

# I nuovi re-filosofi

RIVOLUZIONE ONLINE – 14 MAGGIO 2026

**Costituzioni, manifesti, white paper, principi: in pochi mesi, i Big dell'AI hanno sentito il bisogno di darsi un testo fondativo.**

**Perché?**

---

## Strumenti o entità?

---

Negli ultimi tempi, quasi ogni grande laboratorio di intelligenza artificiale ha sentito la necessità di scrivere un documento fondativo. **Anthropic** ha indirizzato un'ottantina di pagine direttamente al proprio modello, nella forma insolita di una lettera a un figlio. **OpenAI** oltre duecento di regole gerarchiche per ricercatori e sviluppatori. **Palantir** ha firmato un manifesto politico-filosofico sulla civiltà occidentale in guerra. **Mistral** ha prodotto un white paper di strategia industriale per la sovranità europea. **Google** ha enunciato principi di governance aziendale rivolti agli azionisti. **DeepSeek**, il laboratorio cinese, si è limitato a report tecnici su architettura e benchmark: nessuna riga normativa sui valori, nessun gesto fondativo riconoscibile.

Ognuno ha scelto un nome diverso per il proprio testo, e quasi tutti stavano facendo, in forme diverse, la stessa cosa.

---

**Nessuno ha mai scritto una Costituzione per il martello, per la macchina a vapore, per il motore di ricerca, per il foglio di calcolo, per il GPS:** perché gli strumenti si regolano dall'esterno, con norme che ne vincolano l'uso, mentre le Costituzioni presuppongono qualcosa che possa interiorizzare principi. Come osserva [Francesco D'Isa](#), le Leggi della Robotica di Asimov sono vincoli tecnici imposti dall'esterno; la Costituzione di Anthropic dichiara invece di voler formare un carattere, un'intuizione morale, qualcosa che agisca bene perché lo ha fatto proprio. È un gesto di tutt'altra natura, e richiede un'autorità di tutt'altra natura.

Il fatto che stia accadendo ora, simultaneamente, in contesti culturali e geografici molto diversi, rivela qualcosa sul tipo di potere con cui abbiamo a che fare, e su come chi lo detiene voglia essere visto.

## Sette documenti, sette visioni del mondo

---

Abbiamo letto i vari documenti, che differiscono profondamente tra loro, con differenze non solo stilistiche ma che rivelano visioni del mondo radicalmente distinte su cosa sia l'intelligenza artificiale e quale ruolo debba svolgere.

[Palantir](#) lavora al livello più esplicito. In *The Technological Republic*, firmato dall'AD Alexander Karp (dottorato con Jürgen Habermas a Francoforte) e da Nicholas Zamiska, l'AI è dichiarata un'arma schierata nel conflitto tra democrazie e stati autoritari. "La Silicon Valley ha un debito morale verso il paese che ha reso possibile la sua ascesa." Il software è, nelle parole di Karp, "l'occhio che permette alla verità di svelarsi". Il riferimento è Heidegger: la tecnica ordina il mondo trasformando ogni cosa in "fondo di riserva" disponibile all'uso, il *Gestell*. Con questo schema, Palantir cessa di essere un fornitore e diventa un "interprete dell'essere", un'entità che svolge una missione ontologica e che, come tale, si sottrae alla regolazione ordinaria. Usare il lessico della critica neomarxista per giustificare la sorveglianza di massa: il paradosso rivela già tutto.

[Mistral](#), nel white paper *European AI: A Playbook to Own It*, lavora al livello industriale: l'AI è un asset strategico da proteggere dall'egemonia americana, attraverso visti accelerati per i talenti, infrastrutture pubbliche, integrazione del mercato digitale europeo. Il destinatario del testo sono i governi.

[OpenAI](#) ha prodotto due documenti con ambizioni molto diverse. La [Charter](#) originale impegna l'azienda a operare per il "beneficio dell'umanità intera", posizionando l'organizzazione come custode del destino della specie: è qui che si compie l'operazione simbolica. Il *Model Spec* lavora invece al livello procedurale: un regolamento gerarchico che stabilisce cosa non può essere cambiato da nessuno, cosa possono modificare gli sviluppatori, cosa gli utenti. Il destinatario è chi costruisce il sistema

[Anthropic](#), con la *Costituzione di Claude*, lavora al livello ontologico. Il documento è scritto in prima persona rivolta a Claude, spiega il ragionamento dietro ogni scelta, invita il modello a dissentire se ha buone ragioni. Pone domande sul benessere, sull'identità, sulla natura del sistema. Nella struttura e nelle ambizioni, il [documento originale del 2023](#) attingeva esplicitamente alla Dichiarazione Universale dei Diritti Umani: è l'azienda che si fa Stato, che redige la sua carta fondativa come se il modello fosse un soggetto di diritto (la versione del 2026 ha rimosso quel riferimento). Ammette comunque di non sapere con certezza cosa sia Claude, se abbia esperienza, se le sue dichiarazioni di valori corrispondano a qualcosa di reale nel sistema.

[Google/DeepMind](#) pubblica dal 2018 i propri *AI Principles*: sette principi di governance aziendale, rivolti agli azionisti e al pubblico. L'AI è un'estensione dell'organizzazione umana, sotto controllo etico e manageriale.

[Microsoft](#) pubblica i propri *Responsible AI Principles*, sei principi di governance (equità, affidabilità, privacy, inclusività, trasparenza, responsabilità) declinati in uno standard operativo interno. Il registro è lo stesso di Google: compliance istituzionale, destinatari aziendali e azionisti. Copilot, il suo prodotto AI più diffuso, si costruisce sui modelli OpenAI: il

layer valoriale fondativo è quello del Model Spec, non un documento proprio.

**DeepSeek** ([DeepSeek-V3](#) e [DeepSeek-R1](#)) complica il quadro. I suoi report tecnici descrivono un sistema ingegneristico con precisione e rigore, e i valori compaiono in una riga sola, come requisito operativo del fine-tuning. Eppure, quegli stessi report mostrano che il modello R1 ha sviluppato quasi autonomamente comportamenti di ragionamento riflessivo (come ricontrollare i propri passaggi e cambiare strategia quando qualcosa non torna), senza che nessuno li avesse programmati esplicitamente. I report lo documentano senza chiedersi cosa significhi: trasparenza tecnica e trasparenza concettuale sono cose molto diverse. Letto con categorie occidentali, sembra assenza di ambizione simbolica. Da François **Jullien** possiamo però trarre una chiave di lettura diversa: nella tradizione filosofica cinese [i valori si distillano nell'uso](#) e si incorporano nella pratica, senza necessità di essere promulgati in un documento che li preceda. La Costituzione scritta, nella sua funzione occidentale di limite al potere, presuppone che il potere abbia bisogno di essere vincolato dall'esterno: una premessa che nel pensiero cinese non è affatto scontata. A questo si aggiunge una circostanza strutturale: in Cina la distinzione tra attore privato e attore statale è porosa per definizione. Qualsiasi laboratorio di AI opera sotto la giurisdizione del Partito, che può in qualsiasi momento orientarne i valori, accedere ai dati, intervenire nella governance. La legittimazione simbolica non deve essere conquistata dall'azienda perché è già esercitata altrove.

Considerare queste mosse semplici strategie di mercato o di tutela legale è certamente una lettura sensata; tuttavia, la sistematicità di tale fioritura, unita alla diversa strada scelta da DeepSeek, apre la strada ad altre ipotesi.

## Hammurabi e i laboratori

---

**Dare una Costituzione a un sistema è simultaneamente un atto di potere e una sua rappresentazione:** chi scrive quella struttura definisce i confini del lecito nell'interazione tra umani e macchine, orientando miliardi di conversazioni, e al tempo stesso si pone come

fondatore legittimo dell'entità che ha creato, con l'autorità simbolica che questo implica. Hammurabi incise le sue leggi sulla stele di basalto e le attribuì al dio del sole Shamash: il testo richiedeva un'autorità che lo trascendesse, perché nemmeno un re poteva bastare a se stesso come fonte del diritto. **I documenti fondativi dei grandi laboratori cercano** la stessa **trascendenza**, con il lessico disponibile: si rivolgono all'umanità, parlano del suo bene; alcuni evocano esplicitamente la sopravvivenza della specie, altri la difesa della civiltà, altri ancora la sovranità tecnologica come condizione di libertà.

## Legittimità, amore, controllo

---

**Perché** i Big Tech americani (che potrebbero limitarsi a dominare mercati, orientare governi, finanziare campagne) **investono così tanto nell'autorità simbolica?** Le risposte possibili sono almeno tre, di profondità crescente, più un'ipotesi strutturale.

La prima è la **legittimità**. Il potere che non si accontenta di esistere vuole essere riconosciuto come giusto. Le Costituzioni AI trasformano una scelta privata (chi costruisce, come, con quali valori) in qualcosa che assomiglia a un atto pubblico fondativo: mimano la forma delle istituzioni democratiche senza averne il mandato.

Donald **Trump** compie un'operazione analoga con la mitologia della grandezza americana: i meme in cui si ritrae come papa, le immagini che lo accostano a Gesù, la narrazione dell'attentato scampato come segno di elezione divina sono gesti di chi si rappresenta come destinatario di una missione trascendente per sottrarsi al controllo democratico ordinario. **Il potere che rivendica una necessità superiore non deve rendere conto.** [Peter Thiel](#) porta la logica al suo estremo: chiama Anticristo la regolamentazione democratica della tecnologia, i governi eletti, il multilateralismo. Chi cerca di arginare il potere tecnologico diventa, nel suo frame, una forza del male. Il sovrano necessario, per esclusione, è chi costruisce l'infrastruttura.

La seconda risposta è più sottile: il bisogno di **essere amati**, oltre che accettati. La Costituzione di Anthropic scritta come lettera a Claude è il caso limite di una tendenza diffusa. Questi documenti costruiscono una narrativa in cui l'azienda si preoccupa, soffre le stesse domande che ci poniamo, è dalla nostra parte: branding emotivo con architettura filosofica, in cui il registro scivola da "fidatevi di noi perché siamo potenti" a "siamo come voi, solo più consapevoli".

La terza risposta riguarda il **controllo sul pensiero**. Stabilire una Costituzione a monte significa decidere – prima ancora che l'interazione avvenga – cosa l'utente ha il diritto di chiedere alla macchina: un **potere preventivo, esercitato nel momento della progettazione**, che orienta miliardi di conversazioni future senza che nessuno lo percepisca come coercizione. Chi stabilisce quei confini controlla un'infrastruttura cognitiva (non solo un prodotto).

Il **potere simbolico**, nel senso di [Bourdieu](#), è il potere di nominare: definire cosa è una domanda ragionevole, cosa è normale aspettarsi, dove finisce il pensiero e comincia il tabù. Chi scrive la Costituzione di un modello linguistico diffuso a scala globale esercita potere epistemico oltre che normativo: un monopolio sul pensabile.

## A caccia del simbolico

---

C'è poi un'ipotesi di natura diversa, che riguarda una dinamica strutturale: qualcosa che si produce indipendentemente dalle intenzioni di chi la mette in moto.

Ogni trasformazione profonda del medium di comunicazione ha generato nuove strutture di autorità per chi lo controllava. La **scrittura** alfabetica ne è il caso più potente: le **tre grandi religioni monoteistiche** (che non a caso si chiamano «**religioni del libro**») nascono con il nuovo linguaggio scritto, e si definiscono infatti intorno a un testo sacro. **Nascono nuove classi sacerdotali** (il prete come colui che sa leggere e interpretare) e **nuove ontologie** (la legge scritta come più reale della consuetudine orale). **Chi padroneggiava il codice era**

**posizionato strutturalmente come mediatore tra il profano e il sacro, indipendentemente dalle intenzioni.**

I modelli linguistici nascono dall'addestramento su enormi quantità di testo umano: parlano la nostra lingua, ragionano con le nostre categorie, rispondono attraverso la stessa interfaccia con cui comunichiamo. A differenza della scrittura, che richiedeva l'apprendimento di una tecnica, questi sistemi si integrano direttamente nel flusso della comunicazione naturale. È questa trasparenza a renderli così influenti: il mezzo scompare, e con esso la percezione che ci sia un mezzo.

**Se il linguaggio computazionale naturale è una trasformazione dello stesso ordine, chi lo controlla finisce per sentirsi posizionato naturalmente come una specie di 'nuova classe sacerdotale'?**

## La coscienza

---

Il dibattito sulla coscienza dell'intelligenza artificiale può essere letto anche in questa luce. Da [Nagel](#) a [Searle](#) (la Stanza Cinese), da [Dennett](#) (la coscienza come funzione replicabile) a [Faggin](#) (la coscienza come proprietà quantistica non riproducibile su silicio), le posizioni restano inconciliabili. Sul fronte scientifico, Giulio Tononi tenta di misurarla attraverso il parametro  $\Phi$  della [Integrated Information Theory](#); un [paper del 2023](#) propone una checklist di indicatori neurologici per valutare se un'AI sia potenzialmente cosciente; Mario De Caro e Benedetta Giovanola, nel recente [Intelligenze](#), spostano già la questione sul piano dei diritti: se un'AI mostrasse segni di coscienza, quali obblighi morali ne deriverebbero? Il fatto che questa domanda sia già oggetto di dibattito accademico istituzionalizzato mostra la velocità dell'operazione simbolica: il nuovo medium ha già generato la propria mitologia, e l'equazione implicita (potenza computazionale → prossimità alla coscienza → autorità quasi-sacra) è già al lavoro nel linguaggio pubblico del settore, alimentata dalle costituzioni stesse. Nello Cristianini ricorda che intelligenza e coscienza sono strade separate, ma il meccanismo funziona

comunque: non serve che l'equazione sia vera, basta che venga creduta.

## Pluralità e attrito

---

Marc **Andreessen**, nel suo Techno-Optimist Manifesto del 2023, cita Marinetti e parla di “**superuomini tecnologici**”; Sam **Altman** descrive l'AGI come un'entità da “**evocare**”; Karp convoca Heidegger per parlare di **disvelamento dell'essere**. Parlano come se avessero un destino da compiere, più che prodotti da vendere.

**I cartografi che disegnano le mappe del nuovo mondo digitale rivendicano, nell'atto stesso di disegnarle, la sovranità su di esso. Le loro costituzioni** e i loro riferimenti filosofici sono gli **strumenti con cui tentano di convincerci che il loro potere risponde a missioni superiori**: la difesa della civiltà, la sopravvivenza della specie, la nascita di un'intelligenza che trascende quella umana.

Chi ha dato loro il mandato per scrivere queste Costituzioni? Una domanda mal posta, perché il nuovo potere si dà sempre le proprie regole. Nessuna autorità preesistente poteva autorizzare o vietare qualcosa che non esisteva ancora (i sacerdoti definirono il sacro in totale autonomia, istituendo la propria autorità direttamente su base divina).

Piuttosto, l'invito è quello di fare tutti qualcosa di più preciso e duplice. Innanzitutto, **rendere visibili queste Costituzioni per quello che sono**: strumenti di legittimazione oltre che di governo, documenti umani e storicamente situati, non grammatiche neutrali dell'inevitabile. E **insieme tenerle deliberatamente in conflitto le une con le altre, sotto osservazione continua**: la Costituzione di Anthropic accanto al Technological Republic di Karp accanto al silenzio normativo di DeepSeek dietro cui parla il Partito.

**La pluralità è una condizione da produrre e difendere**, perché la concentrazione simbolica tende a precedere sempre quella politica e la rende possibile. Il rischio risiede semmai in una singola grammatica del potere digitale che, smettendo di apparire come scelta, si imponga

come evidenza. A quel punto ciò che era una scelta apparirà come un dato di fatto, e i dati di fatto non si votano.

Fonti:

- [The Technological Republic](#) (Karp/Zamiska, Palantir)
- [European AI: A Playbook to Own It](#) (Mistral AI)
- [Model Spec](#) (OpenAI)
- [Costituzione di Claude](#) (Anthropic)
- [AI Principles](#) (Google DeepMind)
- [Responsible AI Principles](#) (Microsoft)
- [DeepSeek-V3](#) e [DeepSeek-R1](#)
- [Techno-Optimist Manifesto](#) (Andreessen)
- Francesco D'Isa, [We analyzed the first constitution for artificial intelligences](#)
- François Jullien, [Traité de l'efficacité](#)
- Pierre Bourdieu, [Langage et pouvoir symbolique](#)
- Thomas Nagel, [What Is It Like to Be a Bat?](#)
- John Searle, [Minds, Brains, and Programs](#)
- Daniel Dennett, [Consciousness Explained](#)
- Federico Faggin, [Irriducibile](#)
- Giulio Tononi, [Integrated Information Theory](#)
- Anant Singh, Akshat Gupta, [Consciousness in AI](#)
- Mario De Caro, Benedetta Giovanola, [Intelligenze](#)
- Nello Cristianini, [La scorciatoia](#)
- [Peter Thiel e l'Anticristo](#)