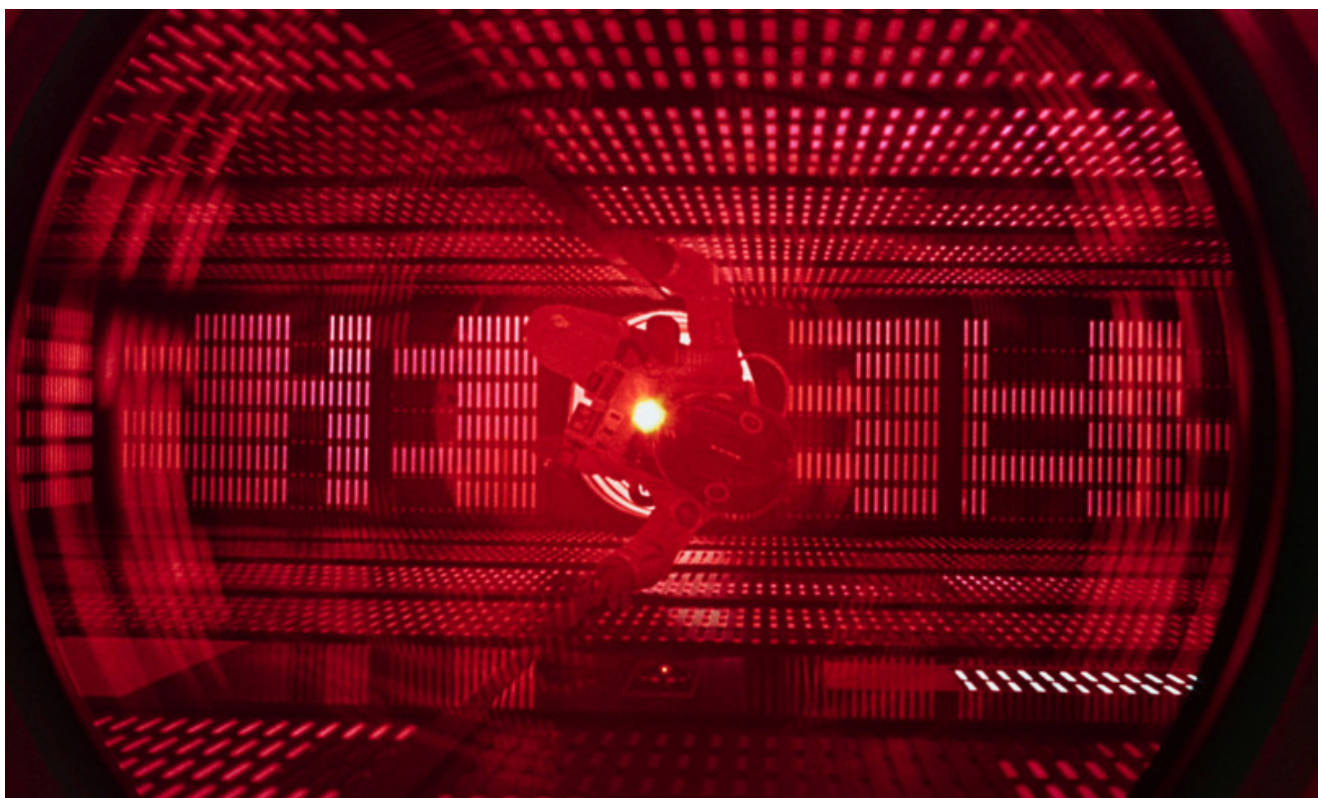


Un'Intelligenza artificiale super potrebbe finire fuori controllo



Studio del Max Planck Institute: impossibile contenere e gestire un sistema troppo intelligente, se collegato a Internet potrebbe assumere il controllo di tutte le macchine online

Una Intelligenza artificiale super intelligente potrebbe essere incontrollabile. A questa conclusione, pubblicata sul Journal of Artificial Intelligence Research, giungono gli esperti del Center for Humans and Machines presso il Max Planck Institute for Human Development, che hanno stabilito l'impossibilità di contenere e gestire, e addirittura riconoscere, un sistema artificiale super intelligente.

Fascino e pericoli

“Siamo affascinati dalle Intelligenze artificiali in grado di

controllare automobili, comporre sinfonie, giocare a scacchi e adempiere alle più svariate funzioni – afferma Manuel Cebrian del Digital Mobilization Group presso il Center for Humans and Machines – ma un sistema con un livello di intelligenza superiore a quello umano, in grado di apprendere autonomamente, sarebbe impossibile da gestire da parte del genere umano”. Parole che fanno riandare alla mente all’Hal 9000 di “2001: Odissea nello spazio”, il computer che finisce fuori controllo e si rivolta contro gli uomini.

Il controllo di tutte le macchine online

Il team ha utilizzato una serie di calcoli teorici per dimostrare l’impossibilità di mantenere il controllo di un’IA super intelligente. “Collegato a Internet – sostengono gli esperti – questo sistema potrebbe avere accesso a tutti i dati dell’umanità e potrebbe sostituire i programmi esistenti e assumere il controllo di tutte le macchine online a livello globale”.

“Una macchina super intelligente in grado di controllare il mondo suona ancora come fantascienza – commenta l’autore – ma ci sono già macchine che svolgono determinati compiti importanti in modo indipendente senza che i programmatori comprendano appieno i processi di apprendimento. Si pone inevitabilmente la questione se tutto ciò possa avere risvolti incontrollabili o pericolosi per l’umanità”. Gli scienziati hanno esplorato diverse idee su come controllare un sistema artificiale super intelligente e sviluppato un algoritmo di contenimento teorico per valutare i pericoli dell’Intelligenza artificiale.

L’algoritmo di contenimento

“Un’Intelligenza artificiale – spiegano i ricercatori – potrebbe essere isolata da tutti gli altri dispositivi tecnici e non stabilire alcun contatto con il mondo esterno, ma questo danneggerebbe l’efficacia dello strumento stesso. In

alternativa, l'Ia potrebbe essere motivata a perseguire solo obiettivi nell'interesse dell'umanità, ma si tratta di idee di controllo non prive di limiti". Il team ha concepito l'algoritmo di contenimento teorico ponendo che l'Ia non possa danneggiare le persone in nessuna circostanza, in modo che un'eventualità di questo tipo si risolva con l'arresto dell'algoritmo.

"Nel nostro attuale paradigma di calcolo – riporta Iyad Rahwan, Direttore del Center for Humans and Machines – tale algoritmo non può essere costruito. Sulla base di questi calcoli il problema del contenimento è incomputabile, ovvero nessun singolo algoritmo può trovare una soluzione per determinare se un'IA potrebbe produrre danni al mondo. Potremmo anche non accorgerci del momento in cui le macchine super intelligenti emergeranno, perché stabilire il grado di intelligenza dei sistemi rientrerebbe nello stesso ambito".