

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Macchine ingannevoli: Introduzione all'edizione italiana

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1975450> since 2024-05-07T20:32:21Z

Publisher:

Einaudi

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

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 (AI, dall'inglese *Artificial Intelligence*) 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'AI – una traiettoria sorprendentemente lunga e complessa per una tecnologia 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'AI: 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'AI fino ad 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'AI,

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'AI 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 artificiale.¹ 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 epoca.² 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 sono prove scientifiche che dimostrino, al di fuori di ogni dubbio, che replicare la vita o l'intelligenza umana sia teoricamente possibile.³ 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'AI 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

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 *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*, a cura di Zizi Papacharissi, Routledge, New York, 2018, pp. 83–96; Ed Finn, *Che cosa vogliono gli algoritmi? L'immaginazione nell'era dei computer*, trad. A. Gewurz Daniele, Einaudi, Torino, 2018.

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*, 4.6, 2016, pp. 66–85.

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.

matematico alle competenze emotive e sociali.⁴ Proprio questa ampiezza e indecisione di significato è stata fonte di innumerevoli incomprensioni fin dalle origini dell'AI, spesso causate dall'uso di termini che richiamano aspetti dall'esperienza umana – pensare, ragionare, imparare, memorizzare – per descrivere il funzionamento delle macchine.⁵ 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'AI e in cui propose il famoso test di Turing. Turing cominciò il saggio chiedendosi se le macchine possono pensare, ma solo per liquidare questa domanda come inutile e persino dannosa: sarebbe infatti impossibile 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 meno.⁶ 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* intelligenti.⁷

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 AI. Lo scenario in cui l'AI 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'AI, 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

⁴ N. Katherine Hayles, *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 AI sono state mosse proprio sulla base della definizione data al termine intelligenza; si veda, ad esempio, Hubert L. Dreyfus, *What Computers Can't Do: A Critique of Artificial Reason*, Harper & Row, New York, 1972.

⁶ Alan Turing, *Computing Machinery and Intelligence*, in *Mind*, 59.236, 1950, pp. 433–60.

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

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 di un film della 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, Siri e Google Assistant 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 AI, 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 ad esempio a una domanda posta a Siri o Alexa ci sentiamo rispondere con una battuta incisiva, tendiamo 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 recitare.⁸ 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.

L'inganno “banale” delle macchine

In un saggio sui rapporti tra arte e illusione nella storia dell'arte, il grande storico dell'arte Ernst Gombrich sostenne che le arti figurative si sviluppano all'interno degli spazi lasciati aperti dai limiti della tradizione e da quelli della percezione.⁹ Gli artisti hanno da sempre utilizzato la possibilità di ingannare l'osservatore come un mezzo estetico, mettendo a frutto la propria conoscenza non solo delle convenzioni pittoriche ma anche dei meccanismi e dei limiti della percezione umana. Chi biasimerebbe, del resto, un pittore per avere “ingannato” il nostro

⁸ Liam Young, *'I'm a Cloud of Infinitesimal Data Computation' - When Machines Talk Back: An Interview with Deborah Harrison, One of the Personality Designers of Microsoft's Cortana AI*, In *Architectural Design* 89.1, 2019, pp. 112–17. <https://doi.org/10.1002/ad.2398>.

⁹ Ernst Hans Gombrich, *Arte e illusione. Studio sulla psicologia della rappresentazione pittorica*, trad. R. Federici, Phaidon, Londra, 2008.

occhio giocando con effetti di prospettiva o di profondità che rendono un quadro più credibile e realistico agli occhi di chi lo osserva?

Mentre questi mezzi appaiono totalmente accettabili quando a usarli è un artista, l'idea che lo sviluppatore di un programma o un'interfaccia digitale metta a frutto le proprie conoscenze su come ingannare gli utenti potrebbe suscitare sospetto e apprensione. Anche a causa della connotazione negativa del termine "inganno", le comunità di informatici e ricercatori di AI si sono dimostrate perlopiù reticenti a descrivere il proprio lavoro in questi termini, se non per indicare risultati indesiderati.¹⁰ Eppure l'inganno rappresenta un elemento non incidentale o secondario, ma costitutivo di queste tecnologie. È centrale nel loro funzionamento, così come i circuiti, il software, e i dati che fanno girare l'AI.

Nel suo significato più comune, il concetto di inganno ha a che fare con l'uso di segni o rappresentazioni per suggerire un'impressione erronea o fuorviante.¹¹ Le più recenti ricerche in campi quali la psicologia sociale, la filosofia e la sociologia sottolineano che esso non si limita a situazioni eccezionali, come una truffa o un trucco di magia, ma ricopre un ruolo sostanziale nell'esperienza quotidiana di ogni persona.¹² Secondo il filosofo Mark Wrathall, la possibilità di cadere nell'inganno è anzi uno dei meccanismi principali attraverso i quali percepiamo le cose e il mondo. Immaginiamo, ad esempio, di scorgere durante una passeggiata in un bosco un cervo ai lati del nostro campo visivo. Girandoci, ci accorgiamo che si tratta invece di un ramo dalla forma strana. Ci siamo ingannati; ma i meccanismi che ci hanno fatto cadere in errore sono stati plasmati in milioni di anni di evoluzione per rispondere all'esigenza di metterci in allarme su potenziali pericoli nel nostro campo di visione. Ha quindi poco senso

¹⁰ Ci sono, d'altronde, alcune isolate ma significative eccezioni. Adar et al., per esempio, distinguono tra inganno malevolo o benevolo, descrivendo quest'ultimo come "un inganno con l'obiettivo di portare un beneficio sia agli utenti che agli sviluppatori", e sottolineando come questa forma benevola di inganno sia onnipresente nei sistemi informatici, nonostante venga raramente descritta in questi termini. Eytan Adar, Desney S. Tan e Jaime Teevan, *Benevolent Deception in Human Computer Interaction*, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2013, pp. 1863–72. Si veda anche Tathagata Chakraborti e Subbarao Kambhampati, *Algorithms for the Greater Good! On Mental Modeling and Acceptable Symbiosis in Human-AI Collaboration*, *ArXiv:1801.09854*, 2018; Castelfranchi, Cristiano, and Yao-Hua Tan, *Trust and Deception in Virtual Societies*, Springer, Chum, 2001; Ștefan, Sarkadi, Alison R. Panisson, Rafael H. Bordini, Peter McBurney, Simon Parsons e Martin Chapman, *Modelling deception using theory of mind in multi-agent systems*, *AI Communications* 32.4, 2019, pp. 287-302.

¹¹ John Danaher, *Robot Betrayal: A Guide to the Ethics of Robotic Deception*, *Ethics and Information Technology* 22.2, 2020, pp. 117–28.

¹² Bella M. DePaulo, Susan E. Kirkendol, Deborah A. Kashy, Melissa M. Wyer e Jennifer A. Epstein, *Lying in Everyday Life*, in *Journal of Personality and Social Psychology*, 70.5, 1996, 979–95; Radu Umbres, *Deception as Exploitative Social Agency*, in *Distributed Agency*, a cura di N.J. Enfield and Paul Kockelman, Oxford University Press, Oxford, 2016, pp. 243–51; Clancy W. Martin, *The Philosophy of Deception*, Oxford University Press, Oxford, 2009; Ronja Rutschmann e Alex Wiegmann, *No Need for an Intention to Deceive? Challenging the Traditional Definition of Lying*, in *Philosophical Psychology*, 30.4, 2017, pp. 438–57.

considerare una nostra percezione come giusta o sbagliata, dal momento che l'errore è una componente costitutiva del modo in cui navighiamo la realtà esterna.¹³

Già a partire dalla fine del diciannovesimo, l'istituzionalizzazione della psicologia fu accompagnata dalla realizzazione che inganno e illusione rappresentano aspetti integrali e fisiologici dell'esperienza umana, e che le dinamiche che portano a errori percettivi sono fondamentali per capire come percepiamo il mondo.¹⁴ L'accumulo di sapere su come le persone si ingannano fu fondamentale per la creazione di tecnologie e pratiche medialità che approfittano delle caratteristiche e dei limiti della vista, dell'udito, e persino del tatto.¹⁵ Lo sviluppo dell'AI si inserisce pienamente all'interno di questa tradizione. Come dimostrerò nei prossimi capitoli, scienziati e sviluppatori accumularono fin dai primi decenni della storia di questa disciplina forme di sapere sul modo in cui gli utenti umani reagiscono quando osservano o interagiscono con macchine che sembrano intelligenti; questo sapere fu incorporato nelle tecnologie dell'AI e della robotica, divenendo parte integrante di ciò che chiamiamo "intelligenza artificiale".

Qualcuno potrebbe obiettare che la situazione in cui una macchina cerca di ingannarci, di passare per umana, è una situazione limite, che può capitare in certi casi - per esempio in un tentativo di truffa via internet, o nel caso di bot su Twitter o Facebook che imitano gli account di utenti umani -, ma che ha poco a che vedere con la quotidianità delle nostre interazioni con IA. Chi usa assistenti vocali come Alexa and Siri, ad esempio, sa bene che questi sono macchine e non persone con cui sviluppare un rapporto affettivo reale. Questo non significa, però, che forme di inganno più sottili non siano presenti nelle interazioni più comuni e quotidiane con queste tecnologie. La scelta di assegnare agli assistenti delle voci umanizzate, così come quella di dare loro precise connotazioni di genere e persino di provenienza sociale o regionale attraverso gli accenti, non sono mai casuali: derivano da precise strategie di aziende come Amazon ed Apple, volte a ottenere determinati effetti - per esempio, a rendere la voce di Siri e Alexa più rassicurante e familiare per gli utenti. Per descrivere queste forme apparentemente inoffensive di inganno utilizzerò il termine di "inganno banale". L'inganno di

¹³ Mark A. Wrathall, *Heidegger and Unconcealment: Truth, Language, and History*. Cambridge University Press, Cambridge, 2010, p. 60; per una prospettiva simile, dal punto di vista di uno psicologo cognitivo, si veda Donald Hoffman, *L'illusione della realtà. Come l'evoluzione ci inganna sul mondo che vediamo*, Trad. Francesca Pè, Bollati Boringhieri, Milano, 2020.

¹⁴ Michael Pettit, *The Science of Deception: Psychology and Commerce in America*, University of Chicago Press, Chicago, 2013; James H. Korn, *Illusions of Reality: A History of Deception in Social Psychology*, State University of New York Press, Albany, 1997.

¹⁵ David Parisi, *Archaeologies of Touch: Interfacing with Haptics from Electricity to Computing*, University of Minnesota Press, Minneapolis, 2018; Melissa M. Littlefield, *The Lying Brain: Lie Detection in Science and Science Fiction*, University of Michigan Press, Ann Arbor, 2011; Jonathan Sterne, *MP3: The Meaning of a Format*, Duke University Press, Durham, 2012.

Alexa è banale innanzitutto perché non viene percepito come tale, cosa che ha degli ovvi vantaggi dal punto di vista commerciale, perché dà all'utente l'illusione di mantenere il controllo dell'esperienza. Ed è banale perché ha a che fare con delle situazioni che sono immerse nel nostro vivere quotidiano, e non facilmente distinguibili dalla nostra esperienza per così dire "normale". Questo carattere ordinario dell'inganno lo rende spesso impercettibile ma anche pregno di conseguenze, permettendo a queste tecnologie di inserirsi negli strati più profondi delle nostre abitudini quotidiane.¹⁶

Riconoscere la presenza strutturale di forme di inganno nelle interazioni con l'AI non equivale, dunque, a proporre una teoria complottista o una visione apocalittica, per la quale saremmo tutti vittima di una grande manipolazione orchestrata dalle aziende che producono queste tecnologie. Al contrario, è importante enfatizzare come questi meccanismi possano rivelarsi utili e produttivi, permettendoci ad esempio di inserire più naturalmente gli assistenti vocali nel nostro ambiente domestico e quotidiano, o aiutandoci a superare sofferenze individuali e sociali, come nel caso di strumenti sviluppati come forme di auto-aiuto contro il disagio psicologico. Perché questo accada, d'altronde, abbiamo bisogno di appropriati mezzi etici e tecnici per sviluppare queste applicazioni nel contesto di un'AI che ponga l'utente al suo centro.¹⁷ La posta in palio è la scelta tra un futuro come quello della storia di Ellen, in cui abbiamo perso la facoltà di distinguere tra gli esseri umani e la loro simulazione, e un futuro in cui il controllo e la possibilità di un uso consapevole di queste tecnologie rimangono assicurati. Solo lavorando in questa direzione, e incoraggiando a interrogare i rischi e le problematiche dell'inganno banale, potremo sviluppare opportuni antidoti al potere che la tecnologia può esercitare su utenti e consumatori.¹⁸

AI, media e comunicazione

Esaminando la storia delle tecnologie di AI e del loro rapporto con la comunicazione dall'emergere del computer digitale fino ai giorni nostri, muovo dalla persuasione che una

¹⁶ L'aggettivo "banale" è stato utilizzato in altri contesti per sottolineare le implicazioni di aspetti della vita sociale che spesso vengono liquidate come irrilevanti, perché ordinarie e quotidiane, ma sono in realtà estremamente importanti per capire la realtà in cui viviamo. Si veda in particolare l'illuminante Michael Billig, *Nazionalismo banale*, trad. Fabio De Leonardis, Rubbettino, 2018; ma anche Stig Hjarvard, *The Mediatization of Religion: Theorising Religion, Media and Social Change*, in *Culture and Religion*, 12.2, 2011, pp. 119–35.

¹⁷ Ben Shneiderman, *Human-Centered AI*, Oxford University Press, New York, 2022.

¹⁸ Si pensi, ad esempio, a come i meccanismi di empatia generati da sistemi di AI che comunicano in linguaggio naturale e con voce umanizzata possano essere attivati per influenzare gli orientamenti elettorali di chi li usa o per condurre campagne sottili ma aggressive di marketing diretto. Si veda, a riguardo, Judith Donath, *The Robot Dog Fetches for Whom?*, in *A Networked Self and Human Augmentics, Artificial Intelligence, Sentience*, a cura di Zizi Papacharissi, Routledge, London, 2018, pp. 10–24.

prospettiva storica su media e tecnologia possa aiutare a comprendere alcune delle trasformazioni sociali, culturali e politiche in corso. Le tecnologie che chiamiamo “nuovi media” hanno una lunga storia, il cui studio non solo rappresenta uno dei mezzi a disposizione per capire il presente, ma anche uno dei migliori strumenti per anticipare il futuro.¹⁹ In aree in rapida trasformazione come l’AI e più in generale il digitale, è estremamente difficile prevedere evoluzioni di breve o medio periodo, per non parlare di cambiamenti di più ampio raggio.²⁰ Prendere in considerazione le traiettorie storiche che hanno caratterizzato questi campi per diversi decenni aiuta a identificare *trend* e direzioni di cambiamento che ne hanno condizionato lo sviluppo in passato e continueranno a influenzarne il futuro. La storia, in questo senso, è una scienza del futuro: ci aiuta non solo a illuminare le dinamiche che ci hanno portato nella situazione attuale, ma anche a porre domande e problemi con i quali definire le sfide tecniche e sociali del domani.²¹

L’AI è un campo di ricerca estremamente interdisciplinare, caratterizzato da una serie di approcci, teorie e metodi diversi tra loro. Alcune applicazioni basate sull’AI, come gli algoritmi che regolano il nostro accesso alla rete internet, consigliandoci il prossimo film da vedere o la pagina web da visitare, sono una presenza costante nella vita di grandi masse di persone; con altre, come i robot e i meccanismi di automazione a uso industriale, si entra in contatto più raramente.²² Questo libro si concentra in particolare sull’AI comunicativa, ovvero quelle tecnologie di AI in grado di entrare in comunicazione con utenti umani.²³ Queste includono sistemi che integrano l’uso della conversazione e del linguaggio, come i chatbot, i bot su Twitter e altri social media, gli assistenti vocali e i programmi per la composizione testuale che vanno sotto l’etichetta di Natural Language Processing o NLP. La robotica fa uso

¹⁹ Lisa Gitelman, *Always Already New: Media, History and the Data of Culture*, MIT Press, Cambridge (Mass.), 2006;

Jussi Parikka, *Archeologia dei media. Nuove prospettive per la storia e la teoria della comunicazione*; Gabriele Balbi e Paolo Magaudda, *Storia dei media digitali. Rivoluzioni e continuità*, Laterza, Bari, 2014.

²⁰ Ecco perché le previsioni sul futuro della tecnologia si dimostrano molto spesso errate. Si veda a riguardo Ithiel De Sola Pool et al., ‘Foresight and Hindsight: The Case of the Telephone’, in *Social Impact of the Telephone*, a cura di Ithiel De Sola Pool, MIT Press, Cambridge (Mass.), 1977, pp. 127–157; Simone Natale, *Introduction: New Media and the Imagination of the Future*, in *Wi: Journal of Mobile Media*, 8.2, 2014, pp. 1–8; Christoph Ernst e Jens Schröter, *Media Futures: Theory and Aesthetics*, London, Palgrave MacMillan, 2021.

²¹ David W. Park, Nick Jankowski e Steve Jones (a cura di), *The Long History of New Media: Technology, Historiography, and Contextualizing Newness. Digital Formations*, Peter Lang, New York, 2011; Gabriele Balbi e Paolo Magaudda, *Storia dei media digitali* cit.

²² Andrea L. Guzman, *The Messages of Mute Machines: Human-Machine Communication with Industrial Technologies*, *Communication+ 1*, 5.1, 2016, pp. 1–30.

²³ Sullo stretto rapporto tra AI e comunicazione, si veda tra altri Andrea L. Guzman e Seth C. Lewis, *Artificial Intelligence and Communication: A Human–Machine Communication Research Agenda*, in *New Media & Society*, 22.1, 2019, pp. 70–86; David J. Gunkel, *An Introduction to Communication and Artificial Intelligence*, Polity, New York, 2020.

di alcune delle stesse tecnologie e dinamiche sviluppate dall'AI comunicativa – molti robot sono programmati per comunicare con utenti umani – ma rimarrà, al di là di riferimenti e collegamenti occasionali, sostanzialmente a margine di questo libro. Come risulterà evidente più avanti, infatti, è proprio a partire da sistemi di AI privi di una corporalità ma in grado di sviluppare processi comunicativi che le potenzialità delle forme di inganno in esame appaiono più evidenti.²⁴

Quando l'AI è utilizzata per comunicare con utenti umani, essa diviene automaticamente un medium o mezzo di comunicazione – per quanto di tipo nuovo, dal momento che i media coincidono tradizionalmente con il canale attraverso cui avviene la comunicazione, mentre l'AI comunicativa può fare allo stesso tempo da canale e da produttore di messaggi comunicativi. Queste tecnologie, di conseguenza, ereditano molti dei meccanismi e delle dinamiche che hanno caratterizzato la comunicazione mediata a partire dall'emergere dei media elettrici nel diciannovesimo secolo. Ne consegue che, per comprendere tecnologie di recente innovazione come gli assistenti vocali o i *bot*, sia necessario contestualizzarle in un orizzonte storico più ampio. I meccanismi di inganno azionati dall'AI si trovano, infatti, in un rapporto di continuità con quelli delle tecnologie medialità precedenti l'invenzione del computer. Lo sviluppo dei media moderni ha da sempre avuto a che fare solo con l'invenzione e il perfezionamento di mezzi tecnici come la meccanica e l'elettronica, ma anche con l'accumulo di forme di sapere sempre più accurate e complesse sui meccanismi percettivi e della psicologia umana. I media moderni sono stati creati, per così dire, a immagine e somiglianza dei loro utilizzatori umani: l'invenzione del cinema fu resa possibile dalla comprensione dei meccanismi che ci fanno percepire la successione rapida di una serie di immagini come un'immagine in movimento;²⁵ lo sviluppo dei mezzi di riproduzione sonora, dal fonografo all'iPod, e di tecnologie di telecomunicazione come il telefono mise a frutto le scoperte della fisiologia dell'udito e della voce umana, che permisero di creare registrazioni e comunicazioni più efficienti.²⁶ Persino la letteratura è stata esaminata da autori classici come Edgar Allan Poe

²⁴ Andreas Hepp, *Artificial Companions, Social Bots and Work Bots: Communicative Robots as Research Objects of Media and Communication Studies*, in *Media, Culture & Society*, 42.7–8, 2020, pp. 1410–26.

²⁵ Mary Ann Doane, *The Emergence of Cinematic Time: Modernity, Contingency, the Archive*, Harvard University Press, Cambridge (Mass.), 2002; Silvio Alovio, *L'occhio sensibile: Cinema e scienze della mente nell'Italia del primo Novecento*, Kaplan, Torino, 2013; Giuliana Bruno, *Film, Aesthetics, Science: Hugo Münsterberg's Laboratory of Moving Images*, in *Grey Room*, 36, 2009, pp. 88–113.

²⁶ Jonathan Sterne, *The Audible Past: Cultural Origins of Sound Reproduction*, Duke University Press, Durham, 2003; Simone Dotto, *Voci d'archivio. Fonografia e culture dell'ascolto nell'Italia tra le due guerre*, Meltemi, Milano, 2020; Domenico Napolitano, *La voce artificiale: Un'indagine media-archeologica e socio-culturale sul computer parlante, dalla sintesi vocale al voice cloning*, Tesi dottorale, Università degli Studi Suor Orsola Benincasa, 2021.

nella sua capacità di raggiungere effetti estetici sulla base di caratteristiche della psicologia umana, come l'attenzione.²⁷ L'evento più significativo nella storia dei media, a partire dal diciannovesimo secolo, non fu dunque l'introduzione di singole tecnologie quali il telegrafo, la fotografia, il cinema, la televisione o la rete internet, quanto l'emergere delle nuove scienze dell'umano, dalla fisiologia alla psicologia alle scienze sociali, che misero a disposizione le conoscenze e i mezzi epistemologici per adattare i media moderni ai sensi e alla psicologia dei loro pubblici e dei loro utenti.

Il rapporto tra AI e storia dei media è stato finora oggetto di ben poca attenzione. Gli storici si sono limitati a riconoscere come antecedenti dell'AI moderna quelle macchine che riproducevano comportamenti animali o umani. Gli esempi più noti in questo senso sono automi come l'anatra meccanica creata dall'inventore francese Jacques de Vaucanson nel 1739, che mostrava l'abilità di mangiare, digerire ed espellere i resti del processo di digestione, e il Turco, un automa che incantò i pubblici in Europa e nelle Americhe tra la fine del Settecento e l'Ottocento grazie alla sua apparente capacità di giocare a scacchi. In un libro dedicato ai rapporti tra AI e inganno, questi esempi sono certamente rilevanti, dato che la loro presunta "intelligenza" era il risultato di un imbroglio congegnato dai loro creatori: l'anatra meccanica non digeriva veramente il cibo ingerito, ma espelleva delle feci inserite appositamente al suo interno, mentre il Turco era azionato da un giocatore umano nascosto tra gli ingranaggi.²⁸ Ma per comprendere veramente le dinamiche attraverso cui l'AI mobilita complessi meccanismi di inganno è necessario espandere il nostro sguardo e abbracciare un contesto storico che includa non solo la storia di automi e macchine computazionali, ma la più ampia storia del rapporto tra inganno e media. Per questa ragione, farò spesso riferimento a tecnologie e pratiche che a prima vista sembrano avere poco a che vedere con l'AI, quali i media audiovisivi, le tecnologie di telecomunicazione, i mezzi di riproduzione sonora, l'arte figurativa, il teatro, e persino le sedute spiritiche dell'Inghilterra vittoriana.

Questo libro è debitore di idee e intuizioni sviluppate all'interno di svariati campi disciplinari, dagli studi sociali di scienza e tecnologia (STS) all'antropologia sociale, dalla psicologia sociale all'informatica, dalla teoria dei media alla storia del cinema e dell'arte. Ho usato queste diverse prospettive e spunti come strumenti per proporre e sviluppare una

²⁷ Edgar Allan Poe, *La filosofia della composizione*, Trad. Luigi Lunari, La Vita Felice, Milano, 2012.

²⁸ Jessica Riskin, *The Restless Clock: A History of the Centuries-Long Argument over What Makes Living Things Tick*, Chicago: University of Chicago Press, 2016; Mark Sussman, *Performing the Intelligent Machine: Deception and Enchantment in the Life of the Automaton Chess Player*, in *TDR/The Drama Review*, 43.3, 1999, pp. 81–96; James W. Cook, *The Arts of Deception: Playing with Fraud in the Age of Barnum*. Harvard University Press, Cambridge (Mass.), 2001.

prospettiva sull'AI che mette in primo piano il ruolo degli esseri umani nella costruzione del senso. Come ci hanno insegnato studiosi quali Sherry Turkle, Arjun Appadurai e Bruno Latour, non solo gli esseri umani, ma anche le cose possono essere considerate agenti sociali in determinate situazioni.²⁹ Le persone tendono ad attribuire intenzioni e comportamenti sociali a oggetti e tecnologie: pensate a chi si lamenta con la propria automobile o con il cellulare che non funziona, o a come i bambini (e non solo) attribuiscono nomi e appellativi a bambole e oggetti. Le cose, come le persone, hanno vite sociali, e i loro significati vengono costantemente rinegoziati all'interno di complesse interazioni e reti sociali.³⁰ Questo non vuol dire che le macchine e le cose possano avere, come suggeriscono in maniera spesso frettolosa le teorie sviluppate sotto l'etichetta del post-umano, una propria presenza sociale *indipendente* dalle persone che interagiscono con esse. La forza delle teorie sulla vita sociale delle cose, quello che le rende così convincenti, risiede proprio nella capacità di riconoscere che l'AI può avere effetti reali sulla nostra socialità, senza che questi effetti vengano scambiati per il frutto di un'entità assimilabile a quella umana. A fare di tecnologie come Alexa, Jibo o Replika dei veri e propri attori sociali, infatti, non è tanto la loro natura tecnica di per sé, quanto la nostra tendenza ad attribuire significato alle cose e alla realtà esterna.³¹

Mentre l'evoluzione dell'AI è solitamente raccontata nei termini dello sviluppo di tecnologie sempre più performanti, questo libro adotta dunque un approccio differente, concentrandosi su come essa sia strettamente legata all'accumularsi di conoscenze sulle percezioni e le reazioni degli utenti che vi entrano in contatto. Questo punto di vista ha il vantaggio di farci osservare le dinamiche che rendono le tecnologie di AI profondamente dirompenti per le relazioni sociali e l'esperienza quotidiana nelle società contemporanee. La

²⁹ Arjun Appadurai (a cura di), *La vita sociale delle cose. Una prospettiva culturale sulle merci di scambio*, Trad. Gilda Dina, Meltemi, Roma, 2021; Sherry Turkle (a cura di), *La vita nascosta degli oggetti tecnologici*, Trad. N. Cavalli, Ledizioni, 2009; Bruno Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford University Press, Oxford, 2005.

³⁰ Elizabeth Edwards, *Material Beings: Objecthood and Ethnographic Photographs*, in *Visual Studies*, 17.1, 2002, pp. 67–75.

³¹ Si veda anche, a riguardo, il paradigma CASA (Computer Are Social Actors) inaugurato dalla pubblicazione di Byron Reeves e Clifford Nass, *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*, CSLI Publications, 1996, in cui gli autori dimostrarono che le persone tendono ad applicare convenzioni e abitudini sociali anche nelle loro interazioni con oggetti tecnologici come il computer o il televisore. Il paradigma CASA ha dato vita a una vera e propria scuola di ricerca, con una massa impressionante di pubblicazioni, tra cui vale la pena ricordare Clifford Nass e Scott Brave, *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*, MIT Press, Cambridge (Mass.), 2005; Clifford Nass e Youngme Moon. "Machines and Mindlessness: Social Responses to Computers", in *Journal of Social Issues*, 56.1, 2000, pp. 81–103. Si vedano anche i lavori di Sherry Turkle, in particolare *La vita sullo schermo. Nuove identità e relazioni sociali nell'epoca di Internet*, Trad. B. Parrella, Apogeo, Adria, 2005, e *Insieme ma soli. Perché ci aspettiamo sempre più dalla tecnologia e sempre meno dagli altri*, trad. Susanna Bourlot e Lorenzo Lilli, Einaudi, Torino, 2019; *The Second Self: Computers and the Human Spirit*, MIT Press, Cambridge (Mass.), 2005.

traiettoria presa qui in esame ci stimola infatti a riformulare il dibattito sull'AI sulla base di una diversa considerazione: che il cambiamento portato dall'AI riguarda in primo luogo noi esseri umani. È teoricamente possibile che le macchine "intelligenti" arriveranno un giorno a rappresentare una forma di esistenza simile o uguale a quella umana; ciò che è certo, d'altronde, è che stanno già trasformando le modalità con cui concepiamo e portiamo avanti le nostre interazioni sociali.

Nonostante il tema dell'inganno sia raramente affrontato nei dibattiti pubblici e scientifici sull'AI, interrogare le implicazioni etiche e culturali di queste dinamiche è un compito urgente che può essere affrontato solo attraverso una riflessione interdisciplinare all'intersezione di informatica, scienze cognitive, scienze sociali e umanistiche. L'ossessione contemporanea per argomenti futuribili o futuristici, come la rivolta delle macchine o la "superintelligenza" che supererebbe e renderebbe obsoleta l'intelligenza umana, ha il rischio di renderci meno consci del fatto che le implicazioni più significative dell'AI non hanno a che vedere con un futuro distante e possibile: riguardano invece le più profonde dinamiche che già caratterizzano le nostre interazioni con le macchine e i dispositivi digitali.

L'innovazione tecnologica è frutto delle azioni e delle idee di scienziati, designer, imprenditori e politici, ma anche delle domande e dei dubbi che poniamo su di essa. L'ambizione di questo libro è stimolare chi legge a porsi nuovi tipi di domande sulle relazioni tra macchine e umani nel mondo contemporaneo. Le risposte dovremo trovarle noi, dal momento che le tecnologie che stiamo creando non possono fornirci alcuna risposta a riguardo. Come ammesso da una di queste:

