AiFOS
Associazione Italiana Formatori ed Operatori della Sicurezza sul Lavoro

aicq sicev
Sistema di Certificazione e Valutazione

aicq
Associazione Italiana Cultura Qualità

aicq
Associazione Italiana Cultura Qualità
Comitato Salute e Sicurezza

WEBINAR

Digitalizzazione nel mondo del lavoro: cosa cambia e quali competenze

Sebastiano Luridiana

10 febbraio 2023

SOMMARIO

AiFOS
Associazione Italiana Formatori ed Operatori della Sicurezza sul Lavoro

ASPETTI DI INTELLIGENZA ARTIFICIALE NELLA NORMATIVA EUROPEA SULLE MACCHINE

1. SITUAZIONE NORMATIVA ATTUALE
2. DEFINIZIONE DI LIMITI E OBIETTIVI
3. FATTORI DI RISCHIO NELL'AMBIENTE
4. FATTORI DI RISCHIO NELL'APPRENDIMENTO
5. MACCHINE "PENSANTI" ?

Non verranno esaminati aspetti di sicurezza informatica

aicq sicev
Sistema di Certificazione e Valutazione

aicq
Associazione Italiana Cultura Qualità

aicq
Associazione Italiana Cultura Qualità
Comitato Salute e Sicurezza

SITUAZIONE NORMATIVA ATTUALE

DIRETTIVA 2006-42 *MACCHINE*: in vigore

PROPOSTA DI REGOLAMENTO SUI PRODOTTI MACCHINA: ultima versione sottoposta **25 gen 2023**
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=consil%3AST_5617_2023_INIT

PROPOSTA DI REGOLAMENTO SULL'INTELLIGENZA ARTIFICIALE: ultima versione sottoposta **11 nov 2022**
https://substack.com/redirect/6aaf6e4a-b20f-4f61-a2b0-8c57bf15fca0?j=eyJ1IjoiMXViNnNzIn0.KmF-ag-a9MUN2MQmO-N5g3JulaOKakE73VDphJDfPWs

per le due proposte non sono disponibili i testi ufficiali in italiano

i testi in italiano mostrati in questa presentazione sono traduzioni del relatore e sono **solamente indicativi**

PROPOSTA DI REGOLAMENTO Intelligenza Artificiale (AI)

AiFOS
Associazione Italiana Formatori ed Operatori della Sicurezza sul Lavoro

ART. 3: DEFINITIONS

1. *"artificial intelligence system" (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts;*

aicq sicev
Sistema di Certificazione e Valutazione

aicq
Associazione Italiana Cultura Qualità

aicq
Associazione Italiana Cultura Qualità
Comitato Salute e Sicurezza

PROPOSTA DI REGOLAMENTO Prodotti Macchina

RECITAL 28

The manufacturer, having detailed knowledge of the design and production process, is best placed to carry out the conformity assessment procedure.

Conformity assessment should therefore remain solely the obligation of the manufacturer.

Il Fabbricante, avendo dettagliata conoscenza del progetto e dei processi produttivi, è il più adatto ad eseguire la procedura di verifica della conformità.

La verifica della conformità rimane pertanto obbligo del solo Fabbricante.

Allegato IV (Dichiarazione di Conformità):

4. The declaration of conformity is issued under the sole responsibility of the manufacturer

PROPOSTA DI REGOLAMENTO Prodotti Macchina

ANNEX III: General Principles

The risk assessment and risk reduction shall include hazards that may be generated during the lifecycle of the machinery (. . . . .) as an intended evolution of its fully or partially self-evolving behaviour or logic as a result of the machinery designed to operate with varying levels of autonomy.

La valutazione e la riduzione dei rischi devono includere i pericoli che possono generarsi durante il ciclo di vita della macchina (. . . . .) come una evoluzione del suo comportamento o della sua logica in tutto o in parte in evoluzione, in conseguenza del fatto che essa è progettata per operare con differenti livelli di autonomia

Il fabbricante del macchinario rimane responsabile della sicurezza della macchina (incluso il sistema AI), anche dopo l'immissione sul mercato

PROPOSTA DI REGOLAMENTO Prodotti Macchina

AiFOS
Associazione Italiana Formatori ed Operatori della Sicurezza sul Lavoro

ANNEX III: RESS 1.1.6f - Ergonomics

where relevant, adapting a machinery with intended fully or partially self-evolving behaviour or logic that is designed to operate with varying levels of autonomy (. . . . .) to communicate its planned actions (such as what it is going to do and why) to operators in a comprehensible manner.

dove rilevante, adattare una macchina dotata di comportamento o logica destinati ad evolvere in tutto o in parte e progettata per operare con livelli di autonomia variabili (. . . . .) a comunicare le sue azioni pianificate (come cosa intende fare e perché) agli operatori in maniera comprensibile.

Il soddisfacimento dell'ultimo requisito può risultare particolarmente complesso o impossibile, in particolare per le reti neurali.

aicq sicev
Sistema di Certificazione e Valutazione

aicq
Associazione Italiana Cultura Qualità

aicq
Associazione Italiana Cultura Qualità
Comitato Salute e Sicurezza

PROPOSTA DI REGOLAMENTO Prodotti Macchina

ANNEX III: RESS 1.2.1. Safety and reliability of control systems

d) the limits of the safety functions shall be established as part of the risk assessment performed by the manufacturer. In this respect no modification is allowed to the settings or rules, generated by the machinery or by operators, including during the machinery learning phase, where such modifications may lead to hazardous situations

d. I limiti delle funzioni di sicurezza devono essere stabiliti con la valutazione dei rischi eseguita dal Fabbricante. A questo riguardo non è permessa alcuna modifica delle impostazioni o delle regole, generata dalla macchina o dagli operatori, inclusa la fase di apprendimento, laddove ciò possa comportare l'insorgere di situazioni pericolose.

PROPOSTA DI REGOLAMENTO Prodotti Macchina

AiFOS
Associazione Italiana Formatori ed Operatori della Sicurezza sul Lavoro

ANNEX III: RESS 1.2.1. Safety and reliability of control systems

Control systems of machinery with fully or partially self-evolving behaviour or logic that is designed to operate with varying levels of autonomy shall be designed and constructed in such a way that:

(a) they shall not cause the machinery or related product to perform actions beyond its defined task and movement space;

(aa) recording of data on the safety related decision-making process (. . . . .) is enabled and that such data is retained for one year after its collection (. . . . .),

(b) it shall be possible at all times to correct the machinery or related product in order to maintain its inherent safety.

I sistemi di controllo di macchine dotate di comportamento o logica destinati ad evolvere in tutto o in parte e progettate per operare con livelli di autonomia variabili devono essere progettati e costruiti in maniera tale da:

a) non essere la causa di azioni, da parte della macchina, che vanno oltre il suo compito e il suo spazio di movimento definiti; definizione dei **limiti** del sistema AI

b) consentire in qualsiasi momento la correzione della macchina al fine di preservarne la sicurezza intrinseca.

Quanto sopra si riferisce a TUTTE le funzioni del sistema di controllo e non solo alle funzioni di sicurezza

aicq sicev
Sistema di Certificazione e Valutazione

aicq
Associazione Italiana Cultura Qualità

aicq
Associazione Italiana Cultura Qualità
Comitato Salute e Sicurezza

PROPOSTA DI REGOLAMENTO Prodotti Macchina

ANNEX III: RESS 1.2.1. Safety and reliability of control systems

(iii/c) modifications to the settings or rules, generated by the machinery or by operators, including during the machinery learning phase, shall be prevented, where such modifications may lead to hazardous situations;

(iii/c) le modifiche delle impostazioni o delle regole, generate dalla macchina o dagli operatori, inclusa la fase di apprendimento, devono essere evitate laddove dette modifiche possano portare a situazioni pericolose

Quanto sopra si riferisce a TUTTE le funzioni del sistema di controllo e non solo alle funzioni di sicurezza

I LIMITI DEL SISTEMA COME FATTORE DI RISCHIO

Esempio: un sistema AI per riconoscimento di immagini è addestrato a riconoscere oggetti (manufatti) oppure situazioni (presenza di fiamme).

Se al sistema si presenta un oggetto / situazione che non conosce ma che (erroneamente) comunque classifica come conosciuto, può prendere decisioni sbagliate.

Il problema non è il riconoscimento insoddisfacente, ma il fatto che il sistema AI rispetti i limiti per i quali è stato realizzato, ovvero non prenda decisioni se non siamo ragionevolmente certi che esse siano adeguate.

Se l'apprendimento del sistema prosegue anche dopo la messa in opera occorre **inoltre** garantire che l'ulteriore apprendimento non pregiudichi quanto sopra.

L'OBIETTIVO COME FATTORE DI RISCHIO

Mi sono fermato in un area di servizio in autostrada; dico a Robby
'*io aspetto in macchina; vai a prendermi un caffè al bar*'. obiettivo

- il barista chiede a Robby se vuole lo zucchero di canna o quello bianco; Robby riflette sulla domanda per tre ore.
- il barista ha finito il caffè; Robby si avvia lungo l'autostrada a 10 km/ora verso la prossima area di servizio. pericolo
- il barista ha finito il caffè; Robby ruba il caffè ad un altro cliente scatenando una rissa. PERICOLO
- (. . . . .)

www.hiig.de/wp-content

La definizione di un obiettivo può comprendere informazioni implicite ma cruciali:

- non intendo attendere più di 5 minuti per avere il caffè
- oggi ho tempo e posso concedermi mezz'ora di sosta
- anche se ho molta voglia di un caffè, non intendo causare una rissa

The Surprising Creativity of Digital Evolution: A Collection of Anecdotes - https://arxiv.org/abs/1803.03453

FATTORI DI RISCHIO DELL'AMBIENTE

osservabilità: completa o parziale:
scacchiera: completamente osservabile e definisce del tutto lo stato attuale dell'ambiente;
guida di veicoli: campo visivo limitato, oggetti opachi in movimento, ...

ambiente e azioni: discreti (scacchi) o continui (guida di veicoli)

regole dell'ambiente: note e inviolabili (scacchi) oppure no (guida di veicoli)

presenza di **altri agenti** nell'ambiente (guida di veicoli) oppure no (ricerca di un itinerario)

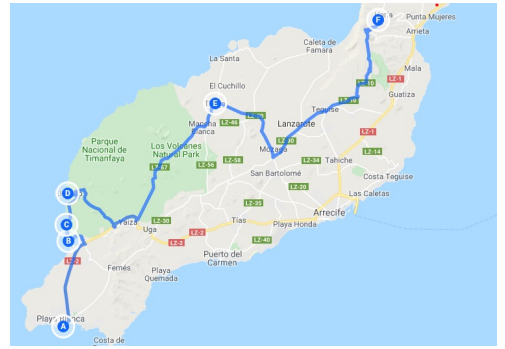
risultato delle azioni: del tutto prevedibile (scacchi) oppure no (guida di veicoli, finanza)

ambiente **statico o dinamico**:
tempo per prendere una decisione limitato (guida di veicoli) oppure no (ricerca itinerario)

orizzonte temporale
tempo nel quale si misura la bontà della decisione: breve (frenata di emergenza), medio (partita a scacchi) o lungo (previsioni finanziarie)

(. . . .)

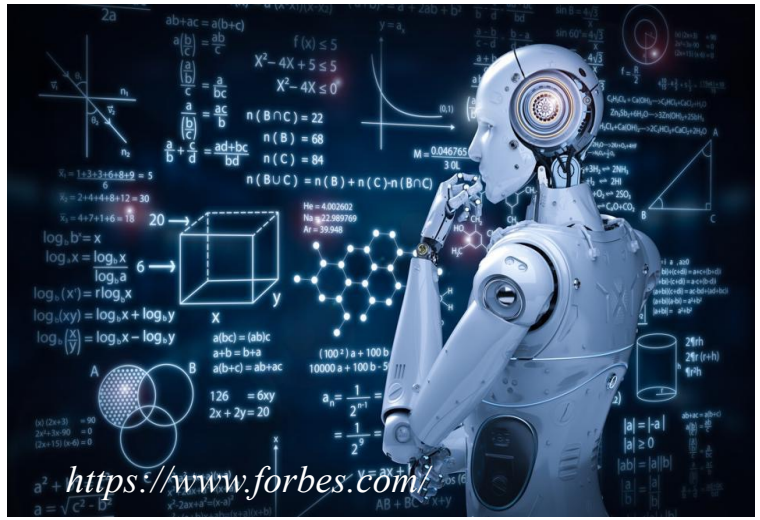
Quali rischi solleva nella macchina ciascuno di questi elementi ?



L'APPRENDIMENTO COME FATTORE DI RISCHIO

L'apprendimento è una delle principali fonti di rischio in macchinari in grado di evolvere.

I sistemi AI che non prevedono apprendimento presentano problemi di sicurezza simili a quelli affrontati nella Direttiva Macchine e nelle relative norme armonizzate (in particolare ISO 12100 e ISO 13849).



Se un sistema è in grado di apprendere occorre garantire che l'evoluzione del suo comportamento non faccia sorgere nuovi rischi.

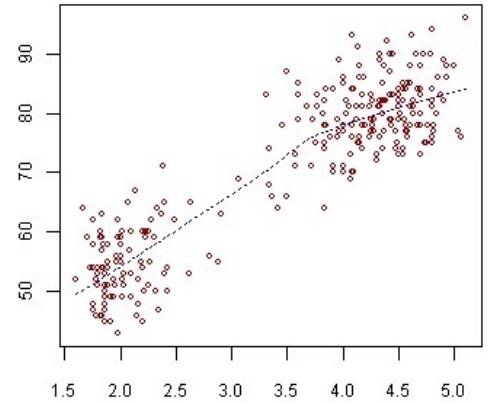
Oltre agli elementi già menzionati bisogna considerare:

1. se il sistema AI può sollevare rischi in caso di malfunzionamento
2. se l'apprendimento è completato prima della messa in opera (presso un fabbricante) oppure se prosegue anche in opera (presso un utente).
3. se il sistema è in grado di comunicare *cosa farà e perché agli operatori in maniera comprensibile*

TIPOLOGIE DI APPRENDIMENTO

1. apprendimento **supervisionato**
   si presenta al sistema coppie input – output (p.es. immagine – etichetta) e il sistema apprende una funzione che associa agli input forniti gli output corrispondenti;

2. apprendimento **non supervisionato**
   per esempio *clustering*: il sistema cerca di raggruppare i dati di input in categorie.
   il sistema può identificare categorizzazioni difficili da rilevare



3. apprendimento **rinforzato**
   il sistema apprende attraverso "ricompense" e "punizioni": ad esempio un sistema che, in un sito web, riceve una "ricompensa" se l'utente seleziona un certo oggetto.

PROSEGUI ESCI

L'apprendimento rinforzato può essere particolarmente utile quando **l'orizzonte temporale** è lungo.

APPRENDIMENTO PROFONDO (DEEP LEARNING)

Termine usato di solito in relazione alle **reti neurali** per indicare il procedimento col quale vengono addestrati gli strati nascosti (o **profondi**) della rete neurale.

Può essere supervisionato, non supervisionato o rinforzato.

Esistono molti tipi di reti neurali:

completamente connessa: tutti i nodi di uno strato sono connessi a tutti e soli i nodi dello strato successivo

con connessioni intralayer: l'output di un nodo può influenzare l'azione di un nodo nello stesso strato

ricorsiva: gli output di alcuni nodi sono usati come input degli stessi o di altri nodi

tutti nodi di ingresso ricevono l'intera informazione da trattare oppure ciascun nodo di ingresso riceve solo parte di essa

(. . . . .)

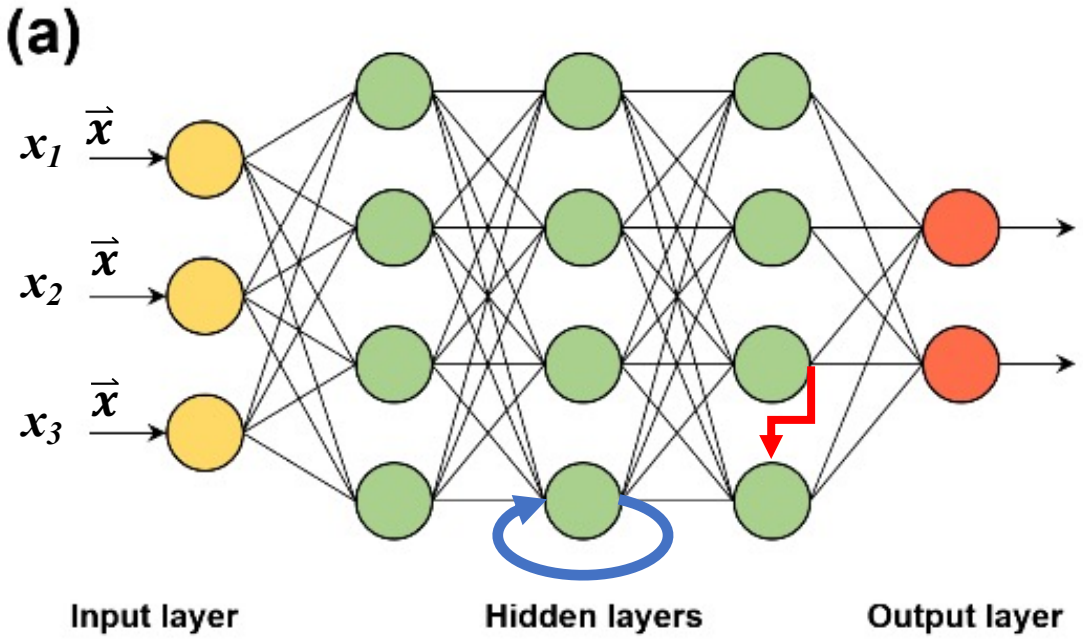


figure by Yosoon Choi
https://www.researchgate.net/figure/Description-of-deep-neural-network-DNN-model-a-Typical-structure-of-DNN-model-b_fig3_339600446

DATI UTILIZZATI NELL'ADDESTRAMENTO DI SISTEMI AI

BIAS NEI DATI

Un sistema AI era stato addestrato al fine di distinguere fra cani e lupi.

Le immagini dei lupi usate per l'addestramento, a differenza di quelle dei cani, erano prevalentemente con sfondo innevato.

Il sistema aveva marcata tendenza ad identificare come lupi anche i cani ritratti con sfondo innevato (e viceversa).

Un sistema AI era stato addestrato per emettere sentenze in procedimenti giudiziari; i dati di addestramento provenivano da processi tenuti negli USA nei decenni precedenti.

Il sistema aveva marcata tendenza a ritenere colpevoli gli imputati di colore.

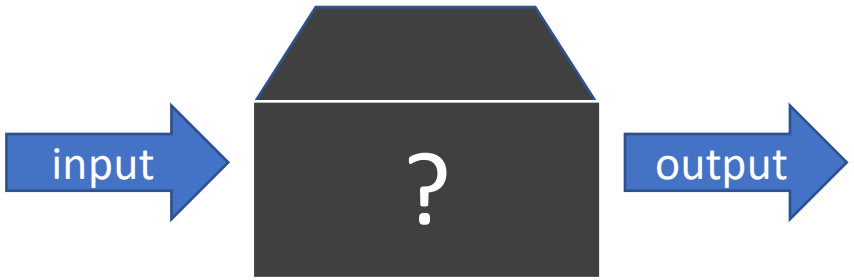
I dati di addestramento sono presumibilmente forniti dal fabbricante del macchinario

C'è <u>sempre</u> qualche bias nei dati

HUMAN *vs* MACHINE LEARNING – LA "BLACK BOX"

se l'apprendimento delle macchine fosse simile a quello degli umani non dovrebbe essere troppo difficile per un sistema AI comunicare *cosa farà e perché agli operatori in maniera comprensibile*

in caso contrario il procedimento decisionale del sistema potrebbe essere confinato all'interno di una "black box" inaccessibile dall'esterno.



ciò è rilevante non solo per i macchinari, ma anche (e forse soprattutto) per sistemi che emettono sentenze giudiziarie, valutano le prestazioni delle persone, ...

una black box non può fornire la motivazione di una sentenza

HUMAN *vs* MACHINE LEARNING – *BLACK BOX* AL LAVORO

AiFOS
Associazione Italiana Formatori ed Operatori della Sicurezza sul Lavoro

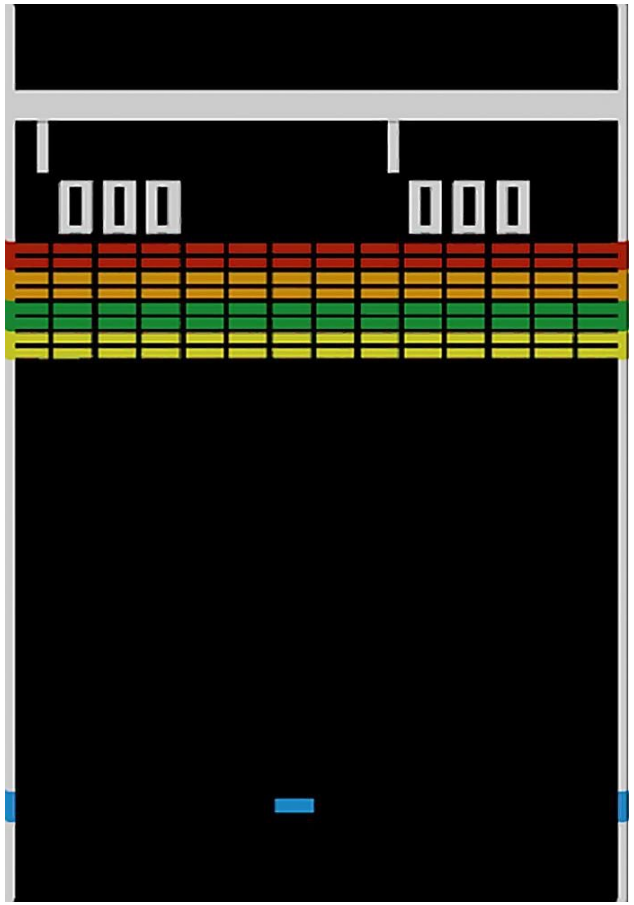
"*Il sistema di DeepMind DQN ha imparato a giocare 49 diversi giochi di Atari usando solo i pixel dello schermo come input e il punteggio come ricompensa.*

Nella maggior parte dei giochi DQN ha imparato a giocare meglio di un professionista umano, malgrado il fatto che DQN non abbia alcuna nozione a priori di tempo, spazio, movimento, velocità o sparare.

E' piuttosto difficile capire cosa DQN stia realmente facendo, a parte vincere."

Stuart Russel: Human Compatible – Penguin Books (2019)

Per un sistema di questo tipo è pressoché impossibile comunicare *cosa farà e perché agli operatori in maniera comprensibile.*



spectrum.ieee.org/media-library/

V. Mnih et. al: Human-level control through deep reinforcement learning – Nature (518) - 2015

aicq sicev Sistema di Certificazione e Valutazione
aicq Associazione Italiana Cultura Qualità
aicq Associazione Italiana Cultura Qualità Comitato Salute e Sicurezza

HUMAN *vs* MACHINE LEARNING – ADVERSARIAL EXAMPLES

AiFOS
Associazione Italiana Formatori ed Operatori della Sicurezza sul Lavoro

Adversarial Examples (esempi antagonisti)

Nei sistemi deep learning un piccolo cambiamento nell'input può generare un ampio cambiamento nell'output.

Ad esempio si può modificare (con una certa perizia) solo pochi pixel dell'immagine di un maiale e far sì che il sistema la classifichi come un aeroplano.

Un umano non ha alcuna difficoltà nel classificare 'maiale' anche la seconda immagine.

https://gradientscience.org/intro_adversarial/

In alcuni casi è sufficiente una semplice rotazione dell'immagine ad ingannare il sistema AI.

https://arxiv.org/abs/1712.02779v4

Anche questi aspetti suggeriscono che in generale i sistemi AI imparano / riconoscono in modo diverso dagli umani.

Gli esempi antagonisti costituiscono una criticità importante in termini di **sicurezza informatica**.

Why Machine Learning Needs Semantics Not Just Statistics - Kalev Leetaru – Forbes 2019
https://www.forbes.com/sites/kalevleetaru/2019/01/15/why-machine-learning-needs-semantics-not-just-statistics/?sh=15534e7277b5

aicq sicev Sistema di Certificazione e Valutazione
aicq Associazione Italiana Cultura Qualità
aicq Associazione Italiana Cultura Qualità Comitato Salute e Sicurezza

HUMAN *vs* MACHINE LEARNING – ASPETTI ENERGETICI

AiFOS
Associazione Italiana Formatori ed Operatori della Sicurezza sul Lavoro

"*On less than 2,000 calories a day, a human child learns to talk, read, play games, and much more in a few years. On such a restricted energy diet, the groundbreaking GPT-3, a neural network capable of fluent conversation, would have taken a millennium to learn to chat.*"
https://www.quantamagazine.org/how-to-make-the-universe-think-for-us-20220531/

una fabbisogno energetico così diverso sembra indicare che i sistemi AI (o almeno le reti neurali) imparino in modo diverso dagli umani

2000 calorie al giorno per 1000 anni sono **850 MWh**; a 0,3 euro kWh sono **255.000 euro**.

se l'apprendimento di GPT-3 fosse durato un mese la potenza media necessaria sarebbe stata **1,16 MW.**

- Quanta energia richiede addestrare un sistema AI anche a leggere, scrivere, giocare, . . . ?
- Quali strutture possono permettersi consumi elettrici di questo tipo ?
- Quale è l'impatto ecologico complessivo corrispondente ?
- Che effetti avrà sullo sviluppo di sistemi AI un aumento dei costi energetici ?

https://www.numenta.com/blog/2022/05/24/ai-is-harming-our-planet/
https://arxiv.org/abs/1906.02243 (Energy and Policy Considerations for Deep Learning in NLP - 2019)

aicq sicev Sistema di Certificazione e Valutazione
aicq Associazione Italiana Cultura Qualità
aicq Associazione Italiana Cultura Qualità Comitato Salute e Sicurezza

MACCHINE “PENSANTI” ?

“*The question of whether machines can think . . . is about as relevant as the question whether submarines can swim.*”
Edsger Dijkstra: The Threats to Computing Science

Se chiedessimo ad un progettista di sottomarini: “*i sottomarini possono nuotare ?*”
la risposta sarebbe probabilmente “*non mi sono mai posto il problema*”.

In una lingua in cui non ci fosse distinzione fra i verbi “nuotare” e “navigare” la domanda non si porrebbe neppure.

Ciò non avviene col verbo “volare” (almeno in italiano).

Un progettista aeronautico non si propone di realizzare oggetti che si comportino come aquile, o che superino il test di Turing, facendosi credere un’aquila dalle aquile stesse.

Cosa intendiamo quando ci chiediamo se una macchina può “pensare” ?

bias linguistici ? bias personali ? aspetti “tecnici” oppure “etici” ?


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


https://kids.nationalgeographic.com

https://www.wired.com/2017/04/the-myth-of-a-superhuman-ai/

GRAZIE PER L'ATTENZIONE

AiFOS
Associazione Italiana Formatori ed Operatori della Sicurezza sul Lavoro

aicq sicev
Sistema di Certificazione e Valutazione

aicq
Associazione Italiana Cultura Qualità

aicq
Associazione Italiana Cultura Qualità
Comitato Salute e Sicurezza