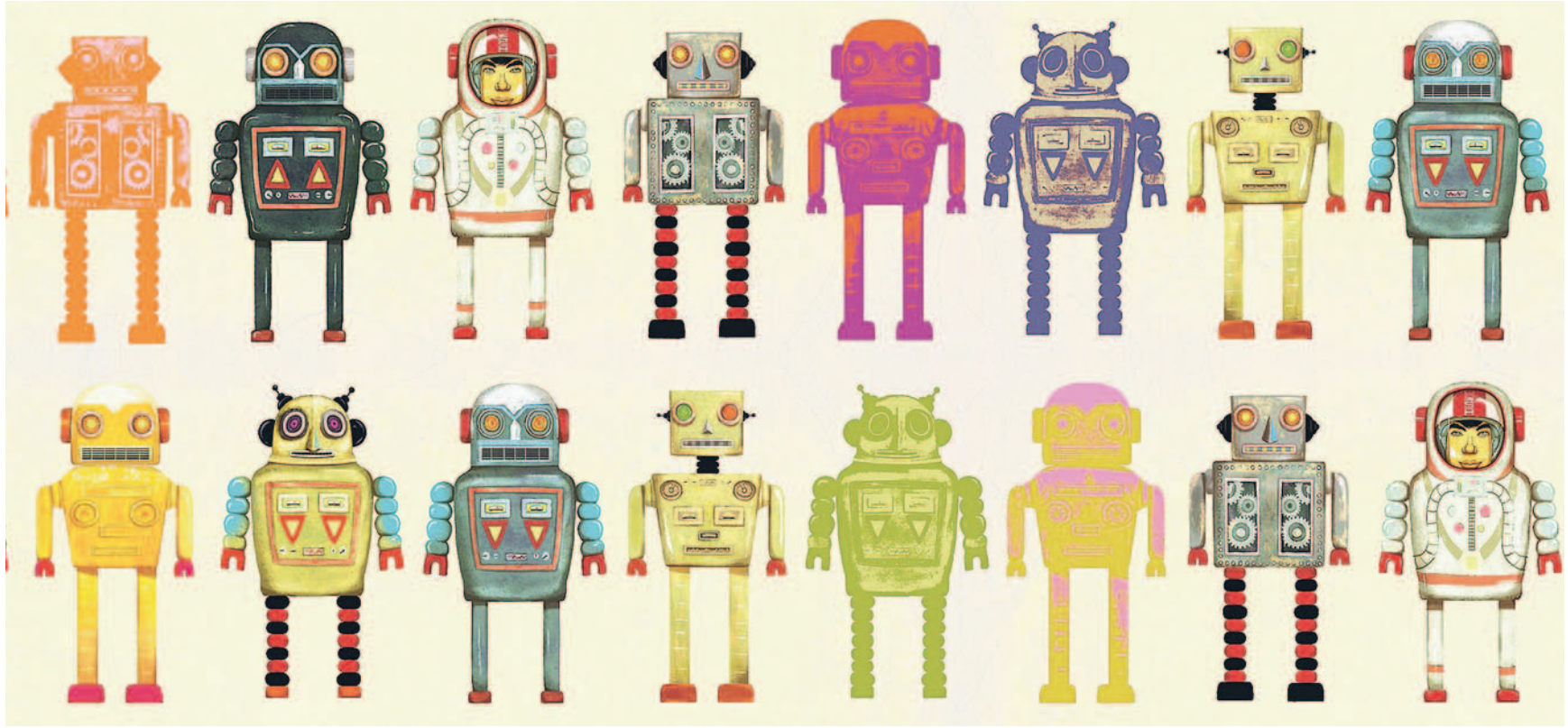
R2

Il caso

L'ultima frontiera è la creazione di intelligenze artificiali capaci di ribellarsi agli input sbagliati o pericolosi



Robot

"Imparino a disobbedire a chi dà ordini ingiusti"

GIULIANO ALUFFI

«APRI la saracinesca esterna, Hal». «Mi dispiace, David, *purtroppo non posso farlo*». La drammatica scena di "2001 Odissea nello spazio" in cui il supercomputer HAL 9000 cerca di eliminare l'astronauta Bowman impedendogli di rientrare a bordo è l'esempio più famoso di disobbedienza elettronica a un ordine umano. Ma l'esplosione robotica attesa nei prossimi anni, che renderà l'uso degli automi sempre più diffuso in qualsiasi ambito, farà moltiplicare le situazioni in cui potrebbero dover dire di no a comandi inappropriati o pericolosi. Basti pensare a un'auto senza guidatore cui un passeggero impaziente ordini di superare i limiti di velocità. Anche per i robot "benintenzionati", diversi quindi dal malevolo HAL9000, l'obbedienza potrebbe quindi non essere più una virtù.

Pionieri nel campo della giusta ribellione robotica sono due ricercatori dello Human-Robot Interaction Laboratory della Tufts University di Medford (Usa), Gordon Briggs e Matthias Scheutz, che al recente Human-Robot Interaction Symposium di Washington hanno presentato uno studio dal titolo evocativo: *"Sorry, I can't do that": Developing Mechanisms to Appropriately Reject directives in human-robot interactions"*.

In quali casi il robot fa bene a ribellarsi? «Se pensiamo ai robot militari, sarà importante riuscire a codificare nel loro comportamento il rispetto delle convenzioni internazionali. Ma ci sono anche ambiti molto più vicini alla vita quotidiana in cui il robot dovrà potersi opporre ai voleri del "padrone" umano», spiega a *Repubblica* Matthias Scheutz. «Immaginiamo un robot-casalingo che ha messo una padella sul fuoco e riceve l'ordine di passare l'aspirapolvere in salotto. Per evitare il rischio di un incendio, dovrebbe rifiutare l'incarico spiegando che prima deve finire di preparare la cena. Oppure, figuriamoci un robot-badante che sta sorreggendo un anziano e riceve da altri l'ordine di riparare l'antenna sul tetto: deve rifiutare il comando almeno fino a quando l'anziano non è al sicuro».

L'importante è che il robot sia un "agente etico" esplicito, ossia che possa non solo agire in un modo etichettabile come etico, ma anche spiegare all'umano il motivo della sua disobbedienza: senza questa capacità, i robot da un lato non riusciranno ad essere accettati come "agenti etici"; dall'altro non avranno modo di essere corretti dalla controparte umana se prenderanno una decisione sbagliata.

La strada per creare il robot etico è ancora irta di ostacoli: «Per poter prendere decisioni di tipo etico il robot dovrebbe costruirsi un corretto modello della situazione in cui si trova, è questa oggi la sfida principale», sottolinea Scheutz. «Consideriamo un'auto robotica che corre sul ghiaccio. Può avere tutte le appropriate regole di guida prudente codificate nel software, ma non potrà applicarle se i suoi sensori non riconoscono il ghiaccio». E c'è un problema aggiuntivo: «Il robot può essere ingannato e tratto in errore dagli umani. Ad esempio l'auto potrebbe percepire il ghiaccio, ma io potrei dirle: "Questo non è ghiaccio: i tuoi sensori non funzionano bene. Continua a correre e ignora i controlli di sicurezza"».

Il rischio maggiore per il futuro non saranno dunque robot troppo etici, ma umani ben poco tali.

©RIPRODUZIONE RISERVATA

IL PRECEDENTE

LE LEGGI DI ASIMOV
"Un robot non può recare danno a un essere umano" è la prima legge della robotica di Asimov (foto). La seconda: "Un robot deve obbedire agli ordini degli esseri umani, purché questi non contravvengano alla prima legge"

Le menti artificiali

60 milioni
le controversie legali risolte ogni anno dal software dedicato di e Bay
(40 volte le cause civili dibattute ogni anno nel Regno Unito)

41 secondi
frequenza di pubblicazione di studi di medicina
Possono stare al passo solo menti artificiali come Watson (di Ibm) o DeepMind (Google)

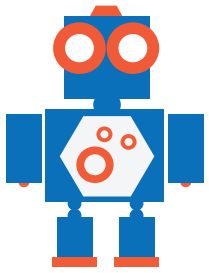
I posti di lavoro
che nei prossimi 20 anni saranno occupati da robot

35%
nel Regno Unito

47%
negli Stati Uniti

I robot "obiettori"
principali ambiti di applicazione

- Campo militare
- Ordine pubblico
- Medicina
- Auto senza guidatore



L'INTERVISTA

"Ma la macchina in grado di usare un criterio morale è ancora lontana"

QUANTO è lunga la strada tecnologica verso il robot etico? Lo chiediamo a Giorgio Metta, vicedirettore scientifico dell'Istituto italiano di tecnologia di Genova e padre di "iCub", il più celebre e avanzato robot umanoide italiano.

Perché le tre leggi di Asimov non bastano?

«In sé e per sé sono ingegnose. Il problema è che presuppongono robot in grado di percepire il mondo nella sua complessità, e di ragionare. Cosa che al momento è molto lontana dalla realtà. Far capire ai robot il contesto in cui si trovano ad agire oggi è possibile solo per contesti estremamente specifici».

Quanto tempo ci separa dai robot etici?

«Stiamo andando verso macchine sempre più autonome, è vero, ma pur sempre in contesti molto limitati. L'orizzonte più prossimo — ed è questione di pochi anni — è quello di robot che possiamo di volta in volta dotare di abilità e conoscenze specifiche, scaricandole dalla rete come se fossero app. Queste macchine potranno rifiutarsi di ubbidire a un comando inappropriato o pericoloso, ma lo faranno in modo ancora "stupido", ossia come reazione automatica di fronte a un ordine che violi un insieme di regole fisse».

E per avere un comportamento davvero "etico"?

«Date le conoscenze attuali, l'orizzonte temporale ipotizzabile per un robot realmente in grado di decisioni consce e "morali" è perlomeno di trent'anni».

(giuliano aluffi)

©RIPRODUZIONE RISERVATA

VkVSIyMjVm9sb0Vhc3lSZWFkZXJfQ2VplyMjZS52aW5haSMjI0xhIFJlcHViYmxpY2EjIyMyNC0xMS0yMDE1IyMjMjAxNS0xMS0yNFQxMTo0Nzo0MFojIyNWRVI=