L’INTELLIGENZA ARTIFICIALE può essere spiegabile?

Data 16 gennaio 2022
Categoria Medicinadigitale

In un precedente numero di pillole abbiamo descritto l’”opacità” della intelligenza artificiale (IA), il cosiddetto modello black box [b]pillole.org/public/aspnuke/news.asp?id=7761[/b] La necessità di maggiore chiarezza e trasparenza è stata colta da varie istituzioni. La Commissione Europea ha prodotto un libro bianco volto alla creazione di un quadro normativo per un ecosistema digitale di fiducia in una IA affidabile, tra i cui requisiti etici fondamentali individuati vi è la trasparenza e la spiegabilità . Si tratta In pratica di realizzare modelli che consentano agli umani di capire, di fidarsi e quindi di governare effettivamente la generazione emergente di macchine dotate di IA, mantenendo un alto livello di performance.
Sta per questo emergendo una nuova disciplina, la eXplainable AI (XAI), intelligenza artificiale spiegabile, definibile sinteticamente come un insieme di strumenti e di tecniche utilizzate per rendere sempre più trasparente e facile da capire il funzionamento dei sistemi di IA, la loro “logica” interna.
Gli studi riguardanti la XAI hanno sviluppato diverse dimensioni concettuali, con punti di osservazione ed obiettivi differenti, anche se connaturati al concetto generale di spiegabilità: interpretabilità, affidabilità, accuratezza, causalità, trasferibilità, informatività, fiducia, equità, accessibilità, interattività, privacy. A tale complessa concettualizzazione corrisponde una ancora più complessa realizzazione pratica, attuabile. I metodi utilizzati per svelare il processo di funzionamento della IA black box, dall’inserimento dei dati all’esito finale, variano infatti in funzione di molteplici fattori, quali la tipologia di input, la scelta degli algoritmi, il livello di spiegazione richiesto (parziale o totale). Secondo alcuni esperti, i modelli esplicabili dovrebbero essere evitati, mentre in alcuni ambiti, per esempio nella giustizia, nell’assistenza sanitaria e nella computer vision, potrebbero sostituire quelli a scatola nera.
La spiegabilità deve essere inoltre modulabile in funzione del target, dell’utilizzatore, che può variare dall’esperto data scientist al comune cittadino. Ognuno di questi soggetti è infatti dotato di competenze ed interessi molto differenti, in funzione del livello di interazione nei confronti della tecnologia e degli obiettivi del suo utilizzo. Una modalità efficace per “aprire” le scatole nere è infine quella di lavorare in modo interdisciplinare, combinando le competenze di ambiti diversi, dalle tecnologie informatiche e ingegneristiche alle cosiddette “humanities” antropologiche, sociali, psicologiche.

Riflessioniconclusive

Stiamo passando da un tempo in cui l’intelligenza umana codificava gli algoritmi ed era responsabile della validità e affidabilità dei risultati ad un periodo storico in cui l’IA impara autonomamente e fornisce risposte con modalità non comprensibili nemmeno per i progettisti. E’ per questo fondamentale che gli algoritmi “black box” diventino accessibili, per valutare la loro efficacia e sicurezza ed allinearli al sistema valoriale umano, oltre che per preservare l’autonomia e la consapevolezza nelle decisioni in ambiti critici come la salute, la giustizia e la sicurezza.
Possiamo concludere che la famosa frase di Andy Rubin, cofondatore di Android, “l’intelligenza artificiale è come il cervello: non si può tagliare la testa e vedere come funziona” è in realtà non completamente vera: il cervello si potrà “vedere”e per giunta con modalità meno cruente, anche se non meno inquietanti, come descritto in un precedente articolo . Allo stesso modo anche l’IA, per poter essere considerata affidabile e responsabile, dovrà svelare i suoi segreti, non solo in nome della trasparenza e dell’etica, ma anche del progresso e della ricerca scientifica.
L’alternativa è l’affermarsi di una “società black box” , governata da misteriosi algoritmi protetti dal segreto industriale, in grado di rendere invisibili, e quindi impossibili da limitare, gli errori e le frequenti discriminazioni, intenzionali o involontarie.

Giampaolo Collecchia e Riccardo De Gobbi

Bibliografia

1) https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_it.pdf
2)Carobene A. Scelte profonde d’intelligenza artificiale. Il Sole 24 Ore 2020, 9 febbraio, pag. 12
3) Deluzarche C, Deep learning, le grand trou noir de l’intelligence artificielle, Maddyness, 2017 https://www.maddyness.com/2019/08/20/ia-deep-learning-trou-noir-intelligence-artificielle/
4) Collecchia G. Neurotecnologie e neurodiritti digitali: la privacy mentale. Recenti Prog Med 2021; 112: 1-4
5) Frank Pasquale. 2015. The Black Box Society: The Secret Algorithms that Control Money and Information. Harvard UniversityPress.

Perapprofondire:
Collecchia G. De Gobbi R.: Intelligenza Artificiale e Medicina Digitale Il Pensiero Scientifico Ed. Roma 2020
pensiero.it/catalogo/libri/pubblico/intelligenza-artificiale-e-medicina-digitale