

## LA MORTE DELLA VERITÀ

- **Capitolo 1:** caduta della ragione a ondate ricorrenti, max con Trump, le opinioni estremiste diventano dominanti. 19% di elettori di Trump credono che il governo USA abbia avuto un ruolo nell'attacco dell'11 settembre, 11% crede alla teoria complottista sulle lampadine al neon che servono per manipolare meglio le persone. "Se i cittadini non si procurano almeno un'informazione di base sui problemi che toccano la loro vita, piaccia o no, rinunciano al controllo su quei problemi", studioso Nichols. "Paragono chi attacca la scienza a chi guida a fari spenti", Mike MacFerrin, glaciologo.
- **Capitolo 2:** internet non ha solo portato a innovazione [...] ha portato anche a una valanga di disinformazione e relativismo, come evidenziato dall'attuale epidemia di fake news. È stato dimostrato che i russi abbiano usati profili social falsi per amplificare le divisioni tra gli americani e favorire l'elezione di Trump. I repubblicani, un tempo paladini della lotta nella guerra fredda contro i russi, hanno fatto spallucce di fronte a ciò. Postmodernisti radicali mettono sotto attacco la scienza e la conoscenza del passato definendola come una semplice costruzione ideologica al servizio di interessi particolari.
- **Capitolo 3:** ascesa della soggettività (bisogno di catturare l'attenzione degli altri con lo spettacolo di sé, amplificata con i social) e svalutazione della verità oggettiva, celebrazione dell'opinione rispetto alla conoscenza, tutto ciò ha contribuito a favorire l'ascesa di Trump. Secondo l'autore a partire dagli anni '70-'80 la scrittura americana si spostò progressivamente verso le autobiografie permeate anche di finzione. Conclusione postmodernista: ci sono molti modi legittimi per rappresentare un avvenimento (sfruttata da coloro che sostengono teorie offensive o infondate). Assurdo che su dibattiti come cambiamento climatico e vaccini si dia lo stesso spazio mediatico alla quasi totalità della comunità scientifica e a una minuscola minoranza del tutto priva di prove.
- **Capitolo 4:** la presidenza di Trump rappresenta una sorta di culmine della distorsione della realtà: sempre più difficile distinguere i titoli della testata satirica "The Onion" e quelli della CNN (ahah). Non solo Trump ha mentito consapevolmente e senza vergogna, ma quelle centinaia e centinaia di menzogne si sono unite a creare racconti altrettanto falsi che facevano appello alle paure della gente (= Salvini). Gli algoritmi dei social network (suggerendo le notizie più popolari e di tendenza anziché quelle accurate o importanti) contribuiscono a promuovere le teorie complottiste. Internet non si limita più a riflettere la realtà, la crea.
- **Capitolo 5:** i regimi autoritari nel corso della storia hanno adottato il linguaggio quotidiano nel tentativo di controllare non solo come la gente comunica ma anche come pensa (vedi Hitler, Trump, Putin e anche in 1984 di Orwell). Tutta l'amministrazione di Trump mente spudoratamente per lui, come il portavoce della Casa Bianca che con le evidenze fotografiche che dimostravano il contrario sosteneva che la folla al discorso inaugurale di Trump fosse la più ampia mai vista. I social dei vari dipartimenti sotto l'amministrazione Trump sono stati sottoposti a censura (controllati da amministratori di Trump).
- **Capitolo 6:** silos di persone: pareri sempre più contrapposti, fine del confronto di idee. Spaccatura del paese in due tribù coerenti [...] che lottano non solo per far vincere la propria parte, ma per provocare, condannare e sconfiggere l'altra. Gli algoritmi di ricerca su Google potrebbero dare risultati opposti riguardo alla ricerca "prove del cambiamento climatico" a un attivista ambientale o a un dirigente petrolifero. È importante quindi assicurarci, se non siamo noi a scegliere cosa guardare ma degli algoritmi, che ci mostrino anche altri punti di vista.
- **Capitolo 7:** quando si tratta di diffondere fake news e minare la fiducia nell'obiettività, la tecnologia (internet) si è dimostrata un acceleratore altamente infiammabile. Mentre la fiducia del pubblico nei media declinava, sempre più persone cominciavano a informarsi tramite Facebook e Twitter, alimentando il mostro delle fake news. In futuro potremo

ritrovarci video realistici di politici che dicono cose che non hanno mai detto (grazie alle nuove tecnologie).

- **Capitolo 8:** paragonati i modi di far propaganda menzognera di Hitler, Lenin, Trump e tutte le politiche populiste e estremiste. Propaganda in Russia sotto Putin: stancare il popolo tramite una tale profusione di menzogne da indurlo a smettere di opporre resistenza e a ritirarsi nell'ambito della vita privata ("idrante delle falsità"). Tale metodo di propaganda si diffonde in tutto il mondo.
- **Capitolo 9:** sfiducia generale dopo la crisi del 2008, perché prendersela con le bugie di Trump se tutti i politici mentono? I portavoce della Casa Bianca dicono che Trump sta solo scherzando o viene frainteso quando fa affermazioni offensive, razziste, sessiste.
- **Epilogo:** Postman sosteneva che già a fine XX secolo si stesse realizzando ciò che Huxley scriveva in *Mondo Nuovo*: Huxley temeva che la verità finisse affogata in un mare di irrilevanza, sommersa da informazioni e fake news. Con l'elezione di Trump anche la paura di Orwell, ovvero che la verità ci venisse nascosta, diventa attuale. Appello di Washington ai cittadini: dovere di difendere le istituzioni create dai padri fondatori, considerate come i pilastri che reggono la casa della democrazia. Altri due capisaldi della democrazia, indispensabili per creare un'opinione pubblica informata e capace di scegliere saggiamente i propri leader: istruzione e stampa libera e indipendente.

A proposito: il 6 settembre 2021 Bolsonaro emette un decreto che impedisce la rimozione delle fake news dai social. La misura è stata adottata dopo che Bolsonaro e il suo staff hanno subito la rimozione di contenuti per diffusione di informazioni false sul covid-19, il presidente l'ha definita "censura".

## ARTIFICIAL YOU

- **Intro:** due argomentazioni centrali del libro: la coscienza, l'esperienza, le sensazioni interiori; le possibili concretizzazioni malevoli della tecnologia IA. Se un'IA è un essere cosciente, obbligarla a servirci sarebbe assimilabile alla riduzione in schiavitù. La coscienza potrebbe essere un inevitabile sottoprodotto della costruzione di un sistema intelligente oppure rivelarsi impossibile da ottenere nell'IA.
- **Capitolo 1:** IAG (Intelligenza Artificiale Generale) = capace di mettere in relazione idee differenti su più argomenti, svolgere vari compiti, superare in intelligenza l'uomo. Miliardi di dollari investiti in tutto il mondo in questo. Nel lungo periodo non c'è competizione: L'IA sarà di gran lunga più capace e durevole di noi, ha margini di miglioramento illimitati. Può basarsi sugli algoritmi del cervello e migliorarli, può non basarsi sul funzionamento del cervello, può essere scaricata su più dispositivi ecc. Se potenziamo la nostra mente mediante tecnologie IA, forse possiamo tenere il passo rispetto a un'intelligenza artificiale.
- **Capitolo 2:** problema della coscienza dell'IA: l'elaborazione di un'intelligenza artificiale sarebbe in un qualche modo percepibile interiormente? Avrebbe per l'IA una qualità percepita? Due fazioni di pensiero opposte: naturalismo biologico, sostiene che anche le forme più sofisticate di IA saranno prive di esperienza interiore; tecno-ottimismo, sostiene che la coscienza sia computazionale e che quindi i sistemi computazionali sofisticati potranno avere un'esperienza interiore. Il ragionamento della stanza cinese di Searle non riesce a fornire una prova a favore del naturalismo biologico.
- **Capitolo 3:** l'autrice è più dell'idea di "aspettiamo e vediamo", è troppo presto per dire se l'IA potrà avere una coscienza. Non è una questione da decidere a tavolino: dobbiamo eseguire test sulle macchine. Le IA potrebbero pure superare la coscienza, proprio come gli umani quando imparano ad eseguire un compito meccanicamente, ovvero inconsciamente. Se un'IA fosse cosciente, l'azienda che la produce potrebbe affrontare

accuse di schiavitù. Se un'IA fosse cosciente la sua coscienza potrebbe influire sul lavoro che deve compiere, magari si appassiona e lo fa meglio, magari non ha voglia di farlo.

- **Capitolo 4:** come scoprire se un'IA ha una coscienza? Ma già oggi i programmatori hanno difficoltà a capire perché i sistemi di deep learning fanno quello che fanno. Non può esistere un test generale per la coscienza dell'IA; un'opzione migliore consiste in una batteria di test da applicare in base al contesto. Coscienza *fenomenale*: quella coscienza che porta a esperienze interiori, sentimenti; coscienza *cognitiva (o funzionale)*: permette di rendersi conto delle cose, la memoria di lavoro, l'attenzione, è forse il passo prima della coscienza fenomenale. Test ACT: serie sempre più impegnativa di interazioni in linguaggio naturale per vedere con che prontezza la IA riesce a comprendere e usare concetti basati sulle esperienze interiori che noi associamo alla coscienza. Una macchina potrebbe non passare il test di Turing, ma passare il test ACT. Test del chip: se un certo tipo di chip supera il test quando viene implementato in un sistema biologico, ciò ci suggerisce di cercare la coscienza nelle IA dotate di tale chip. Teoria Informazione Integrata (IIT): l'informazione è integrata quando c'è una fitta rete di feedback tra gli stati del sistema; il livello di info integrata è misurabile e se conosciamo il valore possiamo determinare se e in che misura un sistema è cosciente. Principio di precauzione: prima di usare una tecnologia che potrebbe avere un impatto catastrofico sulla società, chi vuole sviluppare quella tecnologia deve dimostrare l'infondatezza di tale preoccupazione.
- **Capitolo 5:** fondersi con l'IA, dovremmo intraprendere questo viaggio o no? Anche lievi potenziamenti potrebbero rivelarsi rischiosi. Approccio computazionale: mente come programma software (pattern), come persone siamo assimilabili alla forma che assume l'acqua di un ruscello che scorre sopra le rocce. Le molecole di acqua cambiano ogni millisecondo, ma la forma persiste per ore o addirittura anni (proprio come le molecole del corpo si sostituiscono di continuo ma noi rimaniamo gli stessi per anni).
- **Capitolo 6:** replicare esattamente una mente umana per i patternisti è come creare una stessa persona, le due creature sono la stessa persona. Eppure sono due, esattamente uguali, ma non una sola entità, hanno due coscienze separate. L'identità di pattern non è sufficiente per stabilire l'identità della persona. C'è da capire quanti e quali cambiamenti del pattern sono accettabili prima di parlare di morte dell'identità. In sostanza anche la posizione patternista non è molto chiara, ha bisogno di molte delucidazioni.
- **Capitolo 7:** nell'universo potrebbero già esserci tantissime altre civiltà che hanno compiuto il passo da intelligenza biologica a sintetica. Se davvero la vita è presente e se le civiltà avanzate tendono a svilupparsi e a sopravvivere ad altre, allora i membri delle più avanzate civiltà aliene sono probabilmente IA super intelligenti. Le super intelligenze potrebbero avere una natura imprevedibile perché elaborano pensieri estremamente alieni a noi.
- **Capitolo 8:** mente come programma software? Sottoporre il cervello alla criogenizzazione aspettando la tecnologia per poterlo scaricare su una macchina, significa rinascita? Se il cervello viene scansionato in condizioni ideali, il processo di scansione effettua una copia della tua configurazione neurale (il tuo programma). Si può sopravvivere al caricamento finché sopravvive il proprio pattern. Problemi sollevati dall'autrice sulla teoria mente come software: quando si parla della mente come programma non ci si può riferire a un programma in senso generico, se le menti sono programmi allora sono programmi di un tipo altamente sofisticato in grado di avere esperienze coscienti. Inoltre sostiene che la mente abbia degli stati fisici che ci spingono a prendere determinate decisioni, mentre i programmi (linee di codice) sono solamente delle astrazioni, per cui dire che la mente è un programma attribuisce a delle astrazioni un significato fisico che non hanno. Così come è discutibile che un essere umano possa ottenere l'immortalità caricando e scaricando la sua mente, altrettanto dobbiamo dire per un'IA. L'IA potrebbe essere l'oggetto che esegue un programma, per cui un trasferimento creerebbe un altro oggetto, per cui un'altra entità.

- **Conclusioni:** la scienza delle tecnologie emergenti può sfidare e espandere la nostra comprensione filosofica della mente; la filosofia affina la nostra capacità di capire quali sono gli orizzonti di queste tecnologie. Non dare per scontato che le IA saranno coscienti. Ma qualsiasi diritto negato a un'IA cosciente potrebbe ritorcersi contro di noi, poiché potrebbero trattarci come noi le abbiamo trattate.

## ARTICOLO "THE EXILE", WIRED

Tutto è iniziato quando una dirigente di Google ha ordinato a Gebru (ricercatrice di google) di ritirare la sua ultima ricerca, sulle problematiche dei large language models, un nuovo tipo di software AI. Gebru ha chiesto le motivazioni e non le ha mai avute, ha detto che o sarebbero stati più limpidi sul perché questa ricerca era stata considerata da eliminare oppure avrebbe dato le dimissioni. Ha inoltre mandato una mail indirizzata a varie donne del suo dipartimento, accusando Google di silenziare le voci marginali (ovvero soprattutto di donne). Il giorno dopo è stata licenziata, ma Google ha sostenuto e sostiene tutt'ora che in realtà abbia dato le dimissioni lei. Anche la collega di Gebru (sempre del dipartimento di ricerca di etica legata all'IA) Mitchell è stata licenziata dopo che per fare luce sulla faccenda di Gebru aveva creato uno script per recuperare delle note che aveva lasciato Gebru nel proprio account lavorativo Gmail. Tutta questa faccenda illustra che l'IA è profondamente sregolata e che i ricercatori che studiano i possibili danni sociali collegati sono a rischio di esilio. Gebru: "Non sono preoccupata delle macchine che prendano il controllo del mondo, ma piuttosto dell'arroganza e ignoranza nella comunità IA".

Gebru condusse uno studio nel 2018 che evidenziava che dei servizi di riconoscimento del genere delle facce offerti da IBM e Microsoft era quasi perfetto per uomini bianchi ma molto inaccurato per donne di colore; ciò è dovuto probabilmente al fatto che le foto trovate sul web per allenare il riconoscimento facciale rappresentavano in maggioranza uomini bianchi, che statisticamente hanno maggior accesso al web. Questo studio era una dimostrazione di come la IA possa perpetuare ingiustizie sociali e di come quindi è importante uno studio etico a riguardo. Prima di lavorare a Google Gebru lavorò per Microsoft e Apple e con i suoi studi elaborò un framework utilizzabile dagli sviluppatori IA che per la prima volta permetteva di elaborare strutture IA capaci di rispettare etica e giustizia.

Prima di entrare a Google in vari le sconsigliarono di unirsi all'azienda in quanto ancora ritenuta ostile a persone di colore e donne. Accettò il lavoro solo quando Mitchell le offrì la posizione di coleader del Ethical AI Team. La sua speranza era di cambiare questo aspetto di Google, la sua preoccupazione era di non sopravvivere a quell'ambiente. Un esempio del lavoro di ricerca etica del team: trovarono un problema nella funzionalità Gmail Smart Reply (che consiglia brevi risposte per le mail): assunzioni di genere, consigliava "lui" quando nella mail era presente la parola "ingegnere". Il team ottenne grossi successi e riconoscimenti in tutto il mondo accademico, tuttavia le frustrazioni di Mitchell e Gebru crescevano: non venivano invitate a meeting a cui avrebbero dovuto partecipare, non incluse in email threads. Gebru partecipò a molte proteste, tra lei e Mitchell lei era la più attiva.

Un anno dopo Google sviluppò BERT, un nuovo tipo di sistema di machine learning che permette agli algoritmi di allenarsi su enormi volumi di testo presi da internet. Quando un reporter ha chiesto se BERT potesse quindi imparare linguaggi sessisti e razzisti è stato detto di sì, come tutti gli altri tipi di machine learning. Il problema è che i data set di testi da cui imparano sono infinitamente vasti e quindi impossibili da controllare.

Il team di Gebru iniziò a studiare i lati negativi di BERT e il risultato venne pubblicato nella faticosa ricerca: "On the dangers of stochastic parrots: can language models be too big?", che suggeriva che questi software, come i pappagalli potessero imparare a ripetere cose gravi senza capirne minimamente la pericolosità.

Non è chiaro chi ha deciso che questa ricerca andava eliminata né per quale motivo, né per quale motivo dopo una (prevedibile) protesta di Gebru è stato deciso di licenziarla.

Il licenziamento di Gebru ha suscitato molto scalpore in tutto il mondo tech. Ora molti si chiedono se sia possibile svolgere della buona ricerca etica sulla IA in aziende tipo Google. È come se Google avesse detto a chiunque voglia fare questo tipo di ricerca: noi non lo tollereremo.

Gebru ora sta raccogliendo fondi per lanciare un proprio istituto di ricerca indipendente.