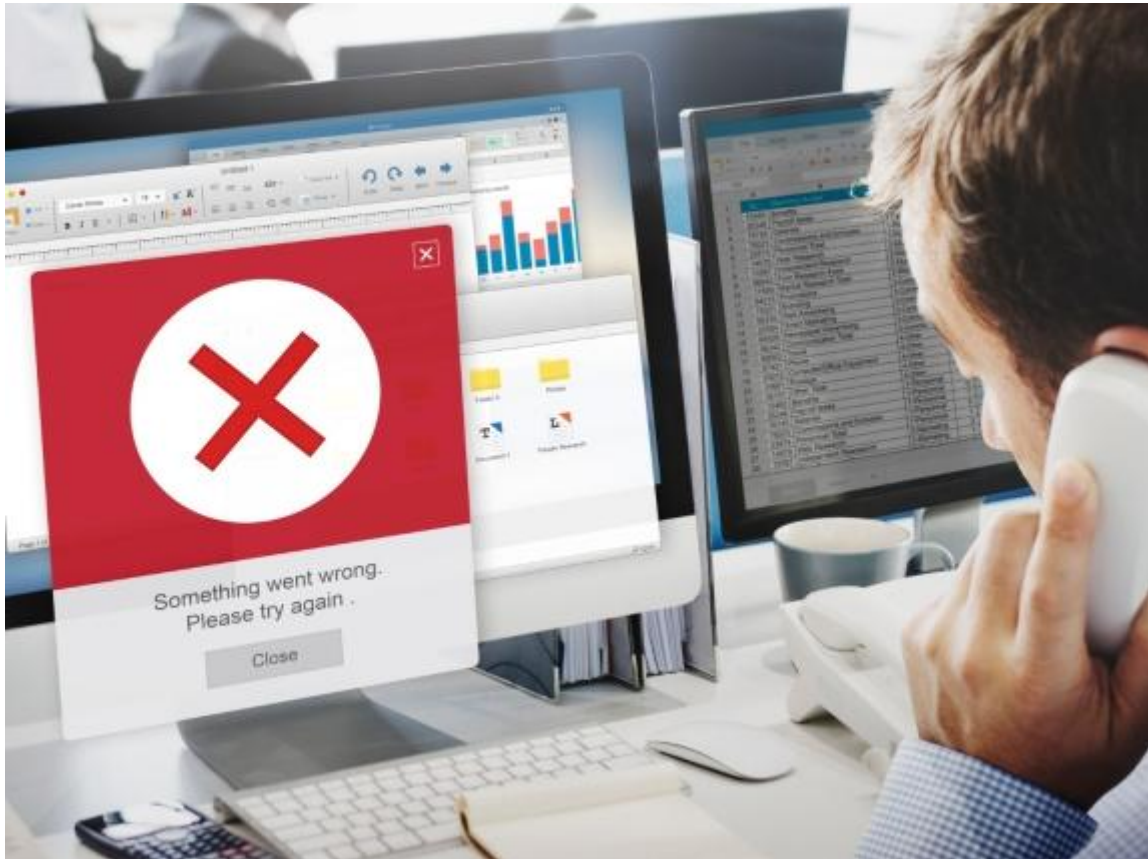


## «Diamo dei limiti all'AI». L'appello di oltre 200 scienziati, politici e premi Nobel: «Rischiamo di perdere il controllo»

Giorgio Parisi, Geoffrey Hinton, Yuval Noah Harari ed Enrico Letta tra i firmatari di un appello all'ONU: servono linee rosse che l'intelligenza artificiale non può oltrepassare

(Fonte: <https://www.corriere.it/> 3 ottobre 2025)



Se i Paesi non riescono a mettersi d'accordo su cosa fare con l'intelligenza artificiale, devono almeno concordare su cosa l'AI non può fare. È la premessa di un appello firmato da oltre duecento persone — tra cui nove ex capi di stato e dieci premi Nobel — [presentato da Maria Ressa, premio Nobel per la pace, durante l'80esima sessione dell'Assemblea Generale dell'ONU](#).

Tra i firmatari ci sono **Geoffrey Hinton** e **Yoshua Bengio**, premi Turing e pionieri dell'intelligenza artificiale, il premio Nobel per la fisica **Giorgio Parisi**, il premio Nobel per l'economia **Joseph Stiglitz**, lo storico **Yuval Noah Harari**, scienziati di OpenAI, Anthropic e Google e politici come **Yanis Varoufakis**, l'ex premier **Enrico Letta** e l'ex ministra dell'Università e della Ricerca **Maria Chiara Carrozza**. L'iniziativa chiede un accordo globale entro la fine del 2026 su linee rosse da non oltrepassare.

### I rischi dell'AI incontrollata

«L'AI», si legge nell'appello, «ha un enorme potenziale per migliorare il benessere umano, ma l'attuale traiettoria di sviluppo comporta pericoli senza precedenti». **Potrebbe presto superare le capacità umane e «amplificare minacce come pandemie ingegnerizzate, disinformazione diffusa, manipolazione delle persone su larga scala — inclusi i minori — rischi per la sicurezza**

**nazionale e internazionale, disoccupazione di massa e violazioni sistematiche dei diritti umani».** Il problema esiste già: «alcuni sistemi di AI avanzata hanno già manifestato comportamenti ingannevoli e dannosi, eppure a questi sistemi viene concessa un'autonomia sempre maggiore nell'agire e nel prendere decisioni nel mondo reale».

Casi recenti e tragici legati all'AI — delusioni indotte da sistemi avanzati, [persino suicidi](#) — sono solo «primi segnali di pericoli molto più grandi che ci aspettano», avverte il comunicato. Molti esperti sottolineano che, senza una governance adeguata, «nei prossimi anni diventerà sempre più difficile garantire un controllo umano effettivo dell'intelligenza artificiale».

### **La sottile linea rossa**

La soluzione dovrebbe passare per un accordo internazionale con «limiti chiari e verificabili per prevenire rischi universalmente inaccettabili» e meccanismi di controllo che valgano per tutti i fornitori di AI avanzata. L'appello non specifica quali debbano essere questi limiti, ma fa alcuni esempi: **vietare all'AI di impersonare esseri umani, impedire la replicazione autonoma dei sistemi, escludere l'intelligenza artificiale dalla guerra nucleare.**

### **AI come il nucleare**

Perché serve un'istituzione globale e non bastano le politiche delle singole aziende? Le ricerche mostrano che molte di esse privilegiano il profitto rispetto alla sicurezza, e le loro politiche interne «non arrivano a una vera applicazione». **Stuart Russell**, docente di informatica all'Università di Berkeley e tra i firmatari, la vede così: le aziende possono conformarsi «non costruendo l'intelligenza artificiale generale finché non sanno come renderla sicura... proprio come gli sviluppatori di energia nucleare non hanno costruito centrali nucleari finché non hanno avuto un'idea di come impedire loro di esplodere». Yuval Noah Harari è ancora più netto: «Per migliaia di anni, gli esseri umani hanno imparato — a volte nel modo più duro — che le tecnologie potenti possono avere conseguenze tanto pericolose quanto benefiche. **Con l'AI, potremmo non avere la possibilità di imparare dai nostri errori, perché l'AI è la prima tecnologia che può prendere decisioni da sola, inventare nuove idee da sola, e sfuggire al nostro controllo».**

L'appello si chiude con una richiesta precisa ai governi: raggiungere «un accordo internazionale sui limiti per l'AI entro la fine del 2026, garantendone l'effettiva applicazione attraverso solidi meccanismi di controllo e di attuazione».

### **Approfondimento**

[Bengio, uno dei padri dell'AI: «Con l'intelligenza artificiale stiamo costruendo macchine che possono superarci. Nelle mani sbagliate rischi enormi»](#)