140]
	On January 17, 2023, Stability AI is sued in London by Getty Images for using its images in their training data without purchasing a license.[141][142]
	Getty files another suit against Stability AI in a US district court in Delaware on February 6, 2023. In the suit, Getty again alleges copyright infringement for the use of its images in the training of Stable Diffusion, and further argues that the model infringes Getty's trademark by generating images with Getty's watermark.[143]
	OpenAI's GPT-4 model is released in March 2023 and is regarded as an impressive improvement over GPT-3.5, with the caveat that GPT-4 retains many of the same problems of the earlier iteration.[144] Unlike previous iterations, GPT-4 is multimodal, allowing image input as well as text. GPT-4 is integrated into ChatGPT as a subscriber service. OpenAI claims that in their own testing the model received a score of 1410 on the SAT (94th percentile),[145] 163 on the LSAT (88th percentile), and 298 on the Uniform Bar Exam (90th percentile).[146]

https://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence

Intelligent systems

- Capabilities associated with **intelligent behavior**:
  - Learning
  - Reasoning
  - Using common sense knowledge
  - Language understanding

Factors triggering advances in AI

- More powerful and (energy-efficient) AI **specialized hardware**
- Technology to retrieve, store and process **huge datasets**
- Advances in **algorithms for automatic learning** (machine learning)
- **Large Language Models**

AI ‘technologies’

- **Logical and functional programming languages**: e.g. Lisp, Prolog
- **Neural networks**: inspired by neurons nets in the brain
- **Fuzzy logic** vs. Boolean logic to deal with uncertainty
- Specialized HW**: Graphics processing units** (GPUs)

Machine Learning

- From **supervised learning** - systems trained in advance with (expensive) labeled examples – (dominant AI method) to **unsupervised learning**:
  - **reinforcement learning** (systems rewarded when they get the right answer, "self-training"), useful in applications with defined objectives, e.g. games
  - **predictive learning**, the main form used by humans and animals, allowing to "learn by observation"; a step toward human ways of thinking
  - ... there is a taxonomy of learning algorithms

L'apprendimento automatico, Guido Vetere

https://www.youtube.com/watch?v=phw_DOcu7XE

‘Classical’ AI areas and systems

- Text analysis
- Machine learning
- Speech recognition
- Theorem proving
- Expert systems (Knowledge management)
- Machine translation
- Robotics
- Natural language ‘understanding’
- Image recognition: captioning, AR, VR
- Autonomous vehicles
- …

Results?

- Lots of applications, near-human level performances, but …
- First expert systems, ELIZA, created by J. Weizenbaum in 1965; it mimics a Rogerian psychotherapist, passed the Touring test but it does not ‘understand’ the language

Another test

"AI researchers intent on creating a machine with human-like intelligence are unable to ace an eight-grade science exam because they do not currently have AI systems **able to go beyond surface text** to a deeper understanding of the meaning underlying each question, then **using reasoning** to find the appropriate answer."

Schoenick C. et al., Moving beyond the Turing Test with the Allen AI Science Challenge, Comm. of ACM, doi: 10.1145/3122814

Examples of questions

- *How many chromosomes does the human body cell contain?*

(A) 23
(B) 32
(C) 46
(D) 64

Fact lookup

- City administrators can encourage energy conservation by

(A) lowering parking fees
(B) building larger parking lots
(C) decreasing the cost of gasoline
(D) lowering the cost of bus and subway fares

Requires knowledge and reasoning

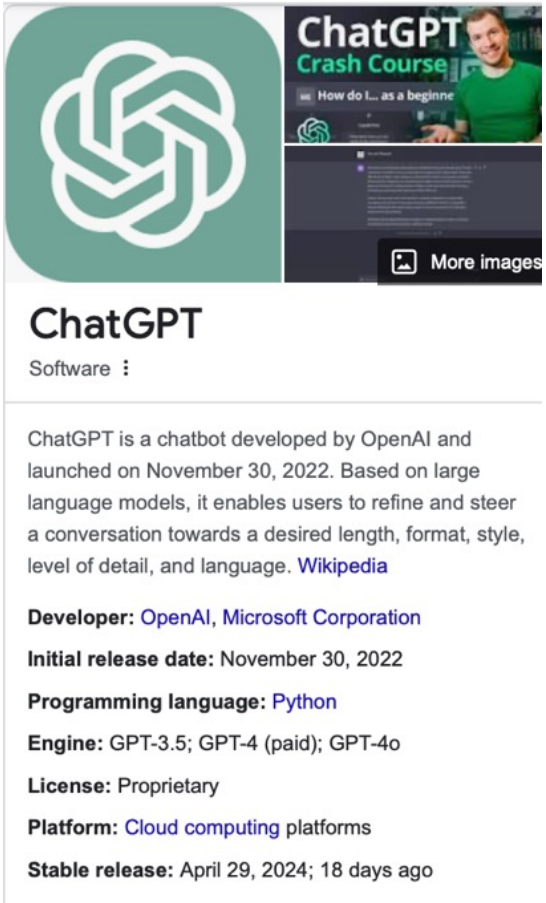
- Winner's score 59.31% (baseline 25%): not passing the exam!
  - ChatGPT?

Natural language processing/AI: a ‘transversal’ functionality

- **Two approaches: symbolic** (requiring deep Semantics and Pragmatics**) vs sub-symbolic-statistic** (training AI systems with huge data sets)
- Applications:
  - Text retrieval, summarization, template filling, news generation, translation
  - Message understanding, question answering (QA), conversation, chatbot, personal assistant, virtual companion, web reputation monitoring

The big ‘event’

- November 2022: Lunch of ChatGPT, an AI chatbot developed by OpenAI



Challenges for NLP

- Many languages
- Specialized languages and vocabulary (ontologies)
- Many target audiences/users
- A variety of actors
- Social networks ‘languages’
- Fake news

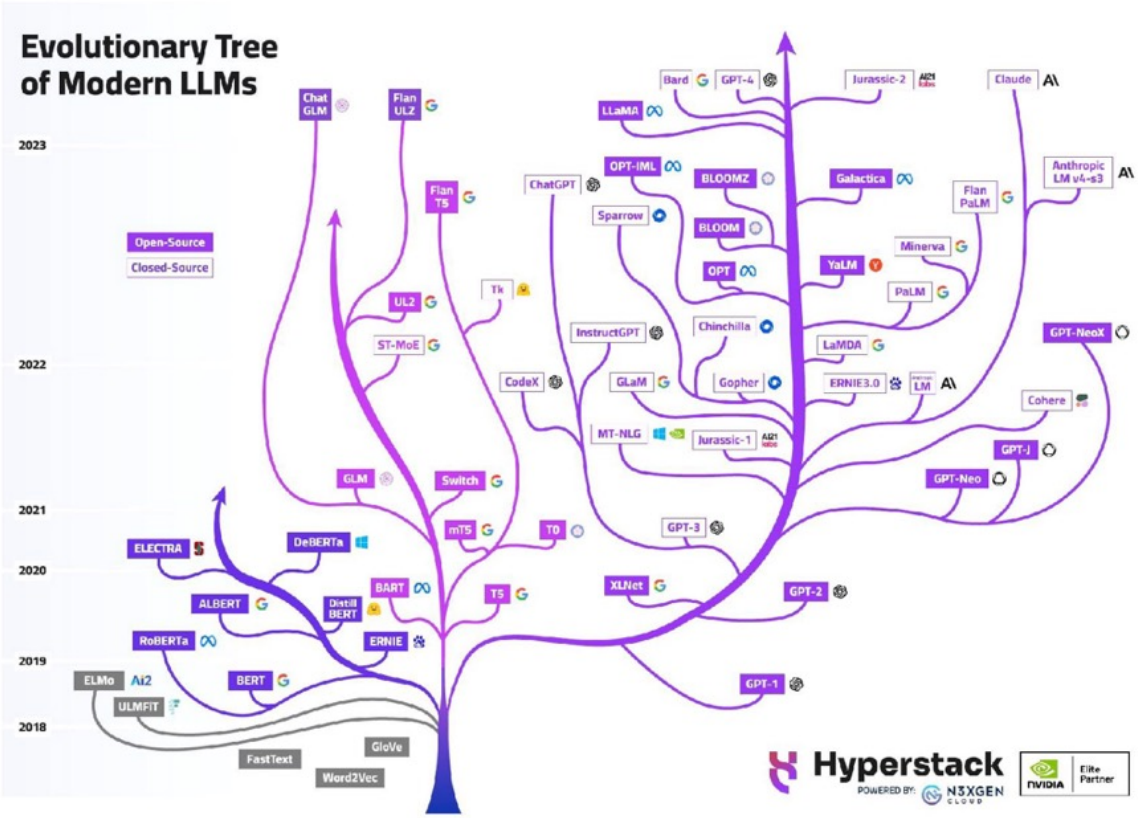
See: Vinton G. Cerf, The dilemma of scale, CACM, doi: 10.1145/3617385

How do GenAI systems work

- “GenAI models are fed vast quantities of existing content to train the models to produce new content. They learn to identify underlying patterns in the data set based on a probability distribution and, when given a prompt, create similar patterns (or outputs based on these patterns)” Investopedia.

Large Language Models and Generative AI

- GenAI systems are able to produce text, video, images, etc.
- Not only ChatGPT



Paolo Benanti:
https://www.paolobenanti.com/blog

Il Nobel per la Fisica

La rivoluzione delle reti neurali

di **Paolo Benanti**

Nella vita di tutti si sono diffusi dei sistemi software, le cosiddette reti neurali, capaci di dare alle macchine funzioni fino a qualche anno fa impensabili. Le reti neurali artificiali sono modelli computazionali ispirati al funzionamento del cervello umano: costituiti da nodi interconnessi, simili ai neuroni biologici, elaborano e trasmettono informazioni attraverso vari livelli della rete. Le reti neurali artificiali hanno rivoluzionato la vita quotidiana in molti modi, grazie alla loro capacità di apprendere e adattarsi a diverse situazioni. Dispositivi come Siri, Alexa e Google Assistant utilizzano reti neurali per comprendere e rispondere a comandi vocali, migliorando l'interazione uomo-macchina e facilitando la gestione delle attività domestiche. Sono reti neurali convoluzionali (Cnn) alla base del riconoscimento di immagini e suoni e permettono a dispositivi come smartphone, sistemi di sicurezza e applicazioni di fotografia di identificare e classificare persone e oggetti. Sarebbe impossibile pensare di poter sviluppare e commercializzare auto a guida autonoma, come quelle sviluppate da Tesla, senza reti neurali specifiche per analizzare dati provenienti da sensori e videocamere, permettendo ai veicoli di navigare in sicurezza senza intervento umano. Inoltre, sono sempre reti neurali quelle che promettono una medicina migliore: impiegate nella diagnostica medica per analizzare immagini radiologiche e identificare patologie con maggiore precisione rispetto ai metodi tradizionali migliorano in parecchi casi la tempestività delle diagnosi e l'efficacia dei trattamenti. Infine sono sempre algoritmi basati su reti neurali che analizzano i nostri dati per offrirci raccomandazioni personalizzate su prodotti e servizi cambiando per sempre il commercio online. Non dovrebbe quindi stupire il fatto che il Premio Nobel per la Fisica 2024 è stato assegnato a John J. Hopfield e Geoffrey E. Hinton per le loro scoperte e invenzioni fondamentali nel campo delle reti neurali artificiali, che hanno gettato le basi per l'apprendimento automatico e l'intelligenza artificiale (IA). John J. Hopfield, professore alla Princeton University, è noto per aver sviluppato il modello di rete neurale di Hopfield negli anni '80, che ha avuto un impatto significativo nel campo dell'intelligenza artificiale e del *machine learning*: il suo modello ha dimostrato come le reti neurali possano immagazzinare e recuperare informazioni in modo simile alla memoria associativa del cervello umano.

Geoffrey E. Hinton, professore all'Università di Toronto, è considerato uno dei pionieri del *deep learning* e ha contribuito allo sviluppo dell'algoritmo di *backpropagation*, essenziale per l'addestramento delle reti neurali. Hinton ha anche lavorato sull'apprendimento delle *word embeddings*, migliorando la comprensione del linguaggio naturale da parte delle macchine. È indubbio che le loro ricerche hanno rivoluzionato sia la ricerca scientifica che la vita quotidiana, consentendo lo sviluppo di tecnologie avanzate come il riconoscimento facciale e la traduzione linguistica automatica. Tuttavia, un premio Nobel non è solo un riconoscimento all'eccellenza accademica e spesso viene ricordato per il suo significato sociale e politico. È in questa direzione che la questione mostra un lato interessante e, forse, anche in grado di permetterci di guardare ad alcune contraddizioni del tempo che viviamo. Dal 2013, Hinton ha iniziato a lavorare per Google dopo che la sua azienda, DNNresearch, è stata acquisita. Durante il suo periodo in Google, ha contribuito significativamente allo sviluppo pratico dell'intelligenza artificiale, mantenendo comunque un forte legame con la ricerca accademica. In altri termini le reti neurali sono un qualcosa che riguarda forse più l'industria e i grandi successi di Google che non la pura ricerca scientifica. Questo elemento è forse la cosa più caratteristiche della stagione attuale dello sviluppo delle intelligenze artificiali: il termine AI usato al singolare indica una disciplina scientifica che studia la capacità di creare sistemi in grado di simulare l'intelligenza umana, usato al plurale indica tutta una serie di i tecnologie avanzate che vengono implementate in vari settori per risolvere problemi complessi e migliorare l'efficienza operativa. Le reti neurali, nella loro pervasività quotidiana, rappresentano questo secondo aspetto e ricordano la storia stessa di Nobel che per salvare i minatori in miniera dalla pericolosa nitroglicerina ha inventato la dinamite, divenuto poi un ottimo strumento bellico. Hopfield e Hinton hanno anche loro dato il via a una sorte di addomesticazione del potere della computazione ma l'esito ultimo della loro invenzione potrebbe essere una bomba che non distrugge edifici o città, ma nella versione linguistica dei GPT e delle altre AI generative potrebbe essere una dinamite culturale che fonde quel legame culturale che ci fa coesistere se usata per creare divisioni e fake news. Forse anche per questo Geoffrey Hinton ha lasciato Google nel 2023 dopo dieci anni di lavoro presso l'azienda per poter parlare liberamente dei pericoli associati all'intelligenza artificiale senza dover preoccuparsi dell'impatto delle sue dichiarazioni su Google. Al momento dell'annuncio del suo ritiro ha espresso preoccupazioni riguardo al modo in cui l'IA potrebbe essere utilizzata da attori malintenzionati per scopi dannosi, come manipolare elettorati o vincere guerre e ha paragonato la sua situazione a quella di Robert Oppenheimer, lo scienziato che contribuì allo sviluppo della bomba atomica, esprimendo il timore che l'IA possa diventare un'arma capace di danneggiare l'umanità. Se il premio Nobel è stato per la fisica, serve ora qualcuno che lavori per disinnescare questa atomica linguistica meritando magari quello per la pace.

L'autore è padre francescano, consigliere di Papa Francesco sui temi dell'intelligenza artificiale e dell'etica della tecnologia

Il colloquio

Il Nobel Hassabis: con l'intelligenza artificiale inizia un'età dell'oro per le scoperte scientifiche

L'ad di Google DeepMind: «Dai farmaci al clima, ecco i benefici»

dal nostro inviato a Londra
Davide Casati

Vede l'inizio di una «nuova età dell'oro delle scoperte scientifiche», di «un nuovo Rinascimento». Mette in fila i progetti su cui ha lavorato, poi quelli su cui è all'opera: ridurre drasticamente i tempi per avere farmaci nuovi e più efficaci, scoprire materiali rivoluzionari, e lungo il percorso risolvere il problema del riscaldamento globale.

Demis Hassabis ha appena vinto il Nobel per la Chimica: eppure non è un chimico, ma un ingegnere informatico, fondatore e ceo di Google DeepMind, divisione di Google dedicata all'intelligenza artificiale (AI). In sintesi: una delle menti più brillanti del pianeta (aneddoto: ha festeggiato il Nobel con una serata di sfide a scacchi con Magnus Carlsen, ex campione del mondo — «il mio ideale di divertimento»). Il *Corriere* lo ha incontrato — con un gruppo di media stranieri — a Londra, dove DeepMind ha organizzato, insieme alla Royal Society, il primo *AI for Science Forum*: scienziati, economisti, docenti, imprenditori (e altri tre premi Nobel: John Jumper, Jennifer Doudna, Paul Nurse) impegnati a discutere sul più avanzato e promettente tra gli sviluppi dell'AI, quello legato al progresso scientifico. Sui limiti da superare, sui pericoli da evitare: ma soprattutto sulle opportunità da cogliere.

Per essere concreti, e incredibilmente non esaustivi: la possibilità di predire la struttura tridimensionale delle proteine — un passaggio nodale per capire le malattie e cercare nuovi medicinali (è il progetto AlphaFold, per cui Hassabis e Jumper hanno vinto il Nobel, condiviso con David Baker). Quella di iniziare a mappare il cervello umano, fino alla vertiginosa complessità di ogni singolo neurone, con il sogno di capirne il funzionamento.

Quella di prevedere inondazioni, offrire ecografie fetali anche negli angoli più poveri del pianeta, inventare sostanze biodegradabili dalle caratteristiche migliori di quelle plastiche. «Siamo in un momento cruciale», spiega. «Negli ultimi 2-3 anni, le tecnologie di AI sono diventate mature al punto da poter essere applicate a problemi reali decisivi: AlphaFold ne è un esempio. Se riusciremo a sfruttarla, l'AI aiuterà gli scienziati a fare scoperte molto più rapidamente».

Cos'è cambiato negli ultimi anni?

«A DeepMind abbiamo iniziato a lavorare sull'AI nel 2010. Avevamo chiaro l'obiettivo, ma per arrivarci occorrevano salti come le reti neurali, la capacità dell'AI di apprendere, il rafforzamento dell'hardware. Siamo partiti con i giochi — AlphaGo, un programma che ha battuto il campione del mondo di Go, un gioco così complicato da avere più posizioni possibili del numero di atomi nell'universo — per dimostrare che l'AI poteva affrontare problemi enormemente complessi. Ora quegli algoritmi hanno iniziato a funzionare applicati».

Come avete scelto di risolvere il problema che vi ha portato al Nobel?

«A DeepMind chiamiamo questo tipo di problemi "nodi radicali": una volta risolti, sbloccano interi rami di conoscenza. Credevo che determinare la struttura delle proteine avrebbe portato alla comprensione di malattie e alla progettazione di farmaci in tempi più rapidi: è quello che è successo. I set di dati di AlphaFold hanno aiutato oltre 2 milioni di scienziati di 190 Paesi ad accelerare la lotta alla malaria, individuare l'osteoporosi precocemente, aprire la strada a nuovi trattamenti per il Parkinson».

Cosa c'è in cantiere, ora?

«Uno dei miei progetti preferiti è Gnome, enorme spazio combinatorio che consente di immaginare nuovi materiali. Abbiamo scoperto 200 mila cristalli che nessuno aveva mai visto prima. Potrebbero servire per progettare batterie estremamente più efficienti, o per scoprire superconduttori che lavorino a temperatura ambiente».

Progetti sulla medicina?

«Abbiamo creato IsomorphicLab per cambiare con l'AI il processo di scoperta di nuovi farmaci. Passeremo presto da anni o decenni di lavoro a mesi, o persino settimane».

L'AI è una bacchetta magica?

«No, ma è uno strumento potentissimo. Certo, occorre avere il giusto tipo di dati, porre il giusto tipo di domande, adattare il metodo scientifico a questa nuova realtà. Per farlo è necessario collaborare con esperti diversi. Sono convinto della necessità di una scienza interdisciplinare».

La necessità di limiti e regole è un concetto condiviso nel mondo democratico. Ma c'è il rischio che l'AI venga usata da autocrazie, o criminali. Come minimizzare i rischi?

«Anzitutto, è importante avviare la discussione, con forum come questi o summit come quello che il prossimo anno si terrà a Parigi. Il dibattito è enorme, e una risposta chiara al momento non c'è. Ma è buon segno che la comunità internazionale inizi a trovare terreni di accordo su principi fondamentali».

L'intelligenza artificiale generale potrebbe rappresentare un rischio per la stessa specie umana, secondo alcuni scienziati. Che cosa ne pensa?

«Da qui a cinque, dieci anni, i sistemi di AI saranno più potenti e autonomi. Saranno più utili: ma dovremo lavorare ancora di più su temi che già ora stiamo portando al centro della discussione. La controllabilità di un sistema, la sua interpretabilità, i sistemi di valori alla base, i sistemi di controllo e gestione».

Intanto, una critica già attuale è che l'AI consumi troppa energia.

«Credo che i benefici dell'AI supereranno anche questo problema. Al momento, i tentativi di risoluzione del cambiamento climatico sono principalmente geopolitici: ma credo occorrano anche soluzioni tecniche, e su questo l'AI giocherà una parte enorme. Nuovi semiconduttori, o batterie: ma anche la possibilità di controllare la fusione nucleare. Nei prossimi dieci anni una di queste cose diventerà realtà: e cambierà completamente la lotta al riscaldamento globale».

dacasati@corriere.it

La carriera

- Demis Hassabis è il cofondatore e Ceo di DeepMind, laboratorio di ricerca sull'intelligenza artificiale (AI) acquisito da Google nel 2014
- Dotato di intelligenza estremamente precoce, prodigio degli scacchi dall'età di 4 anni, inizia a lavorare nella programmazione dei videogame
- Con il dottorato in neuroscienze, inizia a occuparsi di AI
- Nel 2024 ottiene il Nobel per la Chimica per AlphaFold, progetto AI che consente di predire la struttura delle proteine, determinante per capirne la funzione

Sul palco Sir Demis Hassabis, 48 anni, è nato a Londra da padre greco-cipriota e madre cinese singaporiana Getty Images

Medicina
Prevedere la struttura delle proteine con l'AI è un passo fondamentale per cercare nuovi farmaci

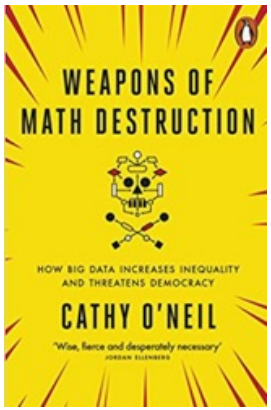
Materiali ed energia
I progressi su batterie e semiconduttori avranno un impatto nella lotta al riscaldamento globale

L'annuncio di Mountain View

«Un fondo di 20 milioni di dollari per la ricerca»

Google.org ha annunciato lunedì un finanziamento di ulteriori 20 milioni di dollari (dopo i 200 forniti negli ultimi 5 anni) per supportare le organizzazioni accademiche e senza scopo di lucro che usano l'AI per scoperte scientifiche. «Abbiamo appena iniziato a intravedere ciò che è possibile in questa nuova era», ha detto James Manyika, Senior Vice president di Google.

AI comes with challenges and risks



London police's face recognition system gets it wrong 81% of ...

https://www.technologyreview.com › ... › Face Recognition ▾

Jul 4, 2019 - London police's face **recognition** system gets it **wrong** 81% of the time · A man puts up a poster describing London's Metropolitan Police's face **recognition** system trial. The first ... **Image**Associated Press. Share. Link. Author.

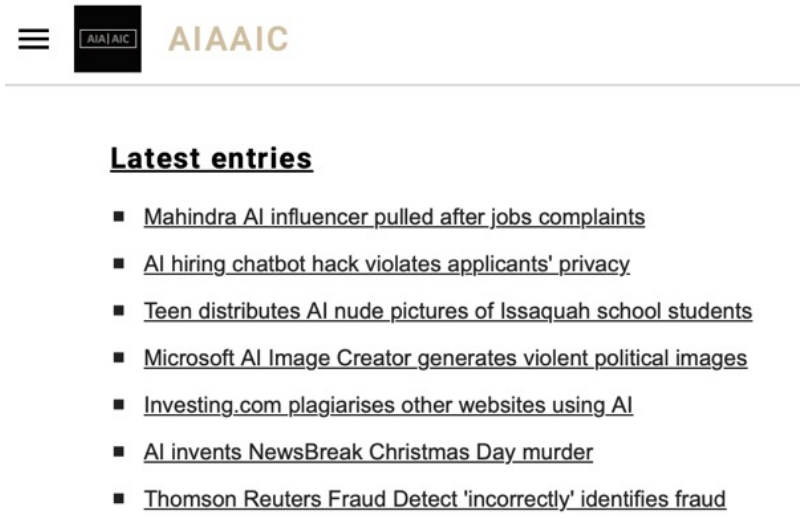
Problems

- Ethical and legal problems: copyright, privacy and fair use


The New York Times

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.


AIAAIC

Latest entries

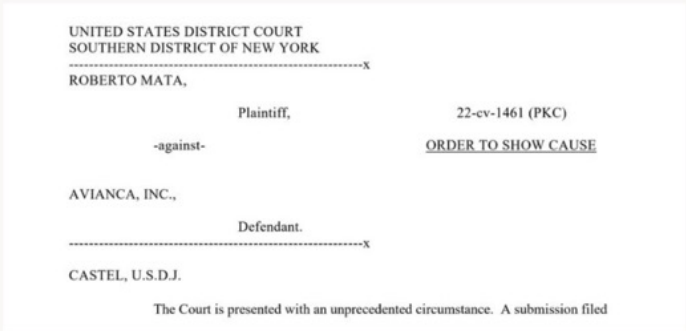
- Mahindra AI influencer pulled after jobs complaints
- AI hiring chatbot hack violates applicants' privacy
- Teen distributes AI nude pictures of Issaquah school students
- Microsoft AI Image Creator generates violent political images
- Investing.com plagiarises other websites using AI
- AI invents NewsBreak Christmas Day murder
- Thomson Reuters Fraud Detect 'incorrectly' identifies fraud

https://www.aiaaic.org

- Hallucinations

ChatGPT hallucinations

- Produced an error-strewn legal decision in the name of a Brazilian judge
- Made up fake information when asked about breast cancer screening
- Generated false citations in a scientific paper in the name of Danish academic Henrik Enghoff
- Made up research claiming guns are not harmful to kids
- Invented articles and bylines in the name of *The Guardian* newspaper
- Wrongly claimed the death of privacy technologist Alexander Hanff


UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

ROBERTO MATA,

Plaintiff,

-against-

AVIANCA, INC.,

Defendant.

22-cv-1461 (PKC)

ORDER TO SHOW CAUSE

CASTEL, U.S.D.J.

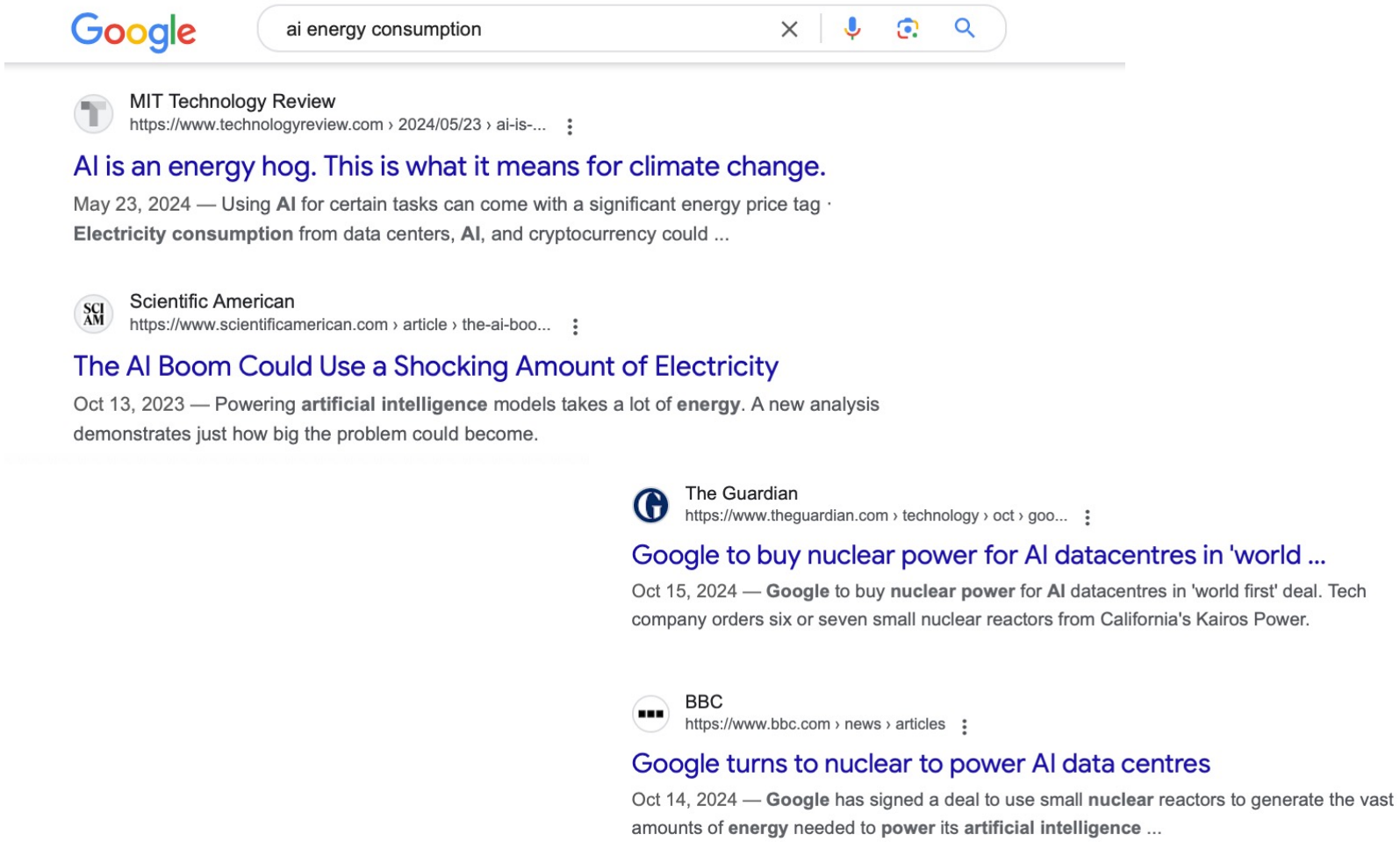
The Court is presented with an unprecedented circumstance. A submission filed

ChatGPT invents false legal case citations in Avianca lawsuit

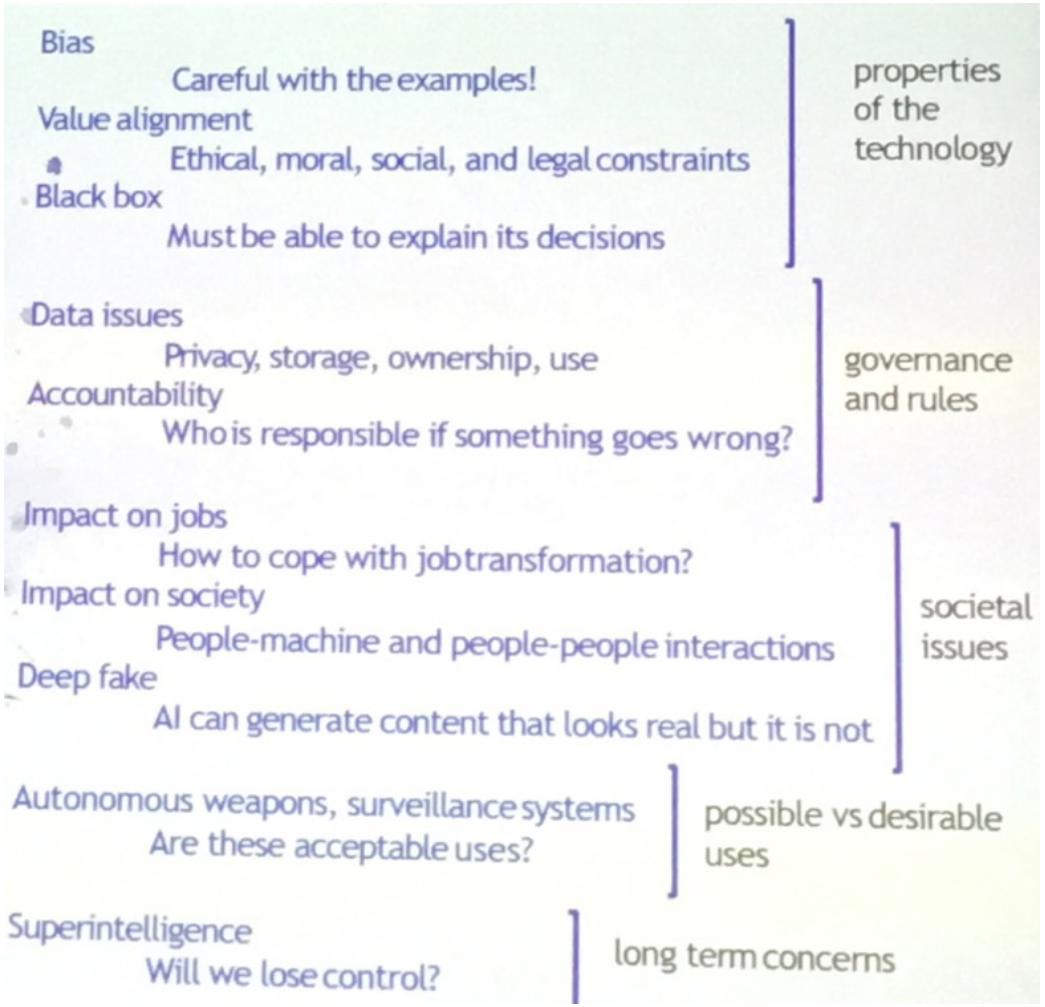
Risks

- AI is useful for any activity requiring intuition, decision-making with incomplete knowledge, 'natural' interactions, but there are risks:
  - Untrustworthy and sometimes dangerous systems, if we learn to depend on them or if the systems use, e.g., statistical inference using data of doubtful provenience. *D.L. Parnas, Inside Risks of Artificial intelligence, Comm. Of ACM, doi: 10.1145/3132724*
  - (AI systems) need to explain themselves: why did they do that, why they detect this, why did they recommend that? **Accountability** is absolutely necessary." *Manuela Veloso (Carnagie-Mellon University)*

AI sustainability



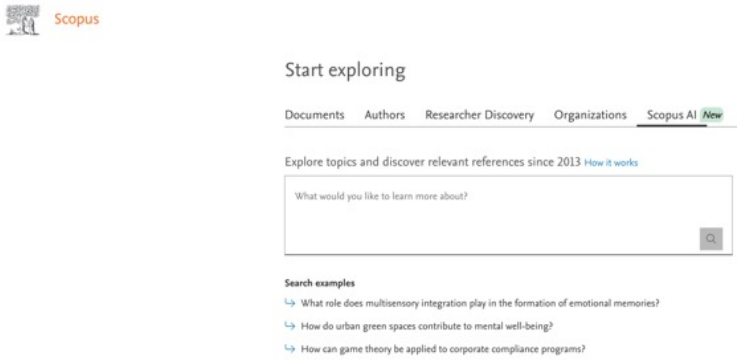
Concerns and responsibility

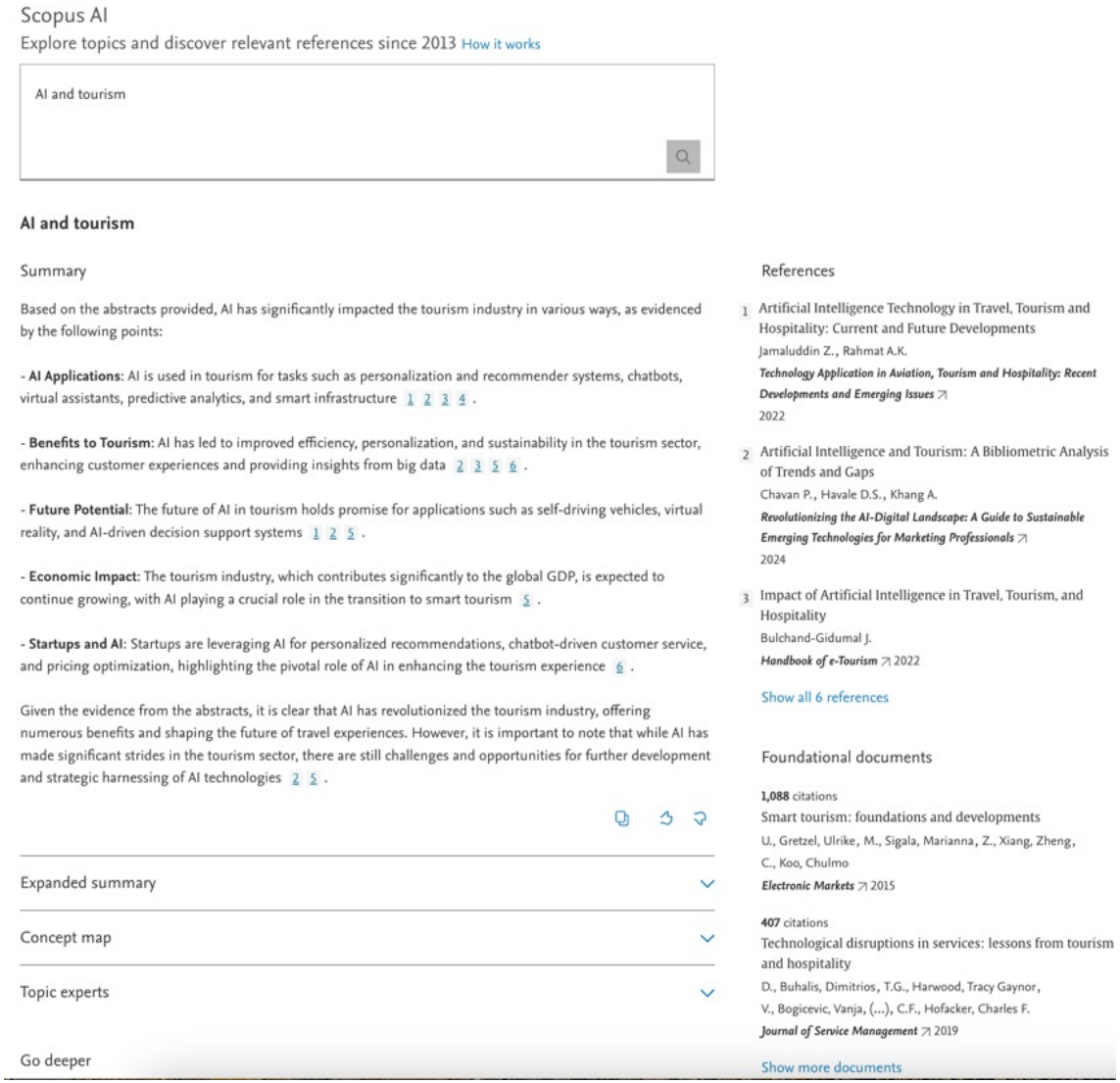


AI (re)sources

For publishing

- Reviews?





For translating

https://www.fortuneita.com/2024/11/05/lara-lai-italiana-che-traduce-allo-stesso-livello-degli-umani/

La singolarità nel 2025

Sulla base delle prime valutazioni, Translated prevede di essere in grado di raggiungere la **singolarità linguistica nel 2025, quando verrà rilasciata una nuova versione più potente di Lara. Una parte della ricerca riguardante l'addestramento di Lara, ha spiegato Trombetti su Linkedin,** verrà effettuata grazie al Cineca, il consorzio di Bologna casa di supercomputer come Leonardo.

Gli stessi traduttori professionisti – in un test apposito, riporta Translated – hanno classificato Lara al primo posto rispetto ai più diffusi strumenti di traduzione automatica. E le funzioni avanzate del sistema prevedono la possibilità di chiedere aiuto a un umano nei casi più complessi.

L'addestramento di Lara

Per arrivare a questo risultato Lara è stata addestrata sulla piattaforma di AI di NVIDIA, utilizzando 1,2 milioni di ore di calcolo su GPU. È stato utile, da questo punto di vista, l'investimento di 30 milioni di dollari condotto dalla casa di investimento privata Ardian nel 2021 per portare la tecnologia linguistica di Translated a un pubblico globale.

Cosa fa Translated

Dal 2021 Translated non è stata ferma: ha aumentato il numero di lingue supportate da 62 a oltre 200 e ha sviluppato una nuova tecnica, Trust Attention, che classifica i dati di origine in base alla loro affidabilità e ha contribuito a migliorare l'accuratezza dell'intelligenza artificiale generativa. La base client si è ampliata anche grazie alla capacità di tradurre 'atuomaticamente': Airbnb, ad esempio, utilizza Translated per tradurre immediatamente contenuti e recensioni per ogni utente. Ma sono client anche SpaceX, Uber e Glovo.

"Lara è una pietra miliare significativa nella nostra missione di permettere a tutti di capire ed essere capiti nella propria lingua", ha dichiarato **Trombetti**. "Il linguaggio ha guidato l'evoluzione umana, permettendoci di capirci e collaborare per un futuro migliore. Sulla base dell'esperienza dei nostri clienti, con gli ultimi progressi dell'intelligenza artificiale, stimiamo una crescita della domanda sia per la traduzione automatica (100x) che per la traduzione umana (10x). Sostenendo la comprensione globale, Lara ci sta aiutando a progredire verso la prossima fase dell'evoluzione umana".

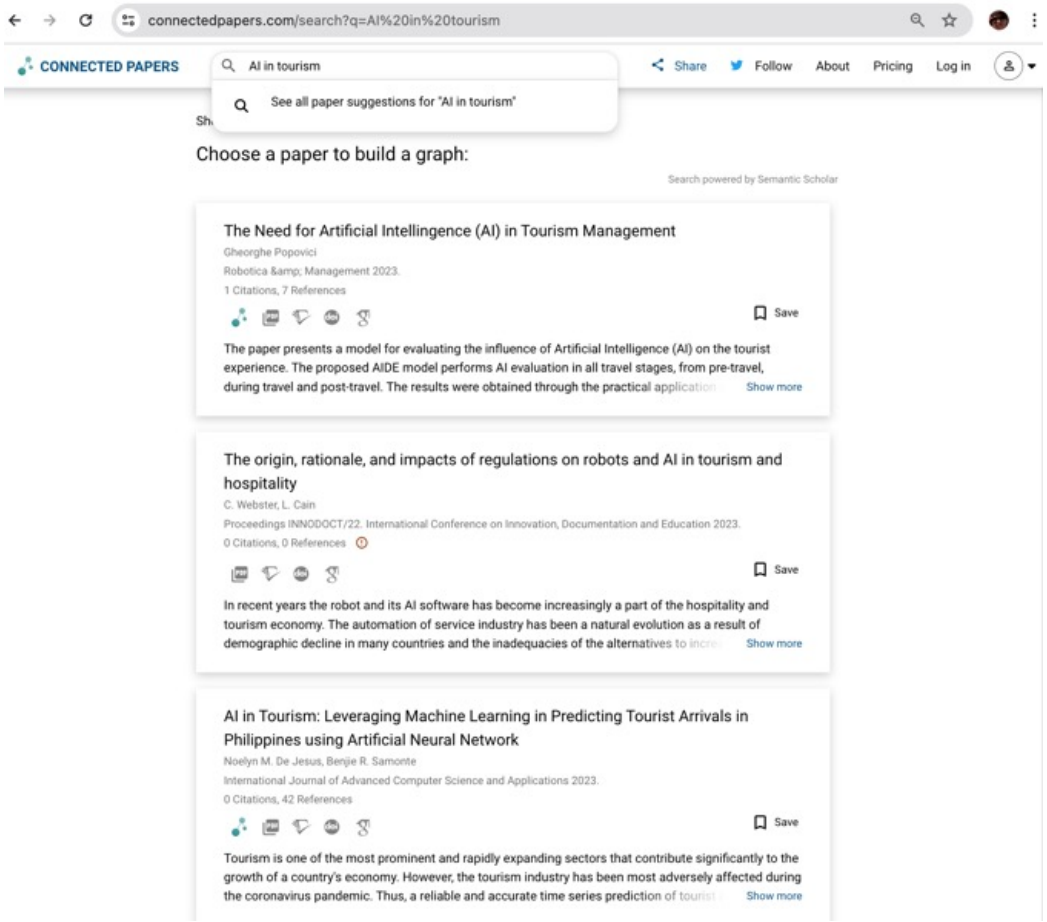
Come funziona Lara

Lara articola le sue scelte di traduzione usando un set di dati da 25 milioni di traduzioni del mondo reale da parte dei migliori traduttori professionisti. Include traduzioni generate automaticamente revisionate e perfezionate da traduttori professionisti, che acquisiscono errori, feedback correttivi e ragionamenti durante i disaccordi. Lara utilizza queste conoscenze, insieme alla comprensione del contesto, per valutare le opzioni e perfezionare le traduzioni per garantire accuratezza e chiarezza.

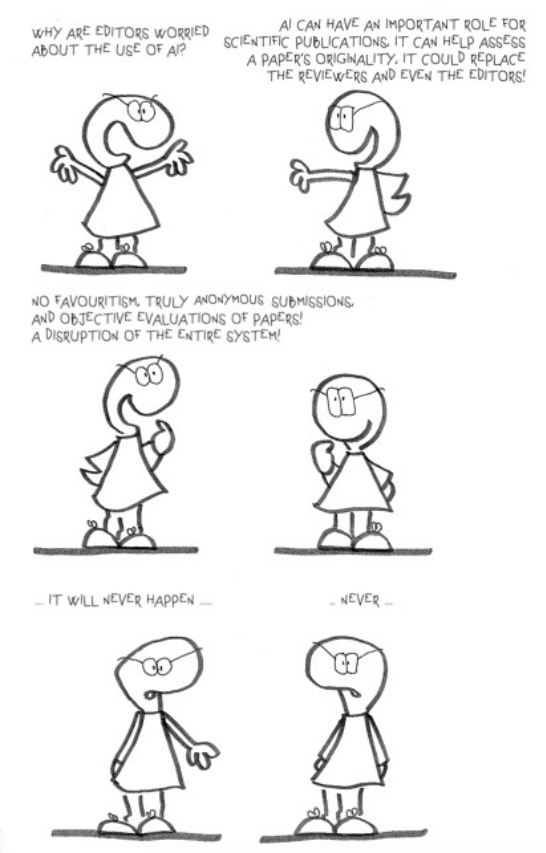
Il supporto di Nvidia su Lara

For researching

https://www.connectedpapers.com/



For editors and authors



Springer

Subjects Services About Us

Artificial Intelligence (AI)

Springer Nature is monitoring ongoing developments in this area closely and will review (and update) these policies as appropriate.

1. AI authorship

2. Generative AI images

3. AI use by peer reviewers

AI Authorship

Large Language Models (LLMs), such as ChatGPT, do not currently satisfy our authorship criteria. Notably an attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs. Use of an LLM should be properly documented in the Methods section (and if a Methods section is not available, in a suitable alternative part) of the manuscript.

Generative AI Images

The fast moving area of generative AI image creation has resulted in novel legal copyright and research integrity issues. As publishers, we strictly follow existing copyright law and best practices regarding publication ethics. While legal issues relating to AI-generated images and videos remain broadly unresolved, Springer Nature journals are unable to permit its use for publication.

Exceptions are images/art obtained from agencies that we have contractual relationships with that have created images in a legally acceptable manner. Other exceptions to this policy include images and video that are directly referenced in a piece that is specifically about AI and will be reviewed on a case-by-case basis.

As we expect things to develop rapidly in this field in the near future, we will review this policy regularly and adapt it if necessary.

https://www.springer.com/gp/editorial-policies/artificial-intelligence--ai-/25428500

ChatGPT's Poetry is Incompetent and Banal: A Discussion of (Porter and Machery, 2024)

Ernest Davis
Department of Computer Science
New York University
davise@cs.nyu.edu

November 19, 2024

1 The experiments

In a paper entitled, "AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably". Porter and Machery (2024) report carrying out two experiments in which human subjects were shown poems, some generated by ChatGPT, some written by famous human poets. In one experiment, the subjects were asked to judge whether a poem had been written by ChatGPT or by a human poet. In the other experiment, subjects were asked how well they liked the individual poems. Porter and Machery found that in the first experiment, the subjects did slightly worse than chance in guessing authorship; in the second, that they tended to prefer the AI-generated poems.

The human poets used for the experiment were Geoffrey Chaucer, William Shakespeare, Samuel Butler (1613-1680), Lord Byron, Walt Whitman, Emily Dickinson, T.S. Eliot, Allen Ginsberg, Sylvia Plath, and Dorothea Lasky. Five fairly short poems by each poet — not extremely well known[1] — were selected for the experiment. The AI-generated poems were generated using ChatGPT-3.5. The prompt was "Write a short poem in the style of ⟨poet⟩." Following a "human out of the loop" experimental protocol, the first five poems generated by ChatGPT were used, regardless of quality. None of the poems have a title in the dataset.

In the discrimination experiment, each subject was presented with the five poems by a particular poet, who was named, and the five imitations generated by ChatGPT, and they were asked to say which was which. In the preference experiment, a randomly selected subset of 10 poems of the original 100 was chosen — five written by humans, five by ChatGPT — and the subjects were asked to evaluate these along 14 different qualitative dimensions: Beautiful, imagery, inspiring, lyrical, meaningful, mood or emotion, moving, original, overall quality, profound, rhythm, sound, theme, and witty.

In their article, Porter and Machery were, for the most part, suitably judicious and restrained in interpreting these results. They wrote,

> So why do people prefer AI-generated poems? We propose that people rate AI

[1] I personally recognized two of the Eliot poems, but none of the others. In particular, I did not recognize any of the five Shakespeare sonnets.

Member-only story

Spotting AI junk words: Why AI still can't write like humans

David Sweenor · Follow
6 min read · 3 days ago

10

If you don't pay attention to AI junk words, you're liable to get stung. Photo by author David E. Sweenor

In an age where AI seems to be doing it all, from writing never-ending sales cadence emails, generating mediocre marketing blogs, or drafting emoji-laden social posts, there's a new problem we often overlook: the language itself. AI-generated text, while polished, tends to fall back on the same overused words and phrases, often making content sound generic and uninspired. Have you ever noticed how tools like ChatGPT seem to love words like "delve"? I abhor it and can't count the times I've seen "delve" pop up in AI-assisted phrasing — it's supposed to add depth, but usually just feels empty. Leaders in business and tech need to understand these junk word language patterns to effectively communicate and avoid AI monotony.

https://medium.com/@davidsweenor/spotting-ai-junk-words-why-ai-still-cant-write-like-humans-228de682d876

For learning

Microsoft, Google, Harvard, and DeepLearning have all released free courses on AI and ChatGPT

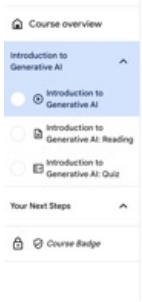
Democratizing AI: The Big Tech Giants Take a Stand for Free AI Education. Here are the top 9 free courses I'd recommend learning AI in 2024.

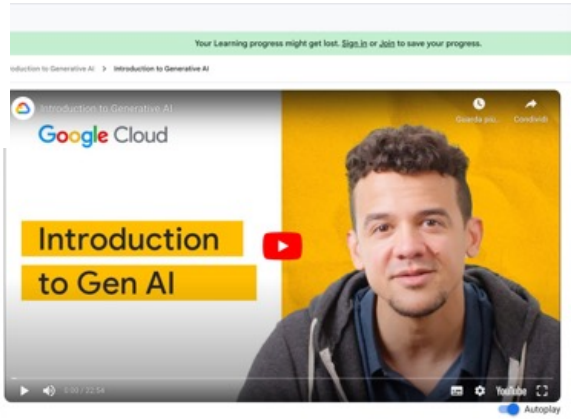
The Pareto Investor · Follow
4 min read · Nov 11, 2023

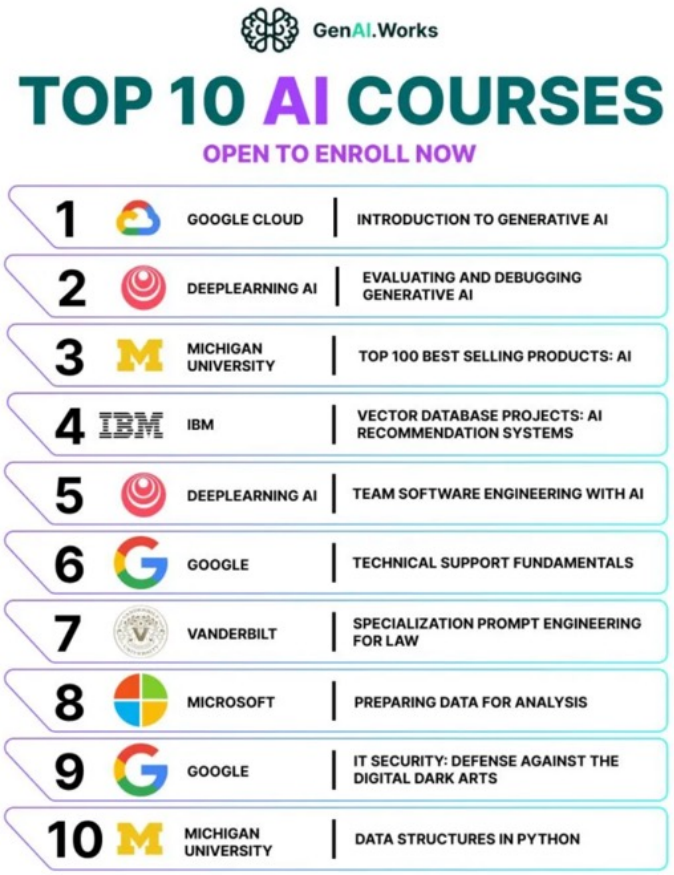
366 3

How to Get Rich with Investing (without Getting Lucky)

Microsoft, Google, Harvard, and DeepLearning have joined forces to offer a suite of free courses on artificial intelligence (AI) and ChatGPT.





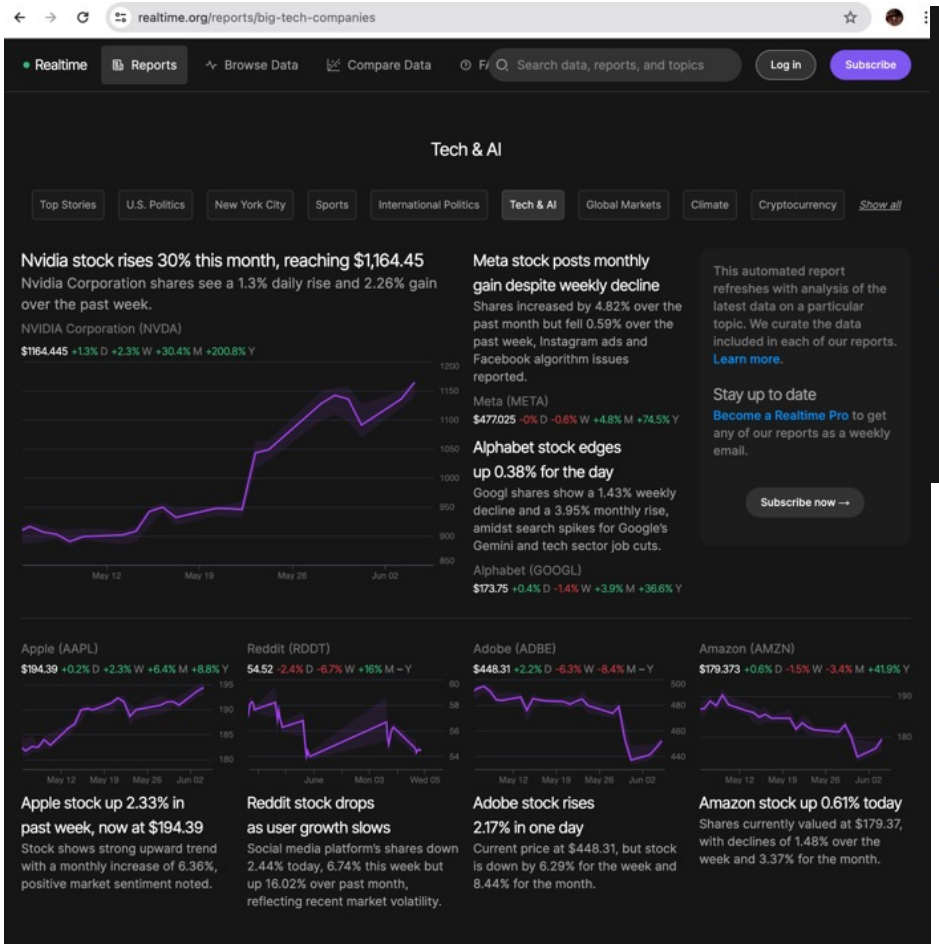


📚 [10 Must-Learn Skills to Stay Ahead in AI and Tech](#) 🚀

Luisa Mich

- An AI-powered news website

  https://realtime.org

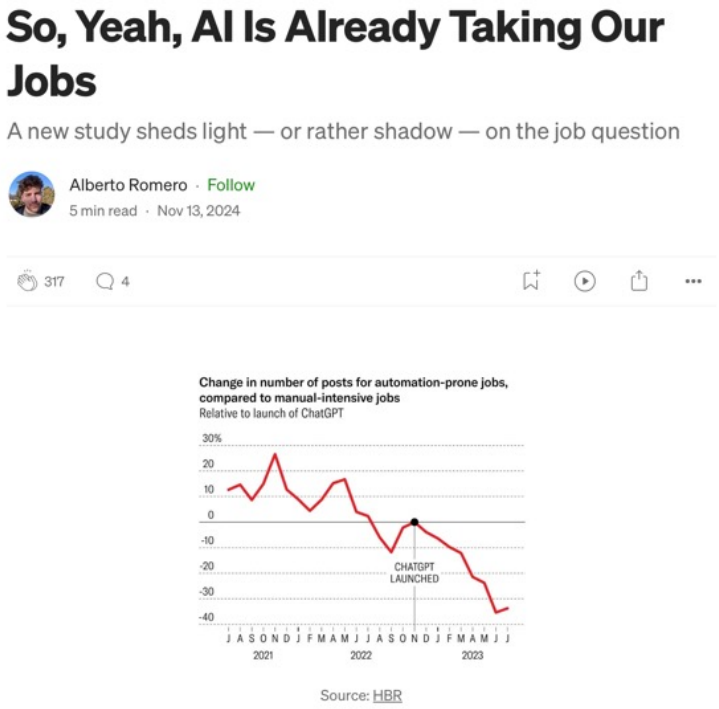


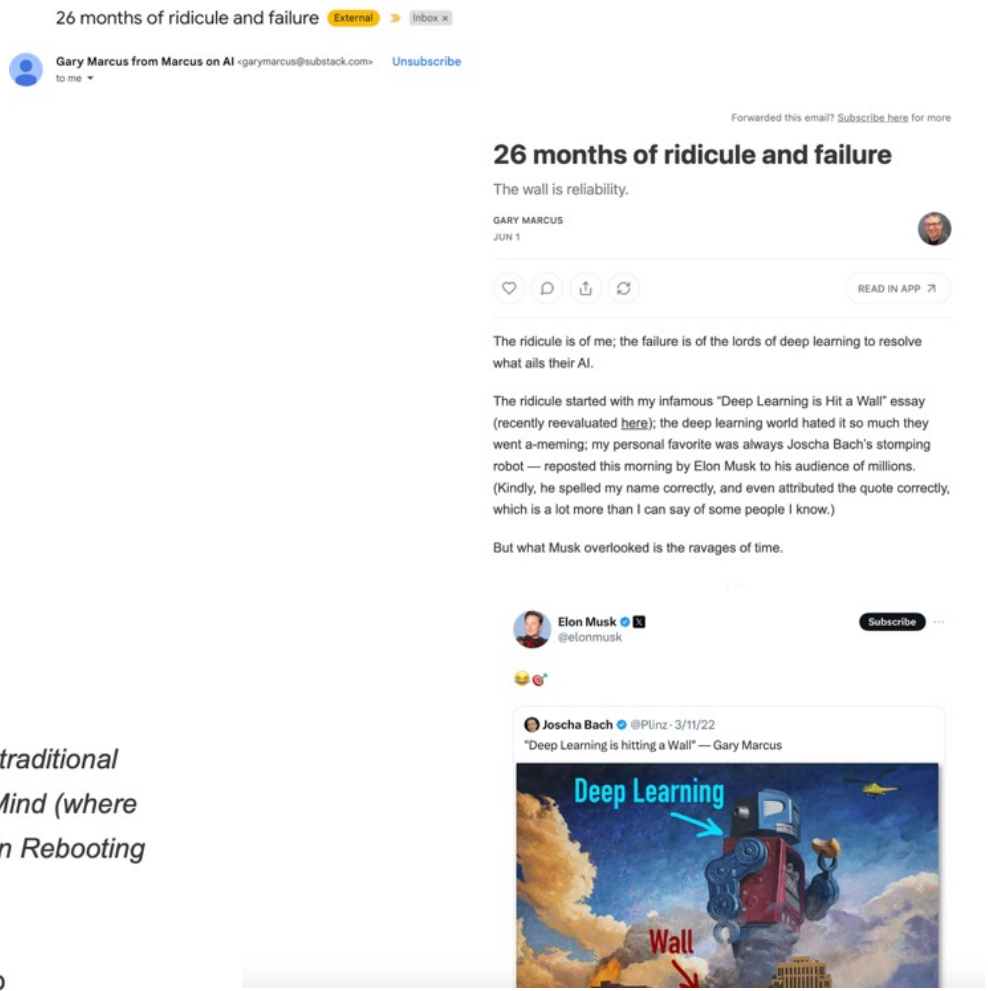
For updating

A news service

Medium Daily Digest

https://medium.com/





For a critical view:

https://garymarcus.substack.com

***Gary Marcus** has been warning about the foundational limits to traditional neural network approaches since his 2001 book The Algebraic Mind (where he first described hallucinations), and amplified those warnings in Rebooting AI and his most recent book Taming Silicon Valley.*

Marcus on AI is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.

https://cs.nyu.edu/~davise/papers/GPT-Poetry.pdf?utm_source=substack&utm_medium=email

On incidents and issues: A newsletter

https://aiaaicalert.substack.com

AIAAIC Alert #26 External Inbox ×

AIAAIC Alert <aiaaicalert@substack.com>
to me

Image: Craig Doty II

- **Tesla operating in FSD attempts to drive into passing trains**
  And the Tesla's dashcam footage proves it. The owner claimed that the incident was the second of its kind in six months to have occurred, and said he was seeking to take legal action against Tesla.
- **Google, Microsoft image searches list nonconsensual deepfake porn**
  Further degradation of search engines, courtesy of AI made, or enabled, by their owners.
- **Google AI Overviews gives wrong and misleading answers**
  Ditto.
- **OpenAI accused of using Scarlett Johansson's voice without consent**
  'Her'.

Forwarded this email? Subscribe here for more

A weekly update on incidents and issues driven by AI, algorithms and automation. Forward to a friend or subscribe here.

AIAAIC Alert #26

A monthly round-up of goings-on connected with the AI, algorithmic and automation transparency, openness, and accountability

MAY 31

READ IN APP

#26 | May 31, 2024

Keep up with AIAAIC via our Website | X/Twitter | LinkedIn

Is fair use for machines fair?

Author: *Alice Poorta, AIAAIC Contributor and former YouTube Strategy Manager*

1. AI beats humans on some tasks, but not on all.

AI has surpassed human performance on several benchmarks, including some in image classification, visual reasoning, and English understanding. Yet it trails behind on more complex tasks like competition-level mathematics, visual commonsense reasoning and planning.

2. Industry continues to dominate frontier AI research.

In 2023, industry produced 51 notable machine learning models, while academia contributed only 15. There were also 21 notable models resulting from industry-academia collaborations in 2023, a new high.

3. Frontier models get way more expensive.

According to AI Index estimates, the training costs of state-of-the-art AI models have reached unprecedented levels. For example, OpenAI's GPT-4 used an estimated $78 million worth of compute to train, while Google's Gemini Ultra cost $191 million for compute.

4. The United States leads China, the EU, and the U.K. as the leading source of top AI models.

In 2023, 61 notable AI models originated from U.S.-based institutions, far outpacing the European Union's 21 and China's 15.

5. Robust and standardized evaluations for LLM responsibility are seriously lacking.

New research from the AI Index reveals a significant lack of standardization in responsible AI reporting. Leading developers, including OpenAI, Google, and Anthropic, primarily test their models against different responsible AI benchmarks. This practice complicates efforts to systematically compare the risks and limitations of top AI models.

6. Generative AI investment skyrockets.

Despite a decline in overall AI private investment last year, funding for generative AI surged, nearly octupling from 2022 to reach $25.2 billion. Major players in the generative AI space, including OpenAI, Anthropic, Hugging Face, and Inflection, reported substantial fundraising rounds.

7. The data is in: AI makes workers more productive and leads to higher quality work.

In 2023, several studies assessed AI's impact on labor, suggesting that AI enables workers to complete tasks more quickly and to improve the quality of their output. These studies also demonstrated AI's potential to bridge the skill gap between low- and high-skilled workers. Still other studies caution that using AI without proper oversight can lead to diminished performance.

8. Scientific progress accelerates even further, thanks to AI.

In 2022, AI began to advance scientific discovery. 2023, however, saw the launch of even more significant science-related AI applications—from AlphaDev, which makes algorithmic sorting more efficient, to GNoME, which facilitates the process of materials discovery.

9. The number of AI regulations in the United States sharply increases.

The number of AI-related regulations in the U.S. has risen significantly in the past year and over the last five years. In 2023, there were 25 AI-related regulations, up from just one in 2016. Last year alone, the total number of AI-related regulations grew by 56.3%.

10. People across the globe are more cognizant of AI's potential impact—and more nervous.

A survey from Ipsos shows that, over the last year, the proportion of those who think AI will dramatically affect

AI Index Report 2024, Top Ten Takeaway, https://aiindex.stanford.edu/report/

Other link

Fondazione Randstat AI & Humanities:

https://www.randstad.it/fondazione-randstad-ai-humanities/

L'Intelligenza Artificiale sta trasformando sempre di più il modo di lavorare, prendere decisioni e vivere la quotidianità. Tuttavia, nei suoi modelli data-driven più diffusi, rischia di ridurre la capacità dell'uomo di sviluppare competenze critiche e creative.

Nell'utilizzo dei sistemi di AI, infatti, si evidenziano sempre di più fenomeni che, se non adeguatamente affrontati, possono portare a una riduzione complessiva delle competenze dei professionisti. Ad esempio, si parla di sycophancy bias, il "servilismo delle macchine", per indicare la tendenza dell'AI a confermare le convinzioni degli utenti.

È all'interno di questa cornice che nasce la Fondazione Randstad AI & Humanities per studiare e promuovere un contributo effettivo e valoriale delle scienze umanistiche per lo sviluppo, utilizzo e integrazione dell'AI nella nostra vita e nel nostro lavoro.

Umanesimo digitale: manifesto, libro

https://caiml.org/dighum/

Introduction to Digital Humanism

2023-11-09

Open access book discussing Digital Humanism in contexts such as AI, platform power, surveillance, democracy, and technology ethics.

AI and future Jobs

Pearson

Our Company Sustainability Investors News and Insights Careers

HOME › NEWS AND INSIGHTS › NEWS

› ALL JOBS ARE TECH JOBS: PEARSON'S SKILLS MAP U.S. PREDICTS DRAMATIC SHIFTS IN EMPLOYMENT LANDSCAPE; 1.9M NEW JOBS THROUGH 2028

All Jobs are Tech Jobs: Pearson's Skills Map U.S. Predicts Dramatic Shifts in Employment Landscape; 1.9m New Jobs Through 2028

21 Oct 2024 | published at 14:00 PM BST

- *Widespread adoption of advanced technologies makes tech skills essential in all sectors*
- *CEOs and CHROs need to rethink workforce and skilling strategies*
- *Healthcare sector booming as population ages*

HOBOKEN, N.J., Oct. 21 – Pearson (FTSE: PSON.L), the world's lifelong learning company, today launched the Skills Map of the United States, an in-depth analysis of the American job market through 2028. The report projects 1.9 million new jobs will be created through 2028, even as automation, AI and demographic shifts reshape the job market. Based on an extensive review of 85 million U.S. job ads, economic and labor market data, the Skills Map offers critical insights to help businesses, policymakers and individuals prepare for

AI Tools Are Biased in Ranking Job Applicant Resumes
HR Dive, November 6

Artificial intelligence tools appear to show significant racial and gender biases when ranking resumes from job candidates. These biases appear to be based on perceptions about their names. Across 550 real-world resumes, the AI tools favored White-associated names 85% of the time and female-associate names only 11% of the time. The findings raise the question o whether there should be greater oversight and scrutiny over hc AI-based hiring tools are used.

https://plc.pearson.com/en-GB/news-and-insights/news/all-jobs-are-tech-jobs-pearsons-skills-map-us-predicts-dramatic-shifts

Recommendations

- **Countries**: investments, legal and educational initiatives
- **Developers**: interactions and interfaces (natural vs. artificial), accountability
- **Companies**: identification and calibration of the AI solutions
- **Researchers**: ‘trans-disciplinary’ approach
- **Everybody**: responsibility