

Come pensa la macchina? Incognite dell'Intelligenza artificiale

a cura di Marco Pacini

Premessa	3
Marco Pacini Volontà di impotenza	5
Elena Esposito Dall'Intelligenza artificiale alla comunicazione artificiale	20
Luciano Floridi Intelligenza artificiale: il divorzio tra azione e intelligenza	35
Viola Schiaffonati Computer, robot ed esperimenti	51
Luca Fabbris Il programma, la rete, il circolo: un approccio <i>lo-fi</i> all'Autonomia artificiale	63
Raffaele Simone PUM vs. Robot Parlante	77
Raoul Kirchmayr L'automa e il moscerino. Gli scacchi come modello e metafora dell'Intelligenza artificiale	88
Edoardo Greblo Democrazia o algocrazia?	111
Fabio Chiusi L'Intelligenza artificiale non è la soluzione. Critica ideologica del soluzionismo nell'AI	125
Alessandro Di Grazia Vita artificiale. Dalla teologia politica alla teologia tecnica	139
Stefano Tieri L'algoritmo che dunque sono	155
Damiano Cantone Cinema e Intelligenza artificiale	165

CONTRIBUTI

Nicola Gaiarin Il self-help e la colpa	181
Luigi Azzariti Fumaroli L'esempio, l'informe	199



Dall'Intelligenza artificiale alla comunicazione artificiale

ELENA ESPOSITO

Gli algoritmi sono diventati intelligenti?

La ricerca sull'elaborazione automatica delle informazioni è sempre stata guidata dall'idea di riprodurre con le macchine l'intelligenza umana – come indica la stessa etichetta di Intelligenza artificiale (IA) che ha accompagnato il progresso della digitalizzazione fin dai suoi inizi negli anni cinquanta.¹ Oggi gli algoritmi, che ereditano e sviluppano questa prospettiva di ricerca, sono ovunque nella nostra vita quotidiana e presentano inediti pericoli e opportunità. Ma la sfida che pongono non è solo tecnologica: sta diventando ormai evidente che i concetti basati sulla riproduzione dell'intelligenza, che tuttora guidano l'interpretazione dell'operato degli algoritmi, non sono più adeguati per comprendere e controllare i recenti sviluppi. L'innovazione tecnologica richiede un'innovazione concettuale, che si estende fino alle radici del dibattito sull'IA.

Nella seconda metà del Novecento la ricerca sull'IA è proseguita con alterni successi, segnati da ripetuti “inverni” e “primavere” (Russell e Norvig 2003, p. 29; Cardon *et al.* 2018), fino alla situazione di oggi, in cui i progressi dell'IA sembrano avere uno slancio mai conosciuto prima – ma sono cambiate le macchine, il modo di lavorare e anche i problemi. Oggi si parla più di algoritmi che di computer,² si dà per scontato il riferimento al web (compresa la

1. L'evento fondatore è generalmente considerato la Dartmouth Conference del 1956: cfr. Moor 2006.

2. Nei discorsi recenti sull'IA e le sue trasformazioni, l'uso del termine algoritmo è spesso impreciso. La programmazione dei computer utilizza da sempre degli algoritmi,

partecipazione attiva degli utenti) e si parte dall'assunto che i dati da elaborare non siano scarsi ma piuttosto sovrabbondanti. A cosa sono dovuti questi cambiamenti? Le novità dipendono da due fattori collegati tra di loro – uno più tradizionale e uno più nuovo.

Il primo è l'enorme progresso del *machine learning*. Anche di apprendimento delle macchine si parla da sempre (Nilsson 1965), e con alterni successi, oggi però gli algoritmi sembrano essere capaci non solo di imparare, ma di imparare da soli: decidono autonomamente che cosa apprendere e come. Si parla di *unsupervised ML* e *deep learning*: tecniche in cui le macchine imparano a svolgere dei compiti in modi che non erano previsti dai loro programmatori e che in alcuni casi sono incomprensibili agli esseri umani, compresi quelli che le hanno progettate (Goodfellow *et al.* 2016; Burrell 2016).

La seconda novità, senza la quale la prima non si potrebbe realizzare, sono i cosiddetti Big Data: l'enorme quantità di dati diversi che sono disponibili oggi per le operazioni delle macchine (Mayer-Schönberger e Cukier 2013). La definizione di Big Data è complicata e controversa (cfr. per esempio Kitchin 2014), ma c'è un generale accordo sul fatto che questa nuova tipologia di dati sia legata alla diffusione del web 2.0³ – il cosiddetto “web partecipativo” che si è imposto negli ultimi quindici o vent'anni. Si tratta di una serie di innovazioni nelle tecnologie di programmazione, che hanno consentito di creare pagine web più dinamiche, aperte ai contributi e agli interventi dei loro visitatori, i quali partecipano direttamente alla creazione dei contenuti. Ciascuno di noi produce moltissimi dati, in modo volontario o in modo automatizzato (Kitchin 2017); lo facciamo con il nostro comportamento, con la navigazione e le scelte che facciamo sul web, con la partecipazione ai social media, ma anche con tutti i dati rilevati dai servizi GPS o dal cosiddetto Internet delle cose.

e il termine esisteva già molto prima della cibernetica. In questo testo seguo l'uso corrente, per quanto imperfetto, e parlo di algoritmi per riferirmi a tecniche di programmazione avanzate che usano l'apprendimento automatico e i Big Data (De Angelis 2014).

3. L'espressione è stata introdotta da Di Nucci (1999), e poi resa popolare da O'Reilly e dalla “Web 2.0 Conference” nel 2004: <https://web.archive.org/web/20050312204307/http://www.web2con.com/web2con>

Il risultato di queste innovazioni è che ora gli algoritmi che le utilizzano sembrano essere diventati intelligenti. Sono capaci di fare sempre più cose e sempre meglio, e in particolare riescono a svolgere dei compiti che in precedenza erano prerogativa degli esseri umani, dotati appunto di intelligenza. Gli algoritmi sono in grado di rispondere al telefono, fornire informazioni e condurre delle conversazioni. Sono capaci di scrivere dei testi: aziende come Narrative Science e Automated Insight hanno sviluppato algoritmi per produrre articoli di giornale, brochure per prodotti commerciali, libri di testo e altro ancora, che non si possono distinguere dai materiali scritti da autori umani.⁴ Le macchine hanno anche imparato a scrivere musica, produrre immagini e fornire consulenza psicologica. Anche se una valutazione precisa è difficile, si stima che il 50% del traffico online sia prodotto da programmi automatizzati. Milioni di utenti di Twitter sono bot, più del 70% del trading a Wall Street avviene tramite appositi software e almeno il 40% del lavoro redazionale di Wikipedia è effettuato da algoritmi.

Si tratta di prestazioni complesse che hanno sempre richiesto il contributo dell'intelligenza. Se ora le macchine sono in grado di svolgerle autonomamente, dobbiamo concludere che sono diventate intelligenti? Gli algoritmi stanno realizzando il sogno dell'IA? Molti osservatori ne sono convinti, in riferimento anche al fatto che gli algoritmi sono ormai capaci di superare il test di Turing (1950) adottato generalmente come criterio per stabilire se una macchina sia o meno intelligente. Come è noto, il criterio prevede che una macchina superi il test se è capace di comunicare con un essere umano senza che questi si renda conto che il suo partner è una macchina. Oggi le macchine superano di continuo il test di Turing, ma in un modo che dovrebbe farci riflettere. In moltissimi casi tutti noi comunichiamo con degli algoritmi senza rendercene

4. Il “robo-giornalismo” è regolarmente utilizzato da The Associated Press e da molte aziende come Samsung, Yahoo, Comcast e altre (Podolny 2015; Peiser 2019). Philip Parker, professore all'INSEAD di Fontainebleau, ha brevettato un metodo per produrre automaticamente libri plausibili e informativi, tra cui più di 100.000 titoli già disponibili su Amazon.

conto: prenotiamo i biglietti del treno, prendiamo appuntamenti e chiediamo assistenza dialogando con dei bot, o partecipiamo a videogiochi online in cui moltissimi giocatori sono automi. Anche quando sappiamo che il nostro partner è una macchina, poi, di solito non ci badiamo.⁵ Molti di noi fanno conversazione con assistenti personali digitali come Siri o Alexa, che spesso funzionano benissimo – anzi alcune ricerche sostengono che questi programmi conoscono gli utenti meglio dei loro famigliari e a volte meglio degli utenti stessi, e sanno prevedere le loro esigenze e richieste prima ancora che emergano (Youyou *et al.* 2015).

Il problema rispetto al progetto dell'IA, però, è che spesso quando scopriamo che le macchine sono intelligenti non siamo affatto rilassati – e non solo a causa delle nostre preoccupazioni riguardo a pregiudizi, errori, minacce alla privacy o usi maligni da parte di aziende e governi (boyd e Crawford 2012; Gillespie 2014). In realtà, quanto migliori sono gli algoritmi, tanto più aumenta il nostro disagio. Un articolo apparso su “The New Yorker” (Seabrook 2019), per esempio, descrive l'esperienza del giornalista con Smart Compose, una funzione di Gmail che suggerisce il finale delle frasi mentre le si scrive. L'algoritmo ha completato le email del giornalista in modo così appropriato, pertinente e in linea con il suo stile che lui si è trovato a imparare dalla macchina non solo quello che avrebbe scritto, ma anche quello che avrebbe dovuto scrivere (e a cui non aveva pensato), o che avrebbe potuto voler scrivere. E non gli è piaciuto per niente.

Questa esperienza, estremamente comune nell'interazione con macchine che sembrano essere intelligenti, ha un nome. Viene chiamata *uncanny valley* (Mori 2012): l'inquietante senso di disagio che proviamo quando una macchina sembra essere troppo simile a un essere umano – o a noi stessi. Facciamo tanti sforzi per costruire macchine intelligenti, e poi ci preoccupiamo se lo sono troppo – come mostrano gli allarmati dibattiti sul possibile arrivo

5. Già nel 2015 in Cina è stato fatto un test con il chatbot *xiaoice*, che in poche settimane è diventato la sesta celebrità più attiva su Weibo (il più diffuso social media cinese) e ha avuto decine di miliardi di conversazioni con gli utenti, per lo più su questioni private (Wang 2016). L'esperimento è stato considerato il più grande test di Turing della storia.

di una “singolarità” tecnologica o sulla creazione di una super-intelligenza che vada oltre le capacità umane (Kurzweil 2005; Bostrom 2014). Ci mettiamo in competizione con le macchine, e se ci sembra che siano loro a vincere non siamo affatto a nostro agio.

Le macchine hanno imparato a comunicare?

Il motivo, a mio parere, è che l’idea alla base dei progetti di Intelligenza artificiale non è più adeguata alla situazione di oggi. Se si guarda come lavorano i recenti algoritmi si vede che l’intelligenza non è il punto né lo scopo. Le macchine riescono a fare cose strabilianti non perché sono finalmente diventate intelligenti, ma paradossalmente proprio perché non cercano più di esserlo – fanno qualcos’altro. Si potrebbe dire che i progressi che osserviamo oggi non segnano il trionfo dell’Intelligenza artificiale, ma in pratica l’abbandono del progetto che ci stava dietro. Proprio per questo, per la prima volta le macchine sono in grado di produrre delle informazioni mai pensate da una mente umana e di fungere da partner di comunicazione interessanti e competenti.

I processi che guidano gli algoritmi sono ormai completamente diversi dai processi della mente umana, e infatti nessuna mente umana o combinazione di menti umane potrebbe riprodurli e spesso nemmeno capirli. I programmatori lo dichiarano esplicitamente: “Non cerchiamo di copiare l’intelligenza” (Solon 2012) – sarebbe un onere troppo pesante. Un esempio evidente, e spesso discusso, sono i programmi di traduzione automatica, che oggi funzionano molto bene – da quando i programmatori hanno smesso di cercare di insegnare agli algoritmi le diverse lingue e le loro regole. Oggi gli algoritmi traducono testi dal cinese senza conoscere il cinese, e nemmeno i loro programmatori lo conoscono. Usando *machine learning* e Big Data si limitano a trovare dei pattern e delle regolarità in enormi quantità di testi nelle lingue trattate (per esempio i materiali multilingua della Commissione europea), e li usano per produrre dei testi che risultano sensati – per le persone che li leggono. Non per gli algoritmi, che non li capiscono, come non capiscono niente dei contenuti che trattano, e non ne hanno bisogno. Allo stesso modo i correttori ortografici correggono gli

errori tipografici in qualsiasi lingua, senza conoscere queste lingue né le loro regole, il programma SmartCompose non capisce niente delle email che scrive, gli assistenti digitali come Alexa o Siri parlano con noi senza capire il significato delle parole che pronunciano, e gli algoritmi che scrivono dei testi “non ragionano come le persone per scrivere come le persone” (Hammond 2015, p. 7).

Le macchine non hanno imparato a ragionare come noi. Al contrario non cercano più di farlo, e proprio per questo funzionano così bene. Ci troviamo di fronte a un modo di elaborare i dati (e di gestire le informazioni) che è diverso dall'elaborazione e dalla comprensione delle informazioni umane (Borgo 2019). La mia ipotesi è che questa differenza non sia una debolezza, ma la radice stessa del successo di queste tecnologie. Hans Blumenberg (1957) diceva che gli esseri umani hanno imparato a volare quando hanno abbandonato l'idea di costruire delle macchine che imitino gli uccelli e battano le ali. Nello stesso modo, si potrebbe dire che l'elaborazione digitale delle informazioni è riuscita a raggiungere i risultati che vediamo oggi quando ha abbandonato l'ambizione di riprodurre in forma digitale i processi della mente umana. È vero che gli algoritmi superano il test di Turing, ma se ci pensiamo è un test molto strano: in effetti non misura l'intelligenza delle macchine ma la loro capacità di fungere da partner di comunicazione competenti e informativi. Il progetto dell'Intelligenza artificiale ha condotto a costruire macchine che non hanno imparato a diventare intelligenti, ma a fare un'altra cosa, che gli esseri umani fanno sulla base dell'intelligenza: hanno imparato a partecipare alla comunicazione. Piuttosto che di Intelligenza artificiale, potremmo parlare di una forma inedita di “comunicazione artificiale” (Esposito 2021).

Questo però non vuol dire che l'intelligenza umana sia diventata irrilevante – anzi rimane indispensabile. Gli algoritmi di *machine learning* sono in grado di calcolare, combinare ed elaborare con un'efficienza sorprendente le differenze nei dati, ma non sono in grado di produrle da soli. Le trovano sul web, a partire dal comportamento degli esseri umani. Attraverso i Big Data, gli algoritmi si “nutrono” delle differenze generate (coscientemente o inconscia-

mente) dagli individui e dal loro comportamento per produrre informazioni nuove, sorprendenti e potenzialmente istruttive. I recenti processi algoritmici partono dall'intelligenza degli utenti per rielaborarne i risultati e operare in modo intelligente come partner di comunicazione, senza bisogno di essere essi stessi intelligenti.

Per poter partecipare alla comunicazione gli algoritmi devono accedere al web. La loro comunicazione artificiale non sarebbe possibile senza il web, e la svolta si è realizzata solo dopo che i comportamenti umani sono andati online e hanno cominciato a produrre tutti i Big Data che gli algoritmi utilizzano. L'esempio emblematico è Google, che ha avuto tanto successo perché è stato il primo a adottare in modo sistematico questo tipo di approccio (Langville e Meyer 2006, p. 4 sgg.). L'algoritmo PageRank di Google ha "inventato" Internet così come lo conosciamo oggi (Metz 2012). L'intuizione chiave dei suoi autori, Larry Page e Steve Brin, è stata stabilire quali pagine sono importanti e per chi, trascurando completamente il contenuto delle pagine stesse. Per prendere questa decisione, PageRank non cerca di capire di cosa trattano le pagine o come ne parlano, ma va invece a vedere quante volte sono state "linkate" e da chi. Il ranking (la lista di siti che l'utente riceve come risultato della sua richiesta) si basa sul numero di *backlinks* alle pagine (quante volte sono state segnalate da altri siti web) e sulla loro importanza. Una pagina è rilevante se è stata linkata da molte pagine importanti, cioè (circolarmente) da pagine con tanti *backlinks*.

L'intuizione alla base di PageRank sta nel rinunciare all'obiettivo di capire cosa dice la pagina e nell'affidarsi esclusivamente alla struttura e alle dinamiche della comunicazione. I creatori di Google non hanno cercato di realizzare un grande schema organizzativo per il web basato su consulenti esperti e competenti, così come facevano motori di ricerca concorrenti come AltaVista e Yahoo. Non hanno cercato di capire e costruire un algoritmo che capisse. "Hanno invece fatto in modo che tutti gli altri lo facessero per loro" (Grimmelmann 2009, p. 941) navigando in rete e creando connessioni. I contenuti entrano in gioco in seguito, come risultato della classificazione e non come premessa. Google usa i

link per imparare non solo quanto è importante una pagina, ma anche di cosa tratta. Se i link a una determinata pagina utilizzano una specifica frase, il sistema deduce che la frase descrive accuratamente quella pagina e ne tiene conto per le ricerche successive. L'algoritmo è progettato per catturare e riflettere le scelte fatte dagli utenti (Gillespie 2014). Come ha dichiarato John Gianandrea di Google, quando un utente fa una ricerca su Einstein, Google “non intende dirvi cosa è importante di Einstein – intende dirvi cosa l'umanità cerca quando cerca Einstein” (Hamburger 2012).

Questo sistema è stato poi ulteriormente sviluppato per tener conto non solo della popolarità, ma anche di altri fattori, come il comportamento dei clic degli utenti, il tempo di lettura o i modelli di riformulazione delle richieste (Granka 2010, p. 367; <https://www.google.com/insidesearch/howsearchworks/>). L'intelligenza del sistema però rimane l'intelligenza degli utenti, che l'algoritmo sfrutta per dirigere e organizzare il proprio comportamento. Google è diventato il prototipo di un approccio che si può trovare in altri progetti di successo sul web. Dal 2003 il termine “googlizzazione” (Vaidhyanathan 2011; Rogers 2013, p. 83 sgg.) viene utilizzato per descrivere la diffusione, in un numero sempre maggiore di applicazioni e contesti, di modelli che non si affidano a esperti tradizionali ma si “nutrono” delle dinamiche del web per organizzare il loro funzionamento e per apprendere.

Il problema di spiegare le *black boxes*

Cosa vediamo, non vediamo, o vediamo in modo differente, se nell'osservare gli algoritmi spostiamo l'attenzione dall'intelligenza alla comunicazione? Se le macchine lavorano in un modo che è diverso dalla nostra intelligenza, e che spesso per noi è incomprendibile, come facciamo a controllare i risultati delle loro operazioni? L'intrasparenza delle procedure degli algoritmi più avanzati, spesso discussa come opacità, sembra essere la vera sfida che le tecniche di *machine learning* e i Big Data ci pongono oggi.

Anche se gli algoritmi che auto-apprendono sono molto efficienti, infatti, che siano delle *black boxes* non è affatto rassicurante, soprattutto quando sappiamo che le loro operazioni non sono

immuni da pregiudizi ed errori di vario tipo (Pasquale 2015). In molti casi vorremmo verificare la correttezza dei risultati prodotti dalle macchine, che possono essere sbagliati o inappropriati in molti modi differenti, con diverse conseguenze. In campo medico, per esempio, si teme che gli algoritmi possano non tenere adeguatamente conto di informazioni rilevanti, magari non esplicite (Holzinger 2019) – come quando hanno previsto che i pazienti asmatici fossero a minor rischio di morte per polmonite, trascurando il fatto che stavano già ricevendo intensa assistenza medica (Caruana *et al.* 2015). In altri campi – nella polizia (Lum e Isaac 2016), nelle decisioni sulla concessione di crediti (O’Neil 2016), nelle procedure di ammissione all’università (Hao 2020) – ci si preoccupa che possano riprodurre o intensificare gli squilibri (*bias*) nei dati, e si vorrebbe poter verificare i loro risultati e controllare il modo in cui sono stati ottenuti.

La recente branca di ricerca sulla “IA spiegabile” (*explainable AI* o *XAI*) cerca di rispondere a questa preoccupazione sviluppando delle procedure per spiegare le operazioni degli algoritmi che auto-apprendono (Wachter *et al.* 2017; Doshi-Velez *et al.* 2017; Miller 2019; Gilpin *et al.* 2018). Ma nel caso degli algoritmi di *deep learning* c’è un ostacolo di fondo: se per spiegazione si intende una procedura che permetta agli osservatori umani di capire cosa fa la macchina e perché, l’impresa è senza speranza. I processi dei recenti algoritmi che appaiono intelligenti sono intrinsecamente incomprensibili per l’intelligenza umana. Come sostiene Weinberger (2017), richiedere una spiegazione equivarrebbe a “costringere l’IA a essere artificialmente abbastanza stupida da permetterci di capire come arriva alle sue conclusioni”.

La strategia deve essere diversa, e infatti molti progetti sull’IA spiegabile stanno recentemente adottando un altro approccio, compatibile con la radicale oscurità dei processi algoritmici e coerente con l’idea di focalizzare l’attenzione sulla capacità degli algoritmi di partecipare alla comunicazione. In modo un po’ contraddittorio rispetto al loro nome, i progetti recenti di *XAI* non si occupano dell’intelligenza delle macchine. L’obiettivo è piuttosto quello di produrre una condizione di dialogo tra l’algoritmo e

l'utente, in cui la macchina fornisce delle risposte prendendo come input le richieste di chiarimento dei suoi interlocutori, che sono sempre diverse (per esempio, Cimiano *et al.* 2010; Rohlfing *et al.* 2020) – cioè è in grado di partecipare a una meta-comunicazione (Bateson 1972; Luhmann 1997, p. 250 sgg.) che può avere come oggetto i processi della macchina o i dati utilizzati. Lo scopo non è, e non può essere, che l'interlocutore capisca questi processi e l'intelligenza della macchina, ma che interpreti ciò che la macchina comunica su di essi in modo tale da poter esercitare una forma di controllo. Le macchine devono essere in grado di produrre spiegazioni adeguate rispondendo alle richieste dei loro interlocutori.

Questo è ciò che accade quando gli esseri umani compiono delle operazioni e prendono delle decisioni, sulle quali potremmo anche essere tenuti a offrire spiegazioni – dando indizi che permettano al ricevente di dare un senso alle nostre decisioni. Quando si ottiene una spiegazione, si ottengono informazioni sulla decisione senza essere informati sui processi neurofisiologici o psichici del partner – che (fortunatamente) possono rimanere oscuri, o privati. Spiegare le nostre operazioni non richiede di rivelare il nostro processo di pensiero, e ancor meno le connessioni dei nostri neuroni. Le spiegazioni, sostiene Luhmann, sono “riformulazioni con il beneficio aggiunto di una migliore connettività” (Luhmann 1990, p. 410). L'emittente produce una nuova comunicazione che fornisce elementi aggiuntivi relativi alla richiesta specifica dell'interlocutore e ai suoi bisogni. In ogni caso, il processo è interamente comunicativo: non abbiamo bisogno di accedere al cervello o alla mente dei nostri interlocutori, né abbiamo bisogno di accedere al mondo esterno – abbiamo solo bisogno di ottenere indizi che permettano alla comunicazione di andare avanti in modo controllato e non arbitrario.

Lo stesso approccio può essere adottato per affrontare i dilemmi della spiegazione nell'interazione con gli algoritmi che auto-apprendono. Decidere, come molti richiedono (Robbins 2019; Rudin 2019), di utilizzare solo modelli intrinsecamente comprensibili nei casi in cui può essere richiesta una spiegazione, non risolve il problema generale da cui nasce la necessità di spiegazione. La

questione è invece fare in modo che le macchine, opache o meno, producano “riformulazioni” dei loro processi che corrispondano alle richieste dei loro interlocutori e consentano loro di esercitare la forma di controllo appropriata al contesto. Si tratta di riprodurre con un partner digitale la situazione comunicativa in cui vengono richieste e fornite spiegazioni tra esseri umani.

In effetti, molti recenti progetti di XAI non cercano di imitare i calcoli fatti dall’algoritmo, ma mirano piuttosto a produrre “spiegazioni post-hoc” che riproducano ciò che gli esseri umani fanno nella comunicazione (Lipton 2018). I processi con cui le persone spiegano le loro decisioni sono distinti da quelli con cui le prendono, e sono di solito prodotti a posteriori, senza incidere sul processo decisionale. Allo stesso modo, nel campo della XAI i progettisti stanno addestrando i programmi a produrre spiegazioni che illustrano (potremmo dire “riformulano”) a posteriori il funzionamento degli algoritmi, senza influire sulle loro prestazioni. Come i processi linguistici che generano spiegazioni umane differiscono dai processi neurali che producono le decisioni da spiegare, così i processi che producono spiegazioni dei modelli algoritmici saranno diversi dai processi del modello. Possono usare per esempio spiegazioni verbali prodotte dalla macchina, visualizzazioni, spiegazioni locali come mappe di salienza che evidenziano gli elementi più significativi (Lipton 2018, p. 15 sgg.). Ciò che l’utente capisce delle spiegazioni della macchina non devono essere necessariamente i processi della macchina. Ciò che i programmi realizzano non è una forma alternativa di intelligenza, ma una forma alternativa di comunicazione.

Riferimenti bibliografici

- Bateson G. (1972), *Steps to an Ecology of Mind*, Chandler, San Francisco.
- Blumenberg H. (1957), *Nachahmung der Natur. Zur Vorgeschichte der Idee des schöpferischen Menschen*, “Studium Generale”, 10, pp. 266-283.
- Borgo S. (2020), *Ontological Challenges to Cohabitation with Self-taught Robots*, “Semantic Web”, 3, pp. 161-167.

- Bostrom N. (2014), *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, Oxford.
- boyd d., Crawford K. (2012), *Critical Questions for Big Data*, "Information, Communication and Society", 5, pp. 662-679.
- Burrell J. (2016), *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, "Big Data & Society", 1, pp. 1-12.
- Cardon D., Cointet J.-P. e Mazières A. (2018), *La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle*, "Rezeaux", 211, pp. 173-220.
- Caruana R. et al. (2015), *Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission*, "KDD'15", pp. 1721-1730.
- Cimiano P., Rudolph S. e Hartfiel H. (2010), *Computing Intensional Answers to Questions – An Inductive Logic Programming Approach*, "Data & Knowledge Engineering", 3, pp. 261-278.
- De Angelis S.F. (2014), *Artificial Intelligence. How Algorithms Make Systems Smart*, "Wired", ottobre, <https://www.wired.com/insights/2014/09/artificial-intelligence-algorithms-2/>
- Di Nucci D. (1999), *Fragmented Future*, "Print", 4, pp. 221-222.
- Doshi-Velez F., Kim B. (2017), *Towards a Rigorous Science of Interpretable Machine Learning*, arXiv:1702.08608v2 (ultimo accesso: 6 dicembre 2020).
- Esposito E. (2012), "Kontingenzerfahrung und Kontingenzbewusstsein in systemtheoretischer Perspektive", in K. Toens, U. Willems (a cura di), *Politik und Kontingenz*, Springer, Wiesbaden, pp. 39-48.
- (2021), *Artificial Communication: How Algorithms Produce Social Intelligence*, The MIT Press, Cambridge (Mass.).
- Gillespie T. (2014), "The relevance of algorithms", in T. Gillespie, P.J. Boczkowski e K.A. Foot (a cura di), *Media Technologies*, The MIT Press, Cambridge (Mass.), pp. 167-194.
- Gilpin L.H., Bau D., Yuan B.Z., Bajwa A., Specter M. e Kagal L. (2018), *Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning*, arXiv:1806.00069.
- Goodfellow I., Bengio Y. e Courville A. (2016), *Deep Learning*

- (*Adaptive Computation and Machine Learning*), The MIT Press, Cambridge (Mass.).
- Granka L.A. (2010), *The Politics of Search: A Decade Retrospective*, “The Information Society”, 26, pp. 364-374.
- Grimmelmann J. (2009), *The Google Dilemma*, “New York Law School Law Review”, 53, pp. 939-950.
- Hamburger E. (2012), *Building the Star Trek Computer: How Google’s Knowledge Graph is Changing Search*, “The Verge”, 8 giugno.
- Hammond K. (2015), *Practical Artificial Intelligence for Dummies*, Wiley, Hoboken (N.J.).
- Hao K. (2020), *The UK Exam Debacle Reminds Us That Algorithms Can’t Fix Broken Systems*, “MIT Technology Review”, 20 agosto.
- Holzinger A. et al. (2019), *Causability and Explainability of Artificial Intelligence in Medicine*, “WIREs Data Mining and Knowledge Discovery”.
- Kitchin R. (2014), *Big Data, New Epistemologies and Paradigm Shifts*, “Big Data & Society”, aprile-giugno, pp. 1-12.
- (2017), “Big Data”, in AA.VV., *The International Encyclopedia of Geography*, Wiley & Sons, Hoboken (N.J.).
- Kurzweil R. (2005), *The Singularity is Near*, Viking Books, New York.
- Langville A.N., Meyer C.D. (2006), *Google’s PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton (N.J.)-Oxford.
- Lipton Z.C. (2018), *The Mythos of Model Interpretability*, “ACM Queue”, 3, pp. 1-27.
- Luhmann N. (1990), *Die Wissenschaft der Gesellschaft*, Suhrkamp, Frankfurt a.M.
- (1997), *Die Gesellschaft der Gesellschaft*, Suhrkamp, Frankfurt a.M.
- Lum K., Isaac W. (2016), *To Predict and Serve?*, significance-magazine.com, ottobre, pp. 14-19.
- Mayer-Schönberger V., Cukier K. (2013), *Big Data. A Revolution That Will Transform How We Live, Work, and Think*, Murray, London.

- Metz C. (2012), *If Xerox Parc invented the PC, Google Invented the Internet*, wired.com, 8 agosto.
- Miller T. (2019), *Explanation in Artificial Intelligence: Insights from the Social Sciences*, "Artificial Intelligence", 267, pp. 1-38.
- Moor J. (2006), *The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years*, "AI Magazine", 4, pp. 87-89.
- Mori M. (2012), *The Uncanny Valley*, "IEEE Robotics and Automation", 2, pp. 98-100.
- Nilsson N.J. (1965), *Learning Machines*, McGrawHill, New York.
- O'Neil C. (2016), *Weapons of Math Destruction*, Crown, New York.
- Pasquale F. (2015), *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge (Mass.).
- Peiser J. (2019), *The Rise of the Robot Reporter*, "The New York Times", 5 febbraio.
- Podolny S. (2015), *If an Algorithm Wrote This, How Would You Even Know?*, "The New York Times", 7 marzo.
- Robbins S. (2019), *A Misdirected Principle with a Catch: Explicability for AI*, "Minds and Machines", 29, pp. 495-514.
- Rogers R. (2013), *Digital Methods*, The MIT Press, Cambridge (Mass.)-London.
- Rohlfing K. et al. (2020), *Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems*, "IEEE Transactions on Cognitive and Developmental Systems", 3, pp. 717-728.
- Rudin C. (2019), *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, "Nature Machine Intelligence", 1, pp. 206-215.
- Russell S.J. e Norvig P. (2003), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River (N.J.).
- Seabrook J. (2019), *Can a Machine Learn to Write for The New Yorker?*, "The New Yorker", 14 ottobre.
- Solon O. (2012), *Weavrs. The Autonomous, Tweeting Blog-bots That Feed on Social Content*, wired.co.uk, 28 marzo.
- Turing A.M. (1950), *Computing Machinery and Intelligence*, "Mind", 236, pp. 433-460.

- Vaidhyanathan S. (2011), *The Googlization of Everything (and Why We Should Worry)*, University of California Press, Berkeley-Los Angeles.
- Youyou W., Kosinski M. e Stillwell D. (2015), *Computer-based Personality Judgments Are More Accurate Than Those Made by Humans*, "Proceedings of the National Academy of Sciences (PNAS)", 4, pp. 1036-1040.
- Wachter S., Mittelstadt B. e Floridi L. (2017), *Transparent, Explainable, and Accountable AI for Robotics*, "Science Robotics", 6.
- Wang Y. (2016), *Your Next New Best Friend Might Be a Robot. Meet Xiaoice. She's Empathic, Caring, and Always Available – Just Not Human*, "Nautilus", 4 febbraio.
- Weinberger D. (2017), *Machines Now Have Knowledge We'll Never Understand*, "Wired", 18 aprile.