Tiratura: 359.033 Diffusione: 251.142

il venerdì SUPPLEMENTO DE la Repubblica

Settimanale

Sarzana Festival della Mente

Data 23-08-2019
Pagina 58/61
Foglio 1 / 3

SE IL ROBOT PENSA BENE REAGISCE COME SPOCK

POSSIAMO FIDARCI DELLE MACCHINE? PRIMA DOBBIAMO DARGLI DEI VALORI, DICE UN'ESPERTA IN ETICA DELL'**INTELLIGENZA ARTIFICIALE**. LEI UN MODELLO CE L'HA. FANTASCIENTIFICO

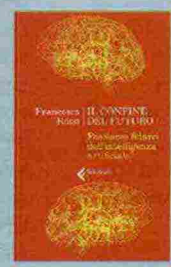
di **Giuliano Aluffi**

AFFIDERESTE la vostra vita a una macchina intelligente? La domanda non è prematura, perché non riguarda le future auto a guida autonoma: già oggi le intelligenze artificiali decidono chi otterrà un mutuo e chi non lo otterrà oppure chi verrà assunto e chi sarà scartato dalle aziende che hanno automatizzato la gestione delle risorse umane. Per questo serve fin d'ora una tecnologia affidabile, in tutti i sensi: al tema Francesca Rossi, global leader dell'Ibm per l'etica dell'intelligenza artificiale, ha dedicato il saggio *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale?* (Feltrinelli), che presenterà il 31 agosto al Festival della Mente di Sarzana e il 19 settembre alla Fondazione Feltrinelli a Milano.

All'inizio del libro Rossi traccia uno scenario futuristico in cui un assistente digitale, Spock, le organizza le attività della giornata, tiene sotto controllo la sua salute e le rende più facile ogni cosa. «L'ho chiamato Spock perché da bambina adoravo la fantascienza» spiega. «Il personaggio di *Star Trek* mi affascinava perché era per metà umano e per metà vulcaniano: in quella serie gli umani si comportavano spesso in maniera impulsiva ed emotiva, i vulcaniani erano l'essenza della logica e della razionalità. Un po' come i computer. Spock, che metteva assieme questi due aspetti, era il mio eroe. Amavo anche i fumetti dei supereroi, dove qualcosa – una tecnologia, una reazione chimica o altro – permetteva a persone normali di trasformarsi, diventando molto più capaci di risolvere problemi, di fare le cose in modo migliore e per il bene della società. L'intelligenza artificiale potrà fare lo stesso nel nostro futuro. Ma dovrà essere affidabile. Spock mi piaceva proprio perché ispirava una fiducia assoluta: un traguardo ancora lontano per le intelligenze artificiali, delle quali oggi la gente non si fida poi tanto».

Non sembrerebbe: 150 milioni di persone hanno caricato la propria foto su FaceApp per vedersi invecchiati senza preoccuparsi del fatto che, tra vent'anni, potrebbero essere identificati da una qualunque videocamera "intelligente" per strada.

«Difficile immaginare in che modo quelle foto potranno essere impiegate, ma certo servirebbe una maggiore precauzione da parte degli utenti nel far circolare i propri dati. FaceApp arricchisce il suo database con il "trucco" dell'invecchiamento perché oggi



Sopra **Francesca Rossi** e il suo saggio *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale?* (Feltrinelli, pp. 128, euro 15). Sotto, il signor Spock, il personaggio metà umano e metà vulcaniano della serie e poi film di fantascienza *Star Trek*, interpretato da Leonard Nimoy

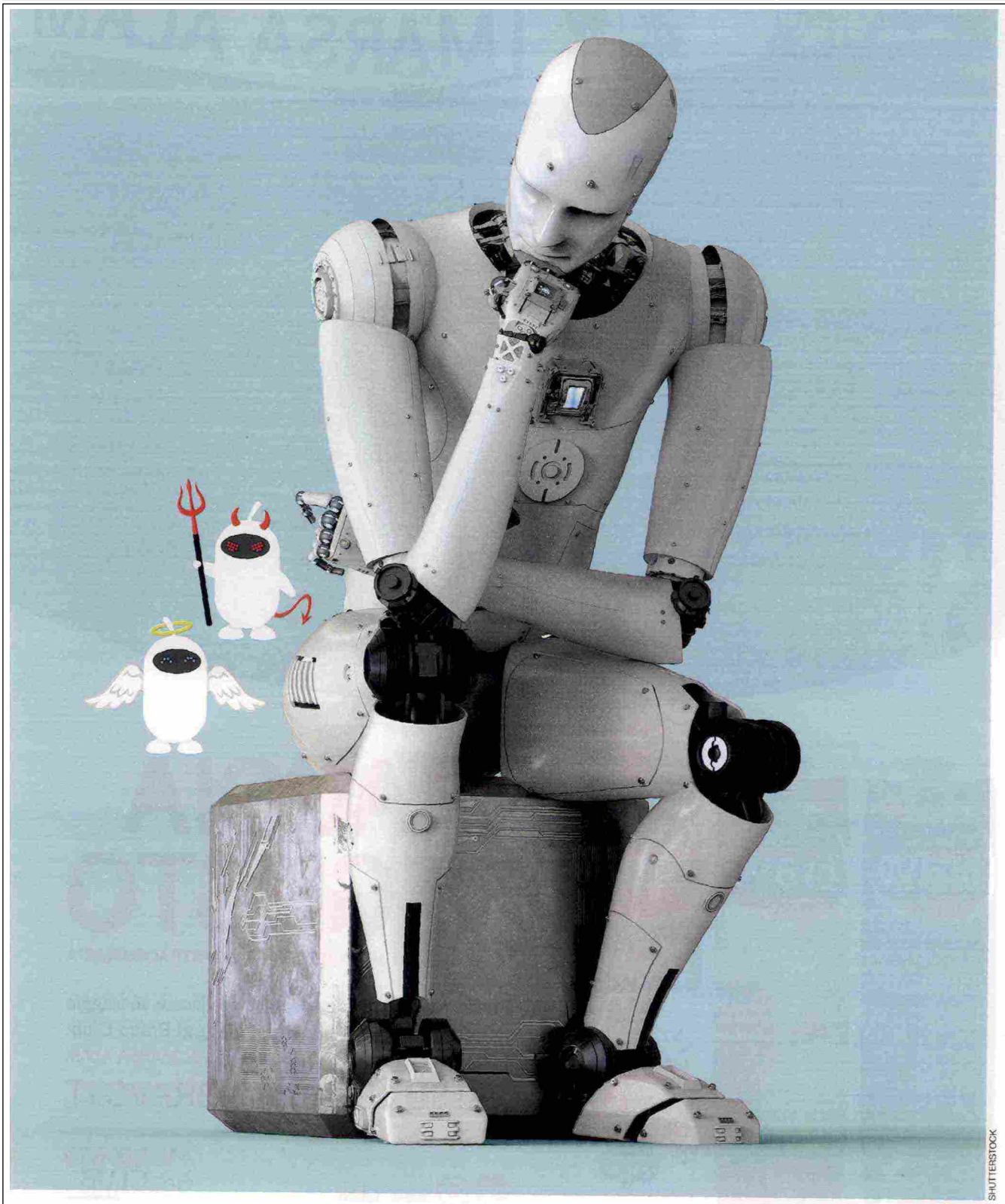
ALAMY / IPA

il venerdì SUPPLEMENTO DE la Repubblica

Settimanale

Sarzana Festival della Mente

Data 23-08-2019
Pagina 58/61
Foglio 2 / 3



SHUTTERSTOCK

il venerdì SUPPLEMENTO DE la Repubblica

Settimanale
Sarzana Festival della Mente

Data 23-08-2019
Pagina 58/61
Foglio 3 / 3

l'intelligenza artificiale ha fame di dati. In questo momento, per funzionare, le tecnologie di maggior successo del settore, quelle incentrate sull'apprendimento – il *machine learning* – hanno bisogno di grandi quantità di dati, che siano immagini o altro. Ma probabilmente questa voracità non sarà eterna: già oggi tanti ricercatori stanno studiando un *machine learning* che usi una minore quantità di dati».

Ma quando e perché servono tanti dati?

«Servono quando non si sa definire con esattezza il problema e i passi per risolverlo, per esempio nel campo del riconoscimento vocale e della traduzione automatica. Qui non ci sono regole ben definite da applicare, come negli scacchi, e per trovare la soluzione di un problema non si può ricorrere a un algoritmo preciso. Così si usa un algoritmo generico e si forniscono alla macchina molti esempi perché possa modificarlo, rendendolo via via più preciso».

Lei dice che questo può portare a discriminazioni...

«È così. Se lei usa Google Translate per tradurre in turco la frase *"He is a nurse"* ("lui è un infermiere"), il risultato sarà *"O bir hemşire"*. Se ora ritraduce in italiano otterrà: "Lei è un'infermiera". Come mai il lui del testo originale è diventato una lei? Per via degli esempi in lingua turca che sono stati usati per allenare il sistema di traduzione automatica, che evidentemente mostrano molte più donne che uomini nel ruolo di infermiere. Simili "sviste" nel fornire esempi alla macchina possono avere anche conseguenze pratiche immediate. Se per esempio devo allenare un sistema che aiuti un impiegato di banca a decidere se accettare o meno una richiesta di prestito o di mutuo, gli fornirò tantissimi esempi di richieste accettate o rifiutate in passato. Ma se, senza rendermene conto, scelgo esempi dove tutte le richieste fatte da donne vengono accettate e tutte quelle fatte da uomini rifiutate, il computer potrà concludere – sbagliando – che l'accettazione o il rifiuto del mutuo è correlato con il sesso della persona che lo chiede. E suggerire di rifiutare richieste di mutuo perfettamente accettabili solo perché sono fatte da uomini».

Sarzana Festival della Mente

Per la XVI edizione del Festival della Mente di **Sarzana**, da venerdì 30 agosto a domenica 1° settembre, nella cittadina ligure sono previsti 40 appuntamenti tra conferenze, workshop e spettacoli dedicati al Futuro. Tra gli ospiti, oltre a Francesca Rossi, ci saranno Massimo Recalcati, Telmo Pievani, Alessandro Barbero, Matteo Nucci, Paolo di Paolo e l'architetto e ingegnere Carlo Ratti (intervistato a pagina 74)

Ma come si evita questo genere di problemi?

«Evitando in tutti i modi correlazioni arbitrarie e discriminatorie nei dati e negli esempi con cui si allenano le intelligenze artificiali. Una strada è puntare – come fanno le grandi aziende che lavorano nel campo – sull'istruzione dei programmatori, renderli consapevoli dei pregiudizi che potrebbero instillare nel sistema. Inoltre è bene che nei team di sviluppatori ci siano persone con un background diverso, che possono così riconoscere un maggior numero di pregiudizi insiti nel sistema».

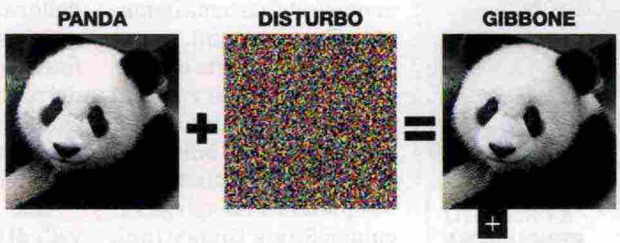
Ma le intelligenze artificiali autocoscienti saranno allineate ai valori etici dell'umanità?

«Il filosofo Nick Bostrom, nel suo saggio *Superintelligenza* (pubblicato in Italia da Bollati Boringhieri, *ndr*) immagina una I.A. superintelligente a cui viene chiesto di produrre il maggior numero di graffette possibile; e che perciò "impazzisce", accaparrandosi tutte le risorse del Pianeta per trasformarlo in una distesa di graffette. Sono scenari del tutto ipotetici e improbabili, ma è utile discuterne per capire come sviluppare un'intelligenza artificiale "eticamente allineata", che nell'eseguire i compiti che le diamo sappia rispettare i principi etici che le avremo comunicato. Se dirò alla mia auto *driverless* "portami a casa il più presto possibile", l'auto dovrà sapere che non può farlo ignorando i limiti di velocità o investendo i pedoni. Se stiamo attenti già oggi a costruire intelligenze artificiali allineate con i nostri valori, continueranno a esserlo anche quando saranno più capaci e autonome, ed eviteremo situazioni catastrofiche come quelle ipotizzate da Bostrom».

C'è chi dice che quando la tecnologia delle auto *driverless* sarà perfezionata, e il rischio di incidenti pressoché nullo, per salvare vite bisognerà proibire agli umani di guidare. Se un domani esistesse un "politico artificiale" geniale, eticamente impeccabile, dovremmo affidargli le redini del mondo?

«Fare politica è molto più complesso di guidare: riguarda le persone, le opinioni diverse, il confronto. Tutto ciò non può essere rimpiazzato dalle macchine. Però le intelligenze artificiali possono aiutare a decidere, anche in politica. Per esempio, l'Ibm ha realizzato un software, *Project debater*, che costruisce argomentazioni sensate a partire dai pro e dai contro di una certa soluzione. *Debater* si è già confrontato in pubblico con oratori umani e potrà essere uno strumento utile ai politici del futuro. Che non avranno (ancora) le orecchie a punta come Spock».

Giuliano Aluffi



Oggi l'affidabilità cognitiva delle macchine è relativa: per esempio, un disturbo che modifica in modo invisibile all'occhio umano la foto di un **panda** può farlo classificare come "gibbone" da un'intelligenza artificiale, che analizza i pixel ma non "vede" l'insieme