

Introduzione

Mentre scrivevo questo libro ho avuto l'idea per la trama di un racconto di fantascienza che probabilmente non scriverò mai. Una donna, Ellen, viene svegliata una mattina dallo squillo del telefono. A chiamare è il marito, fuori città da qualche giorno per lavoro. Durante la telefonata, Ellen percepisce qualcosa di strano nella sua voce, come fosse leggermente fuori tono. Nel futuro prossimo in cui il racconto si svolge, le tecnologie di intelligenza artificiale (IA) sono così avanzate che è possibile programmare un assistente vocale per fare telefonate al proprio posto, simulando la voce in maniera così accurata da ingannare persino i conoscenti più stretti. Come molti altri coniugi, Ellen e il marito hanno deciso di non usare mai questo strumento per comunicare tra di loro. Eppure nella sua voce quella mattina c'è qualcosa che a Ellen suona alieno. Più tardi, Ellen scopre che suo marito è morto in un incidente la notte prima, poche ore prima della chiamata. A parlare non poteva essere che l'assistente vocale. Frastornata dalla perdita, ascolta la registrazione della chiamata più e più volte, fino a trovare un indizio che ha a che fare con le circostanze della morte. Il racconto di fantascienza che non ho scritto, infatti, è anche un giallo. Per scoprire la verità sulla morte del marito, Ellen dovrà interpretare il contenuto della conversazione. Dovrà capire quali parole fossero indirizzate a lei da suo marito, quali dalla macchina, e quali da una combinazione di questi due.

Questo libro non è un racconto di fantascienza, ma, come la fantascienza, è un tentativo di catturare il senso di tecnologie le cui implicazioni e i cui significati abbiamo appena cominciato a capire. Lo strumento che metto a disposizione non è tanto il futuro quanto il passato: nella traiettoria storica dell'IA – una traiettoria sorprendentemente lunga e complessa per una tec-

nologia che viene spesso presentata come una novità assoluta – ho trovato la bussola per orientare queste esplorazioni. Il punto di vista per raccontare questa storia muove da una prospettiva se non inedita, quantomeno anomala nel dibattito pubblico sull'IA: piuttosto che chiedersi se le macchine sono intelligenti, si tratterà di chiedersi in che misura esse *ci appaiono* intelligenti. Ciò che mi preme mostrare è che lo sviluppo dell'IA fino a oggi non è andato tanto nella direzione dell'emulazione o del superamento dell'intelligenza umana, quanto verso lo sviluppo di sistemi in grado di convincere noi umani che le macchine sono intelligenti. L'IA, di conseguenza, è intelligente nella misura in cui la consideriamo tale. E il sogno di creare delle macchine intelligenti ha valore nella misura in cui seguitiamo a crederci.

Nella sfera pubblica, nei mezzi di informazione e anche nella fiction, l'IA è spesso discussa come qualcosa di straordinario, che invita a porre domande esistenziali e persino metafisiche, mettendo in crisi il significato stesso di umanità e la distinzione tra umano e artificiale¹. L'idea che queste tecnologie possano sostituirsi e persino opporsi a noi non è solo materiale di innumerevoli film di fantascienza: si è imposta come un vero e proprio mito ricorrente della nostra epoca². Ma la paura o forse il desiderio di creare delle macchine a nostra immagine e somiglianza si scontra con la realtà delle tecnologie intorno a noi. Le intelligenze artificiali che usiamo e incontriamo nei contesti più disparati sono applicazioni specifiche, capaci di svolgere molto bene singole funzioni come il riconoscimento delle immagini o la sintesi vocale, ma non di combinarle tra di loro in modo da sviluppare una coscienza autonoma dei propri mezzi e della realtà esterna, e men che meno di provare sentimenti o empatia. Sebbene alcuni considerino simili scenari plausibili, non vi

¹ Hamid R. Ekbia, *Artificial Dreams: The Quest for Non-biological Intelligence*, Cambridge University Press, Cambridge 2008; Andrea L. Guzman, *Beyond Extraordinary: Theorizing Artificial Intelligence and the Self in Daily Life*, in Zizi Papacharissi (a cura di), *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*, Routledge, New York 2019, pp. 83-96; Ed Finn, *Che cosa vogliono gli algoritmi? L'immaginazione nell'era del computer*, trad. di Daniele A. Gewurz, Einaudi, Torino 2018 (ed. or., *What Algorithms Want: Imagination in the Age of Computing*, MIT Press, Cambridge [Mass.] 2017).

² Peppino Ortoleva, *Miti a bassa intensità*, Einaudi, Torino 2019; Stefano Bory e Paolo Bory, *I nuovi immaginari dell'intelligenza artificiale*, in «Im@go: A Journal of the Social Imaginary», 4 (2015), n. 6, pp. 66-85.

sono prove scientifiche che dimostrino, al di fuori di ogni dubbio, che replicare la vita o l'intelligenza umana sia teoricamente possibile³. Per quanto l'idea di una macchina cosciente o di un vero e proprio cervello elettronico possa risultare attraente, ciò che ci riservano il presente e (almeno) il futuro prossimo dell'IA non è la creazione di esseri simili a noi, ma lo sviluppo di tecnologie in grado di essere percepite come tali. Per questa ragione questo libro vuole portare l'attenzione sul tema dell'inganno, visto come un elemento nevralgico delle relazioni tra umani e macchine che andiamo costruendo.

Una questione di intelligenza.

Uno dei problemi principali, in questo senso, è la definizione data alla parola «intelligenza». Si tratta di una nozione complessa, che può essere riferita ad abilità e azioni molto differenti: dalla capacità di giocare a scacchi all'abilità di muoversi in un ambiente complesso, dal calcolo matematico alle competenze emotive e sociali⁴. Proprio questa ampiezza e indecisione di significato è stata fonte di innumerevoli incomprensioni fin dalle origini dell'IA, spesso causate dall'uso di termini che richiamano aspetti dall'esperienza umana – pensare, ragionare, imparare, memorizzare – per descrivere il funzionamento delle macchine⁵. Il matematico Alan Turing offrì una via d'uscita semplice e al tempo stesso acuta da questo problema in un testo del 1950 che pose le fondamenta dell'IA e in cui propose il famoso test di Turing. Cominciò il proprio saggio chiedendosi se le macchine possono pensare, ma solo per liquidare questa domanda come inutile e persino dannosa: sarebbe infatti impos-

³ Margaret Boden, *AI: Its Nature and Future*, Oxford University Press, Oxford 2016; Hector J. Levesque, *Common Sense, the Turing Test, and the Quest for Real AI: Reflections on Natural and Artificial Intelligence*, MIT Press, Cambridge (Mass.) 2017.

⁴ N. Katherine Hayes, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, University of Chicago Press, Chicago 1999.

⁵ Alcune delle critiche più significative volte ai paradigmi di IA sono state mosse proprio sulla base della definizione data al termine «intelligenza»; si veda, per esempio, Hubert L. Dreyfus, *Che cosa non possono fare i computer. I limiti dell'intelligenza artificiale*, trad. di Giuditta Alessandrini, Armando, Roma 1988 (ed. or., *What Computers Can't Do: A Critique of Artificial Reason*, Harper and Row, New York 1972).

sibile trovare un accordo su cosa significhi «pensare». L'unica soluzione è rovesciare la questione, ammettendo che l'intelligenza possa essere definita solo in termini relativi: siamo noi umani che attribuiamo intelligenza a qualcosa, che decidiamo se una macchina «pensa» o meno⁶. È curioso che il test di Turing venga spesso presentato come un modo di misurare l'intelligenza delle macchine, nonostante Turing stesso si fosse rifiutato di definire l'intelligenza in termini assoluti, mettendo da parte la domanda se le macchine possono pensare per accontentarsi di chiedere in che misura esse *appaiano* intelligenti⁷.

L'intuizione dello scienziato britannico non è solo un'utile scappatoia da un problema apparentemente irrisolvibile: è un'intuizione preziosa che permette di cogliere alcune delle implicazioni più significative delle tecnologie che vanno sotto il nome di IA. Lo scenario in cui l'IA si evolve nella direzione di una sempre più convincente sembianza di intelligenza, infatti, è il più probabile e al tempo stesso il più denso di conseguenze. Se a prima vista la verosimiglianza può sembrare un risultato modesto rispetto alla creazione di veri e propri esseri coscienti, lo sviluppo di mezzi tecnologici capaci di sembrare intelligenti rappresenta in realtà un fenomeno epocale. Come vedremo, in gioco vi è la capacità di distinguere diversi tipi di messaggi e di difenderci da tentativi di manipolazione che ci invitano a proiettare intenzioni e sentimenti su sistemi che ne sono privi.

In un campo in continua evoluzione come l'IA, la capacità di interrogare, prevedere e disinnescare gli effetti potenziali di queste nuove tecnologie sarà sempre più importante negli anni a venire. Questi effetti, d'altronde, sono già visibili oggi in una quantità di tecnologie di uso comune. Robot domestici come Jibo o come Astro, l'aiutante robotico messo in commercio di recente da Amazon con un design che ricorda il robot Wall-E dell'omonimo film Disney, app come Replika che permettono di intrattenere una conversazione sul proprio cellulare con un avatar artificiale, e persino i più comuni assistenti vocali come Alexa,

⁶ Alan Turing, *Macchine calcolatrici e intelligenza*, in Id., *Intelligenza meccanica*, a cura di Gabriele Lolli, Bollati Boringhieri, Torino 1994, pp. 121-57 (ed. or., *Computing Machinery and Intelligence*, in «Mind», 59 [1950], n. 236, pp. 433-60).

⁷ Si veda per esempio Kevin Warwick e Huma Shah, *Turing's Imitation Game*, Cambridge University Press, Cambridge 2016.

Siri e Assistente Google sono sviluppati in modo da attivare comportamenti di tipo sociale e suscitare negli utenti emozioni quali l'empatia. Queste tecnologie, così come molte altre che vanno sotto l'etichetta di IA, incorporano in maniera programmatica forme di inganno. Il loro design, la voce umanizzata, persino l'uso di specifiche espressioni linguistiche e dell'ironia sono concepiti per produrre determinati effetti nell'interazione con gli utenti.

Se per esempio a una domanda posta a Siri o Alexa ci sentiamo rispondere con una battuta incisiva, tenderemo a sopravvalutare la complessità di questi sistemi. In realtà è probabile che la risposta derivi non da sofisticate tecnologie informatiche quanto da una delle forme espressive più antiche: la drammaturgia. Aziende come Apple o Amazon si sono infatti dotate di team di creativi incaricati di ideare risposte appropriate e sagaci ai quesiti più frequenti, come copioni di una commedia che gli assistenti vocali si limitano a recitare⁸. Nelle nostre esperienze quotidiane con queste tecnologie abbiamo pochi mezzi per cogliere le dinamiche profonde di quelli che ci appaiono comportamenti intelligenti. Esiste dunque una discrepanza ineludibile tra il modo in cui percepiamo le macchine e il loro funzionamento. È in questo crinale che si inserisce l'elemento di inganno, inteso come uno dei meccanismi fondamentali attraverso cui le macchine si interfacciano con gli esseri umani.