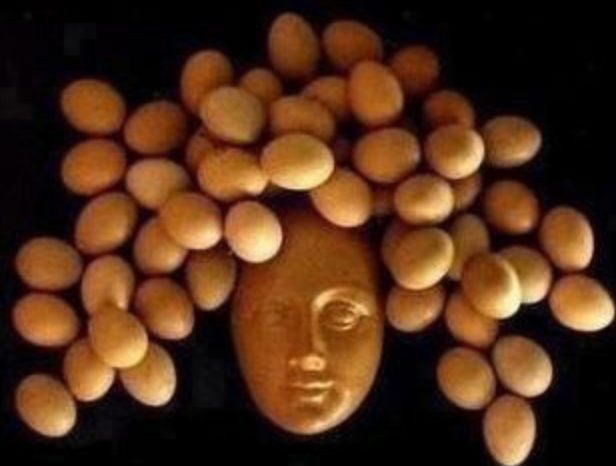


S&F_n. 30_2023



scienzaefilosofia.it

S&F_n. 30_2023

ISSN 2036_2927

COMITATO SCIENTIFICO

PAOLO AMODIO	Docente di Filosofia Morale _ Università degli Studi di Napoli Federico II
GUIDO BARBUJANI	Docente di Genetica _ Università degli Studi di Ferrara
EDOARDO BONCINELLI	Docente di Biologia e Genetica _ Università "Vita-Salute San Raffaele" di Milano
ROSSELLA BONITO OLIVA	Docente di Filosofia Morale _ Università degli Studi di Napoli - L'Orientale
BARBARA CONTINENZA	Docente di Storia della scienza e delle tecniche _ Università degli Studi di Roma "Tor Vergata"
MASSIMILIANO FRALDI	Docente di Scienza delle costruzioni _ Università degli Studi di Napoli Federico II
ORLANDO FRANCESCHELLI	Docente di Teoria dell'evoluzione e Politica _ Università degli Studi di Roma "La Sapienza"
ELENA GAGLIASSO	Docente di Filosofia e Scienze del vivente _ Università degli Studi di Roma "La Sapienza"
GIANLUCA GIANNINI	Docente di Filosofia Morale _ Università degli Studi di Napoli Federico II
† PIETRO GRECO	Giornalista scientifico e scrittore, Direttore del Master in Comunicazione Scientifica della Scuola Internazionale Superiore di Studi Avanzati (SISSA) di Trieste
GIUSEPPE LISSA	Professore Emerito di Filosofia Morale _ Università degli Studi di Napoli Federico II
GIUSEPPE O. LONGO	Docente di Teoria dell'informazione _ Università degli Studi di Trieste
ROBERTO MARCHESINI	Centro Studi Filosofia Postumanista_Direttore della Scuola di Interazione Uomo Animale (SIUA)_Docente di Scienze Comportamentali Applicate
DAVIDE MAROCCO	Docente di Psicometria _ Università degli Studi di Napoli Federico II _ Già Lecturer in Cognitive Robotics and Intelligent Systems, Centre of Robotics and Neural Systems, School of Computing and Mathematics, University of Plymouth, UK
ALESSANDRO MINELLI	già Docente di Zoologia _ Università degli Studi di Padova
MAURIZIO MORI	Docente di Bioetica _ Università degli Studi di Torino
TELMO PIEVANI	Docente di Filosofia della Scienza _ Università degli Studi di Milano-Bicocca
VALLORI RASINI	Docente di Filosofia Morale _ Università degli Studi di Modena e Reggio Emilia
ROBERTO REVELLO	Docente presso l'Università dell'Insubria
† STEFANO RODOTÀ	Docente di Diritto Civile _ Università degli Studi di Roma "La Sapienza"
GIACOMO SCARPELLI	Sceneggiatore cinematografico e storico della filosofia e delle idee, insegna all'Università di Modena e Reggio Emilia. È Fellow della Linnean Society of London e della Royal Geographical Society
SETTIMO TERMINI	già Docente di Cibernetica _ Università degli Studi di Palermo
NICLA VASSALLO	Docente di Filosofia Teoretica _ Università degli Studi di Genova

INTERNATIONAL ADVISORY BOARD

DAVID BANON	Professeur au Département d'études hébraïques et juives, Université de Strasbourg; Membre de l'Institut Universitaire de France; Prof. invité au département de pensée juive, Université hébraïque de Jérusalem
RENAUD BARBARAS	Professeur à l'Université Paris 1 Panthéon-Sorbonne; Membre de l'institut universitaire de France; Grand prix de philosophie 2014 de l'Académie Française
MYLENE BOTBOL-BAUM	Professeure de philosophie et bioéthique, facultés de Médecine et des Sciences Philosophiques; Professeure à l'Unité d'éthique biomédicale - UCL (Université Catholique de Louvain), Bruxelles, Belgique
COLAS DUFLO	Professeur de Littérature française et d'histoire des idées du dix-huitième siècle, Université Paris-Nanterre, France
EDWARD K. KAPLAN	Kevy and Hortense Kaiserman Professor in the Humanities, Brandeis University, Waltham, Massachusetts
NEIL LEVY	Deputy Director (Research) of the Oxford Centre for Neuroethics; Head of Neuroethics at the Florey Neuroscience Institutes, University of Melbourne
ANNA LISSA	Maitre de Conférence Littérature juive et hébraïque Département d'Etudes hébraïques Université Paris-8, France
DIEGO LUCCI	Professor of History and Philosophy, American University in Bulgaria
MAX STADLER	Professur für Wissenschaftsforschung, Eidgenössische Technische Hochschule, Zürich

REDAZIONE

PAOLO AMODIO (DIRETTORE)	Università degli Studi di Napoli Federico II_ Dipartimento di Studi Umanistici_ Via Porta di Massa, 1 80133 Napoli tel. +390812535582 fax +390812535583 email: paamodio@unina.it
LUCA LO SAPIO (COORDINATORE DI REDAZIONE)	Università degli Studi di Torino
TERESA CAPORALE	Università degli Studi di Napoli_Federico II
VIOLA CAROFALO	Università degli Studi di Napoli_L'Orientale
DIDIER ALESSIO CONTADINI	Università degli Studi di Milano_ Bicocca
ROBERTO CORSO	Università degli Studi di Napoli_Federico II
MARIO COSENZA	Università degli Studi di Napoli_Federico II
CRISTIAN FUSCHETTO	Università degli Studi di Napoli_Federico II
FABIANA GAMBARDILLA	Università degli Studi di Napoli_Federico II
LUCA MANDARA	Università degli Studi di Napoli_Federico II
MATTIA PAPA	Università degli Studi di Napoli_Federico II
DELIO SALOTTOLO	Università degli Studi di Napoli_L'Orientale
ALESSANDRA SCOTTI	Università degli Studi di Napoli_Federico II
VALERIO SPECCHIO	Università degli Studi di Napoli_Federico II
ALDO TRUCCHIO	Università degli Studi di Salerno

INTRODUZIONE

Scienza&Filosofia 30_2023 compie 15 anni. Finalmente in Fascia A per l'Anvur per i settori 11/C1, 11/C2, 11/C3 (e retroattivamente dal 2019), sempre online per scelta, in ordine al dinamismo e all'immediata disponibilità della ricezione, adattandosi volentieri ai tempi e agli spazi che la rete in genere istituisce: vorrebbe essere agile e facilmente fruibile per chi è interessato a prender parte alle nostre discussioni. La sua *mission* non può dunque che essere diretta e senza scolastici orpelli:

Scienza&Filosofia è più di un incrocio disciplinare o di un'area nella quale far convergere temi e interrogativi. Le innovazioni tecnico-scientifiche sono incalzanti: occorre lavorare criticamente sul presente e aprire discussioni che coinvolgano non solo questioni teoretiche, ma anche l'etica applicata e l'analisi delle narrazioni collettive. Ci proponiamo dunque di rispondere alle sfide poste dagli sviluppi della biomedicina e delle nuove tecnologie attingendo alle scienze umane e sociali, di promuovere il confronto sugli scenari sociopolitici che si prospettano e di individuare ed evidenziare il potenziale d'innovazione filosofica tracciato dalla ricerca scientifica.

Da questa *mission* la rivista trova l'articolazione che ci è parsa più efficace. Anche questo numero conterrà perciò le tipiche sezioni:

DOSSIER Il vero e proprio focus tematico scelto intorno al quale andranno a orbitare

STORIA Esposizione e ricostruzione di questioni di storia della scienza e di storia di filosofia della scienza con intenzione sostanzialmente divulgativa;

ANTROPOLOGIE Temi e incroci tra scienze, antropologia filosofica e antropologia culturale;

ETICHE Riflessioni su temi di "attualità" bioetica;

LINGUAGGI Questioni di epistemologia;

ALTERAZIONI Dalla biologia evuzionistica alla cibernetica, temi non direttamente "antropocentrati";

COMUNICAZIONE La comunicazione della scienza come problema filosofico, non meramente storico o sociologico. In altri termini: quanto la comunicazione della scienza ha trasformato la scienza e la sua percezione?;

ARTE Intersezioni tra scienze e mondo dell'arte;

RECENSIONI&REPORTS Le recensioni saranno: *tematiche*, cioè relative al dossier scelto e quindi comprensive di testi anche non recentissimi purché attinenti e importanti; *di attualità*, cioè relative a testi recenti. *Reports* di convegni e congressi.

E infine ospitiamo uno Special Issue dedicato a *Forma e Filosofia: Un Dialogo tra Scienza, Tecnologia ed Estetica*.

Per favorire la fruibilità telematica della rivista, i contributi si aggireranno tra le 15.000 - 20.000 battute, tranne rare eccezioni, e gli articoli saranno sempre divisi per paragrafi. Anche le note saranno essenziali e limitate all'indicazione dei riferimenti della citazione e/o del riferimento bibliografico e tenderanno a non contenere argomentazioni o ulteriori approfondimenti critici rispetto al testo.

A esclusione delle figure connesse e parti integranti di un articolo, le immagini che accompagnano i singoli articoli saranno selezionate secondo il gusto (e il capriccio) della Redazione e non pretenderanno, almeno nell'intenzione - per l'inconscio ci stiamo attrezzando - alcun rinvio didascalico.

Le immagini d'apertura ai singoli articoli - coperte da copyright - che appaiono in questo numero, sono di RICCARDO FISSORE, cui va il nostro ringraziamento (inunmodosurreale@gmail.com)

In rete, dicembre 2023

La Redazione di S&F_

S&F_n. 30_2023



INDICE

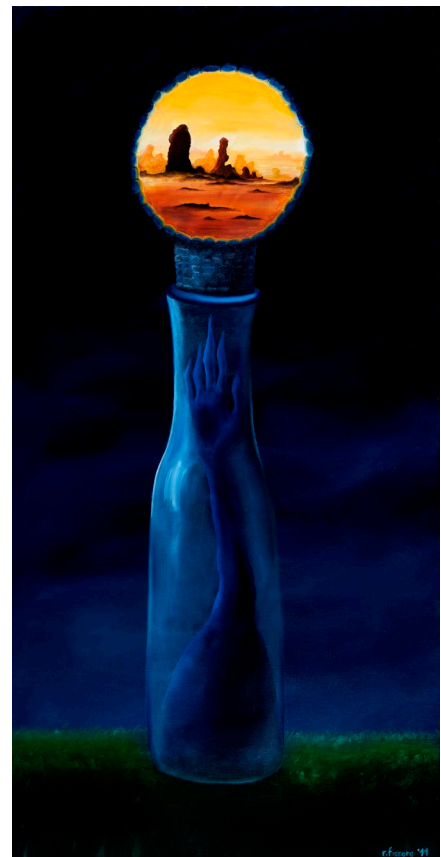
- 1 Introduzione
- 4 Indice

Dossier

- 7 Pluralità dei mondi, vita, cosmo. Il sogno di Giordano Bruno
- 13 Giovanni Covone *Lo stato dell'arte nella ricerca dei pianeti extrasolari di tipo terrestre*
- 26 Stephen Webb *Exploring an extension of Dick's "intelligence principle"*
- 37 Lorenzo De Piccoli *Lezioni dall'Antartide. Per un'ecologia del sistema solare*
- 54 Maria Antonia Delli Gatti *Contingency and necessity in the debate on the abundance of life in the cosmos*
- 69 Valerio Di Lauro - Matteo Bedetti *Are the characteristics of life as we know it universal?*
- 86 Luigi Laino *Into the Others' mind. Remarks on the philosophy of geometry from Kant onwards*

Special Issue - Filosofia delle forme: Nuovi approcci e problemi

- 111 Forma e Filosofia. Un Dialogo tra Scienza, Tecnologia ed Estetica
- 117 Marco Tamborini *La costruzione della forma. Percorsi in una filosofia delle forme bio-tecniche*
- 138 Salvatore Tedesco *La poetica morfologica di Inger Christensen come "derealizzazione" della modernità estetica*
- 158 Maddalena Mazzocut-Mis *Forma, forme e metamorfosi*
- 182 Guido Cusinato *Formazione e trasformazione. Il desiderio di forma in un essere che nasce privo di forma esistenziale*
- 210 Simona Chiodo *Intelligenza artificiale e destino della forma*
- 235 Valeria Maggiore *La Plasticità delle Forme: Catherine Malabou e La configurazione plastica fra trasformazione e distruzione*



Storia

- 273 Giorgio Bertolazzi *La Falsificazione nella scienza moderna*

Antropologie

- 288 Federico Simonetti *Esteriorizzazione e immaginazione nell'era dell'Intelligenza Artificiale Generativa. Le IA generative e la nascita di una immaginazione artificiale*

Etiche

- 320 Lorella Meola *Big data e medicina: gli effetti del nudging personalizzato*

Linguaggi

- 338 Luca Valentino *Lo sfruttamento della libertà come fondamento del neoliberalismo: Una riflessione attraverso la filosofia politica di Byung-Chul Han*

Alterazioni

- 366 Cristiano Calì *Come ci cambia La tecnologia. L'agency delle AI e La capacità cognitiva di prendere decisioni razionali*

Comunicazione

- 387 Alessandra Scotti *Henri Bergson e L'intelligenza nelle cose*

Arte

- 405 Delio Salottolo** *La terza epoca del concetto di "rivoluzione": da Platonov a Latour passando per Koselleck (e oltre)*

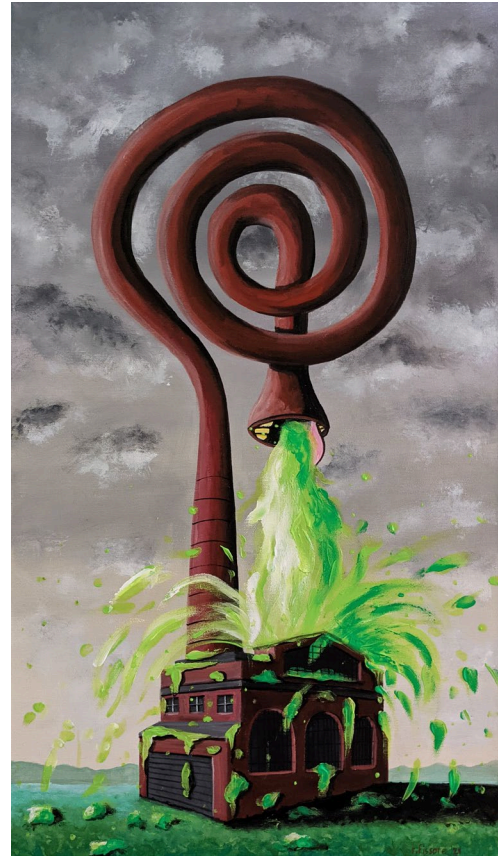
Recensioni&Reports

Report

- 441** *New Pathways for a sustainable future: science and ethics for a renewed society*
Torino 20 novembre 2023
(Valerio Specchio)

recensioni

- 448** Teresa Caporale, *Ateismo religioso e religione atea. Genesi e destinazione dell'antropoteismo feuerbachiano*, Orthotes, Napoli-Salerno 2021
(Antonio Lucci)
- 453** Eugenio Mazzarella, *Contro Metaverso. Salvare La presenza*, Mimesis, Milano-Udine 2022
(Alessia Gifuni)
- 458** Iain McGilchrist, *IL padrone e il suo emissario: I due emisferi del cervello e La costruzione dell'Occidente*, a cura di L. Lunari, testo inglese a fronte, Utet, Torino 2022
(Adil Bellafqih)
- 469** Maurice Merleau-Ponty, *L'istituzione, La passività. Corso al Collège de France (1954-1955)*, Mimesis, Milano-Udine 2023
(Alessandra Scotti)
- 473** David Chalmers, *Reality+. Virtual worlds and the problems of philosophy*, Allen Lane, 2022
(Valerio Specchio)
- 484** Roberto Marchesini, *Technophysiology or how technology modifies the Self*, Cambridge Scholars Publishing, Newcastle upon Tyne 2023
(Josephine Condemi)
- 495** Steven Umbrello, *Oggetti Buoni. Per una tecnologia sensibile ai valori*, Fandango Libri, Roma 2023
(Francesca Varesi)
- 500** Vincenzo Cuomo - Luca Scafoglio (a cura di), *Sensazione e protostoria nel pensiero di Christoph Türcke*, Kaiak, Pompei (NA) 2023
(Piero Carreras)



S&F_n. 30_2023



ALTERAZIONI

CRISTIANO CALÌ

COME CI CAMBIA LA TECNOLOGIA.

L'AGENCY DELLE AI E LA CAPACITÀ COGNITIVA

DI PRENDERE DECISIONI RAZIONALI

1. Introduzione
2. Agenti o Intelligenti?
3. Problemi decisionali "a confronto"
4. Il paradigma della sostenibilità
5. La complementarità tra uomo e macchina nel processo decisionale
6. Interazione dentro la decisione

ABSTRACT: HOW TECHNOLOGY CHANGES US. THE AGENCY OF AI AND THE COGNITIVE ABILITY TO MAKE RATIONAL DECISIONS
The article explores the concept of agency in the context of artificial intelligence, analyzing the approach of Stuart Russell and Peter Norvig. After examining three possible levels of machine agency, the contribution reflects on the challenge of attributing agency to algorithms by proposing two interaction paradigms: "substitutivist" and "complementarist". The latter suggests collaboration between humans and AI, emphasizing a shared decision-making process. The ultimate aim of the contribution is to raise ethical considerations regarding autonomous AI decisions on one hand and the human-machine interaction on the other, especially when such usage involves the cognitive ability to make free decisions.



1. Introduzione¹

A partire dagli anni Cinquanta del secolo scorso, gli algoritmi in grado di compiere operazioni un tempo

appannaggio esclusivo degli esseri umani sono state definite

¹ Per la prima parte di questo contributo mi sono rifatto a una presentazione orale tenuta durante la conferenza annuale della SISEC - Società Italiana di Sociologia Economica, svoltasi a Brescia nel febbraio 2023. Il testo di quella presentazione è rinvenibile in C. Calì, *Razionalità degli agenti e "razionalità" delle macchine. Un tentativo di confronto*, in *Internet nel nuovo. Millennio. Diritto, società e tecnologia*, a cura di A. Alù, A. Ciccarello, *Internet nel nuovo. Millennio. Diritto, società e tecnologia*, Editoriale Scientifica, Napoli 2023, in corso di stampa. Una precedente versione inedita del presente saggio è stata premiata col premio MyLlennium Award 2023, per la sezione MyBook.

intelligenze artificiali (AI); una nozione talmente controversa che Stuart Russell (Docente a Berkeley) e Peter Norvig (Direttore della ricerca in Google), due tra i più importanti studiosi di AI, nella prefazione alla quarta edizione italiana del loro manuale - utilizzato nelle maggiori accademie per introdurre all'intelligenza artificiale - dichiarano come al centro del loro testo vi sia la nozione di *agente*. Questa definizione, a detta loro, dovrebbe prendere il posto di quella molto più comune di *intelligenza artificiale*, dal momento che l'AI sarebbe soltanto «lo studio degli agenti che ricevono percezioni dall'ambiente ed eseguono azioni»². Tale disciplina, tuttavia, non si limiterebbe allo studio degli agenti ma, *in primis*, permetterebbe di *costruire* agenti³. Risulta fondamentale, allora, cogliere cosa i due autori - veri e propri alfieri dell'AI - intendano con *agente*, soprattutto in ragione del fatto che essi assumono questo concetto nella sua accezione più pregnante.

La mia riflessione sarà volta, pertanto, a intercettare se e come sia possibile predicare l'*agency* di alcuni algoritmi di AI, e, quindi, a comprendere le modalità di interazione di quest'*agency* con l'*agency* umana.

A tal fine suggerirò due modalità d'interazione che, facendo ricorso a una sorta di neologismo, chiamerò paradigma *sostituivista* e *complementarista*. Dall'analisi di alcuni casi concreti legati al mondo della giurisprudenza, s'indicherà, infine, quale ritengo essere la sfida nel momento in cui gli algoritmi assumono un ruolo determinante proprio su quella capacità che è comunemente

² S. Russell, P. Norvig, *Intelligenza artificiale. Un approccio moderno* (2016), tr. it. Pearson, Milano-Torino 2021, 1, p. XXIV. Assunta in questa accezione l'intelligenza artificiale potrebbe essere una disciplina che studia anche l'essere umano, al pari dell'antropologia, e sono gli stessi autori del manuale a favorire una tale lettura nel momento in cui si afferma che «ogni agente implementa una funzione che mette in corrispondenza sequenze percettive e azioni», *ibid.*

³ Credo che sia un caso più unico che raro, un vero e proprio *hapax*, il fatto che si parli di *costruzione* dell'agente e non di *costituzione* dell'agente o di semplice *analisi* dell'agente (espressioni sicuramente molto più comuni in diverse branche del sapere scientifico).

riconosciuta come “la cosa più propria che ciascuno di noi ha”, ovvero la capacità di decidere liberamente dei nostri atti.

2. Agenti o intelligenti?

2.1. Tre diversi gradi di “agency”

Russell e Norvig sviluppano implicitamente (e forse anche inconsapevolmente) una riflessione “scalare” sulla nozione di agente, nel senso che a ogni grado di un particolare tipo di agente corrisponde un grado di sofisticatezza dell’AI intesa solo ed esclusivamente come disciplina.

Ad un primo livello (1) un’agente «è semplicemente qualcosa che agisce, che fa qualcosa»⁴. A questa nozione di agente corrisponde un’AI dedita a sviluppare macchine che emettano un *output* a partire da determinati *input* (per esempio un robot che pulisce il pavimento quando è sporco e che si spegne quando invece è pulito, o anche un qualsiasi altro *software* installato sul nostro *personal computer*). Questa dimensione più basilica lascia spazio, tuttavia, ad un secondo livello, per il quale l’agente non è compreso soltanto come “colui che detiene l’*agere*” (nel senso di colui che *fa delle cose*) ma come colui che è “originatore delle proprie azioni”, nel senso di colui che delibera sul da farsi. A questo secondo livello (2) – che è occupato da quello che può essere definito *agente artificiale* – corrisponde anche il secondo livello dell’AI, la quale non avrà più come obiettivo quello di costruire macchine-che-fanno-cose ma di elaborare «macchine in grado di *calcolare come agire* in modo efficace e sicuro in un’ampia varietà di situazioni nuove»⁵.

Si badi bene che in questa affermazione, per quanto ambiziosa, si cela già un presupposto cruciale, per il quale la scelta è essa stessa azione, anche qualora essa non abbia un corrispettivo fenomenico osservabile esternamente (in altre parole: il semplice deliberare è esso stesso una modalità di agire).

⁴ *Ibid.*, p. 6.

⁵ *Ibid.*, p. 3 [corsivo mio].

Il “calcolare come agire”, tuttavia, non è sufficiente, ed è qui che appare il terzo livello (3), quello dell’*agente* (artificiale) *razionale*, ovvero quello di un agente che «agisce in modo da ottenere il miglior risultato o, in condizioni di incertezza, il miglior risultato atteso»⁶. A questo livello dovrà corrispondere una disciplina in grado di progettare macchine che agiscano in vista di un fine (ottimizzandolo).

Si noti però come il parlare di *agenti razionali* non sia di certo più semplice che parlare di *intelligenze artificiali*. Quest’ulteriore definizione, che fa riferimento alla razionalità e non più all’intelligenza, infatti, non è scevra di complicazioni, dal momento che anche la razionalità, così come l’intelligenza, è ben lontana dall’essere raggiunta da parte delle AI.

Se però l’attributo di *razionale*, usato dai due ricercatori, è controverso, il sostantivo, di contro, è ormai pacificamente acquisito nel dibattito. Gli algoritmi sono veri e propri *agenti*, e questo dato – benché possa sembrare inusuale – è molto più evidente di quanto non riconsegnino le altre definizioni.

Gli algoritmi oggi, infatti, *prendono decisioni* e, dal momento che – stando a una tradizione filosofica e giuridica vastissima – «decidere è anche agire e tale verbo va assegnato ed esteso a chiunque agisca, si tratti dell’uomo o dell’intelligenza artificiale»⁷, gli algoritmi non sono soltanto dei semplici agenti ma degli *agenti decisionali*.

Nel momento, allora, in cui si ha questa ambizione: progettare e costruire agenti artificiali razionali – ovvero agenti che, in vista di un fine/risultato (3), deliberano/calcolano (2) e quindi agiscono/riconsegnano *output* (1) – è necessario comprendere come tutto questo avvenga nella macchina-uomo per poi permettere ad una macchina, questa volta di silicio e non di carbonio, di

⁶ *Ibid.*, p. 6.

⁷ C. Canullo, *Chi decide? Intelligenza artificiale e trasformazioni del soggetto nella riflessione filosofica*, in E. Calzolaio (a cura di), *La decisione nel prisma dell’intelligenza artificiale*, Cedam, Milano 2020, p. 35.

implementarlo. Se questa volontà riproduttiva, infatti, può venire bypassata nel caso dell'intelligenza⁸ non sembra potersi dire lo stesso per quanto concerne la deliberazione.

Essendo impossibile in questa sede soffermarsi su cosa comporti, ingegneristicamente parlando, la realizzazione di questo sogno (e se un approccio produttivo o riproduttivo possa essere più congeniale), è mia intenzione soffermarmi sulla nozione di *agente* che - a prescindere da come la intendono i due autori statunitensi - è divenuta un termine chiave nel dibattito odierno, dal momento che stante il fatto che al giorno d'oggi non abbiamo ancora *intelligenze* artificiali, è fuor di dubbio che abbiamo *agenti* artificiali.

2.2. *Artificial agency (AA)*

Tale suggerimento ci giunge da Luciano Floridi, per il quale sarebbe più opportuno non esprimersi più nei termini di *artificial intelligence* (AI) ma di *artificial agency* (AA): «Solo se si capisce che l'AI è una nuova forma di capacità di agire, o di *agency* [...], e non una nuova forma di intelligenza, si può poi capire veramente la sua sfida etica e quindi affrontarla con successo»⁹.

Questa capacità di agire in vista di uno scopo, allora, richiede di essere ben compresa, soprattutto in raffronto alla medesima capacità detenuta dall'essere umano, dal momento che, per la prima volta

⁸ La modalità di approccio definita *riproduttiva* - dal momento che mira a riprodurre le capacità umane su artifici della tecnica - è stata percorso sin dagli albori delle AI ma si è dimostrata del tutto fallimentare per quanto riguarda l'intelligenza, tant'è che i maggiori progressi si sono riscontrati quando si è capito che le macchine non dovevano copiare i *Sapiens* o, per dirla con un'espressione ormai famosa: "L'uomo è riuscito a volare quando ha smesso di imitare gli uccelli". Quel paradigma, pertanto, è stato ormai abbandonato per ciò che attiene all'intelligenza ma non è possibile fare lo stesso per quanto concerne la capacità di deliberazione. Capire come gli esseri umani decidono risulta, infatti, cruciale, dal momento che, sino a non poco tempo fa, essi erano gli unici sul pianeta Terra capaci di decidere razionalmente ma, soprattutto, in grado di considerare, soppesare, attribuire valore a diversi corsi d'azione.

⁹ L. Floridi, *Agere sine intelligere. L'intelligenza artificiale come nuova forma di agire e i suoi problemi etici*, in L. Floridi, F. Cabitza (a cura di), *Intelligenza artificiale. L'uso delle nuove macchine*, in Bompiani, Milano 2021, p. 145.

nella storia del nostro pianeta, l'*agency* viene "scollata" dall'intelletto.

Secondo il docente di Oxford, infatti, «oggi l'AI non è il matrimonio tra ingegneria (artefatti) e biologia (intelligenza almeno animale, se non umana), ma il divorzio dell'agire (*agency*) dalla necessità di essere intelligenti per aver successo»¹⁰.

La cosiddetta AI, infatti, ha operato un divorzio tra *agere* e *intelligere*, tra *agency* artificiale e intelligenza naturale, il che è rivoluzionario:

Non è mai stato possibile prima d'ora avere successo nello svolgimento di un compito senza essere intelligenti almeno come un cane pastore. Abbiamo modificato - prosegue Floridi - una delle equazioni fondamentali su cui si è sempre basata la storia umana e la valutazione morale, quella che identifica l'agire con l'agire naturale, almeno biologico se non umano¹¹.

Desidererei che questo fosse il punto di partenza. Al giorno d'oggi algoritmi in grado di esercitare un certo qual grado di *agency* sono di dominio comune. Cito soltanto i più comuni. I navigatori satellitari - come *Waze*, *Google Maps* o *Mappe* di Apple - non sono altro che algoritmi che selezionano per l'utente due/tre possibili percorsi alternativi tra una miriade di altre opzioni, incrociando dati come il flusso del traffico, la distanza in metri dal luogo d'interesse, le eventuali chiusure stradali.

I *software* in uso presso alcuni istituti di credito per l'assegnazione di mutui sono anch'essi algoritmi "decisionali", dal momento che - sulla base di vari parametri fiscali e non - *decidono* se a quel determinato utente può essere concesso o meno il mutuo. Alcuni dei programmi utilizzati da diversi studi legali e società di consulenza, infine, sono algoritmi che sulla base di ripetizioni statiche di una parola o di tag preimpostati, sottopongono all'utente-umano alcuni dati e non altri.

Tutti questi casi ci debbono far riflettere sul dato che nel momento in cui ci avvaliamo di algoritmi per agevolare alcuni nostri processi decisionali, questi processi sono avvenuti in modo

¹⁰ *Ibid.*, pp. 150-151.

¹¹ *Ibid.*, p. 151.

diametralmente differente da come li avrebbe condotti un essere umano. Questo però - lungi dal destare allarmismo - deve farci interrogare, perché, di fatto, stiamo realizzando una nuova modalità d'interazione che sino al giorno d'oggi non si era mai verificata. Infatti, se al tavolo di una riunione di *team*, sino a qualche tempo fa avrebbero avuto posto a sedere un gruppo di delegati delle varie arie di un'azienda - ciascuno con la possibilità di dire le motivazioni per le quali ha preso una determinata decisione (decisioni che si suppone essere state intelligentemente ponderate in vista di un fine) - oggi a quel medesimo tavolo non sediamo soltanto noi ma anche algoritmi i quali - pur senza avere intelligenza - decidono.

3. Processi decisionali "a confronto"

Per capire come le AI prendono decisioni è necessario volgersi brevemente ai metodi utilizzati per addestrarle.

3.1. Come decidono Le AI

Uno dei modi più comuni per addestrare un'AI è quello di utilizzare l'apprendimento automatico, un processo che consente ai sistemi di imparare da dati-esempio senza essere esplicitamente programmati. Ci sono due tipi principali di apprendimento automatico: l'apprendimento supervisionato e quello non supervisionato¹².

L'apprendimento supervisionato utilizza dati etichettati, in cui ogni esempio è associato a una etichetta. Ad esempio, un'AI può essere addestrata a riconoscere immagini di gatti utilizzando un set di immagini di gatti etichettate come tali. Durante il processo di addestramento, l'AI "impara" a riconoscere le caratteristiche comuni delle immagini di gatti e le utilizza per classificare nuove immagini.

¹² Per un'introduzione, S. Russell, P. Norvig, *op. cit.*, 2, capp. 19-20.

L'apprendimento non supervisionato, invece, utilizza dati non etichettati e l'AI cerca di scoprire relazioni e *pattern* nel *dataset*¹³.

Una intelligenza artificiale decide utilizzando un insieme di algoritmi e di modelli matematici che sono stati addestrati su dati specifici. Per esempio, un'AI addestrata per riconoscere immagini utilizza un algoritmo di apprendimento automatico, come una rete neurale, per analizzare le caratteristiche di un'immagine e classificarla in una categoria predefinita. La decisione di classificare un'immagine in una determinata categoria viene quindi presa in base a una serie di calcoli matematici volti a confrontare i dati provenienti dall'immagine e quelli forniti in fase di addestramento¹⁴.

In generale, le decisioni prese da un'AI sono il risultato dell'elaborazione di grandi quantità di dati e dell'applicazione di algoritmi matematici sofisticati. Tuttavia, è importante notare che le decisioni prese da un'AI possono essere influenzate dai *bias* presenti nei dati di addestramento e dagli errori nell'implementazione dell'algoritmo.

Nel 2016, ad esempio, Microsoft ha introdotto su Twitter un'AI conversazionale di nome *Tay*. Gli utenti hanno iniziato a inviare a

¹³ Un esempio di questo tipo di apprendimento è l'utilizzo di un algoritmo di *clustering* per raggruppare i dati in base alle loro similitudini. Le AI che utilizzano l'apprendimento supervisionato sono spesso utilizzate per compiti di classificazione, in cui la risposta desiderata è una categoria predefinita. Ad esempio, un'AI può essere addestrata a riconoscere se un'immagine contiene un gatto o un cane. Le AI che utilizzano l'apprendimento non supervisionato sono utilizzate per compiti di clustering e rilevamento di anomalie. Ad esempio, un'AI può essere utilizzata per rilevare transazioni fraudolente in un dataset di transazioni finanziarie. Oltre all'apprendimento automatico, ci sono anche altri metodi utilizzati per addestrare le AI, come l'apprendimento per rinforzo e l'apprendimento evolutivo. L'apprendimento per rinforzo utilizza un sistema di premi e punizioni per incoraggiare l'AI a prendere decisioni corrette, mentre l'apprendimento evolutivo utilizza algoritmi di ottimizzazione per far evolvere una soluzione iniziale verso una soluzione migliore.

¹⁴ In altri casi, un'AI potrebbe utilizzare un algoritmo di apprendimento per rinforzo per prendere decisioni in un ambiente di gioco o di simulazione. In questo caso, l'AI esplora diverse azioni e riceve un premio o una punizione in base alla loro efficacia nel raggiungere un obiettivo predefinito. In base a queste informazioni, l'AI adatta la sua strategia per prendere decisioni più efficaci in futuro.

Tay ogni sorta di commenti misogini, razzisti e “Donald Trumpisti” e nel giro di ventiquattro ore Tay è diventato un teorico della cospirazione, misogino, razzista e xenofobo, e questo ha indotto Microsoft a staccare la spina¹⁵.

Un rapporto del 2019 ha rilevato, inoltre, che gli algoritmi di riconoscimento facciale, come quelli che sbloccano i telefoni cellulari o identificano le persone nelle foto caricate sui *social media*, sono intrinsecamente razzisti e sessisti. Questo avverrebbe perché la maggior parte dei programmatori sono maschi e bianchi e, di conseguenza, il set di dati usato per addestrare questi algoritmi era fortemente influenzato¹⁶. Il problema dei *bias* esiste anche in medicina. Schemi di valutazione medica come il SOFA (*Sequential Organ Failure Assess*), ad esempio, denotano forti pregiudizi razziali¹⁷

Questi esempi del mondo reale dimostrano una realtà delle AI: esse tendono a riflettere i pregiudizi dei dati che vengono loro forniti. È cruciale, pertanto, valutare attentamente la validità e l'affidabilità delle decisioni prese da un'AI prima di utilizzarla in contesti critici.

Tuttavia, prescindendo dai *bias*, alla base dell'approccio sinora descritto vi è un pensiero che è molto più antico dell'AI, un pensiero per il quale pensare equivale a calcolare. E a tal proposito già Thomas Hobbes si era espresso in questi termini:

Quando uno ragiona non fa altro che ottenere una somma totale attraverso una addizione di parti, o un resto sottraendo una somma da un'altra. [...] E infatti come l'aritmetica ci insegna a sommare e a sottrarre in termini di numeri, così [...] i logici insegnano le stesse cose nel campo della connessione fra le parole: sommando insieme due nomi si ha un'affermazione, sommando due affermazioni si ha un sillogismo, sommando alcuni sillogismi si ha una dimostrazione; e dalla somma, o conclusione di un sillogismo sottraggono una

¹⁵ Cfr. C.M. Klugman, *Black Boxes and Bias in AI Challenge Autonomy*, in «*The American Journal of Bioethics*», XXI, 2021, p. 34.

¹⁶ Cfr. P. Grother, M. Ngan, K. Hanaoka, *Face Recognition Vendor Test (FRVT). 3: Demographic effects*. National Institute of Standards and Technology, U.S. Department of Commerce, Washington D.C. 2019.

¹⁷ W.D. Miller, M.E. Peek, W.F. Parker, *Scarce Resource Allocation Scores Threaten to Exacerbate Racial Disparities in Health Care*, in «*Chest*», 158, 2020, pp. 1332-1334; D.A. Vyas, L.G. Eisenstein, D.S. Jones, *Hidden in Plain Sight. Reconsidering the Use of Race Correction in Clinical Algorithms*, in «*The New England Journal of Medicine*», 383, 2021, pp. 874-882.

proposizione per trovarne un'altra. [...] Insomma in qualunque campo in cui c'è posto per l'addizione e la sottrazione c'è anche posto per la ragione; dove queste cose mancano la ragione non ha niente da fare. [...] Poiché ragione in questo senso significa nient'altro che calcolo, cioè addizione e sottrazione¹⁸.

In effetti, un primo modo per definire come agisce un agente razionale è quello di ricorrere alle inferenze, questo però non basta perché molte altre azioni razionali vengono compiute senza che prima vi sia stata un'inferenza di qualche tipo. In un primo tempo l'AI - partendo dal presupposto che il pensiero si potesse ridurre a calcolo - si è concentrata sul trasporre algoritmicamente le regole della logica del primo ordine.

In quei casi l'*agere* delle AI era circoscritto alla forma *if-then*. Ci si rese ben presto conto però che la logica era poco utile, se non altro per il fatto che la stragrande maggioranza della popolazione mondiale quando agisce non ricorre di certo al sillogismo aristotelico.

Trovo una certa difficoltà a comprendere come le parole di Hobbes possano essere state così superficiali, dal momento che se è vero che ben si addicono quando parliamo seguendo la logica aristotelica (per la quale da premesse vere si ottengono conclusioni vere), di contro non si attagliano al linguaggio naturale. Questo palese iato appare non appena ci volgiamo a considerare la deliberazione umana. È proprio a quest'ultima che si rifanno i nuovi algoritmi decisionali basati su modello.

3.2. Algoritmi basati su modello e deliberazione umana

Gli agenti basati su modello, infatti, prima di scegliere (e quindi di agire) devono «prendere in considerazione il futuro sotto due aspetti: “cosa accadrà se faccio così e cosà?” e anche “se faccio questo sarò soddisfatto?”. La razionalità si esplicherà ulteriormente non già nell'avere obiettivi ma nel massimizzare l'utilità per raggiungere quel determinato obiettivo». Alla base

¹⁸ T. Hobbes, *Il Leviatano* (1651), tr. it. Utet, Torino 1955, pp. 75-76.

del processo decisionale di un'AI vi è quindi la teoria dell'utilità e quella della probabilità, in base alle quali «un agente [...] può scegliere razionalmente l'azione da intraprendere in base a ciò che crede e desidera»¹⁹. La combinazione delle due teorie permette all'agente di muoversi nel momento in cui le inferenze della logica del primo ordine (un agente puramente logico) non bastano. Infatti, quando l'agente artificiale si trova a dover agire senza che vi sia un obiettivo certamente utile, e quindi si trova in una condizione di incertezza se fare *a*, *non a* o *b*, allora viene introdotta la cosiddetta *funzione di utilità* che assegna un singolo numero per esprimere la desiderabilità di uno stato (in altre parole: viene assegnato un valore di utilità a ciascuna delle azioni che possono essere compiute). In tal caso l'agente realizzerà il principio della massima utilità attesa, ovvero sceglierà quell'azione che massimizza l'utilità attesa dell'agente. In questo *modus operandi*, secondo gli autori del noto manuale, può risiedere l'intelligenza, dal momento che «tutto ciò che un agente intelligente deve fare è calcolare le varie quantità, massimizzare l'utilità sulle sue azioni ed ecco tutto»²⁰.

Nelle parole di Russell e Norvig, però, si nota un ritorno a quel presupposto che essi stessi avevano bandito: il pensiero non è riducibile al calcolo. Per il padre del pragmatismo Charles Sender Peirce, ad esempio, «pensare non è calcolare. Si tratta piuttosto di compiere un balzo, formulando congetture»²¹; o, per dirla con un'immagine: se il calcolo porta a congiungere i puntini noti - come potrebbe avvenire in uno di quei giochini sulla *Settimana enigmistica*, l'analisi, invece, conduce a dare un senso a quei puntini²².

¹⁹ S. Russell, P. Norvig, *op. cit.*, p. 539.

²⁰ *Ibid.*, p. 540.

²¹ Qui in E.J. Larson, *The Myth of Artificial Intelligence*, Harvard University Press, Cambridge/MA 2022, p. 97.

²² Cfr. *Ibid.*, p. 96.

Cito, per far un esempio cogente, una pagina straordinaria di Edgar Allan Poe:

Le facoltà mentali che si sogliono chiamare analitiche sono, di per se stesse, poco suscettibili di analisi. Le conosciamo soltanto negli effetti. [...] Come l'uomo forte gode della sua potenza fisica e si compiace degli esercizi che mettono in azione i suoi muscoli, così l'analista si gloria di quella attività spirituale che serve a «risolvere». E trova piacere anche nelle occupazioni più comuni purché diano gioco al suo talento. [...] E i risultati, abilmente dedotti dalla stessa essenza e anima del suo metodo, hanno veramente tutta l'aria dell'intuito. [...] Il massimo potere della riflessione è più decisamente e utilmente provato dal modesto gioco della dama che non dalla complicata futilità degli scacchi. In quest'ultimo essendo i pezzi dotati di movimenti diversi e bizzarri e di valori diversi e variabili, quello che è soltanto complessità vien preso (errore abbastanza comune) per profondità²³.

Si badi però che le AI, a volte, possono “prendere decisioni” in modo del tutto randomico (si pensi all'aspirapolvere automatico o al GPS).

Una volta compreso, quindi, seppur per grandi linee, come decidono le AI, non ci rimane che comprendere cosa avviene qualora esse si sostituiscano a noi nell'esercizio della decisione o qualora esse interagiscano con noi in tale processo e per far questo indicherò, rispettivamente, questi possibili paradigmi con due neologismi: il paradigma *sostituivista* e quello *complementarista*.

4. Il paradigma della sostituibilità

Per comprendere il primo paradigma che qui suggerisco mi rifaccio alle parole del professor Paolo Moro, per il quale la domanda che oggi dovremmo porci consiste nel chiederci se un robot «dotato di intelligenza e di volontà artificiale del tutto analoghe, se non superiori, a quelle umane, possa essere titolare di libertà, diventando un centro di autodeterminazione e d'imputazione soggettiva di comportamenti interattivi»²⁴. Perché questo scenario si possa realizzare non è necessario raggiungere la cosiddetta *Artificiale General Intelligence*, basterebbe richiamare alla

²³ E.A. Poe, *Racconti*, traduzione di D. Cinelli e E. Vittorini, Mondadori, Milano 1961, pp. 4-5.

²⁴ P. Moro, *Libertà del robot? Sull'etica delle macchine intelligenti*, in *Filosofia del diritto e nuove tecnologie. Prospettive di ricerca tra teoria e pratica*, a cura di R. Brighi e S. Zullo, Aracne, Roma 2015, p. 526.

memoria alcuni casi ormai famosi in cui l'intelligenza artificiale ha scelto in base a dei dati in ingresso da un lato, e a un autoapprendimento (del quale essa stessa è capace) dall'altro, senza che si rendesse necessario il concorso dell'essere umano. Cito alcuni casi famosi.

Nel 2013 il tribunale del Wisconsin (episodio sul quale avrò modo di soffermarmi a breve) ha negato la cauzione all'afroamericano Eric Loomis sulla base di un report dell'algoritmo COMPAS; nel 2014 le decisioni prese dall'algoritmo in uso presso il settore *human resources* di Amazon si sono dimostrate palesemente influenzate da *bias* di genere nell'attribuzione di cariche apicali nel settore ingegneristico dell'azienda; nel 2019 l'algoritmo *Apple Credit Card* da poco brevettato dalla nota casa americana riconsegnava significative differenze nell'attribuzione del massimale di credito tra uomini e donne; nel 2021, infine, il tribunale di Bologna ha sancito la condanna dell'algoritmo di Glovo dal momento che discriminava le *performance* dei *rider* basandosi su fattori che non dipendevano dai *rider* stessi, attribuendo, di fatto, responsabilità decisionale all'algoritmo.

Questi sono tutti esempi in cui le AI, in modo del tutto autonomo, hanno operato delle decisioni, condizionate però da palesi *bias* cognitivi. Non sarebbe possibile addentrarmi in questa sede sulle ripercussioni che l'utilizzo generalizzato di questi algoritmi comporterebbe (e sui più recenti sviluppi dell'*algorithm fairness* o dell'*explainable artificial intelligence*); mi preme far notare, tuttavia, come questo primo paradigma richieda imprescindibili approfondimenti dal momento che le decisioni prese dalla AI hanno un impatto sulla vita dei singoli e della società. Adduco soltanto un esempio per segnalare un'indifferibile riflessione etica in tale ambito.

Si pensi alle conseguenze provocate da un algoritmo che sulla base di alcuni *bias* non concedesse il mutuo ad alcune famiglie in necessità. In tanti potrebbero obiettare che un tale errore potrebbe

essere compiuto anche da un consulente di banca; la differenza, tuttavia, risiede nel fatto che se a un consulente “umano” è possibile analizzare dalle tre alle dieci richieste in una giornata, un’AI, in pochi minuti, è in grado di effettuare un processo decisionale - sulla base di milioni di dati posseduti - riconsegnando un esito negativo a milioni di famiglie, che verrebbero gettate nella totale disperazione.

La soluzione più immediata, allora, anche in considerazione del fatto che siamo ancora lontani da un’intelligenza artificiale in grado di prendere decisioni cogenti, sarebbe quella di, cito il sito di IBM: «creare e offrire [...] una tecnologia affidabile, in grado di aumentare (non sostituire) il processo decisionale manuale»²⁵. Lo scenario prospettato è simile a quello che già si attua all’interno degli aeromobili, i quali potrebbero benissimo essere guidati esclusivamente dai sofisticatissimi computer di bordo installati ma, nondimeno, sono presenti due ufficiali in cabina a controllare²⁶.

Dato questo primo scenario, allora, è possibile abbandonare il paradigma della sostituibilità e rivolgersi al secondo modello che qui suggerisco come occasione di riflessione: quello della complementarità.

5. La complementarità tra umani e macchine nel processo decisionale

A questo paradigma è possibile ricondurre quella che è stata definita come *replacement technology*, ovvero quella tecnologia che *sostituisce* le fasi più pesanti o più tecniche di un lavoro (analisi della giurisprudenza, udienze *online*, piattaforme per la risoluzione delle controversie, soltanto per citare alcune

²⁵ XAI (*eXplainable AI - AI spiegabile*), in <https://www.ibm.com/it-it/watson/explainable-ai> (ultimo accesso: 3-3-2023).

²⁶ Sui recenti tentativi di modificazione della prassi a bordo degli aerei, che fanno leva proprio sull’efficienza dei sistemi di AI, cfr. L. Berberi, *Aerei, le compagnie chiedono un solo pilota a bordo per tagliare i costi*, in https://www.corriere.it/economia/aziende/22_novembre_25/aerei-compagnie-chiedono-solo-pilota-bordo-tagliare-costi-fe5571d4-6ca0-11ed-a41d-76ead3b90d6e.shtml (ultimo accesso: 3-3-2023).

applicazioni in ambito legale). Questi strumenti si collocano “a metà strada”: essendo né semplici supporti né sostitutivi dell’impegno umano ma, appunto, complementari.

Nel processo decisionale, nello specifico, essi «sostituiscono almeno un segmento del processo decisionale»²⁷ compiendo quella raccolta di dati previa alla decisione. In questo scenario «per sfruttare l’incredibile potenziale dell’Intelligenza Artificiale occorre una sapiente collaborazione tra persone e macchine, in cui – nella gran parte dei casi – *saranno sempre e comunque gli esseri umani ad avere l’ultima parola*»²⁸.

Questo approccio, che si rifà alla cosiddetta *AI debole* o *Narrow AI*, non sarebbe poi così preoccupante dal momento che esisterebbe sempre la presenza di un supervisore umano. Osservata da questa prospettiva la macchina non sarebbe nient’altro che una nuova e sofisticatissima leva d’Archimede alla quale l’essere umano demanda la parte più pesante e svilente del lavoro (al pari di un montacarichi o di uno strumento di catalogazione di testi), avocando a sé il controllo o comunque la parte più delicata del processo. Assumendo questo approccio *complementarista*, pertanto, molti errori potrebbero essere evitati. Per tornare agli esempi precedentemente adottati, infatti: il giudice si baserebbe sul *report* dell’algoritmo ma non ne sarebbe determinato; il *recruiter* potrebbe prendere in mano il *report* maschilista dell’algoritmo e reintrodurre comunque le quote rosa; un capo attento potrebbe valutare di caso in caso le *performance* dei *rider*, e così via.

Per rappresentarci meglio il paradigma *complementarista* nella relazione essere umano-AI, all’interno del processo decisionale, recupero le svariate applicazioni di geolocalizzazione già citate, e delle quali si fa ormai uso generalizzato attraverso gli

²⁷ E. Calzolaio, *Intelligenza artificiale ed autonomia della decisione: problemi e sfide*, *La decisione nel prisma dell’intelligenza artificiale*, a cura di E. Calzolaio, cit., p. 2.

²⁸ S. Maggioni, *AI & ETHICS - La sottile linea tra decisioni giuste e decisioni*, in https://www.sas.com/it_it/news/leading-art-innovation/whats-hot/ai-ethics-decisioni.html (ultimo accesso: 3-3-2023) [corsivo mio].

smartphone. *Google Maps*, *Maps* della Apple o *Waze* – solo per citare i più noti – sono tutti algoritmi che, nel momento in cui selezionano la strada per farci raggiungere la nostra meta tra decine o centinaia di percorsi possibili, “decidono per noi”. Proprio questo esempio così pregnante permette di far emergere quello che è l’approccio maggioritario nell’utilizzo dell’intelligenza artificiale: gli algoritmi sono degli *aiuti* all’azione umana.

Ma è proprio nella complementarità tra essere umano e algoritmo che si annida la vera sfida che, secondo me, richiede di essere tematizzata più di molte altre che si rifanno in certo qual modo alla cosiddetta algoretica: nel momento in cui la complementarità tra essere umano e macchina si sposta dal piano della forza motrice a quello della causalità formale, è ancora possibile parlare in termini di complementarità? Cerco di spiegarlo più chiaramente.

Se per quanto riguarda l’attività pratico-manuale (come il sollevare dei blocchi di cemento o il catalogare dei testi) è abbastanza semplice affidarsi a uno strumento meccanico o elettronico che compia il nostro stesso lavoro in minor tempo (e molto probabilmente con più accuratezza), si può dire lo stesso quando quello strumento entra a pieno titolo nel nostro processo decisionale?

Per comprendere la mia posizione è utile richiamare cosa è necessario perché un essere umano pervenga a una decisione deliberata, e quindi libera, e della quale pertanto può essere ritenuto responsabile. Risulta necessario che egli sia in grado di:

1. comprenderne le conseguenze attraverso un ragionamento controfattuale (requisito dell’*agency*),
2. raffigurarsi i diversi corsi di azioni disponibili (requisito delle *alternative possibilities*),
3. assumersi la piena responsabilità di quell’azione, dal momento che essa discende dalla sua volontà e non dalla volontà altrui, come nel caso di una scelta coartata (requisito del *controllo*).

I giuristi sanno bene come qualora mancasse uno di questi tre elementi l’azione non sarebbe pienamente imputabile al soggetto:

come nel caso di demenza, o ignoranza, o qualora ci si trovasse durante stati attenuati di coscienza. In contesti usuali – mancando particolari patologie neurologiche o stati di coscienza alterata – gli esseri agenti hanno tutti e tre i requisiti e sulla base di essi possono operare delle scelte²⁹. A quanti, tuttavia, non è mai capitato di compiere una scelta, a seguito di lunga ponderazione, e poi pentirsene perché si era fatta una valutazione dei rischi e delle conseguenze non pienamente in linea e non pienamente soddisfacente? Questo scenario, se è all'ordine del giorno per gli esseri umani, non lo è per le intelligenze artificiali che hanno raggiunto una capacità di previsione strabiliante, almeno in alcuni settori specifici. Per far comprendere ancora meglio il suggerimento che qui invito ad assumere, riporto un caso che saltò subito agli onori della cronaca.

6. Interazione 'dentro' la decisione

Sofferinarsi in questa sede sul processo decisionale degli esseri umani sarebbe impossibile. Rifacendomi, tuttavia, a due casi che attengono alla giurisprudenza, cercherò di addurre alcune riflessioni conclusive.

Nel 2013 i giornali americani portarono alla ribalta il caso di Eric Loomis, condannato da un giudice di un tribunale locale del Wisconsin, il quale – per l'emissione del suo verdetto – aveva fatto ricorso a un rapporto, contenente la storia dell'indagato (precedenti penali, contesto familiare e sociale, ecc.), all'interno del quale era presente anche una valutazione operata dall'algoritmo COMPAS³⁰ che catalogava l'imputato afroamericano come

²⁹ Per un'approfondita discussione sul tema, mi sia permesso rimandare a C. Calì, *Il libero arbitrio in questione. Una ricerca tra filosofia, scienze e intelligenza artificiale*, Mimesis, Milano 2023, in particolare il cap. 1.

³⁰ Sebbene l'algoritmo di COMPAS sia rimasto segreto, in quanto prodotto da una società privata e tutelato dalle leggi sul diritto d'impresa, è possibile dire su quali dati esso si basasse per giungere al proprio verdetto: precedenti penali, utilizzo pregresso di sostanze stupefacenti, e un insieme di altri dati che lo stesso indagato forniva all'algoritmo mediante un test di 137 domande, volte a indagare la personalità e la storia dell'individuo. Questi dati erano quindi combinati con profili e circostanze similari. Per un approfondimento sulla

ad “alto rischio di recidiva”. Nonostante i due gradi di giudizio ai quali fece ricorso Loomis, nel 2016 la *Corte suprema del Wisconsin* rigettava il ricorso dell'imputato e definiva che il verdetto del primo tribunale non poteva essere considerato invalido o lesivo del diritto al giusto processo perché il *report* dell'algoritmo aveva costituito soltanto un “ausilio” al giudice umano, ma non era stato in alcun modo decisivo ai fini della sentenza definitiva³¹.

Nel momento in cui faccio emergere le criticità di quello che ho definito come *approccio complementarista* è doveroso, tuttavia, far notare come al giorno d'oggi, epoca in cui si parla del dominio dei dati, risulti del tutto impossibile a un agente umano gestire una quantità tale di dati in un tempo conveniente. L'analisi dei dati a nostra disposizione, infatti, costituisce – come già ho avuto modo di accennare – un momento significativo del processo decisionale. Ora è proprio quest'ultimo processo che sembra non potersi più condurre senza l'ausilio di algoritmi specifici. Una tale situazione richiede pertanto una seria riflessione se non, addirittura, un cambiamento del nostro modo di intendere e di gestire il processo decisionale³².

Tornando al caso di Eric Loomis, non mi è possibile entrare qui nella miriade di complessità che questa sentenza porta con sé. È doveroso sottolineare però che nel maggio 2016 (prima della condanna definitiva) in un'inchiesta condotta da *ProPublica* si asseriva, sulla base di ricerche statistiche, che l'algoritmo COMPAS

vicenda e i suoi risvolti etici cfr. C. Calì., *L'imparzialità del giudice. Alcune implicazioni etiche dell'utilizzo dell'intelligenza artificiale in giurisprudenza*, in *La pubblica amministrazione del futuro. Tra sfide e opportunità per l'innovazione del settore pubblico*, a cura di A. Alù, A. Ciccarello, Editoriale Scientifica, Napoli 2021, pp. 121-134

³¹ Cfr. *State of Wisconsin v. Loomis*, cit., 4. 7-8.

³² Per la discussione se le cosiddette *alternative possibilities* amplino o riducano la nostra capacità decisionale, mi sia permesso rimandare a C. Calì, *Algoritmi e processo decisionale. Alle origini della riflessione etico-pratica per le IA*, in «Scienza&Filosofia», 27, 2022, pp. 69-87 e Id., *Il libero arbitrio in questione. Una ricerca tra filosofia, scienze, intelligenza artificiale*, cit., pp. 53-65.

presentava forti pregiudizi nell'attribuire agli afroamericani un tasso di rischio sempre più elevato (*bias* razziale confermato da un'ulteriore inchiesta sul medesimo algoritmo condotta dal Dartmouth College)³³.

Si vede quindi come i giudici elettronici non siano perfetti. Non sembra migliore la situazione con i giudici umani.

Nel 2011 lo studioso israeliano Shai Danziger, insieme ad alcuni colleghi, ha deciso di valutare come i giudici pervenissero a un determinato verdetto: se si limitassero ad applicare argomenti giuridici ai fatti del caso concreto in modo razionale, meccanico e deliberativo (formalismo giuridico) o se il processo deliberativo fosse influenzato da una serie di fattori del tutto esterni alla legge ma interni alla mente del giudicante (realismo giuridico). Attraverso uno studio sperimentale - i cui risultati sono confluiti in un articolo pubblicato sulla rivista scientifica *PNAS* della National Academy of Sciences statunitense - sono state osservate le decisioni dei giudici nell'arco temporale di una giornata lavorativa (in cui i giudici fanno due *break*). L'articolo ha rilevato che:

La percentuale di decisioni favorevoli al reo scendeva gradualmente dal 65% a quasi zero all'interno di ciascun segmento della giornata [in misura

³³ «Was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants. White defendants were mislabeled as low risk more often than black defendants» J. Angwin, J. Larson, S. Mattu, L. Kirchner, *Machine Bias. There's Software Used Across the Country to Predict Future Criminals. And It's Biased against Blacks*, ProPublica, 23 maggio 2016 www.propublica.org. All'inchiesta è seguito un acceso dibattito, al quale hanno partecipato sia la società produttrice di COMPAS con uno studio [cfr. W. Dieterich, C. Mendoza, T. Brennan, *COMPAS Risks Scales: Demonstrating Accuracy Equity and Predictive Parity*, 2016, go.volarisgroup.com] sia un team di ricercatori [cfr. A. Flores, K. Bechtel, C. Lowenkamp, *False Positives, False Negatives, and False Analyses: A Rejoinder to 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks'*, in «Federal probation», 80, 2016, pp. 38-46; J. Jung, C. Concannon, R. Shroff, S. Goel, D.G. Goldstein, *Simple Rules for Complex Decisions*», in «arXiv», 1702, 2017, p. 04690]. A conclusioni simili a quelle edite da ProPublica nel 2016 sono giunti invece, nel 2018, due ricercatori del Dartmouth college, Julia Dressel e Hany Farid, i quali hanno condotto uno studio per mostrare che nel valutare la potenziale recidività di un individuo, COMPAS non è più affidabile di un gruppo di volontari scelti a caso su internet. Cfr. J. Dressel, H. Farid, *The Accuracy, Fairness and Limits of Predicting Recidivism*, in «Science Advances», IV, 2018, p. eaao5580.

direttamente proporzionale a quanto ci si allontanava dalla colazione del giudice]; subito dopo una pausa, la percentuale risaliva bruscamente a 65%³⁴.

In altre parole, il verdetto dipendeva «da quello che il giudice aveva mangiato a colazione»³⁵.

L'apporto di questi due casi in un contesto così delicato com'è quello del giudizio in tribunale, ci consente di comprendere come la vera sfida per la ricerca scientifica e per i decisori politici non sia tanto quella di normare le “macchine intelligenti” o di fornire strumenti etici ai programmatori delle stesse, quanto piuttosto quella di formare le coscienze affinché forse realizzando il sogno spinoziano in merito alla libertà – si prenda atto di alcuni dei meccanismi messi in atto dagli algoritmi.

Un tale compito è legato al fatto, già concreto e non futuro, che l'intromissione delle AI nei processi decisionali cogenti sta già operando un ri-ontologizzazione non del mondo attorno a noi ma di quella che sinora è stata la dimensione più nostra: la capacità cognitiva di prendere azioni libere informate dalla ragione. Questa a mio modo di vedere è la sfida per quanti a diverso titolo progettano gli algoritmi di AI – ingegneri, programmatori, filosofi – ma, ancor di più, per tutti noi che ogni giorno ci avvaliamo di alcuni sistemi – dai *cookies* del *browser* alle *newsletter* per la profilazione del cliente ai navigatori satellitari – per compiere le nostre azioni e informare le nostre decisioni, dalla più banale alla più significativa.

CRISTIANO CALÌ è Assegnista di Ricerca in Filosofia Morale presso l'Università di Torino e Docente a Contratto di Filosofia della Medicina presso L'Università di Roma “La Sapienza”

cristiano.cali@unito.it

³⁴ S. Danziger, J. Levav, L. Avnaim-Pessoa, *Extraneous factors in judicial decisions*, in «PNAS Proceedings of the National Academy of Sciences of the United States of America», 108, 2011, pp. 6889-6892 [traduzione mia].

³⁵ «What the judge ate for breakfast», *Ibid.*, p. 6889.

NORME REDAZIONALI

I testi vanno inviati esclusivamente via email a
redazione@scienzae filosofia.it
in formato Word con le seguenti modalità:

Abstract in inglese (max. 200 parole)

Testo
Carattere: Calibri o Times o Times New Roman
Corpo: 12 Interlinea: 1,5

Le note vanno inserite a fine testo con:
Carattere: Calibri o Times o Times New Roman
Corpo: 10 Interlinea: singola

Per favorire la fruibilità telematica della rivista, i contributi devono aggirarsi tra le 15.000 - 20.000 battute, tranne rare eccezioni, e gli articoli vanno sempre divisi per paragrafi. Anche le note devono essere essenziali, limitate all'indicazione dei riferimenti della citazione e/o del riferimento bibliografico e non dovrebbero contenere argomentazioni o ulteriori approfondimenti critici rispetto al testo. È indispensabile un abstract in lingua inglese (max. 200 parole). A esclusione delle figure connesse e parti integranti di un articolo, le immagini che accompagnano i singoli articoli sono selezionate secondo il gusto (e il capriccio) della Redazione e non pretendono, almeno nell'intenzione - per l'inconscio ci stiamo attrezzando - alcun rinvio didascalico.

Note

Norme generali

a) *Autore*: nome puntato e cognome in Maiuscolo/minuscolo tondo seguito da una virgola. Se si tratta di due o più autori, citarli tutti di seguito inframmezzati da virgole o trattino. Evitare l'uso di Aa.Vv. e inserire il curatore o i curatori come Autori seguito da "(a cura di)"

b) *Titolo*: Maiuscolo/minuscolo corsivo sempre, seguito da virgola.

c) *Editore*: occorre inserire la Casa Editrice.

d) *Città e data*: Maiuscolo/minuscolo tondo, non inframezzate da virgola. Le città straniere vanno in lingua originale.

e) *L'anno di edizione*. Nel caso in cui non si cita dalla prima edizione a stampa, occorre specificare l'edizione con un apice.

Esempio:

¹ G. Agamben, *L'aperto. L'uomo e l'animale*, Bollati Boringhieri, Torino 2002.

² A. Caronia, *IL Cyborg. Saggio sull'uomo artificiale* (1984), Shake, Milano 2008.

³ E. Morin, *IL paradigma perduto. Che cos'è la natura umana?* (1973), tr. it. Feltrinelli, Milano 2001.

⁴ G. Hottois, *Species Technica*, Vrin, Paris 2002.

⁵ P. Amodio, R. De Maio, G. Lissa (a cura di), *La Sho'ah tra interpretazione e memoria*, Vivarium, Napoli 1998.

⁶ G. Macchia, *IL paradiso della ragione*, Laterza, Roma-Bari 1961², p. 12. ["2" sta per seconda edizione].

Nel caso in cui si tratti di uno scritto già precedentemente citato, le indicazioni circa l'opera possono essere abbreviate con le seguenti diciture: "cit." (in tondo), "op. cit." (in corsivo), "ibid." o "Ibid." (in corsivo).

Dopo la prima citazione per esteso si accetta il richiamo abbreviato costituito da: Autore, Prime parole del titolo seguite da puntini di sospensione e dall'indicazione "cit." (invariata anche nel caso di articoli di riviste).

Esempio:

¹² A. Caronia, *IL Cyborg...*, cit.

Casi in cui si usa "cit.":

Quando si tratta di opera citata in precedenza ma non nella Nota immediatamente precedente (per quest'ultimo caso si veda più avanti).

Esempio:

¹ E. Morin, *IL paradigma perduto. Che cos'è la natura umana?*, cit.

- Casi in cui si usa "op. cit." (in corsivo):

Quando si tratta di un Autore di cui fino a quel punto si è citata un'unica opera.

Esempio:

¹ B. Croce, *Discorsi di varia filosofia*, Laterza, Roma-Bari 1942, pp. 232- 233.

² G. Hottois, *Species Technica*, Vrin, Paris 2002.

³ B. Croce, *op. cit.*, p. 230. [Il riferimento è qui chiaramente a *Discorsi di varia filosofia*, poiché nessun'altra opera di Croce era stata precedentemente citata].

Nel caso in cui, invece, siano già state citate due o più opere dello stesso Autore, o nel caso in cui in seguito si citeranno altre opere dello stesso autore, *op. cit.* va usato solo la prima volta, poi si utilizzerà "cit."

Esempio:

¹ B. Croce, *Discorsi di varia filosofia*, Laterza, Roma-Bari 1942, pp. 232- 233.

² G. Hottot, *Species Technica*, Vrin, Paris 2002.

³ B. Croce, *op. cit.*, p. 230.

⁴ Id., *Saggio sullo Hegel*, Laterza, Roma-Bari 1913, p. 44.

⁵ P. Piovani, *Conoscenza storica e coscienza morale*, Morano, Napoli 1966, p. 120.

[Se a questo punto si dovesse citare nuovamente B. Croce, *Discorsi di varia filosofia*, per non creare confusione con *Saggio sullo Hegel*, si è costretti a ripetere almeno il titolo seguito da "cit."; la Nota "6" sarà dunque]:

⁶ B. Croce, *Discorsi di varia filosofia*, cit., pp. 234-235.

In sostanza, "op. cit." sostituisce il titolo dell'opera (è questo il motivo per cui va in corsivo) e comprende anche le indicazioni tipografiche; cit. sostituisce solo le indicazioni tipografiche (è questo il motivo per cui non va mai in corsivo).

- Casi in cui si usa "ibid." o "Ibid." (in corsivo):

a) Quando si tratta di un riferimento identico alla Nota precedente.

Esempio:

¹ B. Croce, *Discorsi di varia filosofia*, Laterza, Roma-Bari, 1942, pp. 232- 233.

² *Ibid.* [Ciò significa che ci riferisce ancora una volta a B. Croce, *Discorsi di varia filosofia*, Laterza, Roma-Bari 1942, pp. 232- 233].

[N.B.: *Ibid.* vale anche quando si tratta della stessa opera, ma il riferimento è ad altra pagina e/o volume o tomo (che vanno specificati)]:

³ *Ibid.*, p. 240.

⁴ *Ibid.*, vol. I, p. 12.

b) Quando ci si riferisce a uno scritto diverso, ma dello stesso autore (ad esempio nelle raccolte moderne di opere classiche. In tal caso, inoltre, la data della prima pubblicazione va tra parentesi).

Esempio:

¹ F. Galiani, *Della moneta (1750)*, in Id., *Opere*, a cura di F. Diaz e L. Guerci, in *ILLuministi italiani*, Ricciardi, Milano-Napoli 1975, t. VI, pp. 1-314.

² Id., *Dialogues sur Le commerce des bleds (1770)*, *ibid.*, pp. 345-612. [*ibid.* in tal caso sostituisce: F. Galiani, *Opere*, a cura di F. Diaz e L. Guerci, in *ILLuministi italiani*, Ricciardi, Milano-Napoli 1975, t. VI].

c) Quando ci si riferisce a uno scritto contenuto in opera generale (l'esempio classico sono i volumi collettanei) citata nella Nota immediatamente precedente:

Esempio:

¹ G. Spini, *Alcuni appunti sui libertini italiani*, in *IL Libertinismo in Europa*, a cura di S. Bertelli, Ricciardi, Milano-Napoli 1980, pp. 117-124.

² P. Rossi, *Discussioni sulle tesi libertine su linguaggio e barbarie*, *ibid.*, pp. 319-350. [*ibid.* in tal caso sostituisce: *IL Libertinismo in Europa*, a cura di S. Bertelli, Ricciardi, Milano-Napoli 1980].

Tutte queste indicazioni valgono non solo quando si tratta di Note diverse, ma anche quando, nella stessa Nota, si cita più di un'opera.

Esempio:

¹ Cfr. G. Spini, *Alcuni appunti sui libertini italiani*, in *IL Libertinismo in Europa*, a cura di S. Bertelli, Milano-Napoli, 1980, pp. 117-124; ma si veda anche P. Rossi, *Discussioni sulle tesi libertine su linguaggio e barbarie*, *ibid.*, pp. 319-350.

Nel caso in cui si tratta dell'edizione moderna di un classico, è indispensabile specificare tra parentesi l'anno di pubblicazione e quindi il curatore, in particolare se si tratta di edizioni critiche.

Esempio:

¹ G. Galilei, *Dialogo sopra i due massimi sistemi (1632)*, a cura di L. Sosio, Einaudi, Torino 1970, pp. 34-35.

Opere in traduzione

Quando si cita dalle traduzioni è consentito omettere il titolo originale, ma occorre sempre specificare la data dell'edizione originale tra parentesi, e l'editore della traduzione preceduto dall'abbreviazione "tr. it.", "tr. fr." ecc.

Esempio:

¹ M. Heidegger, *Essere e tempo (1927)*, tr. it. Utet, Torino 1969, p. 124.

² Id., *Les problèmes fondamentaux de la phénoménologie (1927)*, tr. fr. Gallimard, Paris 1985.

Articoli di riviste

La citazione completa è così composta:

Autore, Titolo del saggio, indicazione "in" seguita dal titolo della rivista tra *virgolette basse*, annata in numeri romani, numero del fascicolo in numeri arabi (sempre preferito all'indicazione del mese), numeri delle pagine.

Esempio:

¹ D. Ferin, *Profilo di Tranquillo Marangoni*, in «Grafica d'arte», XV, 57, 2004, pp. 22-25

Citazioni

Le citazioni nel testo possono essere introdotte in due modi:

- 1) se si tratta di brani molto lunghi o di particolare rilevanza possono essere trascritti con corpo più piccolo rispetto al resto del testo, preceduti e seguiti da una riga vuota e senza virgolette.
- 2) se si tratta di citazioni più brevi o interrotte e spezzettate da interventi del redattore dell'articolo vanno messe nel corpo del testo principale, introdotte da caporali: «XXXXXX»
Nel caso 2) un'eventuale citazione nelle citazione va posta tra virgolette inglesi semplici: «Xxxx "XXXXXX"»

Segno di nota al termine di una citazione

Quando la citazione rimanda a una nota, il richiamo di nota deve venire subito dopo l'ultima parola nel caso 1, subito dopo le virgolette nel caso 2: solo dopo va introdotto il segno di punteggiatura che conclude la frase.

Esempio:

«Conobbi il tremolar della marina»².

Congiunzioni (“d” eufonica)

Si preferisce limitare l'uso della “d” eufonica ai soli casi in cui essa serva a staccare due vocali uguali.

Esempio:

“a essi” e non “ad essi”; “ad anticipare” e non “a anticipare”.

È consentito “ad esempio”, ma: “a esempio”, in frasi del tipo “venire citato a esempio”.

Bibliografie

Evitare le bibliografie, i testi di riferimento vanno in nota.

Avvertenza sulle note

Sempre per garantire una più immediata fruibilità di lettura, le note devono essere essenziali e non introdurre nuovi elementi di analisi critica. Questi ultimi vanno solo ed esclusivamente nel testo.

Titoli e Paragrafi

Sempre per garantire una più immediata fruibilità di lettura, gli articoli vanno titolati e suddivisi in paragrafi. Qualora l'autore non provvedesse, il redattore che cura l'editing dell'articolo è tenuto a dare il titolo all'articolo e a suddividere l'articolo in diversi e brevi paragrafi.

S&F_scienzaefilosofia.it

ISSN 2036 _ 2927

www.scienzaefilosofia.it