

Raffaello Cortina Editore

Luciano Floridi

Etica dell'intelligenza artificiale

Sviluppi, opportunità, sfide



SCIENZA
E IDEE

Collana fondata
da Giulio Giorello

Luciano Floridi

Etica dell'intelligenza artificiale

Sviluppi, opportunità, sfide

Edizione italiana a cura di
Massimo Durante



Raffaello Cortina Editore

www.raffaellocortina.it

Titolo originale
*The Ethics of Artificial Intelligence.
Principles, Challenges, and Opportunities*
© Luciano Floridi, 2022
International Rights Management:
Susanna Lea Associates
(la traduzione è una versione ridotta che comprende:
Preface, Chapters 1-9, 15-16, 18-20)

Traduzione
Massimo Durante

Copertina
Studio CReE

ISBN 978-88-3285-409-1
© 2022 Raffaello Cortina Editore
Milano, via Rossini 4

Prima edizione: 2022

Stampato da
Geca SRL, San Giuliano Milanese (MI)
per conto di Raffaello Cortina Editore

Ristampe

| | | | | | |
|------|------|------|------|------|---|
| 2 | 3 | 4 | 5 | 6 | 7 |
| 2023 | 2024 | 2025 | 2026 | 2027 | |

1.1 INTRODUZIONE: LA RIVOLUZIONE DIGITALE E L'INTELLIGENZA ARTIFICIALE

Nel 1964, la Paramount Pictures distribuì *Robinson Crusoe su Marte*. Il film descriveva le avventure del comandante Christopher “Kit” Draper (Paul Mantee), un astronauta statunitense naufragato su Marte. Se lo guardiamo su YouTube anche solo per pochi minuti, ci rendiamo conto di quanto radicalmente sia cambiato il mondo in pochi decenni. In particolare, il computer che compare all’inizio del film sembra un motore di epoca vittoriana, con leve, ingranaggi e quadranti. Un pezzo di archeologia che avrebbe potuto usare il dottor Frankenstein. Eppure, verso la fine della storia, Friday (Victor Lundin) viene rintracciato da un’astronave aliena attraverso i suoi braccialetti. Un pezzo di futurologia che sembra inquietantemente preveggen- te.

Robinson Crusoe su Marte apparteneva a un’epoca diversa, tecnologicamente e culturalmente più vicina al secolo scorso che al nostro. Descrive una realtà *moderna*, non *contemporanea*, basata sull’*hardware* e non sul *software*. Computer portatili, Internet, servizi web, touch screen, smartphone, orologi intelligenti, social media, shopping online, video e musica in streaming, automobili a guida autonoma, tosaerba robotizzati o assistenti virtuali non esistono ancora. L’intelligenza artificiale è più un progetto che una realtà. Il film mostra una tecnologia fatta di dadi, bulloni e meccanismi che seguono le goffe leggi della fisica newtoniana. È una realtà del tutto analogica, basata sugli atomi piuttosto che sui byte, di cui i Millennials non hanno esperienza, essendo nati dopo i primi anni Ottanta. Per loro un mondo senza tecnologie digitali assomiglia a ciò che era per me (nato nel 1964) un mondo senza automobili: qualcosa di cui ho sentito parlare solo da mia nonna.

Si fa spesso osservare che uno smartphone racchiude molta più potenza di calcolo in pochissimo spazio di quanto la NASA potesse mettere insieme quando Armstrong atterrò sulla Luna cinque anni dopo *Robinson Crusoe su Marte*, nel 1969, e tutto questo a un costo pressoché trascurabile. Molti articoli hanno tracciato alcuni raffronti precisi nel 2019, per il cinquantesi-

mo anniversario dello sbarco sulla Luna, facendo emergere alcuni fatti sorprendenti. L'Apollo Guidance Computer (AGC) a bordo dell'*Apollo 11* aveva 32.768 bit di RAM (*random access memory*) e 589.824 bit (72 KB) di ROM (*read only memory*): uno spazio di memoria su cui non saremmo stati in grado di salvare questo libro. Cinquant'anni dopo, un cellulare possiede, in media, 4 GB di RAM e 512 GB di ROM: vale a dire, un milione di volte in più di RAM e sette milioni di volte in più di ROM. Per quanto riguarda il processore, l'AGC funzionava a 0,043 MHz. Un processore per iPhone funziona, in media, a circa 2490 MHz: vuol dire che è circa 58.000 volte più veloce. Per afferrare meglio il senso di questa accelerazione, forse un paragone può esserci di aiuto. In media, una persona cammina alla velocità di 5 km/h. Oggi, alcuni jet supersonici possono viaggiare alla velocità di 6100 km/h, vale a dire a una velocità più di cinque volte superiore a quella del suono (1235 km/h). Eppure è soltanto poco più di mille volte più veloce che camminare. Immaginiamo di moltiplicarlo per 58.000.

Dove sono andate a finire tutta questa velocità e questa potenza di calcolo? La risposta è duplice: *fattibilità* e *usabilità*. Possiamo fare sempre di più, in termini di applicazioni, e possiamo farlo in modi sempre più semplici, non solo per ciò che concerne la programmazione, ma soprattutto per ciò che riguarda l'esperienza dell'utente. I video, per esempio, sono avidi di potenza di calcolo, così come i sistemi operativi. L'IA oggi è possibile anche perché abbiamo la potenza di calcolo necessaria per far funzionare il suo software.

Grazie a questa crescita sbalorditiva delle capacità di archiviazione ed elaborazione, a costi sempre più contenuti, oggi miliardi di persone sono connessi e trascorrono molte ore online ogni giorno. Nel Regno Unito, per esempio, "nel 2018, il tempo medio trascorso utilizzando Internet è stato di 25,3 ore a settimana. Si è trattato di un incremento di 15,4 ore rispetto al 2005".¹ Ciò è tutt'altro che insolito. E la pandemia ha generato

1. Fonte: <https://www.statista.com/statistics/300201/hours-of-internet-use-per-week-per-person-in-the-uk/>.

un incremento ulteriore. Tornerò su questo punto nel prossimo capitolo, ma un altro dei motivi per cui l'IA è possibile oggi è proprio il fatto che noi esseri umani trascorriamo sempre più tempo in contesti digitali e, pertanto, adattati all'IA.

Più memoria, più velocità e più ambienti e interazioni digitali hanno generato più dati, in quantità immense. Tutti abbiamo osservato diagrammi con curve esponenziali, che indicavano quantità che non sappiamo neppure come rappresentare. Secondo la società di analisi di mercato IDC,² nel 2018 abbiamo raggiunto 18 zettabyte di dati creati, catturati o riprodotti, e questa sorprendente crescita di dati non mostra segni di rallentamento: a quanto pare, diventeranno 175 zettabyte nel 2025. Si tratta di un aspetto difficile da cogliere in termini di quantità, ma due conseguenze meritano un momento di riflessione (sono entrambe esaminate in Floridi, 2014a). La velocità e la memoria delle nostre tecnologie digitali non crescono alla stessa velocità dell'universo dei dati. Ne consegue che stiamo rapidamente passando da una cultura della registrazione a una della cancellazione: la questione non è più che cosa salvare ma che cosa eliminare per fare spazio ai nuovi dati. La maggior parte dei dati disponibili è stata creata a partire dagli anni Novanta, anche se contiamo ogni parola pronunciata, scritta o stampata nella storia dell'umanità e ogni biblioteca o archivio mai esistiti. Per rendersene conto, basta osservare uno dei tanti diagrammi disponibili online che illustrano l'esplosione dei dati: l'aspetto sorprendente non sta solo sulla parte destra, dove la freccia della crescita sale, ma anche sulla parte sinistra, dove la crescita ha avuto inizio: si tratta solo di una manciata di anni fa. Poiché tutti questi dati sono stati creati dalla generazione attuale, stanno anche invecchiando insieme, in termini di supporti e tecnologie obsolete. Per questo motivo la loro conservazione costituirà una questione sempre più urgente.

Una maggiore potenza di calcolo e una maggiore quantità di dati hanno reso possibile il passaggio dalla logica alla statistica. Le reti neurali che erano interessanti solo da un punto di

2. Fonte: <https://www.seagate.com/gb/en/our-story/data-age-2025/>.

vista teorico (ne ho discusse alcune in un libro precedente già alla fine degli anni novanta: Floridi, 1999) sono diventate strumenti ordinari nell'ambito dell'apprendimento automatico. La vecchia IA era per lo più simbolica e poteva essere interpretata come una branca della logica matematica, ma la nuova IA è principalmente connessionista e potrebbe essere interpretata come una branca della statistica. Il principale cavallo di battaglia dell'IA non è più la deduzione logica ma l'inferenza e la correlazione statistica.

La potenza e la velocità di calcolo, le dimensioni della memoria, la quantità di dati, i potenti algoritmi, gli strumenti statistici e le interazioni online sono tutti fattori che stanno crescendo in modo incredibilmente rapido. Ciò accade anche perché (e la relazione causale procede in entrambe le direzioni) il numero di dispositivi digitali che interagiscono tra loro è già notevolmente superiore alla popolazione umana. Perciò, la maggior parte delle comunicazioni avviene da macchina a macchina, senza coinvolgimento umano. Ci sono robot computerizzati su Marte controllati a distanza dalla Terra. Il comandante Christopher "Kit" Draper li avrebbe trovati assolutamente fantastici.

Tutte le tendenze precedenti continueranno a crescere, inarrestabilmente, nel prossimo futuro. Queste tendenze hanno modificato il modo in cui impariamo, giochiamo, lavoriamo, amiamo, odiamo, scegliamo, decidiamo, produciamo, vendiamo, compriamo, consumiamo, pubblicizziamo, ci divertiamo, ci preoccupiamo di qualcosa e ce ne prendiamo cura, socializziamo, comunichiamo e così via. Sembra impossibile trovare un qualsiasi aspetto della nostra vita che non sia stato influenzato dalla rivoluzione digitale. Nell'ultimo mezzo secolo circa, la nostra realtà è diventata sempre più digitale, fatta di zero e uno, gestita da software e dati, piuttosto che da hardware e atomi. Un numero crescente di persone vive sempre più diffusamente *onlife*, sia online sia offline, e nell'*infosfera*, sia digitalmente sia analogicamente. Questa rivoluzione digitale ha anche influito sul modo in cui concepiamo e comprendiamo le nostre realtà, che sono sempre più interpretate in termini computazionali e digitali. Basti pensare alla "vec-

chia” analogia tra il nostro DNA e il nostro “codice”, che ora diamo per scontata. Tutto ciò ha anche alimentato lo sviluppo dell’IA, dal momento che condividiamo le nostre esperienze onlife e gli ambiti dell’infosfera con agenti artificiali, siano essi algoritmi, bot o robot. Per capire in che cosa effettivamente consista l’IA – la mia tesi è che si tratti di una nuova forma di agire, e non di intelligenza – occorre dunque dire qualcosa di più sull’impatto della rivoluzione digitale: è quanto mi propongo di fare nella parte restante di questo capitolo. È solo comprendendo la traiettoria concettuale delle sue implicazioni che possiamo avere una corretta prospettiva sulla natura dell’IA (secondo capitolo), i suoi probabili sviluppi (terzo capitolo) e le sue sfide etiche (seconda parte).

1.2 IL POTERE DI SCISSIONE DEL DIGITALE: TAGLIARE E INCOLLARE LA MODERNITÀ

Le tecnologie, le scienze, le pratiche, i prodotti e i servizi digitali, in breve il *digitale* come fenomeno globale sta profondamente trasformando la realtà. Tutto questo è piuttosto ovvio e pacifico. Le vere domande consistono casomai nel chiedersi *perché, come e con quali conseguenze*, soprattutto in relazione all’IA. In ciascuno di questi casi, la risposta è tutt’altro che banale e sicuramente aperta al dibattito. Per chiarire quali siano le risposte che ritengo più convincenti e introdurre nel prossimo capitolo un’interpretazione dell’IA come una *riserva* crescente di capacità di *agire intelligente*, conviene entrare *in medias res* e partire dal “come”. Diventerà allora più facile fare un passo indietro, per comprendere il “perché”, e poi procedere in avanti, per affrontare le “conseguenze” e collegare le risposte all’emergere dell’IA.

Il digitale “taglia e incolla” le nostre realtà sia ontologicamente sia epistemologicamente. Con questo intendo dire che incolla, scolla o rincolla certi aspetti del mondo – e quindi le nostre corrispondenti ipotesi su di essi – che pensavamo fossero immutabili. Separa e riunisce, per così dire, gli “atomi”

delle nostre esperienze e cultura “moderne”. Modifica il letto del fiume, per usare la metafora di Wittgenstein. Alcuni esempi lampanti possono chiarire l'idea in modo più diretto.

Consideriamo, anzitutto, uno dei casi più significativi di incollamento. La nostra identità e i nostri dati personali non sono mai stati incollati insieme così indistinguibilmente come accade oggi, allorché si parla di *identità personale* dei *soggetti interessati*.³ I conteggi dei censimenti sono molto datati (Alterman, 1969). L'invenzione della fotografia ha avuto un forte impatto sulla privacy (Warren, Brandeis, 1890). I governi europei hanno reso obbligatorio viaggiare con il passaporto durante la Prima guerra mondiale, per motivi di migrazione e sicurezza, estendendo così il controllo dello Stato sui mezzi di mobilità (Torpey, 2000). Ma è soltanto il digitale, con il suo enorme potere di registrare, monitorare, condividere e processare quantità illimitate di dati su Alice, che ha saldato insieme chi è Alice, la sua identità e il suo profilo individuali, con le informazioni personali su di lei. La privacy è diventata una questione urgente anche, se non principalmente, a causa di tale incollamento, e oggi, almeno nella normativa dell'Unione Europea, la protezione dei dati personali è discussa in termini di dignità umana (Floridi, 2016c) e identità personale (Floridi, 2005a, 2006), con i cittadini descritti come *soggetti interessati*.

L'esempio successivo riguarda la *posizione* e la *presenza*, e il loro scollamento. In un mondo digitale, è ovvio che uno può trovarsi fisicamente in un posto, diciamo un bar, ed essere presente interattivamente in un altro, diciamo una pagina su Facebook. Eppure tutte le generazioni passate che vivevano in un mondo esclusivamente analogico hanno concepito e sperimentato posizione e presenza come due lati inseparabili della stessa situazione umana: l'essere situati nello spazio e nel tempo, qui e ora. L'azione a distanza e la telepresenza appartenevano a mondi magici o alla fantascienza. Oggi, que-

3. Il soggetto interessato è, per il Regolamento generale sulla protezione dei dati personali (GDPR), la persona fisica a cui si riferiscono i dati personali. [NdT]

sto scollamento è un semplice tratto dell'esperienza comune in qualsiasi società dell'informazione (Floridi, 2005b). Siamo la prima generazione per la quale "Dove sei?" non è soltanto una domanda retorica. Ovviamente, tale scollamento non ha interrotto tutti i collegamenti. La geolocalizzazione funziona solo se è possibile monitorare la telepresenza di Alice. E la telepresenza di Alice è possibile solo se si trova all'interno di un ambiente fisicamente connesso. Ma questi due aspetti sono ora totalmente distinti e in effetti il loro scollamento ha almeno in parte declassato la posizione a favore della presenza. Perché se tutto ciò di cui Alice ha bisogno e si preoccupa è di essere presente digitalmente e interattiva in un particolare angolo dell'infosfera, non importa in quale parte del mondo si trovi analogicamente, se a casa, in treno o nel suo ufficio. Questo è il motivo per cui banche, librerie, biblioteche e negozi di vendita al dettaglio sono tutti luoghi disegnati per la *localizzazione* oggi alla ricerca di un riutilizzo in termini di *presenza*. Quando un negozio apre una caffetteria, sta cercando di ricongiungere la presenza e la localizzazione dei clienti che è stata separata dall'esperienza digitale.

Consideriamo, inoltre, lo scollamento tra *legge* e *territorialità*. Per secoli, all'incirca dalla pace di Westfalia (1648), la geografia politica ha fornito alla giurisprudenza una pronta risposta alla questione relativa all'ambito di applicazione di una sentenza: la decisione del giudice si applica nel perimetro dei confini nazionali entro cui opera l'autorità della legge. Tale connessione potrebbe essere riassunta nei seguenti termini: "Il mio posto le mie regole, il tuo posto le tue regole". Ora può sembrare scontato, ma ci sono voluti molto tempo ed enorme fatica per raggiungere un approccio così semplice, che funziona ancora oggi molto bene, purché si operi solo all'interno di uno spazio fisico, analogico. Tuttavia, Internet non è uno spazio fisico, e il problema della territorialità si profila a partire dallo scollamento ontologico tra lo spazio normativo del diritto, lo spazio fisico della geografia e lo spazio logico del digitale. Si tratta di una nuova "geometria" variabile che stiamo ancora imparando a gestire. Per esempio, lo scollamento tra legge e territorialità

è diventato tanto palese quanto problematico durante il dibattito sul cosiddetto *diritto all'oblio* (Floridi, 2015a). I motori di ricerca operano all'interno di uno spazio logico online fatto di nodi, collegamenti, protocolli, risorse, servizi, URL, interfacce ecc. Ciò significa che qualsiasi cosa è a portata di clic. Perciò è difficile dare piena attuazione al diritto all'oblio, richiedendo a Google di rimuovere i collegamenti alle informazioni personali di qualcuno dalla sua versione .com negli Stati Uniti in forza di una decisione presa dalla Corte di Giustizia dell'Unione Europea, dal momento che tale decisione può risultare inutile se i collegamenti non sono rimossi da tutte le versioni del motore di ricerca. Si noti però che un tale disallineamento degli spazi non genera solo problemi, ma fornisce anche soluzioni. La non territorialità del digitale fa miracoli per la libera circolazione delle informazioni. In Cina, per esempio, il governo deve compiere uno sforzo costante e sostenuto per controllare le informazioni online. Parimenti, il Regolamento generale sulla protezione dei dati personali deve essere ammirato per la sua capacità di sfruttare la "saldatura" tra identità personale e informazioni personali per aggirare lo "scollamento" tra legge e territorialità, fondando la protezione dei dati personali sul legame tra persona e dato (che è ora fondamentale) piuttosto che sulla geografia (il luogo dove tali dati personali sono trattati non è più rilevante).

Infine, possiamo considerare un incollamento che risulta essere più precisamente un ri-incollamento. Nel suo libro *The Third Wave (La terza ondata)* Alvin Toffler ha coniato il termine *prosumer* in riferimento all'affievolirsi della differenza e alla progressiva fusione tra il ruolo di produttore e quello di consumatore (Toffler, 1980). Egli ha attribuito questa tendenza a un mercato altamente saturo e alla produzione di massa di prodotti standardizzati, che ha stimolato un processo di personalizzazione di massa e quindi un crescente coinvolgimento dei consumatori come produttori dei propri prodotti personalizzati. Questa idea era stata anticipata da Marshall McLuhan e Barrington Nevitt (1972), che hanno attribuito il fenomeno alle tecnologie basate sull'elettricità. In seguito, si è

fatto riferimento con ciò al consumo di informazioni prodotte dalla stessa popolazione di produttori, per esempio su YouTube. Ignaro di questi precedenti, quasi vent'anni dopo Toffler, io ho introdotto la parola *prosumer* per cogliere lo stesso fenomeno.⁴ Tuttavia, in tutti questi casi, la posta in gioco non è tanto un *nuovo* incollamento, quanto piuttosto, per essere precisi, un ri-incollamento. Per la maggior parte della nostra storia (circa il 90% [Lee, Daly, 1999]) abbiamo vissuto in società di cacciatori-raccoglitori, che cercano cibo per sopravvivere. Durante questo lungo periodo, produttori e consumatori normalmente hanno coinciso. I prosumer che cacciavano animali selvatici e raccoglievano piante selvatiche erano, in altri termini, la normalità, non l'eccezione. È solo dopo lo sviluppo delle società agrarie, circa 10.000 anni fa, che abbiamo assistito a una separazione completa, e nel tempo culturalmente palese, tra produttori e consumatori. In alcuni angoli dell'infosfera, quella disgiunzione è ricongiunta. Su Instagram o TikTok, per esempio, consumiamo ciò che produciamo. Si può quindi rimarcare che, in alcuni casi, questa parentesi stia volgendo al termine e che i prosumer siano tornati, ricongiunti dal digitale. Di conseguenza, è perfettamente coerente che il comportamento umano in rete sia stato comparato e studiato in termini di modelli di foraggiamento sin dagli anni Novanta (Pirolli, Card, 1995, 1999; Pirolli, 2007).

Il lettore può facilmente elencare ulteriori casi di incollamento, scollamento e ri-incollamento. Si pensi, per esempio, alla differenza tra *realtà virtuale* (scollamento) e *realtà aumentata* (incollamento); la consueta disgiunzione tra *uso* e *proprietà* nella share economy; la ricongiunzione di *autenticità* e *memoria* grazie alla blockchain; o l'attuale dibattito su un reddito di base universale, un caso di scollamento tra *stipendio* e *lavoro*. Ma è giunto il momento di procedere dalla domanda relativa al “come” a quella relativa al “perché”. Perché il digitale ha questo *potere di scissione*, vale a dire di incollare, scollare o ri-incolla-

4. In Floridi (1999); vedi anche Floridi (2004, 2014b). Avrei dovuto conoscerlo meglio e usare il termine *prosumer* di Toffler.

re il mondo? Perché altre innovazioni tecnologiche sembrano non avere un impatto simile? La risposta, suppongo, sta nella combinazione di due fattori.

Da un lato, il digitale è una *tecnologia di terzo ordine* (Floridi, 2014a). Non è solo una tecnologia che sta tra noi e la natura, come un'ascia (primo ordine); o una tecnologia che sta tra noi e un'altra tecnologia, come un motore (secondo ordine). È piuttosto una tecnologia che sta tra una tecnologia e un'altra tecnologia, come un sistema computerizzato che controlla un robot che dipinge un'automobile (terzo ordine). A causa dell'autonoma potenza di calcolo del digitale, potremmo anche non avere controllo sul (per non parlare di essere parte del) processo.

Dall'altro, il digitale non è semplicemente qualcosa che potenzia o aumenta una realtà, ma qualcosa che la trasforma radicalmente, perché crea nuovi ambienti che abitiamo e nuove forme di agire con cui interagiamo. Non c'è un termine specifico per descrivere questa profonda trasformazione, perciò in passato (Floridi, 2010b) ho utilizzato l'espressione re-ontologizzazione per fare riferimento a una radicale forma di re-ingegnerizzazione, che non consiste soltanto nel disegnare, costruire o strutturare un sistema (come una società, una macchina o un qualche artefatto) in modo nuovo, ma nel trasformare fondamentalmente la sua natura intrinseca, vale a dire la sua ontologia. In questo senso, per esempio, le nanotecnologie e le biotecnologie stanno non semplicemente re-ingegnerizzando, ma re-ontologizzando il nostro mondo. Attraverso la *re-ontologizzazione della modernità*, per dirlo in breve, il digitale sta anche *ridefinendo dal punto di vista epistemologico la mentalità moderna*, cioè molte delle nostre concezioni e idee consolidate.

Considerati unitariamente, tutti questi fattori suggeriscono che il digitale deve il suo potere di scissione al suo essere una tecnologia di terzo ordine re-ontologizzante e ri-epistemologizzante. Questo è il motivo per cui fa ciò che fa e per cui nessun'altra tecnologia neppure si avvicina ad avere un effetto simile.

11

LA SOCIETÀ PER LA BUONA IA

Sommario In precedenza, nei capitoli 4-6, ho esaminato i concetti fondamentali che possono fondare una “società per la buona IA”. Poi, nei capitoli 7-9, ho discusso le sfide, le cattive e le buone pratiche che caratterizzano l’uso dei sistemi di IA. Nel decimo capitolo, ho sostenuto che tale sostituzione è una questione che coinvolge l’etica, la cultura e la politica, non uno scenario di fantascienza. In questo capitolo, e nei due successivi, mi concentro su alcune raccomandazioni costruttive e concrete, per valutare, sviluppare, incentivare e sostenere una buona IA, che in taluni casi possono essere attuate direttamente da soggetti politici nazionali o sovranazionali, mentre in altri casi possono essere guidate da altri soggetti interessati, dalla società civile agli attori privati e alle organizzazioni di settore. La speranza è che, se adottate, tali raccomandazioni possano gettare una solida base per l’istituzione di una società per la buona IA.

11.1 INTRODUZIONE: QUATTRO MODI PER REALIZZARE UNA SOCIETÀ PER LA BUONA IA

L’IA non è un altro dispositivo che deve essere regolamentato una volta che sia stato sviluppato. Abbiamo commesso l’errore di pensare in questi termini quando il web ha iniziato a crescere negli anni Novanta e oggi il pasticcio è evidente. Il web era sin dal principio destinato a essere un nuovo ambiente, parte della nostra infosfera, e non semplicemente un altro tipo di mezzo

di comunicazione di massa, per cui avremmo dovuto regolarlo di conseguenza (le regole per uno spazio pubblico sono diverse da quelle per la comunicazione privata: Floridi, 2014a). Non dovremmo ripetere lo stesso errore. L'IA non è solo un servizio commerciale. È una forza potente, una nuova forma dell'agire smart nell'infosfera, che sta già rimodellando le nostre vite, interazioni e ambienti. Questa nuova forma dell'agire deve essere orientata verso il bene della società, di tutti i suoi membri e degli ambienti che condividiamo. Le forze di mercato saranno irrilevanti o insufficienti. Abbiamo bisogno di un approccio normativo. In questo capitolo, contribuisco all'attuale sforzo internazionale e collaborativo per lo sviluppo di una società per la buona IA proponendo 20 raccomandazioni.¹ Se adottate, possono aiutare tutte le parti interessate a cogliere le opportunità offerte dall'IA, a evitare o quantomeno minimizzare e controbilanciare i rischi, a rispettare i principi discussi nel quarto capitolo e dunque a sviluppare una società per la buona IA.

Come detto nei capitoli precedenti, non si tratta più di discutere se l'IA avrà o no un impatto importante su individui, società e ambienti. Il dibattito attuale verte invece su quanto questo impatto sarà positivo o negativo, per chi, in che modo, in quali luoghi e in quale orizzonte temporale. In altre parole, nel quarto capitolo ho scritto che le domande chiave ora sono *chi, come, dove e quando* sentirà l'impatto positivo o negativo dell'IA. Per inquadrare queste domande in modo più sostanziale e pratico, introduco qui quelle che possono essere considerate le quattro principali opportunità che l'IA offre alla società. Sono quattro perché affrontano i quattro punti fondamentali della nostra antropologia filosofica, cioè della nostra comprensione della dignità e del progresso dell'umanità:

- a) la realizzazione autonoma di noi stessi, ovvero chi possiamo diventare;
- b) l'*agire umano*, ovvero cosa possiamo fare;

1. Sono il risultato di un progetto che ho ideato e presieduto nel 2017, denominato AI4People. Per saperne di più vedi Floridi, Cows, Beltrametti et al. (2018).

LA SOCIETÀ PER LA BUONA IA

- c) le capacità individuali e sociali, ovvero cosa possiamo conseguire;
 d) la *coesione sociale*, ovvero come possiamo interagire gli uni con gli altri e con il mondo.

In ciascun caso, l'IA può essere *utilizzata* per favorire la natura umana e le sue potenzialità, generando in tal modo opportunità; *sottoutilizzata*, creando in tal modo costi opportunità; *sovrautilizzata* o *utilizzata male*, creando in tal modo rischi. La Figura 11.1 offre una rapida panoramica, mentre i paragrafi seguenti offrono una spiegazione più dettagliata.

Come indica la terminologia, il presupposto è che l'*utilizzo* dell'IA sia sinonimo di buona innovazione e applicazioni positive di questa tecnologia. Tuttavia, la paura, l'ignoranza, le preoccupazioni fuori luogo o le reazioni eccessive possono portare una società a *sottoutilizzare* le tecnologie di IA, cioè a impiegarle al di sotto del loro pieno potenziale, a causa di quelle che potrebbero essere ampiamente descritte come le ragioni sbagliate. Ciò può creare rilevanti costi opportunità. Potrebbe includere, per esempio, una regolamentazione farraginoso o mal concepita, investimenti insufficienti o una reazione pubblica negativa

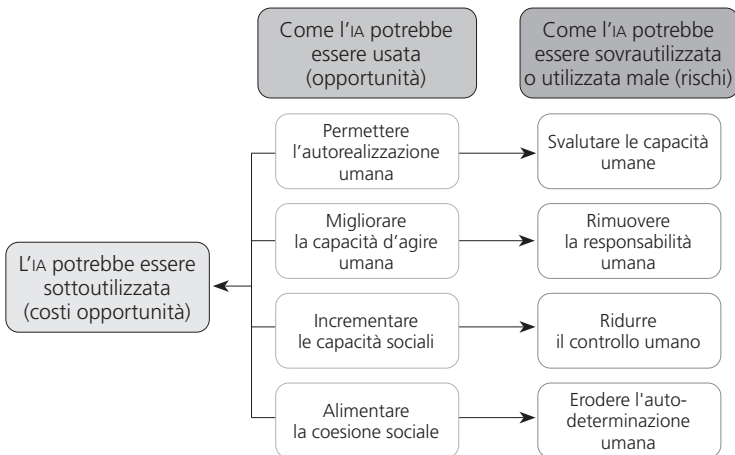


Figura 11.1 Panoramica delle quattro opportunità principali offerte dall'IA, dei quattro rischi corrispondenti e del costo opportunità di un sottoutilizzo dell'IA.

VALUTARE L'INTELLIGENZA ARTIFICIALE

simile a quella che ha investito le colture geneticamente modificate (Imperial College London, 2017) e l'energia nucleare. Di conseguenza, i vantaggi offerti dalle tecnologie di IA potrebbero non essere pienamente realizzati dalla società. Questi pericoli derivano in gran parte da conseguenze non volute e sono di regola legati a buone intenzioni tradite. E, naturalmente, si deve anche tenere in considerazione i rischi associati a un involontario *uso eccessivo* o a un intenzionale *cattivo uso* delle tecnologie di IA, radicato, per esempio, in incentivi orientati male, avidità, geopolitica contraddittoria o intenti malevoli. Come rilevato nei capitoli settimo e ottavo, tutto – dalle truffe via e-mail alla cyberguerra su ampia scala – può essere accelerato o intensificato dall'uso malevolo delle tecnologie di IA (Taddeo, 2017b). E nuovi mali possono essere creati. La possibilità di progresso sociale rappresentata dalle suddette opportunità deve essere soppesata di contro al rischio che una manipolazione malevola sia resa possibile o potenziata dall'IA. Tuttavia, c'è il forte rischio che l'IA possa essere sottoutilizzata per paura di un uso eccessivo o improprio dovuto alla mancanza di un quadro etico e giuridico chiaro, come vedremo nel resto di questo capitolo.

11.2 CHI POSSIAMO DIVENTARE: RENDERE POSSIBILE L'UMANA REALIZZAZIONE DI SÉ, SENZA SVALUTARE LE CAPACITÀ UMANE

L'IA può rendere possibile la realizzazione di sé, ossia la capacità delle persone di prosperare in termini di caratteristiche, interessi, capacità o abilità potenziali, aspirazioni e progetti di vita. Proprio come le invenzioni, per esempio la lavatrice, hanno sollevato le persone (in particolare le donne) dalla fatica di quel lavoro domestico, l'automazione “smart” di altri aspetti mondani della vita può restituire ancora più tempo per attività culturali, intellettuali e sociali e per un lavoro più interessante e gratificante. Più IA può facilmente voler dire più vita umana trascorsa in modo più intelligente. Il rischio in questo caso non risiede tanto nell'obsolescenza di alcune vecchie competenze e nell'emergerne

di nuove, quanto piuttosto nel ritmo con cui ciò sta accadendo e nell'ineguale distribuzione dei costi e dei benefici che ne derivano. Una svalutazione molto rapida delle vecchie competenze e quindi un celere stravolgimento del mercato del lavoro e della natura dell'occupazione possono cogliersi a livello sia dell'individuo sia della società. A livello individuale, i lavori sono spesso intimamente legati all'identità personale, all'autostima e al ruolo o alla posizione sociale, tutti fattori che possono essere influenzati negativamente dalla sovrabbondanza, anche mettendo da parte il potenziale di gravi danni economici. Inoltre, a livello della società, la perdita di competenze in ambiti sensibili e che richiedono elevate capacità, come la diagnostica sanitaria o l'aviazione, può creare pericolose fragilità in caso di malfunzionamento dell'IA o di attacchi avversi. Promuovere lo sviluppo dell'IA a sostegno di nuove capacità e competenze, anticipando e mitigando al contempo il suo impatto su quelle tradizionali, richiederà sia uno studio approfondito sia idee potenzialmente radicali, come la proposta di qualche forma di "reddito di base universale", di cui popolarità e uso sperimentale sono in crescita. Alla fine, abbiamo bisogno di una solidarietà intergenerazionale tra le persone svantaggiate di oggi e quelle avvantaggiate di domani, per garantire che la transizione dirompente tra il presente e il futuro sia la più equa possibile, per tutti. In altre parole, adottando una terminologia che purtroppo abbiamo dovuto tristemente apprezzare durante la pandemia da Covid-19, dobbiamo far diminuire la curva dell'impatto sociale delle tecnologie digitali, inclusa l'IA.

11.3 COSA POSSIAMO FARE: MIGLIORARE L'AGIRE UMANO, SENZA RIMUOVERE LA RESPONSABILITÀ UMANA

L'IA sta fornendo una riserva crescente di capacità di "agire smart". Messa al servizio dell'intelligenza umana, tale risorsa può potenziare enormemente l'agire umano. Possiamo fare di più, meglio e più velocemente, grazie al sostegno fornito dall'IA. Nel senso di "intelligenza aumentata", l'IA potrebbe essere parago-

VALUTARE L'INTELLIGENZA ARTIFICIALE

nata all'impatto avuto dal motore sulle nostre vite. Maggiore è il numero di persone che godranno delle opportunità e dei benefici di una tale riserva di agire smart "a disposizione", migliori saranno le nostre società. La responsabilità è dunque essenziale, in considerazione del tipo di IA che sviluppiamo, di come la utilizziamo e se ne condividiamo vantaggi e benefici con tutti. Naturalmente, il rischio corrispondente risiede nell'assenza di tale responsabilità. Ciò può verificarsi non solo perché disponiamo del quadro di riferimento sociopolitico errato, ma anche a causa di una mentalità da "scatola nera" (*black-box*), per la quale i sistemi di IA che presiedono a un processo decisionale sono concepiti come qualcosa che si colloca al di là della comprensione umana, e quindi del controllo. Queste preoccupazioni si applicano non solo a casi di alto profilo, come i decessi causati dai veicoli autonomi, ma anche a usi più comuni ma parimenti rilevanti, come le decisioni automatizzate sulla libertà vigilata o sull'affidabilità creditizia. Tuttavia, il rapporto tra grado e qualità dell'agire di cui godono le persone e la quantità di agire che deleghiamo ai sistemi autonomi non è a somma zero, né pragmaticamente né eticamente. Infatti, se viene sviluppata in modo equilibrato, l'IA offre l'opportunità di *migliorare* e *moltiplicare* le possibilità dell'agire umano. Consideriamo gli esempi di "moralità distribuita" nei sistemi da umano a umano come il prestito tra pari (*peer-to-peer*) (Floridi, 2013). L'agire umano può in definitiva essere sostenuto, migliorato e ampliato dall'incorporazione di "quadri di facilitazione", progettati per migliorare la probabilità di risultati moralmente buoni, nell'insieme di funzioni che deleghiamo ai sistemi di IA. I sistemi di IA potrebbero, se progettati in modo efficace, amplificare e rafforzare i sistemi morali condivisi.

11.4 COSA POSSIAMO CONSEGUIRE: INCREMENTARE LE CAPACITÀ DELLA SOCIETÀ, SENZA RIDURRE IL CONTROLLO UMANO

L'IA offre molte opportunità per migliorare e incrementare le capacità degli individui e della società più in generale. Che sia

per la prevenzione e la cura delle malattie o per l'ottimizzazione di trasporti e logistica, l'uso di tecnologie di IA presenta innumerevoli possibilità per reinventare la società, migliorando radicalmente ciò di cui gli esseri umani sono collettivamente capaci. Una migliore collaborazione e quindi obiettivi più ambiziosi possono essere sostenuti da una maggiore IA. L'intelligenza umana potenziata dall'IA può individuare nuove soluzioni a problemi vecchi e nuovi, da una distribuzione più equa o più efficiente delle risorse a un approccio più sostenibile al consumo. Proprio perché tali tecnologie hanno il potenziale per essere così potenti e dirompenti, introducono anche rischi della stessa grandezza. Ci saranno sempre più contesti in cui potremmo non doverci trovare più "nel o sul processo" (vale a dire, essere parte del processo stesso o almeno in condizione di controllarlo), se possiamo delegare i nostri compiti all'IA. Tuttavia, se ci affidiamo all'uso delle tecnologie di IA per incrementare le nostre capacità in modo sbagliato, possiamo delegare compiti importanti e, soprattutto, decisioni cruciali a sistemi autonomi che dovrebbero rimanere almeno in parte soggetti a supervisione, scelta e correzione umana. Ciò a sua volta può ridurre la nostra capacità di monitorare le prestazioni di tali sistemi (non essendo più "sul processo") e di prevenire o correggere gli errori o i danni che si verificano ("dopo il processo"). È anche possibile che questi potenziali danni si accumulino e diventino radicati, poiché un numero crescente di funzioni è delegato a sistemi artificiali. È dunque tassativo trovare un equilibrio tra il perseguimento delle ambiziose opportunità offerte dall'IA per migliorare la vita umana e ciò che possiamo conseguire, da un lato, e, dall'altro, assicurarci di mantenere il controllo su questi importanti sviluppi e sui loro effetti.

11.5 COME POSSIAMO INTERAGIRE: COLTIVARE LA COESIONE SOCIALE, SENZA ERODERE L'AUTODETERMINAZIONE UMANA

Dal cambiamento climatico e dalla resistenza antimicrobica alla diffusione del nucleare e al fondamentalismo, i pro-

VALUTARE L'INTELLIGENZA ARTIFICIALE

blemi globali sono caratterizzati da livelli sempre più elevati di complessità di coordinamento, il che significa che possono essere affrontati con successo soltanto se tutte le parti interessate progettano insieme e condividono le soluzioni e cooperano per realizzarle. L'IA, con le sue soluzioni ad alta intensità di dati e basate su algoritmi, può contribuire fortemente ad affrontare tale complessità di coordinamento, supportando una maggiore coesione e collaborazione della società. Per esempio, gli sforzi per affrontare il cambiamento climatico hanno messo in luce l'esigenza di rispondere a tale sfida creando una risposta coesa, sia all'interno delle società sia tra di esse. La portata di questa sfida è tale che potremmo presto dover decidere come trovare un equilibrio tra l'ingegnerizzazione diretta del clima e il design di strutture sociali volte a incoraggiare una drastica riduzione delle emissioni nocive. Quest'ultima opzione potrebbe essere sostenuta da un sistema algoritmico diretto ad alimentare la coesione sociale. Un tale sistema non dovrebbe essere imposto dall'esterno; dovrebbe essere il risultato di una scelta autoimposta, non dissimile dalla scelta di non acquistare cioccolata se abbiamo scelto di stare a dieta, o di programmare la sveglia per svegliarci presto. "L'autonoma spinta gentile" (*self-nudging*) per comportarsi in modi socialmente preferibili è la migliore forma di condizionamento e l'unica che preserva l'autonomia (Floridi, 2015b, 2016f). È il risultato di decisioni e scelte umane, ma può fare affidamento su soluzioni di IA per essere attuata e resa più agevole. Tuttavia, il rischio è che i sistemi di IA possano erodere l'autodeterminazione umana, poiché possono portare a cambiamenti non pianificati e voluti nei comportamenti umani per adattarsi alle routine che agevolano il funzionamento dell'automazione e rendono la vita delle persone più facile. Il potere predittivo e l'incessante spinta gentile dell'IA, anche se non intenzionali, dovrebbero essere posti al servizio dell'autodeterminazione umana e favorire la coesione sociale, senza minare la dignità o il progresso umani (Milano, Taddeo, Floridi, 2019, 2020).

14

CONCLUSIONE: IL VERDE E IL BLU

Sommario Nei capitoli precedenti, ho esaminato l'IA, il suo impatto positivo e negativo, e le questioni etiche che ne discendono. In quest'ultimo capitolo, traggio alcune conclusioni generali e getto uno sguardo su ciò che potrebbe accadere: un passaggio dall'etica dell'agire artificiale alla politica delle azioni sociali.

14.1 INTRODUZIONE: DAL DIVORZIO TRA AGIRE E INTELLIGENZA AL MATRIMONIO TRA IL VERDE E IL BLU

A volte dimentichiamo che la vita senza il contributo di una buona politica, di una scienza affidabile e di una robusta tecnologia diventa presto “solitaria, povera, sgradevole, brutale e breve”, per prendere in prestito le parole del *Leviatano* di Thomas Hobbes. La crisi del Covid-19 ci ha tragicamente ricordato che la natura può essere spietata. Solo l'ingegno umano e la buona volontà possono migliorare e salvaguardare il tenore di vita di miliardi di persone. Oggi, gran parte di tale ingegnosità è impegnata nel realizzare una rivoluzione epocale: la trasformazione di un mondo esclusivamente analogico in un mondo sempre più digitale. Gli effetti sono già diffusi: questa è la prima pandemia in cui un nuovo habitat, l'*infosfera*, ha contribuito a superare i pericoli della *biosfera*. Viviamo *onlife* (sia online sia offline) ormai da tempo, ma la pandemia ha trasformato l'e-

sperienza onlife in una realtà che costituisce un punto di non ritorno per l'intero pianeta.

Nei capitoli precedenti ho sostenuto che lo sviluppo dell'IA è un fattore importante in questa rivoluzione epocale. L'IA dovrebbe essere concepita come l'ingegnerizzazione di artefatti in grado di fare cose che richiederebbero intelligenza se dovessimo farle noi. Con un classico esempio che ho usato più volte, un telefono cellulare può battere quasi chiunque a scacchi, pur essendo intelligente come un tostapane. In altre parole, l'IA segna il divorzio senza precedenti tra la capacità di portare a termine compiti o risolvere problemi con successo in vista di un dato obiettivo e il bisogno di essere intelligenti per farlo. Questo riuscito divorzio è diventato possibile solo negli ultimi anni, grazie a gigantesche quantità di dati, strumenti statistici molto sofisticati, enorme potenza di calcolo e alla trasformazione dei nostri contesti di vita in luoghi sempre più adatti all'IA (avvolti intorno all'IA). Quanto più viviamo nell'infosfera e onlife, tanto più condividiamo le nostre realtà quotidiane con forme di agire ingegnerizzate, e tanto più l'IA può affrontare un numero crescente di problemi e compiti. Il limite dell'IA non è il cielo, ma l'ingegno umano.

In questa prospettiva storica ed ecologica, l'IA è una straordinaria tecnologia che può essere una potente forza positiva, in due modi principali. Può aiutarci a conoscere, comprendere e prevedere di più e meglio le numerose sfide che stanno diventando così impellenti, in particolare il cambiamento climatico, l'ingiustizia sociale e la povertà globale. La corretta gestione di dati e processi da parte dell'IA può accelerare il circolo virtuoso tra maggiori informazioni, migliore scienza e politiche più avvedute. Eppure, la conoscenza è potere solo se si traduce in azione. Anche a questo riguardo, l'IA può essere una notevole forza positiva, aiutandoci a migliorare il mondo, e non soltanto la sua interpretazione. La pandemia ci ha ricordato che fronteggiamo problemi complessi, sistemici e globali. Non possiamo risolverli individualmente. Abbiamo bisogno di coordinarci (non dobbiamo intralciarci), collaborare (ognuno fa la sua parte) e cooperare (lavoriamo insieme) di più, meglio e a livello

CONCLUSIONE: IL VERDE E IL BLU

internazionale. L'IA può consentirci di realizzare queste 3C in modo più efficiente (più risultati con meno risorse), in modo efficace (migliori risultati) e in modo innovativo (nuovi risultati).

Tuttavia, c'è un "ma": sappiamo che l'ingegno umano, senza buona volontà, può essere pericoloso. Se l'IA non è controllata e guidata in modo equo e sostenibile, può esacerbare i problemi sociali, dai pregiudizi alla discriminazione; erodere l'autonomia e la responsabilità umana; amplificare i problemi del passato, dall'iniqua allocazione della ricchezza allo sviluppo di una cultura della mera distrazione, quella del "panem et digital circenses". L'IA rischia di trasformarsi da parte della soluzione a parte del problema. Questo è il motivo per cui iniziative etiche come quelle descritte nel quarto capitolo e, in ultima analisi, buone norme internazionali sono essenziali per garantire che l'IA rimanga una potente forza per il bene.

L'IA per il bene sociale è parte integrante di un nuovo matrimonio, tra il verde di tutti i nostri habitat – naturali, sintetici e artificiali, dalla biosfera all'infosfera, da ambienti urbani a contesti economici, sociali e politici – e il blu delle nostre tecnologie digitali – dai cellulari alle piattaforme sociali, dall'Internet delle Cose ai Big Data, dall'IA ai futuri computer quantistici. Il matrimonio tra il verde e il blu, con i suoi vantaggi, controbilancia il divorzio tra l'agire e l'intelligenza, con i suoi rischi. Siamo noi che abbiamo la responsabilità di disegnare e gestire entrambi con successo. La pandemia ha reso palese che la posta in gioco non risiede tanto nell'innovazione digitale, quanto piuttosto nella corretta governance del digitale. Le tecnologie aumentano e migliorano ogni giorno. Tuttavia, per salvare il nostro pianeta e noi stessi, anche da noi stessi, possiamo e dobbiamo utilizzarle molto meglio; basti pensare alla diffusione di disinformazione relativa al Covid-19 sui social media o all'inefficacia delle cosiddette app per il coronavirus. La pandemia è stata la prova generale di quello che dovrebbe essere il progetto umano per il ventunesimo secolo, un matrimonio stabile e fruttuoso tra il verde e il blu. Possiamo farne un successo insieme e facendo affidamento su più e migliore filosofia, non su meno.

14.2 IL RUOLO DELLA FILOSOFIA COME DESIGN CONCETTUALE

Per un certo tempo, il divario umano/macchina è stata la frontiera del cyberspazio. Oggi ci siamo spostati all'interno dell'infosfera. La sua natura onnicomprensiva dipende anche dalla misura in cui accettiamo la sua natura digitale come parte integrante della nostra realtà e come trasparente per noi, nel senso che non la percepiamo più come presente. Ciò che conta non sono tanto i bit in movimento invece degli atomi – questa è un'interpretazione della società dell'informazione obsoleta e basata sulla comunicazione che deve troppo alla sociologia dei mass media – quanto piuttosto il fatto molto più radicale che il nostro modo di comprendere e concepire l'essenza e il tessuto della realtà sta cambiando. In effetti, abbiamo cominciato ad accettare il virtuale come parzialmente reale e il reale come parzialmente virtuale. La società dell'informazione è concepita più adeguatamente come società neomanifatturiera in cui materie prime ed energia sono state sostituite da dati e informazioni, il nuovo oro digitale e la vera fonte di valore aggiunto. Le chiavi per una corretta comprensione della nostra situazione e per lo sviluppo di un'infosfera sostenibile non risiedono dunque solo nella comunicazione e nelle transazioni, ma nella creazione, nel design e nella gestione delle informazioni. Tale comprensione richiede una nuova narrazione, ossia un nuovo tipo di storia che raccontiamo a noi stessi sulla nostra situazione e sul progetto umano che desideriamo perseguire. Questo può sembrare un passo anacronistico nella direzione sbagliata. Fino a poco tempo fa, le “grandi narrazioni”, dal marxismo al liberalismo, alla cosiddetta “fine della storia”, sono state oggetto di numerose critiche. Ma la verità è che anche tale critica era solo un'altra narrazione, e non ha funzionato. Una critica sistematica delle grandi narrazioni è fatalmente parte del problema che cerca di risolvere. Capire perché ci sono narrazioni, che cosa le giustifica e quali narrazioni migliori possono sostituirle è un modo meno immaturo e più fruttuoso di procedere. Le ICT stanno creando il nuovo

CONCLUSIONE: IL VERDE E IL BLU

ambiente informazionale in cui le generazioni future vivranno la maggior parte del loro tempo. Le precedenti rivoluzioni nella creazione di ricchezza, in particolare quella agricola e industriale, hanno portato a trasformazioni macroscopiche nelle nostre strutture sociali e politiche e negli ambienti architettonici, spesso in modo imprevisto e, di regola, con profonde implicazioni concettuali ed etiche. La rivoluzione dell'informazione – intesa sia in termini di creazione di ricchezza sia in termini di riconcettualizzazione di noi stessi – non è meno decisiva. Avremo grossi problemi, se non prendiamo sul serio il fatto che stiamo costruendo i nuovi ambienti che saranno abitati dalle generazioni future. Alla luce di questo importante cambiamento nel tipo di interazioni, mediate dalle ICT, che intraprenderemo sempre più con altri agenti, sia biologici sia artificiali, e nella comprensione di noi stessi, un approccio etico è essenziale per affrontare le nuove sfide poste dalle ICT. Deve essere un approccio che non privilegi il naturale o l'incontaminato, ma tratti come autentiche e genuine tutte le forme di esistenza e comportamento, anche quelle basate su artefatti artificiali, sintetici, ibridi e ingegnerizzati. Il compito è quello di formulare un quadro etico che possa trattare l'infosfera come un nuovo ambiente degno dell'attenzione morale e della cura degli *infor*g umani che lo abitano. Un simile quadro etico deve affrontare e risolvere le sfide senza precedenti che sorgono nel nuovo ambiente. Come ho sostenuto (in Floridi, 2013), deve essere un'etica *e-cologica* per l'intera infosfera. Questa sorta di *e-cologismo sintetico* (sia nel senso di olistico o inclusivo, sia nel senso di artificiale) richiederà un cambiamento nel modo in cui percepiamo noi stessi e i nostri ruoli rispetto alla realtà, in ciò che reputiamo degno di rispetto e cura, e nel modo in cui potremmo negoziare una nuova alleanza tra il naturale e l'artificiale. Ciò richiederà una seria riflessione sul progetto umano e una revisione critica delle nostre attuali narrazioni, a livello individuale, sociale e politico. Sono tutte questioni urgenti che meritano la nostra piena e totale attenzione. Purtroppo, temo che ci vorrà del tempo e un nuovo tipo di educazione e sensibilità per rendersi conto che l'infosfera è uno spazio comune,

che deve essere preservato a beneficio di tutti. La filosofia come design concettuale (Floridi, 2019c) dovrebbe contribuire a tale cambiamento di prospettiva e dei nostri sforzi costruttivi. Dovrebbe anche aiutarci a mutare la comprensione che abbiamo di noi stessi, in quanto bellissimo errore di Natura.

14.3 IL BELLISSIMO ERRORE DI NATURA

Abbiamo visto nel secondo capitolo che Galileo riteneva che la natura fosse come un libro, scritto in simboli matematici, per essere letto dalla scienza. Poteva essere una forzatura metaforica ai suoi tempi, ma oggi il mondo in cui viviamo è certamente sempre più un libro scritto in cifre, per essere letto ed esteso dall'informatica e dalla scienza dei dati. Le tecnologie digitali hanno sempre più successo al suo interno perché, come i pesci nel mare, sono i veri nativi dell'infosfera. Ciò spiega anche perché le applicazioni di IA sono migliori di noi in un numero crescente di compiti: noi siamo semplici organismi analogici che cercano di adattarsi a un habitat così nuovo vivendo onlife. La trasformazione epocale del nostro ambiente in un'infosfera mista, sia analogica sia digitale, e il fatto di condividere l'infosfera con agenti artificiali sempre più smart, autonomi e sociali, ha profonde conseguenze. Alcune di queste non sono ancora distinguibili. Le scopriremo solo con il tempo. Altre sono appena riconoscibili all'orizzonte. Ed altre ancora sono davanti ai nostri occhi. Cominciamo da queste.

Gli agenti di IA, siano essi software (app, webbot, algoritmi, software di ogni genere) o hardware (robot, auto senza conducente, orologi intelligenti e gadget di ogni genere) stanno sostituendo gli agenti umani in ambiti che pensavamo fossero al di fuori della portata di qualsiasi tecnologia fino a pochi anni fa: catalogare immagini, tradurre documenti, interpretare radiografie, pilotare droni, estrarre nuove informazioni da enormi quantità di dati e molte altre cose che solo i colletti bianchi avrebbero dovuto fare. I colletti marroni in agricoltura e quelli

CONCLUSIONE: IL VERDE E IL BLU

blu nell'industria avvertono da decenni la pressione del digitale; i servizi sono ora il nuovo target. Perciò, spariranno anche i posti di lavoro dei colletti bianchi. Possiamo solo tentare di fare una stima ragionevole di quanti ne spariranno e con quale rapidità, ma è probabile che lo sconvolgimento sia profondo. Ovunque gli esseri umani lavorino oggi come interfacce, per esempio tra un GPS e un'automobile, tra due documenti in lingue diverse, tra alcuni ingredienti e un piatto, tra i sintomi e la malattia corrispondente, quel lavoro è a rischio. Al contempo, appariranno nuovi lavori – li ho chiamati colletti verdi – perché saranno necessarie nuove interfacce, tra i servizi forniti dai computer, tra i siti web, tra le applicazioni di IA, tra i risultati dell'IA e così via. Qualcuno dovrà controllare se una traduzione abbastanza buona sia una traduzione sufficientemente affidabile. Molte attività resteranno troppo costose per le applicazioni di IA, pur ammettendo che siano realizzabili con l'IA. Consideriamo il caso di Amazon. Fornisce "l'accesso a più di 500.000 lavoratori provenienti da 190 paesi",¹ i cosiddetti turchi, definiti anche da Amazon come "intelligenza artificiale [*sic*] artificiale". La ripetizione è indicativa: si tratta di lavori che non richiedono intelligenza pagati pochi centesimi. Non è il tipo di lavoro che potremmo desiderare per i nostri figli, ma è comunque un lavoro di cui molte persone hanno bisogno e che non possono rifiutare. Se non elaboreremo migliori quadri normativi giuridici ed etici, l'IA creerà ulteriori polarizzazioni nella nostra società, specialmente tra i pochi al di sopra delle macchine – i nuovi patrizi – e coloro che stanno al di sotto di esse – la nuova plebe. Con i posti di lavoro, diminuiranno anche le tasse, anche se ciò accadrà un po' più avanti nel futuro. Nessun lavoratore, nessun contribuente, questo è ovvio; e le aziende che sfrutteranno la delegazione di compiti all'IA non saranno generose come i loro ex dipendenti allorché si tratterà di sostenere il benessere sociale. Occorrerà fare qualcosa al riguardo, facendo pagare più tasse alle aziende e ai benestanti. Le norme giocheranno un ruolo importante an-

1. <https://requester.mturk.com/create/projects/new>.

che nel determinare quali lavori dovranno essere mantenuti “umani”. I treni a guida autonoma sono una rarità anche per ragioni normative,² eppure sono molto più facili da gestire rispetto ai taxi o agli autobus senza conducente. Chiaramente le regole contribuiranno in modo significativo al modo in cui disegnamo il futuro della nostra infosfera. Prima di considerare le conseguenze palesi, occorrono due ultime precisazioni. Molti compiti che scompariranno non faranno sparire i lavori corrispondenti: i giardinieri assistiti da uno dei tanti robot tosaerba esistenti avranno semplicemente più tempo per fare altre cose, diverse dal tagliare l'erba. Nessun robot li sostituirà. E molti compiti non scompariranno, verranno semplicemente riaffidati a noi come utenti: premiamo già i pulsanti dell'ascensore (quel lavoro non esiste più), siamo sempre più abituati a scansionare la merce al supermercato (anche il lavoro di cassa sta per sparire) e faremo certamente più lavori in prima persona in futuro.

Che cosa dire delle altre conseguenze appena distinguibili all'orizzonte, quando l'IA non sarà più in mano a tecnici e manager, ma “democratizzata” nelle tasche di miliardi di persone? Al riguardo, posso essere soltanto piuttosto astratto e incerto. L'IA e più in generale gli agenti smart, autonomi e sociali, così come gli strumenti predittivi in grado di anticipare e manipolare le decisioni e le scelte umane offrono un'opportunità storica per ripensare l'eccezionalità umana non come qualcosa di errato quanto piuttosto come qualcosa di mal compreso. I nostri comportamenti intelligenti saranno messi alla prova dai comportamenti smart dell'IA, che possono risultare più efficaci nell'infosfera in termini di adattamento. I nostri comportamenti autonomi saranno messi alla prova dalla prevedibilità e manipolabilità delle nostre scelte razionali e dallo sviluppo dell'autonomia artificiale. E anche la nostra socievolezza sarà messa alla prova dalla sua controparte artificiale, rappresentata da compagni artificiali, ologrammi o semplici voci, servitori

2. <https://www.vice.com/en/article/wnj75z/why-dont-we-have-driverless-trains-yet>.

CONCLUSIONE: IL VERDE E IL BLU

3D o robot sessuali simili a umani, che possono essere attraenti per gli umani e talvolta indistinguibili da loro. Come andrà a finire tutto questo non è chiaro, ma una cosa è certa: lo sviluppo di agenti artificiali non darà luogo ad alcuna allarmante creazione di scenari fantascientifici, che sono irresponsabilmente fuorvianti. Come già detto, non c'è alcun Terminator in vista. L'IA è pressoché un ossimoro: le tecnologie smart saranno tanto stupide quanto le nostre vecchie tecnologie. Ma l'IA ci inviterà a riflettere più seriamente e con minore compiacimento su chi siamo, potremmo essere o vorremmo diventare, e quindi sulle nostre responsabilità e sulla comprensione che abbiamo di noi stessi. L'IA sfiderà profondamente il nostro modo di concepire ciò che intendiamo quando ci percepiamo come “speciali” dopo la quarta rivoluzione (vedi Floridi, 2014a). Non sto sostenendo che il nostro eccezionalismo sia errato. Suggesto piuttosto che l'IA ci farà rendere conto che il nostro essere eccezionali risiede in un modo speciale e forse irriproducibile di essere disfunzionali con successo. Siamo un *hapax legomenon* nel Libro della Natura di Galileo, un po' come l'espressione *gopher* (“legno di cipresso”), che si riferisce al materiale originario con cui è stata costruita l'arca di Noè e che ricorre una sola volta in tutta la Bibbia. Con una metafora più digitale e contemporanea, siamo un bellissimo errore nel grande software dell'universo, non l'app di maggior successo. Resteremo un bug, un errore unico e riuscito, mentre l'IA sarà ancora di più un elemento peculiare nel libro matematico della natura di Galileo. Un così bell'errore sarà sempre più responsabile della natura e della storia. In breve, Shakespeare aveva ragione:

Gli uomini in certi momenti sono padroni dei loro destini.
 La colpa, caro Bruto, non è nelle nostre stelle.
 Ma in noi stessi, se siamo schiavi.
 (Shakespeare, *Giulio Cesare*, I, II)

La questione è che non potremo essere padroni dei nostri destini senza ripensare un'altra forma dell'agire, quella politica. Per questo, *La politica dell'informazione* sarà il tema del mio prossimo libro.