



SI STUDIA COME RENDERE PIÙ TRASPARENTE E ANCHE PIÙ ETICA L'INTELLIGENZA ARTIFICIALE

Missione nell'enigma dell'IA

L'Europa premia due ricercatori italiani

"Sveleremo le logiche della scatola nera"

SILVIA BANDELLONI

Trasparenza è la parola-chiave che unisce i progetti di ricerca nel settore dell'IA, l'Intelligenza Artificiale. Obiettivo: nuovi sistemi efficaci e sicuri, non più a rischio di violazione della privacy e nemmeno di sottostare ai pregiudizi dei programmatori umani.

Tra i protagonisti di questa rivoluzione ci sono i vincitori dei premi appena assegnati dall'Erc, lo European Research Council dell'Ue: si tratta di Fosca Giannotti, direttore di ricerca dell'Istituto di Scienza e Tecnologie dell'Informazione del Cnr di Pisa, e Giuseppe De Giacomo, professore al dipartimento di ingegneria informatica e automatica e gestionale dell'Università La Sapienza a Roma.

Si sente spesso parlare di «macchine pensanti», i cui passaggi logici - gli algoritmi - rimangono a volte sconosciuti anche agli esperti: sono le «Black boxes», le «scatole nere». Ora, dal momento che si vuole contestualizzare lo studio dei sistemi intelligenti, rendendoli funzionali alle esigenze della società, si punta

allo sviluppo di tecniche in grado di spiegare ogni passaggio del «machine learning» e delle metodologie di apprendimento, rendendone trasparenti i «ragionamenti». Il progetto «Explainable AI» di Giannotti e del suo gruppo mira proprio a una forma di Intelligenza Artificiale controllabile e che operi in modi «eticamente giusti».

«Le tecniche di «machine learning» ereditano difetti e pregiudizi che rimangono nascosti negli algoritmi, rischiando di suggerire scelte eticamente sbagliate. Se una macchina deve prendere decisioni, è importante garantire che operi usando il buon senso», spiega Giannotti. Ecco perché si devono creare approcci non solo sicuri ma comprensibili. «I sistemi intelligenti di oggi non sono in grado di motivare i loro suggerimenti e questo non aiuta l'operatore a ottimizzare il processo - prosegue Giannotti - . Pensiamo a un medico chiamato a decidere se eseguire un intervento, sulla base dei risultati che la macchina gli ha fornito. Analogamente, se un sistema mi impedisce di aprire un mutuo in banca, ho il diritto

di conoscere il motivo della mia scarsa affidabilità».

Il progetto è nato al Data Mining Laboratory (Kdd Lab) di Pisa, in cui Giannotti e il suo team hanno dato inizio al percorso che li ha portati all'Erc. «Lavoriamo per un avanzamento tecnologico che pongesse al centro l'essere umano (la cosiddetta «Human AI»), insieme con i suoi diritti e i suoi valori: una tecnologia che garantisca il rispetto della privacy e assicurasse l'uso di algoritmi trasparenti e non discriminatori». Come un esoscheletro consente a una persona paralizzata di camminare, analogamente si tenta di sviluppare una nuova forma di «Human AI» a livello cognitivo, attraverso tecniche di «data mining»: sono centrate sull'estrazione e sull'utilizzo



di informazioni a partire da grandi quantità di dati. E anche il progetto «WhiteMech» («White-box self-programming mechanisms») del gruppo di De Giacomo pone al centro un messaggio di trasparenza. «Il progetto riguarda lo studio di software dotati di IA che siano in grado di riprogrammarsi autonomamente per raggiungere un nuovo obiettivo o per reagire a cambiamenti nel contesto in cui operano - racconta De Giacomo -. Questi sistemi possono quindi sintetizzare «Piani B», non programmati in precedenza. È infatti opinione comune che non sia possibile pre-programmare la risposta a circostanze eccezionali, essendo troppo numerose e imprevedibili. Dare a una macchina la capacità di autoprogrammarsi richiede, però, di rendere trasparenti i suoi passaggi logici».

L'essere umano deve essere

in grado di interrogare il sistema sulle azioni che questo ha compiuto, sul come e sul perché abbia fatto una scelta invece di un'altra e che cosa avrebbe fatto in una determinata situazione. «Un sistema di questo tipo, che per esempio gestisce una fabbrica, è in grado di sostituire in tempo reale un robot difettoso con un altro che svolgeva operazioni diverse. Il sistema è infatti in grado di adattare la macchina sostitutiva alla funzione richiesta. E, benché il «Piano B» sia probabilmente meno efficace di quello originale, eviterà di bloccare la fabbrica», aggiunge De Giacomo.

Alla base di questa ricerca ci sono una serie di discipline, come la rappresentazione della conoscenza e la pianificazione automatica, e altrettanto coinvolte sono quelle dei metodi formali (come la verifica e la sintesi automatica di

nuovi programmi) e delle basi di dati (come lo studio di processi «data-aware»).

Il finanziamento dell'Erc - 2,5 milioni di euro per progetto - consentirà ai vincitori di sviluppare i loro progetti in cinque anni. Investiranno non solo in strumentazioni, ma in ricercatori multidisciplinari, «perché serve la conoscenza dell'informatica, della matematica e della fisica e la capacità di spiegare, che è materia degli psicologi», commenta Giannotti. «Vogliamo dare ai ricercatori un luogo di riferimento, una casa», conclude De Giacomo. Una «casa» che riguardi la società nel suo complesso: non a caso l'Europa punta alla «Human Centered AI»: l'Intelligenza Artificiale dovrà garantire servizi sempre più affidabili in una società complessa e sofisticata. —

BY NC ND ALIQUINI DIRITTI RISERVATI



FOSCA GIANNOTTI
È DIRETTORE DI RICERCA
DELL'ISTITUTO DI SCIENZA
E TECNOLOGIE DELL'INFORMAZIONE
DEL CNR DI PISA



GIUSEPPE DE GIACOMO
È PROFESSORE AL DIPARTIMENTO
DI INGEGNERIA INFORMATICA
E AUTOMATICA DELL'UNIVERSITÀ
LA SAPIENZA DI ROMA

