

## **Le nuove leggi della robotica. Definizione e difesa dell'esperienza umana**

*Intervista di Teresa Numerico a Frank Pasquale, autore di *New laws of robotics: defending human expertise in the age of AI* (Belknap Press, 2020, 330 pp.). Una versione ridotta di questa intervista è apparsa su "il manifesto".*

**Nel tuo libro parli della necessità di non “contraffare l’umanità” ma potresti spiegare come sia possibile definire concretamente cosa sia l’umanità? Seguendo Bernard Stielger, a sua volta è influenzato da André Leroi Gurhan, possiamo dire che il *principium individuationis* dell’umanità è strettamente legato alle tecnologie che riusciamo a mettere al nostro servizio. Non saresti d’accordo con l’idea che definire l’umano sia un’impresa politica?**

Questa domanda va al cuore del progetto del libro. Penso sia difficile sviluppare un’economia politica dell’automazione senza solide fondamenta metafisiche rispetto alla natura dell’umanità e i limiti degli sviluppi futuri rispetto ai robot e alle loro interazioni con gli esseri umani. Dal punto di vista concreto il successo evolutivo e il livello di abbondanza materiale e culturale che molti di noi possiamo sperimentare è causata dall’intensa coltivazione della tecnologia e se includiamo tra le forme tecnologiche l’alfabeto abbiamo un buon argomento per sostenere la tesi che l’umanità sia definita dalle sue tecnologie. Ma io ho due obiezioni.

La prima riguarda i diritti della disabilità. Una persona con una severa demenza che non sia in grado di interagire coerentemente con la tecnologia non perde per questo la sua appartenenza all’umanità. È il fatto di essere incarnata in un corpo, non inteso dal punto di vista del riduzionismo biologico, che le dà il diritto di essere considerata umana. L’alternativa sarebbe considerare come una discontinuità umana quelli che non sanno più o non possono interagire con la tecnologia. E sarebbe ingiusto.

La seconda obiezione è che ci sono tecnologie che non amplificano la nostra umanità, anzi introducono aspetti inumani che ci rendono simili a macchine. Ne ho parlato nel saggio del 2007 *Technology, Competition, and Values*.

**Cosa intendi per tecnologie che ci rendano inumani?**

Il tema è chiarito anche nel libro *Re-engineering humanity* di Brett Frischmann e Evan Selinger. Si può immaginare un farmaco che consenta alle persone di non dormire mai. Ne ha scritto anche Jonathan Crary in *24/7 Il capitalismo all’assalto del sonno* parlando della pressione del capitale a favore di un farmaco simile. Le persone devono dormire, mangiare prendersi cura dei loro cari, avere delle attività piacevoli. Questo non piace al capitale, che vuole estrarre tutto il possibile plusvalore dal lavoro e non vuole negoziare. Da questa prospettiva un farmaco che evita il nostro bisogno di sonno potrebbe essere considerato come l’attivazione di una possibilità utopistica: essere sempre conscio e in contatto col mondo senza bisogno di riposare. Per me è un vero attacco all’umanità, al ritmo del riposo e alla funzione che il mondo dei sogni ha di processare le informazioni in modi misteriosi. Una parte del progetto del libro è tirare una linea tra le tecnologie che amplificano l’umanità e quelle che la erodono.

**Pensi che questa linea si possa tirare una volta per tutte? Non credi che sia il frutto di una negoziazione dinamica e politica su cosa vogliamo preservare?**

Concedo che ci sia un valore politico cruciale nella tecnologia. Prendiamo costantemente decisioni politiche sull'obiettivo della tecnologia nella vita delle persone e ci sono argomenti politici in favore della necessità che la tecnologia non contraffaccia l'umanità come fanno alcuni bots online che si fanno passare per esseri umani, e sono già qui. Ma rischiamo di perdere risorse importanti dal punto di vista della capacità di persuasione e dell'epistemologia se rinunciamo all'idea di verità e accettiamo che ogni cosa sia politica, cioè decisa da coloro che sono al potere. Se non c'è nessuna possibilità di usare nozioni come verità e libertà, ma dobbiamo solo riferirci alle diverse relazioni di potere che esercitano modalità di imposizione del consenso, allora non avremo argomenti per controbattere le scelte politiche abusive. Non voglio trovarmi in quella situazione nemmeno dal punto di vista ontologico, perché penso possiamo lottare per la verità a proposito di come definire l'umanità.

Nel mio ultimo libro si discute di questi temi nei capitoli su medicina, cura, educazione a proposito di quali valori vogliamo conservare e cosa vogliamo abbandonare perché riguarda solo il passato. Nell'ultimo capitolo parlo dell'arte e della creatività sostenendo che dovrebbero essere riservate agli esseri umani. Analizzo il dibattito circa la possibilità che l'IA possa essere creativa. Tale decisione resta politica. Non possiamo chiudere il dibattito una volta per tutte perché possono cambiare i criteri e il consenso sui campi estetici attuali. Per me siamo legati alle opere d'arte perché vengono dalla storia personale degli artisti, ma possiamo pensare a un futuro in cui l'economia politica dell'IA, concentrata nelle mani delle grandi aziende, possa consentire di riprodurre l'arte meccanicamente. È questa la direzione in cui vanno alcuni algoritmi. Ma si tratterebbe solo di un altro processo di manipolazione sui dati personali per aumentare il profitto.

Lo stato terribile in cui versa la sfera pubblica è un esempio di cosa succede quando si prende una prospettiva post-umana e si accetta con piacere che gli algoritmi di Facebook e Google decidano quello che vediamo e quello che leggiamo e cosa sia popolare. Credo dimostri che stare attenti rispetto all'attribuzione di compiti all'IA in campi come la cura, l'educazione degli esseri umani sia un atto politico che, però, riguardi anche la discussione di cosa sia un'azione morale e virtuosa che vada oltre la determinazione che serve per affermare una scelta di governo.

Non voglio vivere in un mondo in cui tutto è deciso "politicamente" e metto questo termine tra virgolette perché le decisioni dipendono dal potere economico e non si argomenta bene contro quello. Voglio mantenere l'autonomia della sfera accademica per essere libero di esprimere le mie idee sulla bellezza e la giustizia che trascendono ogni sistema politico attualmente incarnato nelle nostre società.

**Nel tuo libro analizzi l'effetto mistificatorio di alcuni robot e software di IA che fanno finta di agire come esseri umani. Sostieni che dobbiamo essere protetti da questa tendenza. Ma, come sai, nell'articolo di Turing sul gioco dell'imitazione (1950), si afferma esplicitamente che per passare il test la macchina deve solo fare finta di essere umana tanto da essere capace di ingannare una giuria di giudici inesperti di programmazione.**

Turing è una specie di fantasma che infesta tutto il libro e alla fine prende la forma del personaggio del romanzo di Ian McEwan *Macchine come noi*. Possiamo accettare il test perché è al centro di tanta parte dello sviluppo dell'intelligenza artificiale, ma possiamo anche rifiutare l'idea che i dispositivi debbano essere un'emulazione abbastanza buona dell'essere umano. A questo proposito vorrei citare un libro di Simone Natale che sta per uscire *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test*, che

argomenta contro la tentazione dell'inganno che ha attraversato tanta parte della storia dell'IA. Le relazioni tra le persone sono basate su una millenaria storia socio-biologica e prendono senso in un ambiente culturale di lungo periodo, pensare di farle valere fuori dal contesto, mantenendo solo le parole e i gesti agiti dalla macchina sarebbe scorretto. Sostituire un essere umano con un dispositivo meccanico, come fa notare Walter Benjamin, a proposito della riproduzione meccanica, priva dell'aura l'espressione artistica, ma anche le persone. Se non possiamo distinguere tra androidi e umani la vita perde valore, perché gli androidi non ne hanno. Contraffare la comunicazione, l'affetto, l'espressione emotiva e l'azione le priva di significato in assoluto perché produce un'inflazione delle loro caratteristiche. Inoltre, non voglio entrare in competizione con un'armata di robot per fare una prenotazione al ristorante, considerando che nel tempo di una telefonata loro ne potrebbero fare trecento, aumentando inutilmente i tempi di attesa per gli esseri umani.

L'ultimo punto critico è definire il livello di riproducibilità tecnica che vogliamo accettare in contesti come la medicina, l'educazione, il giornalismo, il design, ecc. Non solo il contributo umano è molto importante, ma spesso le decisioni in questi ambiti riguardano la mediazione tra valori contrapposti e da tale armonizzazione dipende la conservazione della democrazia, cioè la mediazione tra le persone e i loro valori.

**Se la tecnologia è un sistema per organizzare il mondo al fine di farlo corrispondere alle sue aspettative, prima o poi entrerà in contrasto con la legge rispetto a chi tra gli esperti dei due campi abbia l'autorità per stabilire le regole. Qual è la migliore strategia per vincere nel conflitto come regolatori?**

Per vincere dobbiamo allocare meglio le risorse. La lotta è impari se quelli che devono imporre la regolamentazione hanno risorse cento o mille volte inferiori a quelli che dovrebbero essere oggetto delle loro regole. Abbiamo molto da imparare dai paesi di successo in Asia. USA e Europa devono impegnarsi anche attraverso il sistema di tassazione, per ridurre il peso politico delle aziende da regolare. Non si può chiedere ai regolatori di entrare in azione solo dopo che il disastro è avvenuto. Abbiamo bisogno di un mondo in cui le regole siano in grado di funzionare da apripista rispetto all'evoluzione tecnologica per evitare che accadano disastri sociali che abbiamo già sperimentato in passato.

**Sei convinto che maggiore sia l'impiego di intelligenza artificiale e algoritmi di *machine learning*, più abbiamo bisogno di esperti e professionisti. Come mai?**

Voglio mettere in discussione la narrazione standard sul carattere dirompente dell'intelligenza artificiale, come se l'introduzione di un'app per monitorare la pelle potesse sostituire i dermatologi o si potesse rimpiazzare uno psicologo con un'app per fare la terapia, o un insegnante coi corsi online automatizzati e personalizzati.

Prendiamo il caso degli insegnanti, l'idea dei Moocs (Massive open online courses) è farli diventare la prima linea della formazione, a cui si aggiungerebbero strumenti personalizzati per le necessità individuali. Ritengo che sia un approccio sbagliato, perché l'interazione tra studenti e col docente è estremamente importante per apprendere la cooperazione, la comunicazione e la collaborazione democratica. Tutti aspetti essenziali di una convivenza civile che solo la scuola può insegnare. Bisognerebbe mettere questi dispositivi al servizio delle necessità formative della collettività e solo le competenze dei professori potrebbero riuscire a usarli per amplificare le capacità umane.

**Sostieni che sia preferibile una tecnologia capace di potenziare le capacità umane fornendo solo un supporto per la presa di decisione. Non pensi che questo possa produrre una situazione in cui le macchine prendono le decisioni di cui poi le persone si assumono la responsabilità senza comprenderne le motivazioni, a causa del pregiudizio dell'affidabilità di ogni automazione?**

Questo costituisce un rischio, ma ci sono modi per alleviarlo o combatterlo. Bisogna spiegare alle persone che la maggior parte dei loro problemi sono complessi perché ci sono molte dimensioni tutte incommensurabili tra loro e prendere una decisione implica comprendere questa pluralità. Se si deve misurare l'esito di un intervento di protesi al ginocchio possiamo scegliere tanti aspetti: quanto dolore si è sopportato, quanto tempo ci è voluto per riprendere l'attività, quanto per correre, com'è stata l'esperienza di convalescenza, ma anche altre variabili a lungo termine. Se la protesi manda nel sangue tracce del suo metallo bisogna capire come intervenire. Il monitoraggio di tutte gli elementi richiederà un *expertise* umano. Le tecnologie non possono essere preparate in tutto, né gli esperti di informatica possono sostituire quelli di medicina o di ogni altro campo. Ci sarà sempre bisogno di scelte creative e prospettive originali. La mia più grande paura è che i tecnologi si spaccino per esperti in tutto, mentre gli ematologi devono sempre comprendere cosa succede quando piccoli componenti di metallo finiscono nel sangue e altre cose simili. Per questo ritengo che gli esperti non possano solo seguire le indicazioni della macchina.

Si vede la stessa problematica nei tribunali dove c'è stato in Wisconsin negli USA in un caso *Loomis vs Wisconsin*, su cui ho scritto nel 2017. Il giudice umano aveva attribuito al sistema opaco di predizione la scelta sulla probabilità di recidività del condannato. La compagnia che aveva fornito l'algoritmo aveva usato la legge che protegge il segreto industriale per impedire agli outsider di comprendere bene come stesse valutando la situazione. Si trattava di una IA molto primitiva, ma è stato evidente l'entusiasmo che questo metodo ha causato tra le persone dell'ambito tecnico-legale ritenendolo meno discriminatorio rispetto al giudizio delle persone. Eppure, non sarebbe legittimo anche se fosse meno discriminatorio. Non è legittimo per qualcuno essere condannato a più anni di prigione di un altro solo perché un opaco programma di software in qualche situazione ha trovato una particolare correlazione in un particolare dataset.

L'intelligenza artificiale con la sua scatola nera non è una scienza, ma solo il preludio alla scienza, è il modo in cui si possono generare ipotesi partendo dall'analisi delle correlazioni, ma non trova le cause dei fenomeni. L'addestramento degli esperti deve servire anche per non farli soccombere al pregiudizio dell'automazione.

**Il titolo dell'ultimo capitolo riecheggia il titolo di Joseph Weizenbaum *Il potere del computer e la ragione umana*. Cosa ti ha ispirato di questo libro?**

Sono particolarmente empatico con Weizenbaum perché era concentrato sulla questione della terapia della parola. Una terapia in cui le persone potessero parlare dei loro problemi con professionisti capaci di raggiungere un'incredibile consapevolezza filosofica, psicologica, psichiatrica al fine di aiutarle a comprendere cosa stessero attraversando e sostenerli nel trovare la loro soluzione. La professionalità degli operatori nel comprendere il mondo della sofferenza psichica e supportare le terapie li rende insostituibili da presunte app negli *employee assistance programs* perché queste sono basate su presupposizioni circa le ragioni psicologiche di malessere e conseguentemente identificano metodi standard per risolvere i problemi. La sofferenza può essere causata da una stratificazione di motivazioni. Per dirla in

modo semplice: il problema può essere della persona o riguardare il mondo in cui la persona si trova. Se il potere del *machine learning* e dell'IA è addestrato a convincerci che il problema quando siamo depressi o abbiamo delle ansie è solo nostro e non ha nulla a che fare con il mondo non permetterà mai che la soluzione del malessere riguardi un cambiamento dell'ambiente circostante e non in noi stessi.

**Alla fine del libro affermi la necessità della responsabilità personale per una *governance* legittima. Pensi che abbia a che fare con la capacità di empatia propria degli umani o con l'inevitabile pluralismo dei diversi punti di vista che permette l'innovazione, la sorpresa, la serendipità e imprevedibilità dei comportamenti umani?**

Credo riguardi tutti e due gli aspetti. L'empatia è un elemento cruciale capace di tenere insieme una comunità politica, in quanto costituisce la linfa di appartenenza a un corpo comune. Altrettanto conta l'imprevedibilità delle persone: dobbiamo decidere quello che vogliamo conservare e quello che dobbiamo scartare del passato. Parte della tensione costruttiva del libro si riferisce al riconoscimento di tutti i modi in cui la tecnologia possa aprire nuove opportunità e insieme costituire una sorta di tradimento, nella forma di un'erosione dell'insieme di valori e impegni che sono parte della nostra identità.

**Nel libro si parla di robot, ma non possiamo dimenticare le piattaforme che già dominano l'ambiente digitale. Come possiamo tenere sotto controllo il loro potere?**

Questa domanda si riferisce al lavoro sulla *Black Box Society*. Ho rinnovato la mia posizione in un recente articolo *Internet nondiscrimination principles revisited* (2020). Penso che le piattaforme siano simili alle telecomunicazioni che sono state pesantemente regolamentate per favorire la concorrenza. Dobbiamo creare un'infrastruttura solida che le governi su due profili: assicurarci che non facciano circolare e non amplifichino il discorso d'odio e le teorie cospirazioniste, evitare che prendano il controllo dell'economia nel suo insieme, strangolando le aziende più piccole. Si potrebbe separare Facebook da Whatsapp e da Instagram, così come Google da YouTube. Bisognerebbe contenere la posizione dominante di Apple nel suo app store e impedire a Amazon di possedere sia la piattaforma di e-commerce sia di usarla per vendere le merci. Tali azioni regolative aggressive sono necessarie e sta crescendo l'approvazione tra gli esperti del digitale. Aspettiamo solo la volontà politica di correggere le incongruenze.

### **Riferimenti bibliografici citati nel testo**

- Crary J. (2014) *24/7: late capitalism and the ends of sleep*, Verso, London; trad.it. *24/7 Il capitalismo all'assalto del sonno*, Einaudi, Torino, 2015.
- Frischmann B. e Selinger E. (2018), *Re-Engineering Humanity*, Cambridge University Press, Cambridge, Mass.
- McEwan I (2019) *Machines like me*, Jonathan Cape, London, trad.it. *Macchine come me*, Einaudi, Torino, 2019.
- Natale S. (2021) *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test*, Oxford University Press, Oxford.
- Pasquale F. (2007) "Technology, competition, and values" *Minn. JL Sci. & Tech.* 8: 607-622, <https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1271&context=mjlst>
- Pasquale F. (2015) *Black box society*, Harvard University Press, Cambridge, Mass.

- Pasquale F. (2017) “Secret Algorithms Threaten the Rule of Law”, *MIT Technology review* (1/6/2017) <https://www.technologyreview.com/2017/06/01/151447/secret-algorithms-threaten-the-rule-of-law/>
- Pasquale F. (2019) “The Second Wave of Algorithmic Accountability”, LPE Project (25/11/2019) <https://lpeproject.org/blog/the-second-wave-of-algorithmic-accountability/>
- Pasquale, F. (2020) “Internet Nondiscrimination Principles Revisited” (June 24, 2020). Available at SSRN: <https://ssrn.com/abstract=3634625> or <http://dx.doi.org/10.2139/ssrn.3634625>
- Supiot A. (2015) *La gouvernance par le nombres*, Fayard, Paris.
- Turing A. (1950) *Computing machinery and intelligence*, “Mind”, vol.59, pp.433-460; ristampato in Copeland (2004), *The essential Turing*, Clarendon Press, Oxford, pp.441-464 (trad.it. *Macchine calcolatrici e intelligenza*, in Lolli G. (a cura di) *Intelligenza Meccanica*, Bollati Boringhieri, Torino, 1994, pp.121-57).
- Weizenbaum J. (1976) *Computer power and human reason. From judgement to calculation*, W.H. Freeman and Company, New York; trad. it. *Il potere del computer e la ragione umana*, Edizioni Gruppo Abele, Torino, 1987.