AI MINDSET · PENSIERO CRITICO

Le regole del gioco. Chi le ha scritte e perché devi saperlo.

Ogni AI ha una costituzione. Alcune aziende la pubblicano. Altre no. Capire la differenza ti rende un utente più libero e più competitivo.

Stai usando ChatGPT o Claude ogni giorno. Ma hai mai pensato che dietro ogni risposta c'è un documento scritto da esseri umani che decide come quella AI deve comportarsi, cosa può dirti e cosa no, a chi deve obbedienza e in quale ordine?

Questi documenti esistono. Sono pubblici. E cambiano tutto il modo in cui dovresti interpretare quello che l'AI ti dice.

Si chiamano Model Spec e Constitutional AI. Sono due approcci diversi al problema più urgente dell'era digitale: come si insegnano i valori a una macchina.

> ***"Non basta usare l'AI. Bisogna capire da chi ha imparato a risponderti."***

1 Cosa sono e da dove vengono

Model Spec è il documento pubblicato da OpenAI, la società dietro ChatGPT. La prima versione è uscita nel maggio 2024. L'idea è semplice: mettere per iscritto, in modo pubblico e leggibile, come i modelli devono comportarsi, chi può dare istruzioni all'AI e in quale ordine di priorità.

Constitutional AI è invece l'approccio sviluppato da Anthropic, la società che ha creato Claude. Nasce da un paper scientifico del dicembre 2022 e parte da un'idea diversa: invece di usare migliaia di persone per valutare ogni risposta, si dà all'AI un insieme di principi scritti, una vera e propria costituzione, e le si insegna a valutare da sola se le sue risposte rispettano quei principi.

> ✦ **DOVE TROVARE I DOCUMENTI ORIGINALI**
>
> OpenAI Model Spec (versione aggiornata): model-spec.openai.com
> OpenAI Model Spec su GitHub: github.com/openai/model_spec
> Constitutional AI (paper originale Anthropic 2022): arxiv.org/abs/2212.08073
> La Costituzione di Claude: anthropic.com/news/claudes-constitution

2 La differenza che conta davvero

© 2026 @milenatantera_coach | MILENA TANTERA | AI MINDSET & CAREER COACH | Riproduzione vietata

Entrambi i documenti cercano di rispondere alla stessa domanda: come si fa a rendere un'AI utile senza che sia pericolosa? Ma lo fanno in modo molto diverso.

MODEL SPEC: LE REGOLE SCRITTE PER GLI ESSERI UMANI

OpenAI pubblica un documento pensato principalmente per essere letto da persone.

Il Model Spec definisce una gerarchia chiara: prima vengono le regole di OpenAI, poi le istruzioni degli sviluppatori che usano le API, poi le richieste degli utenti. Ogni livello può modificare il comportamento del modello entro certi limiti. È un sistema trasparente, flessibile e aggiornato nel tempo. La versione del dicembre 2025 è rilasciata con licenza Creative Commons CC0: chiunque può usarla, adattarla, costruirci sopra.

CONSTITUTIONAL AI: I PRINCIPI CHE L'AI INTERIORIZZA

Anthropic insegna all'AI a ragionare sui propri comportamenti usando una lista di principi.

Nel Constitutional AI il modello non si limita a seguire regole: viene addestrato a criticare le proprie risposte, a chiedersi se violano i principi della costituzione, e a riscriverle. Il processo si chiama RLAIF (Reinforcement Learning from AI Feedback) e riduce drasticamente la necessità di supervisione umana su ogni singola risposta. La costituzione di Claude attinge alla Dichiarazione Universale dei Diritti Umani, ai principi di DeepMind e a prospettive non occidentali.

3 Cosa cambia per te che la usi

Potresti pensare che siano dettagli tecnici che non ti riguardano. In realtà cambiano moltissimo il modo in cui dovresti leggere le risposte che ricevi.

COSA DEVI SAPERE COME UTENTE

L'AI non è neutrale. Ogni risposta è filtrata da valori scelti da qualcuno.
Sapere chi ha scritto quei valori ti aiuta a capire i limiti di ogni strumento.
Il Model Spec di OpenAI ti dice esplicitamente che l'AI deve obbedire prima a OpenAI, poi agli sviluppatori, poi a te.
La Costituzione di Claude è pensata per far sì che l'AI ragioni sui principi, non solo che li applichi meccanicamente.
Entrambi i documenti sono pubblici e leggibili: leggerli è un vantaggio competitivo concreto.

4 Le curiosità che non ti aspetti

Dietro questi documenti ci sono scelte sorprendenti, contraddizioni interessanti e dettagli che raccontano molto di più di quanto sembri.

IL MODEL SPEC È DI TUTTI

OpenAI ha rilasciato il documento con licenza Creative Commons CC0: pubblico dominio.

Una società privata che pubblica i propri principi etici come bene comune. Chiunque può copiarlo, adattarlo e usarlo nei propri progetti. Vale la pena rifletterci.

CLAUDE HA UNA COSTITUZIONE DI 79 PAGINE

La versione del 2026 spiega non solo cosa Claude deve fare, ma perché.

© 2026 @milenatantera_coach | MILENA TANTERA | AI MINDSET & CAREER COACH | Riproduzione vietata

Mentre le prime versioni erano liste di principi brevi, la costituzione aggiornata di Claude è un documento lungo e articolato che cerca di spiegare il ragionamento dietro ogni scelta. È uno dei pochi casi in cui una società di AI ha reso pubblica non solo la regola, ma la motivazione.

L'AI CHE VALUTA SE STESSA

Nel Constitutional AI il modello critica le proprie risposte prima che tu le veda.

Anthropic ha scoperto che un modello addestrato con la propria costituzione come guida produce risposte più sicure rispetto a uno addestrato solo con feedback umani. In pratica l'AI impara a fare da sola il lavoro che prima facevano migliaia di revisori umani.

IL MODEL SPEC È ARRIVATO DOPO IL PRODOTTO

OpenAI ha pubblicato questo documento quando ChatGPT esisteva già da anni.

Alcuni ricercatori hanno fatto notare che un documento del genere avrebbe dovuto esistere prima del lancio del prodotto. Il fatto che sia arrivato tardi racconta qualcosa sul ritmo con cui si muove questo settore e su quanto sia difficile pensare prima di costruire.

5 Perché tutto questo ti riguarda

Se usi l'AI per lavorare, per creare contenuti, per prendere decisioni o per formarti, stai delegando una parte del tuo pensiero a uno strumento che ha valori incorporati. Non è un problema. È una realtà da conoscere.

Chi conosce questi documenti sa leggere tra le righe delle risposte che riceve. Sa quando l'AI sta applicando un limite imposto da chi l'ha costruita. Sa come formulare meglio le domande. Sa quale strumento scegliere per quale scopo.

In un mercato del lavoro in cui l'AI è ovunque, capire come funziona davvero è già un vantaggio. Non devi diventare una ricercatrice. Devi sapere abbastanza da fare le domande giuste.

> ***"L'AI più potente che puoi usare è quella di cui capisci i limiti."***

Usare l'AI in modo consapevole non è un lusso per addetti ai lavori. È la competenza base del professionista che vuole restare rilevante.

© 2026 @milenatantera_coach | MILENA TANTERA | AI MINDSET & CAREER COACH | Riproduzione vietata