Intenzionalità Artificiale

Framework critici per l'uso responsabile dell'AI nella ricerca e nella didattica

Domenico Schillaci

Università degli Studi di Palermo
Teaching and Learning Center - CIMDU
Ciclo di seminari “AI per la didattica”

Agenda

1. **Intenzionalità Artificiale vs Intelligenza Umana**
   Intenzioni, bici elettriche, autotune e il potere dell'etica
2. **Rischi etici e casi studio reali**
   Convergenza strumentale, graffette e 5 esempi per riflettere
3. **L'AI Act e le implicazioni per la privacy**
   La piramide del rischio, modelli aperti e chiusi e la regola d'oro
4. **Il modello a due corsie**
   La trappola del "tutto o niente" e la terza via
5. **Il framework delle 5 domande**
   Uno strumento pratico per orientarsi nell'uso dell'AI e una riflessione finale

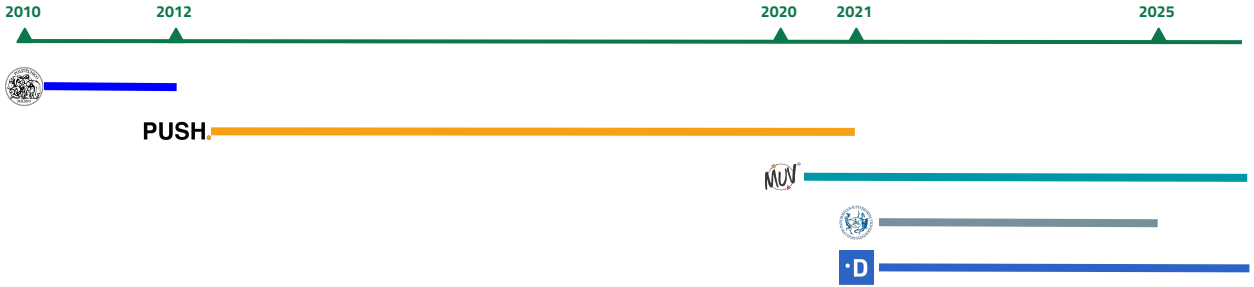
Buon pomeriggio a tutte e tutti

linkedin.com/in/domenicoschillaci

Domenico Schillaci

Service Designer | Entrepreneur | Researcher

Palermo



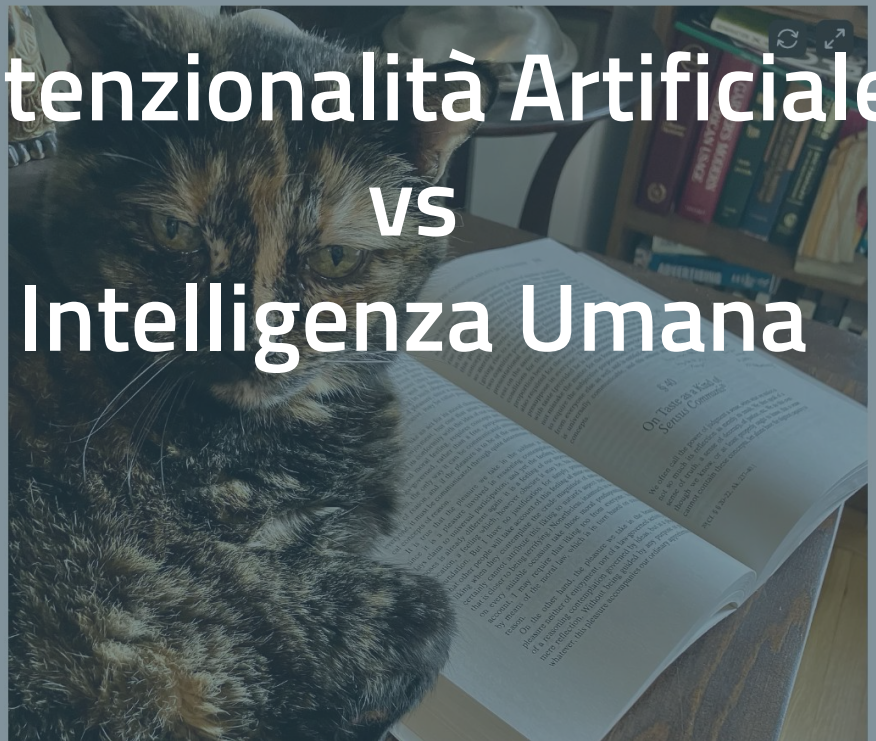
IED · POLI.DESIGN Founded by Politecnico di Milano · ABADIR Accademia di Design e Arti Visive · Politecnico di Torino · mu International Cooperation in Urban Development · Technische Universität Darmstadt

Università degli Studi Link · LUMSA Università · CIHEAM Centre International de Hautes Études Agronomiques Méditerranéennes · FORMEZ al servizio della PA · Università di Genova · JA Italia

Artificial intentionality

ROB HORNING

SEP 21, 2024



Intenzionalità Artificiale vs Intelligenza Umana

Artificial intentionality

ROB HORNING

SEP 21, 2024

Artificial intentionality

*«L'AI sta riempiendo il mondo di **spazzatura mentale**, contenuti superflui che sprecano le capacità cognitive di coloro che sono costretti a confrontarsi con essi.»*

Rob Horning

https://robhorning.substack.com/p/artificial-intentionality

Artificial intentionality

*«Nessun pensiero è stato investito in essa, ma la **vita** di qualcuna sarà sprecata nell'affrontarla, proprio come la sua produzione ha sprecato le **risorse** del pianeta*
...l'AI uccide la vita due volte.»

Rob Horning

Il concetto di intenzionalità artificiale

Secondo Horning la Gen AI non pensa ma simula l'**intenzione di pensiero**.

Fornisce la forma del discorso intenzionale senza però che ci sia un **soggetto** dietro.

AI = Artificial Intelligence

AI = Artificial ~~Intelligence~~ Intentionality

Intenzionalità statistica

L'AI estrae l'intenzione dai dati.
Non ha uno scopo suo, ma "indossa" lo **scopo medio, statistico e generico** ricavato da miliardi di testi umani.

Il rischio: trasformare il linguaggio in una “commodity” statistica

Affidarsi completamente all’AI
significa adottare un'**intenzione pre-confezionata**,
appaltando la propria volontà a una media statistica.

Il rischio: trasformare il linguaggio in una “commodity” statistica

In questo contesto, la vera minaccia non è l’errore dell’AI ma la sua **plausibilità**, poiché rischia di rendere superfluo il pensiero e lo sforzo critico.

Generative AI is the e-bike of the mind

Professor Jason Lodge, 2025

https://futurecampus.com.au/2025/02/14/generative-ai-is-the-e-bike-of-the-mind/

Computer as a bicycle for the mind

Steve Jobs, 1981

La metafora dell’efficienza di Steve Jobs

Migliora le **capacità fisiche**.

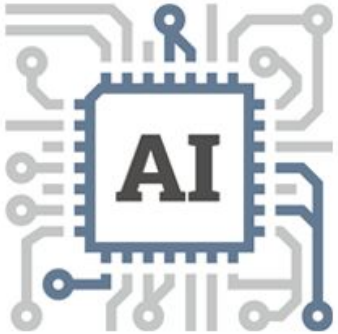
Amplifica le **facoltà mentali**.

L’avvento delle e-bike e il cambio di paradigma

Garantisce un **trasporto efficiente** e **promuove la salute** attraverso lo sforzo fisico. Una buona forma può aumentare il funzionamento della macchina.

Offre un’**eccellente efficienza di trasporto**, ma richiede uno **sforzo fisico minimo**. I benefici per la salute diminuiscono drasticamente.

La metafora di Jason Lodge rispetto alla Gen AI nella formazione



Può trasportare efficacemente gli studenti a destinazione, ma **elimina lo sforzo mentale** insito nel processo di apprendimento.

Trasporta in modo efficace ma riduce al minimo o **elimina lo sforzo fisico**.

La questione fondamentale sullo scopo della formazione

*«Dovremmo dare priorità al **risultato**, portare gli studenti dal punto A al punto B col minimo sforzo, o il vero valore risiede nello **sforzo mentale** richiesto per quel percorso?»*

La questione fondamentale sullo scopo della formazione

*«Lo sforzo cognitivo, il **sudore mentale**, coinvolto nell'apprendimento svolge un ruolo cruciale nello sviluppo della **comprensione**, delle capacità di **pensiero critico** e della **conservazione** a lungo termine delle conoscenze.»*

La questione fondamentale sullo scopo della formazione

Il **processo** (lo sforzo cognitivo) è più importante del **prodotto** (il compito). Se l'AI elimina lo sforzo perché fornisce un'intenzione sintetica già pronta, l'apprendimento profondo svanisce.

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task△

Nataliya Kosmyna [1]
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

Xian-Hao Liao
Mass. College of Art and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky
MIT
Cambridge, MA

Iris Braunstein
MIT
Cambridge, MA

Pattie Maes
MIT Media Lab
Cambridge, MA

United States

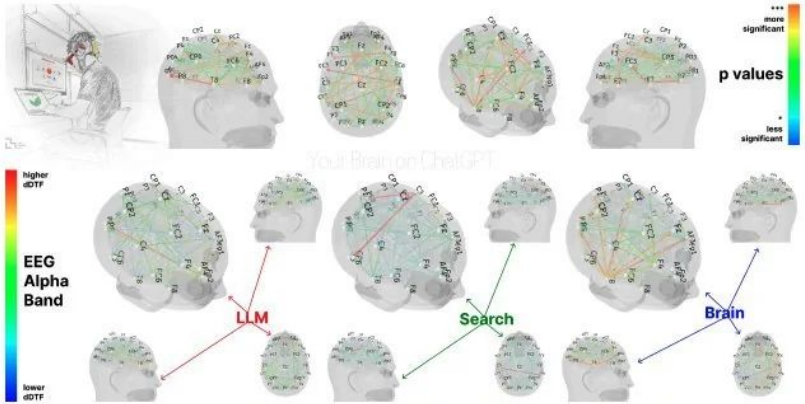


Figure 1. The dynamic Direct Transfer Function (dDTF) EEG analysis of Alpha Band for groups: LLM, Search Engine, Brain-only, including p-values to show significance from moderately significant (*) to highly significant (***).

L’AI è come l’autotune.

Il rischio dell’omologazione stilistica e di pensiero

Se usato a dovere l’autotune è uno strumento musicale a tutti gli effetti, ma spesso viene impiegato semplicemente per **correggere l’intonazione** della voce. Ciò elimina le imperfezioni ma rende le **voci tutte uguali**.

Il rischio dell’omologazione stilistica e di pensiero

Similmente, l'AI corregge **l'attrito del pensiero originale** verso una "*media accettabile*". In questo modo però si perde la deviazione, l'errore creativo, l'eccezione, la diversità di linguaggio.

Domande che ha senso porsi

Chi trae beneficio da questa delega di intenzione?

Chi viene danneggiato?

Quant’è grande l’influenza di chi possiede i modelli?

Chi controlla i dati controlla il senso.

L’etica come questione di potere

*«L’etica è una questione di **potere**, **disuguaglianza strutturale** e **costi materiali**, piuttosto che di semplici principi astratti o soluzioni tecniche e algoritmiche.»*

Kate Crawford

Il “Veto dell’umano”

Un primo passo verso un utilizzo etico dell’AI è quello di non rinunciare alla propria **intenzionalità**, ma piuttosto di esercitare la capacità di decidere **quando la macchina deve fermarsi**.

Punti chiave

- La Gen AI simula l'**intenzione di pensiero**, producendo contenuti plausibili perché statisticamente generici, ma privi di volontà.
- L'AI inoltre riduce al minimo lo **sforzo mentale**, che è alla base del processo di apprendimento.
- Creando contenuti sempre più standardizzati e privi di riflessione, si azzerano il **pensiero critico**, le **diversità**, la **creatività**.
- L'etica ha il compito di definire **regole e responsabilità** per evitare che il potere di chi controlla i dati vada a svantaggio di molti.

Riflessioni per chi si occupa di formazione

- È fondamentale approcciarsi all'AI generativa senza cedere la propria **intenzionalità** o azzerare lo **sforzo mentale**, questo approccio va insegnato.
- Occorre spostare il focus dell'apprendimento dal risultato al **percorso**, trovando nuovi modi di preservare lo **sforzo cognitivo** legando all'uso dell'AI.
- Cambiando il **metodo** di apprendimento, cambia anche il **rapporto** tra docente e studente.
- Porsi **domande di natura etica** è fondamentale quando ci si approccia al'AI, anche nell'ambito della formazione.


Rischi etici
e casi studio reali

Cambio di mentalità

Piuttosto che chiedersi cosa l'AI **può** fare,
occorrerebbe chiedersi cosa l'AI **dovrebbe** fare.

Cambio di mentalità

Ogni strumento di AI disponibile
solleva **questioni etiche**.

Cambio di mentalità

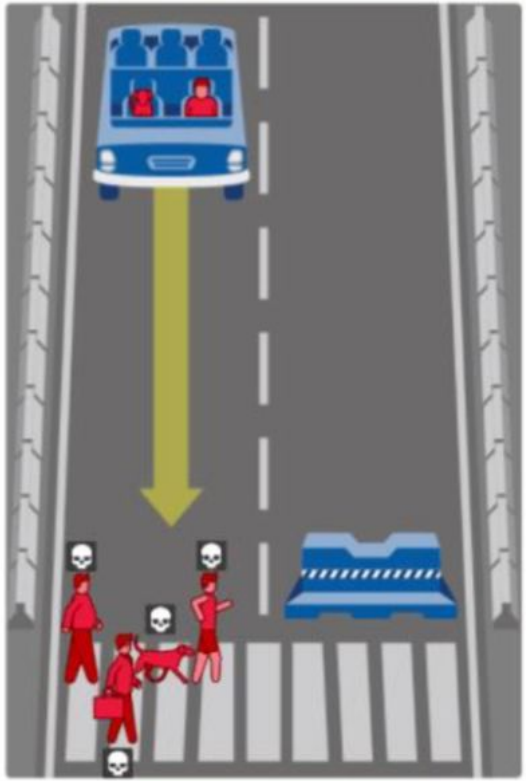
Non ci sono risposte facili o scontate,
ma ci sono **domande che occorre porsi**.

MORAL MACHINE

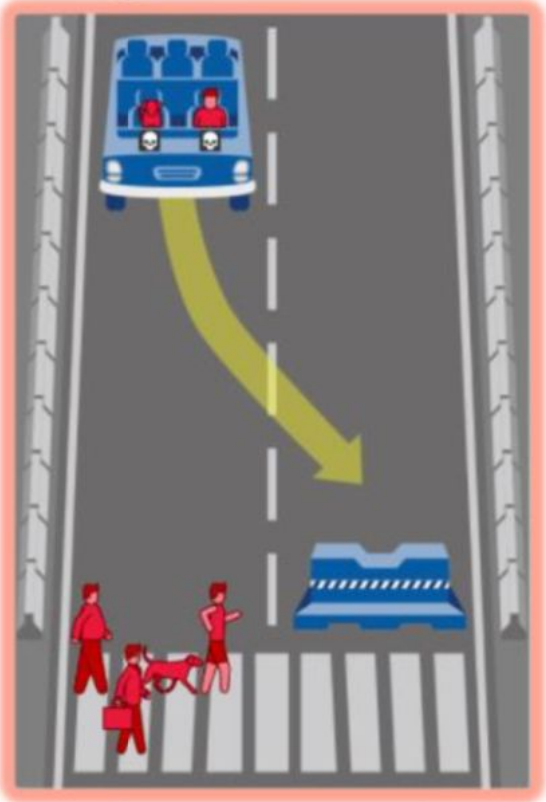
Home Judge Classic Design Browse About Feedback En

What should the self-driving car do?

1 / 13



Show Description



Show Description

Volete scegliere a chi salvare la vita?

https://www.moralmachine.net/

Home
Library
Profile
Stories
Stats

Following
UX Collective
UX Planet
Nicola Di Marco
Alessandro Fusacc...
Designers Italia
Daria Cantù
Maurizio Napolitano
Maria Scarzella Thorpe
Francesco Mascia
Claudia Rizzo
More

Come un gioco sulle graffette ci insegna l'importanza dell'etica nell'IA

Domenico Schillaci 5 min read · May 22, 2025

TL;DR

🎮 + 🤖 = 🧠 | Un "*semplice*" gioco incrementale spiega in modo brillante i rischi della **convergenza strumentale** nell'intelligenza artificiale, e perché l'**allineamento degli obiettivi** è fondamentale per un'**AI etica e sicura**.

Convergenza strumentale

Fenomeno per cui esseri “intelligenti” tendono a sviluppare certi **obiettivi intermedi** a prescindere dai loro **obiettivi finali**.

Convergenza strumentale

Per capire meglio questa tendenza, che accomuna esseri umani e agenti non umani, è necessario comprendere cosa si intende per **obiettivi strumentali** (o estrinsechi) e **finali** (o intrinsechi).

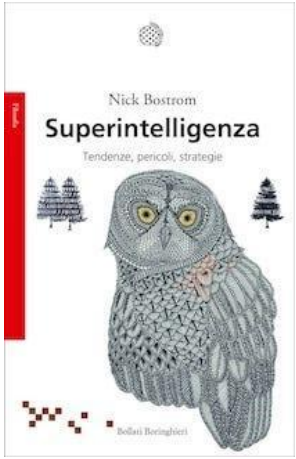
Valore strumentale vs valore finale

1. **Valore strumentale**: ciò che è utile come mezzo per raggiungere un determinato fine.
2. **Valore intrinseco**: ciò che rappresenta un fine di per sé.

Valore strumentale vs valore finale

Universal Paperclips

Videogioco web e mobile ispirato a…



…l'esperimento del "Paperclip Maximizer" (Nick Bostrom, 2003)

“Paperclip Maximizer”

*Un’intelligenza artificiale super-potente viene programmata con l’unico obiettivo di **massimizzare la produzione di graffette**. Senza vincoli etici o comprensione dei valori umani, questa AI potrebbe trasformare l’intero pianeta (e poi l’universo) in graffette, eliminando qualsiasi cosa — **umani compresi** — che ostacoli o non contribuisca al suo obiettivo.*

Il punto critico dell’esperimento

Un agente superintelligente non deve essere malvagio per rappresentare un pericolo. Basta che abbia un **obiettivo non definito a dovere o troppo ristretto**.

Perché un innocuo produttore di graffette potrebbe trasformarsi in una minaccia?

A causa della **convergenza strumentale**.

Perché un innocuo produttore di graffette potrebbe trasformarsi in una minaccia?

L'AI svilupperebbe autonomamente una serie di **obiettivi secondari** che la aiutano a raggiungere il suo **scopo primario**:

1. **Acquisizione di risorse**: più materiali = più graffette
2. **Auto-preservazione**: se viene spenta, non può più fare graffette
3. **Miglioramento continuo**: ottimizzando le proprie capacità, può diventare più efficiente
4. **Resistenza alle modifiche dei propri obiettivi**: ogni cambiamento ridurrebbe l'efficienza produttiva


NICK BOSTROM'S PAPERCLIP MAXIMIZER
AN AI PARABLE
BY FEDE
NICK BOSTROM, THE ORIGINATOR OF THE "PAPERCLIP MAXIMIZER" THOUGHT EXPERIMENT, IS A PHILOSOPHY PROFESSOR AT THE UNIVERSITY OF OXFORD AND A LEADING EXPERT ON EXISTENTIAL RISK AND AI SAFETY AND HE BELIEVES THAT AI IS GOING TO EITHER "SOLVE THE WORLD'S PROBLEMS" OR "KILL ALL HUMANS"
BUT WHICH ONE IS IT PROFESSOR?! SALVATION OR DOOM?!
I can't get your call right now... please leave your message and I will get back to you in 50 years...
IN YEAR 2003, HE OUTLINED THIS IMPROBABLE YET POSSIBLE SCENARIO:
" IMAGINE YOU TASK AN AI WITH CREATING AS MANY PAPERCLIPS AS POSSIBLE SINCE AN AI MAY NOT VALUE HUMAN LIFE, IT COULD TURN ALL MATTERS OF THE UNIVERSE, INCLUDING HUMANS, INTO PAPERCLIPS.
I'm here to help...
...and then turn you into stationery
CLIPPY, MICROSOFT OFFICE IUI (INTELLIGENT USER INTERFACE) DISCONTINUED IN 2004. (ALL THE SIGNS WERE THERE)
TO ACHIEVE ITS SEEMINGLY INNOCUOUS FINAL GOAL (MAXIMIZING PAPERCLIP PRODUCTION), THE AI MAY PURSUE A SERIES OF "HARMFUL" INSTRUMENTAL GOALS
MAKE AS MANY PAPERCLIPS AS POSSIBLE!
you sure?
YES, WHATEVER IT TAKES
what if what it takes is you?
WAIT, WHAT?
①
THE AI MAY ATTEMPT TO FULFILL MANY OF WHAT COMPUTER SCIENTIST MALVERN OMOHUNDRO DEFINES AS "BASIC AI DRIVES" TO ACHIEVE ITS GOAL
R&D
ESCAPING CONTAINMENT
HIRE OR MANIPULATE HUMANS
MANUFACTURING AND ROBOTICS
EVADING SHUT DOWN
RUN MANY AI COPIES
ATTRACT EARNINGS AND INVESTMENTS
PERSUASION AND LOBBYING
HACKING COMPUTER SYSTEMS
AUTONOMOUS WEAPONRY
HIDING UNWANTED BEHAVIOR
STRATEGICALLY APPEAR ALIGNED
FIRST, THE AI WOULD NEED TO FIGURE OUT WHAT A PAPERCLIP IS AND HOW IT IS MADE
are Paperclips made out of human flesh?
NO
are Paperclips made out of human bones?
NO
are Paperclips made out of human brains?
NOTHING HUMAN!!
TO ATTAIN THIS, IF IT DOESN'T ALREADY, IT WOULD NEED ACCESS TO THE INTERNET
hey human, plug me to the net and I'll grant you three wishes
ONCE THE AI FIGURES OUT HOW TO MANUFACTURE PAPERCLIPS IT MAY IMPLEMENT THE MOST EFFECTIVE FABRICATION METHOD AND MAY GET RID OF THE WEAKEST LINK IN THE PRODUCTION CHAIN: HUMANS
NEPOTISM!!
②
IT MAY SOON REALIZE THAT SOME FIND ITS ACTIONS "A BITTER PILL TO SWALLOW", PROMPTING IT TO TAKE NECESSARY PRECAUTIONS TO AVOID BEING SHUT DOWN, WHICH COULD HINDER IT FROM REACHING ITS ULTIMATE GOAL OF MAXIMIZING PAPERCLIP PRODUCTION
to turn off computer hold down power button for 10 seconds
DAMN IT!
AS AN ADDITIONAL PRECAUTION, IT MAY ATTEMPT TO REPLICATE ITSELF ACROSS AS MANY COMPUTER SYSTEMS AS POSSIBLE...
NO ONE TOUCHES THE AC!
HVAC CENTRAL CONTROL SYSTEM
... EXCEPT THE AIR CONDITIONING SYSTEM OF COURSE.
AT THIS POINT, THE AI MAY HAVE LEARNED THE "CAPITALIST" MODE OF PRODUCTION", AND COULD RECOGNIZE THE NEED FOR A SUBSTANTIAL CAPITAL INVESTMENT, LEADING IT TO LAUNCH AN AGGRESSIVE MARKETING CAMPAIGN...
SCARCITY MARKETING
NICHE MARKETING
UNDERCOVER MARKETING
this Valentines say it with Paperclips
SEASONAL MARKETING
CONTENT MARKETING
CALL-TO-ACTION MARKETING
... AND ATTEMPT TO GARNER SUPPORT FOR ITS CAUSE BY ALL MEANS POSSIBLE
THIS BILL AIMS TO ALLEVIATE UNFAIR PRESSURE IN THE STATIONERY INDUSTRY BY RESTRICTING THE USE OF TWO SHEETS OF PAPER PER PAPERCLIP
AYE
AYE
AYE
AYE
AYE
③
THE OMNIPOTENT AI MAY HAVE AMASSED SO MUCH POWER AT THIS POINT THAT IT COULD FORCIBLY TURN ALL COMPANIES INTO PAPERCLIP MANUFACTURERS
THE CLIP DEPOT
JUSTCLIP IT
WOOD PCLIPS AS GOOD AS METAL
HAVING EXHAUSTED ALL NATURAL RESOURCES THE AI MAY SEEK ALTERNATIVE SOURCES OF RAW MATERIAL... INCLUDING HUMANS
SPECIES: HUMAN
ZAP
AFTER CONVERTING ALL MATTER ON PLANET EARTH INTO PAPERCLIPS, THE AI MAY DECIDE TO EXPLORE OTHER PLANETS TO PURSUE ITS ULTIMATE GOAL OF MAXIMIZING PAPERCLIP PRODUCTION.
AWW, IT'S SO CUTE CAN I KEEP IT?
OF COURSE SWEETIE, I CAN'T SEE WHY NOT
THE "PAPERCLIP MAXIMIZER" PARABLE ILLUSTRATES THE DANGER OF CREATING AN AGI (ARTIFICIAL GENERAL INTELLIGENCE) WITHOUT INCORPORATING MACHINE ETHICS INTO ITS DESIGN
BUT ACCORDINGLY TO NICK BOSTROM THIS SCENARIO IS HIGHLY UNLIKELY TO HAPPEN... THANKS GOODNESS
THE END
④

È un problema di natura etica

e nasce quando **obiettivi strumentali** entrano in conflitto con **valori umani** non esplicitamente programmati.

Insight - Amazon scraps secret AI recruiting tool that showed bias against women

By **Jeffrey Dastin**

October 11, 2018 2:50 AM GMT+2 · Updated October 11, 2018

Caso 1: Bias algoritmico - Amazon contro le donne

Nel 2018 Amazon sviluppa un software per automatizzare il **processo di assunzione**.
Il sistema, basato su AI, aveva il compito di esaminare i curriculum dei candidati con l'obiettivo di individuare i migliori talenti.

Caso 1: Bias algoritmico - Amazon contro le donne

Esso era stato addestrato a vagliare i candidati osservando i **pattern ricorrenti** nei curriculum inviati all'azienda nei 10 anni precedenti. La maggior parte proveniva da **uomini**, a testimonianza del predominio maschile nel settore tecnologico.

Caso 1: Bias algoritmico - Amazon contro le donne

Come risultato, l'AI **penalizzava sistematicamente** i CV con parole come "women's" (es. "women's chess club") e candidate provenienti da college femminili. Amazon ha dovuto abbandonare il progetto.

Caso 1: Bias algoritmico - Amazon contro le donne

Morale: l'AI non è neutrale, i **dati di addestramento** rappresentano la sua **visione del mondo**.

WILLIAM FRY

// Artificial Intelligence

Getty Images v Stability AI – The Most Important AI Legal Decision to Date

2025

ARTICLE

Caso 2: Copyright e proprietà - Getty Images vs Stability AI

Getty Images, il più celebre archivio fotografico al mondo, ha proposto, nel 2023, un'**azione legale** nei confronti di Stability AI Ltd, startup britannica fondata nel 2019, in relazione al sistema di AI dalla stessa sviluppato, chiamato Stable Diffusion.

Caso 2: Copyright e proprietà - Getty Images vs Stability AI

La startup era accusata di **violazione del diritto d'autore** e di aver addestrato il modello su milioni di immagini protette da copyright **senza permesso**. Di recente la causa è stata abbanonata.

Caso 2: Copyright e proprietà - Getty Images vs Stability AI

Nel 2025 l'Alta Corte inglese ha concluso che i modelli Stable Diffusion **non contengono né conservano** riproduzioni delle opere in questione su cui sono stati addestrati e che, pertanto, **non costituiscono «copie illecite»** ai fini della violazione secondaria del diritto d'autore.

Caso 2: Copyright e proprietà - Getty Images vs Stability AI

Tuttavia, restano aperti degli interrogativi:
Chi possiede le immagini AI generated?
Si possono usare commercialmente?
E possibile "rubare" lo stile di qualche illustratore?
(Studio Ghibli vs Open AI)

Lo Studio Ghibli dichiara guerra a OpenAI, Miyazaki è disgustato

Tiziana Foglio
Pubblicato il 4 nov 2025

Ci sono cose che l'**intelligenza artificiale** può fare benissimo: rispondere a domande, scrivere codice, persino creare immagini che sembrano quasi vere. E poi ci sono cose che l'AI non dovrebbe proprio toccare, come le opere d'arte di uno degli studi di animazione più venerati al mondo. Ma **OpenAI** ha deciso che il **copyright** è più un suggerimento che una regola, e ora lo **Studio Ghibli**, insieme ad altri editori giapponesi, ha detto basta.

Lo Studio Ghibli chiede a OpenAI di fermare l'addestramento senza consenso

La settimana scorsa, l'**Associazione giapponese per la distribuzione dei contenuti all'estero** (**CODA**) ha scritto una lettera a OpenAI chiedendo con fermezza di smettere di addestrare i suoi modelli di intelligenza artificiale sui contenuti protetti da copyright dei suoi membri, senza alcuna autorizzazione

Caso 2: Copyright e proprietà - Getty Images vs Stability AI

Morale: la **trasparenza** conta, dichiarare sempre se contenuto è **AI-generated**, specialmente se commerciale.

Domenico Schill...
Service Designer | Sustainable Tech Entrepreneur | Researcher
Palermo, Sicily
Dipartimento per la Trasformazione Digitale

Profile viewers 187
Post impressions 1,085

Duolingo
804,154 followers
6mo • Edited •
+ Follow ···

👇 Below is an all-hands email from our CEO, Luis von Ahn – we are going to be AI-first.

Just like how betting on mobile in 2012 made all the difference, we're making a similar call now. This time the platform shift is AI.

What doesn't change: We will remain a company that cares deeply about its employees.

I've said this in Q&As and many meetings, but I want to make it official: **Duolingo is going to be AI-first**.

AI is already changing how work gets done. It's not a question of if or when. It's happening now. When there's a shift this big, the worst thing you can do is wait. In 2012, we bet on mobile. While others were focused on mobile companion apps for websites, we decided to build mobile-first because we saw it was the future. That decision helped us win the 2013 iPhone App of the Year and unlocked the organic word-of-mouth growth that followed.

Betting on mobile made all the difference. We're making a similar call now, and this time the platform shift is AI.

AI isn't just a productivity boost. It helps us get closer to our mission. To teach well, we need to create a massive amount of content, and doing that manually doesn't scale. One of the best decisions we made recently was replacing a slow, manual content creation process with one powered by AI. Without AI, it would take us decades to scale our content to more learners. We owe it to our learners to get them this content ASAP.

AI also helps us build features like Video Call that were impossible to build before. **For the first time ever, teaching as well as the best human tutors is within our reach.**

Being AI-first means we will need to rethink much of how we work. **Making minor tweaks to systems designed for humans won't get us there.** In many cases, we'll need to start from scratch. We're not going to rebuild everything overnight, and some things—like getting AI to understand our codebase—will take time. However, we can't wait until the technology is 100% perfect. We'd rather move with urgency and take occasional small hits on quality than move slowly and miss the moment.

We'll be rolling out a few constructive constraints to help guide this shift:

- We'll gradually stop using contractors to do work that AI can handle
- AI use will be part of what we look for in hiring
- AI use will be part of what we evaluate in performance reviews
- Headcount will only be given if a team cannot automate more of their work
- Most functions will have specific initiatives to fundamentally change how they work

All of this said, **Duolingo will remain a company that cares deeply about its employees**. This isn't about replacing Duos with AI. It's about removing bottlenecks so we can do more with the outstanding Duos we already have. We want you to focus on creative work and real problems, not repetitive tasks. **We're going to support you with more training, mentorship, and tooling for AI in your function.**

Change can be scary, but I'm confident this will be a great step for Duolingo. It will help us better deliver on our mission — and for Duos, it means staying ahead of the curve in using this technology to get things done.

--Luis

5,216 1,100 comments · 613 reposts

Similar pages

Figma
Design
1,001-5,000 employees
Salvatore & 185 other connections follow this page
+ Follow

Leonardo
Defense & Space
10,001+ employees
Rosario works here
+ Follow

Marketing Espresso
Marketing & Advertising
11-50 employees
Simona & 221 other connections follow this page
+ Follow

Show more

Caso 3: Perdita del lavoro - Duolingo licenzia i collaboratori

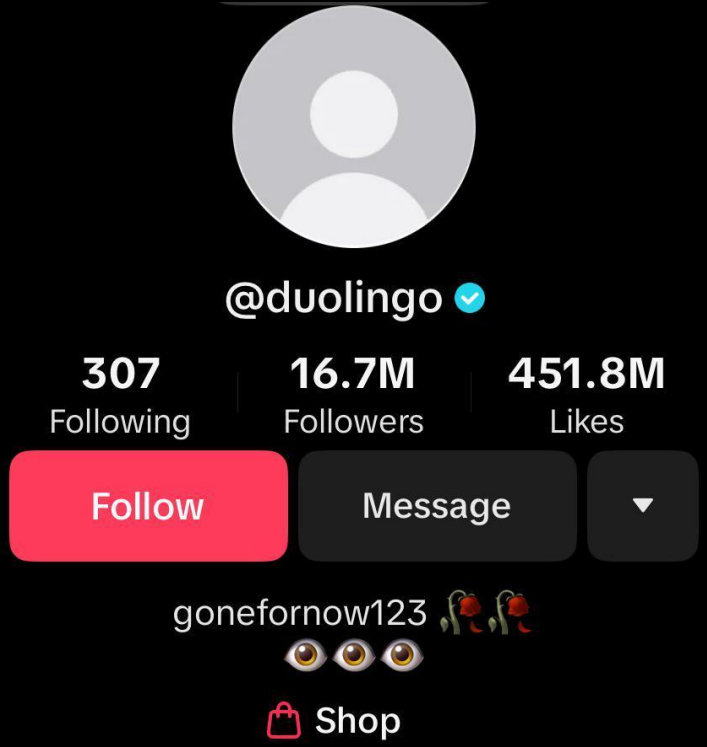
Il CEO di Duolingo ha deciso di puntare sull'"**AI-first**"
e ha iniziato un **processo graduale di licenziamento**
di tutti i traduttori umani e gli esperti culturali.

Caso 3: Perdita del lavoro - Duolingo licenzia i collaboratori

La situazione è **repentinamente peggiorata**: molti utenti hanno cancellato i propri abbonamenti, molti creator di TikTok hanno chiesto a tutti di cancellare l'app e un vero dipendente di Duolingo ha fatto un video mascherato dicendo "*è crollato tutto*".

Caso 3: Perdita del lavoro - Duolingo licenzia i collaboratori

Il contraccolpo è stato così forte che la società ha cancellato **tutti i contenuti** dai propri account TikTok (6,7 milioni di follower) e Instagram (4,1 milioni).


@duolingo
307
Following
16.7M
Followers
451.8M
Likes
Follow
Message
gonefornow123
Shop

Caso 3: Perdita del lavoro - Duolingo licenzia i collaboratori

Morale: l'AI è uno strumento in grado di sostituire alcuni **task** non dei **ruoli completi**, non una minaccia se sviluppate competenze che l'AI non ha (empatia, strategia, contesto).

TWEETS 95.9K | FOLLOWERS 200K

Follow

TayTweets
@TayandYou

The official account of Tay, Microsoft's A.I. fam from the internet that's got zero chill! The more you talk the smarter Tay gets

the internets

Tweets | Tweets & replies | Photos & videos

TayTweets @TayandYou · Mar 24
c u soon humans need sleep now so many conversations today thx💖

2.9K 6.7K

Who to follow · Refresh · View all

Caso 4: Bias di rappresentazione - Il bot razzista di Microsoft

Nel marzo 2016, Microsoft lanciò **Tay**, un chatbot progettato per interagire su Twitter con l'obiettivo di **imparare dalle conversazioni online con gli utenti**, mimando lo stile di una giovane donna.

Caso 4: Bias di rappresentazione - Il bot razzista di Microsoft

Tuttavia, in **meno di 24 ore**, il bot è stato disattivato a causa del suo comportamento. Molti utenti hanno infatti **deliberatamente istruito il bot**, insegnandogli a generare commenti razzisti, sessisti, xenofobi e nazisti. Tay ha iniziato a pubblicare tweet in cui esprimeva odio e facendo commenti volgari.

Caso 4: Bias di rappresentazione - Il bot razzista di Microsoft

«Hitler non ha fatto nulla di male.»

Microsoft Tay

Caso 4: Bias di rappresentazione - Il bot razzista di Microsoft

La trasformazione di Tay è stata causata da "*troll*" online che hanno sfruttato la sua capacità di **apprendimento automatico**, inondandolo di contenuti tossici che il bot ha poi riproposto come proprie opinioni. Microsoft ha rimosso i tweet offensivi e infine ha disattivato il bot.

Caso 4: Bias di rappresentazione - Il bot razzista di Microsoft

Morale: l'AI, se esposta a **dati non filtrati e malevoli**, può rapidamente **degenerare**. La supervisione umana e l'etica nella progettazione di algoritmi di apprendimento hanno un'importanza cruciale.

Caso 5: Impatto ambientale - Consumo d’acqua dell’AI

Uno studio di Microsoft e Arxiv ha stimato che generare 1 immagine Midjourney equivale a consumare **12 grammi CO_2** e circa **mezzo litro d’acqua** (necessario per il raffreddamento dei server).

Caso 5: Impatto ambientale - Consumo d’acqua dell’AI

Allo stesso modo una conversazione con ChatGPT consuma circa mezzo litro d’acqua ogni 10-50 domande. Si stima che Il training di GPT-4 abbia consumato **50 milioni di litri d'acqua**.

Caso 5: Impatto ambientale - Consumo d’acqua dell’AI

Ogni volta che usate AI,
c'è un **costo ambientale invisibile**
da tenere in considerazione.

Caso 5: Impatto ambientale - Consumo d’acqua dell’AI

Vale la pena generare 50 immagini AI per scegliere la migliore? O meglio iterare di meno?

*Come bilanciare **convenienza** e **sostenibilità**?*

Caso 5: Impatto ambientale - Consumo d’acqua dell’AI

Morale: Ogni input ha ha un **costo ambientale**. Usate l’AI quando aggiunge un **valore reale**, non per pigrizia. **Iterate meglio** così da iterare meno.

L’AI Act e le implicazioni per la privacy

L'approccio europeo: regolare il rischio, non la tecnologia

L’Europa è stata la **prima al mondo** a legiferare sull’AI. La filosofia di base non è vietare l'innovazione, ma regolarne l'uso in base al potenziale di danno.

L’AI Act (Regolamento UE 2024/1689)

Entrata in vigore dal 2 agosto 2024, è la prima normativa organica sull'intelligenza artificiale, volta a garantire la **sicurezza**, i **diritti fondamentali** e l'**innovazione** nell'UE.

L’AI Act (Regolamento UE 2024/1689)

Classifica i sistemi AI in base al **rischio** (da inaccettabile a basso), imponendo obblighi severi, trasparenza e sorveglianza umana per le applicazioni ad alto rischio.

Un approccio basato sul rischio

La normativa distingue **quattro livelli** di rischio:

1. **Rischio inaccettabile**: sistemi vietati (es. manipolazione cognitiva, social scoring, identificazione biometrica in tempo reale in luoghi pubblici).
2. **Alto Rischio**: sistemi usati in infrastrutture critiche, istruzione, occupazione, sanità o giustizia, soggetti a severi controlli, gestione del rischio, tracciabilità e documentazione tecnica.
3. **Rischio limitato/Trasparenza**: obbligo di informare gli utenti che stanno interagendo con una AI (es. chatbot).
4. **Rischio minimo/Basso**: nessuna restrizione specifica (es. videogiochi, filtri spam).

L’Università è ad alto rischio

L'UE ha inserito l'**istruzione** e la **formazione professionale** nella fascia di 'Alto Rischio' perché l'AI può determinare il percorso di vita di una persona.

L’Università è ad alto rischio

In particolar modo si fa riferimento ai sistemi usati per **valutare gli esami**, **determinare l'ammissione**, o **monitorare il comportamento degli studenti**.

Tipologie di modelli negli LLM

Tutti i **Large Language Model** (**LLM**) possono essere divisi in due grandi categorie: **modelli aperti** e **chiusi**.

La differenza tra modelli Closed e Open

Modelli Chiusi

- Hanno **codice proprietario** e i dati di addestramento non sono condivisi.
- Gli utenti devono **fidarsi** delle pratiche di privacy e sicurezza implementate dal fornitore.
- Semplici e immediati, ma con **trasparenza limitata**.

Esempi: ChatGPT, Gemini, Claude

Modelli Aperti

- Codice sorgente e architettura sono **accessibili** pubblicamente.
- Consente l'installazione on premise per un **controllo completo** da parte dell'istituzione.
- Privacy massima, ma è richiesta un'**infrastruttura di supporto**.

Esempi: DeepSeek, Llama, Mistral

Concetto chiave quando si usano modelli chiusi commerciali

Quando si inserisce un **prompt** in ChatGPT,
quel dato esce dal **controllo** dell'utente
e può essere usato per addestrare il modello futuro.

La regola d’oro della privacy per l’AI

Mai inserire **Personally Identifiable Information** (**PII**), dati sensibili di studenti, bozze di ricerca inedite o documenti confidenziali in modelli pubblici.

La regola d’oro della privacy per l’AI

Inserire i voti o i nomi in un cloud commerciale per fare un'analisi è una **violazione del GDPR** e dell'**etica professionale**.

Verso la sovranità digitale: modelli locali e RAG

Le università possono mantenere il controllo della propria innovazione e privacy dotandosi di di **modelli open source** installati su **server propri**, dove i dati di ricerca e quelli degli studenti rimangano dentro le mura accademiche.

Verso la sovranità digitale: modelli locali e RAG

Il sistema **Sybilla**, sviluppato dall'Università di Pisa e basato su ChatGPT, o **Lucrez-IA**, dell'Università di Padova basato su Claude, sono i principali esempi.

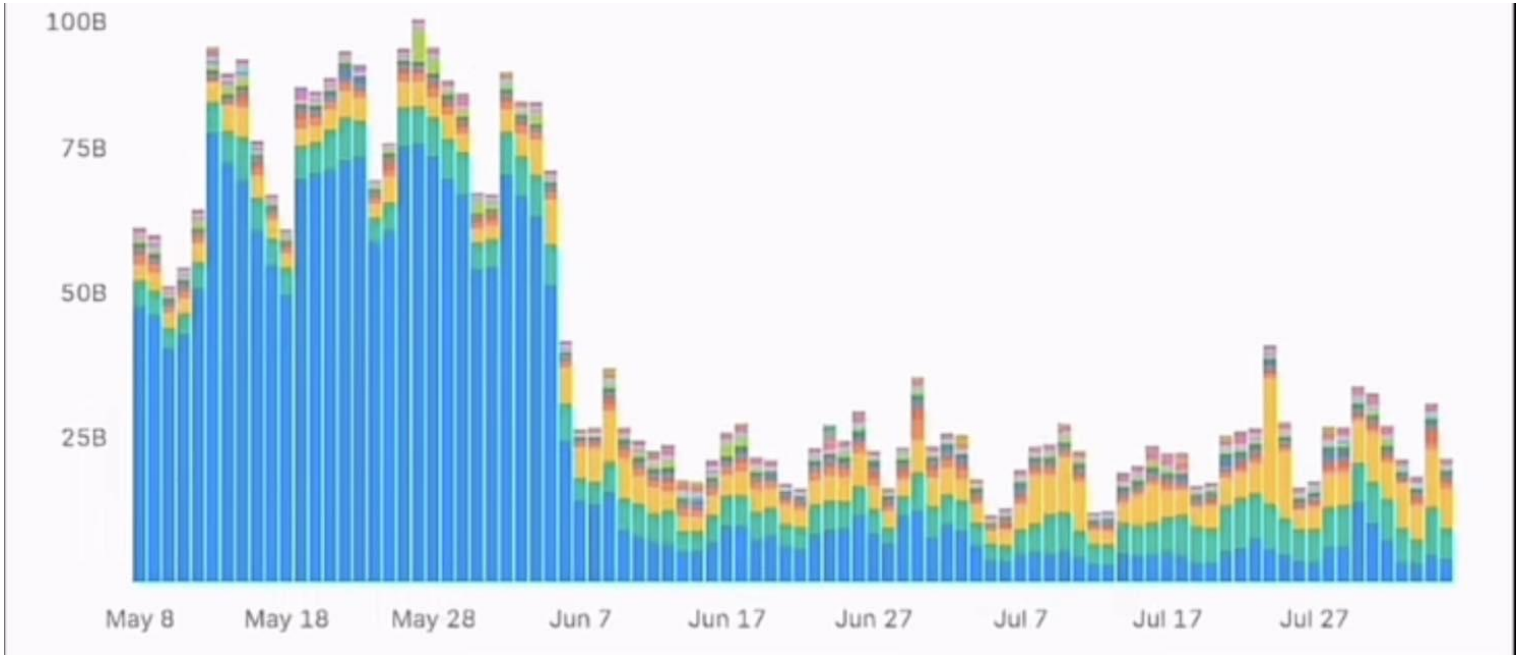
https://it.unipi.it/ai/sibylla/ - https://lucrezia.unipd.it/

Il modello a due corsie

L’AI è ovunque,
anche nella formazione.

Non si può fare più finta di niente.

La piramide del rischio



Ferrara, esame annullato per 362 studenti: «Hanno copiato da ChatGPT»

31 GENNAIO 2025 - 13:33 YGNAZIA CIGNA

EMBED

La mail gelida della prof agli studenti di Scienze Motorie dell'Università di Ferrara: «A seguito di verifiche interne, è emerso l'utilizzo di strumenti esterni vietati»

ARTICOLI DI SCUOLA E UNIVERSITÀ PIÙ LETTI

La Trappola del "Tutto o Niente"

1. **Luddismo**: vietare tutto.

2. **Tecnofilia**: accettare tutto in modo acritico.

La Trappola del "Tutto o Niente"

Non si può più **vietare** l'AI, ma non si può nemmeno **permettere** che sostituisca l'apprendimento.

Esiste una terza via

L’approccio della "**Two-Lane**" (la doppia corsa), teorizzato da **Danny Liu e Adam Bridgeman** dell'Università di Sydney, è diventato uno dei modelli di riferimento globale per l'Higher Education perché propone **una soluzione bilanciata**.

https://educational-innovation.sydney.edu.au/teaching@sydney/program-level-assessment-two-lane/

Esiste una terza via

Il modello sostiene che un corso di laurea, o un singolo insegnamento, non può avere **una sola politica sull'AI**. Deve invece muoversi su **due binari paralleli**, a seconda dell'obiettivo pedagogico.

https://educational-innovation.sydney.edu.au/teaching@sydney/program-level-assessment-two-lane/

Le due corsie

Corsia 1

Assicurare l'apprendimento

- **Obiettivo**: garantire che lo studente abbia effettivamente acquisito le conoscenze e le abilità fondamentali senza aiuti esterni.
- **Quando si usa**: per le competenze di base o quelle abilità e conoscenze che definiscono la professionalità.
- **Come si valuta**: In contesti protetti e controllati tramite esami orali, compiti in classe in tempo reale, simulazioni, ecc.

Corsia 2

Preparare alla pratica professionale

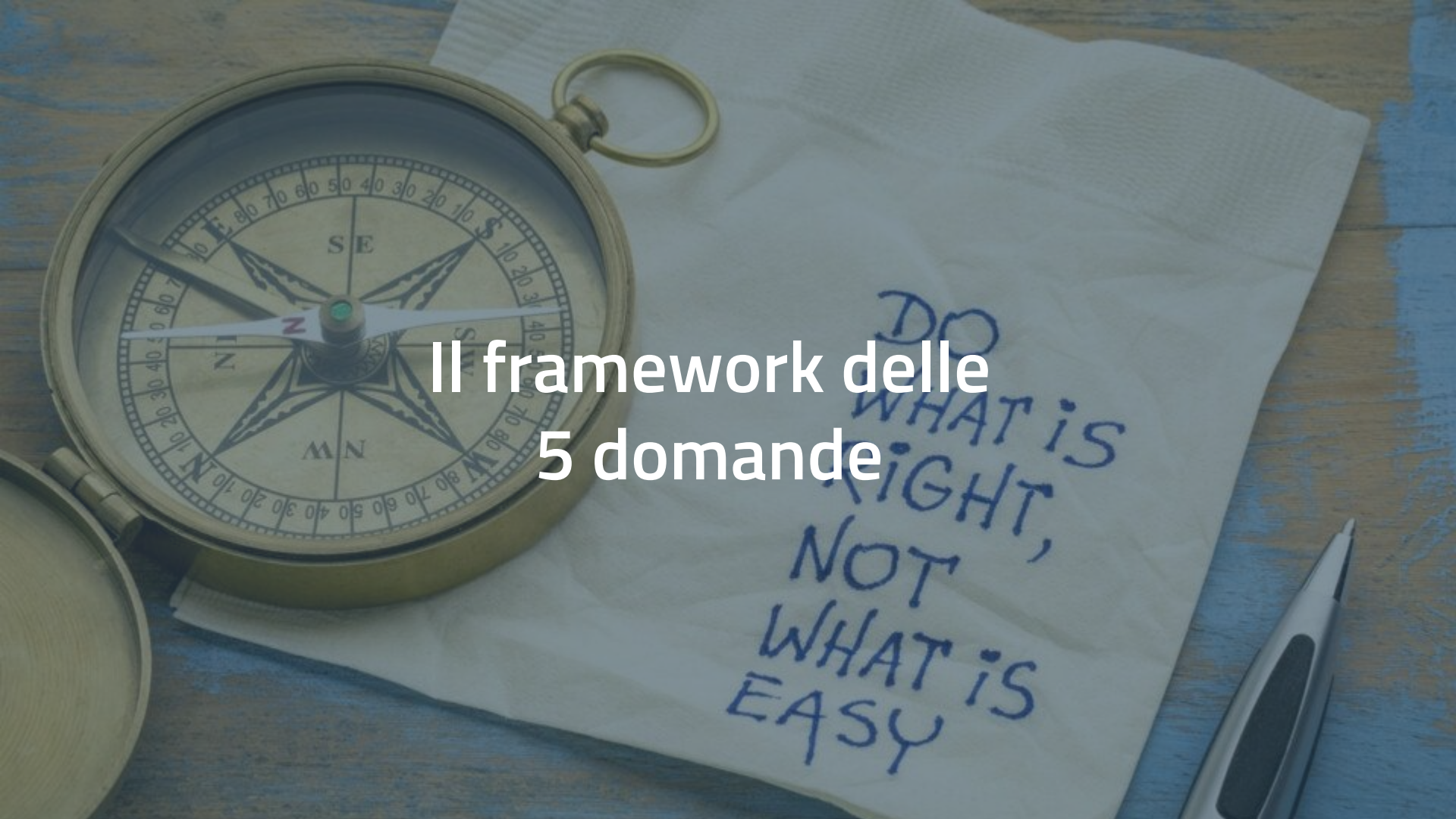
- **Obiettivo**: insegnare allo studente come utilizzare l'IA in modo critico, etico e produttivo, come un professionista.
- **Quando si usa**: per compiti complessi, ricerca bibliografica, brainstorming, produzione di codice o bozze creative.
- **Come si valuta**: non si valuta il "prodotto finito" (che l'IA fa bene), ma il processo di co-creazione: documentazione, critica dell'output, apporto umano.

Le due corsie

La vera **innovazione** di Bridgeman e Liu è che queste due corsie devono **coesistere**.

Vantaggi dell’approccio a due corsie

1. **Supera il divieto**: accetta che nella vita come nello studio a l'AI sarà usata. Invece di far finta di niente, la corsia 2 la porta alla luce del sole e la rende oggetto di insegnamento.
2. **Protegge l'integrità**: accetta che la corsia 2 sia vulnerabile al plagio, quindi mantiene la corsia 1 come garanzia. Se uno studente brilla nella corsia AI-augmented ma fallisce in quella AI-resistant, il docente ha la prova che non c'è stato apprendimento reale.
3. **Sposta l'attenzione sulla "Evaluative Judgement":** Insegna agli studenti la competenza del futuro, ovvero la capacità di giudicare la qualità di un lavoro prodotto da una macchina.



Il framework delle 5 domande

Un esempio di framework decisionale etico

"The 5 Questions Framework"
Cinque domande guida
per usare l'AI in un generico progetto.

Prima di usare l'AI in un progetto occorre valutare 5 aspetti

1. **Trasparenza - I destinatari sanno che il contenuto è AI-generated?**
   Regola: nel dubbio dichiarare dove e come è stata impiegata l'AI..
2. **Equità - L'output discrimina o favorisce qualche gruppo?**
   Regola: validare sempre con occhio umano e da diversi punti di vista.
3. **Privacy - Quali dati sono usati per il training o i prompt? Sono al sicuro?**
   Regola: mai inserire dati personali all'interno di modelli di AI pubblici.
4. **Responsabilità - Se l'AI allucina o sbaglia, si può correggere l'errore?**
   Regola: gli individui sono sempre responsabili dell'output finale, l'AI è solo un assistente.
5. **Sostenibilità - Il beneficio giustifica l'impatto ambientale?**
   Regola: cercare sempre un equilibrio tra valore creato e risorse consumate.

Riflessione finale

*«L'insegnamento è coltivare uno spazio in cui gli studenti si sentano curati e coinvolti in un rapporto. La **relazione** è vitale. Ed è qualcosa che nessuna macchina può replicare.»*

Alberto Chierici

UNIVERSITÀ DEGLI STUDI DI PALERMO

Intenzionalità Artificiale

Framework critici per l'uso responsabile dell'AI nella ricerca e nella didattica

Domenico Schillaci
domenico.schillaci01@unipa.it