

Dall'azione alla regolamentazione:
autonomia pubblica, automazione
e intelligenza artificiale

*From action to regulation: public
autonomy, automation and artificial
intelligence*

ALESSANDRO VOLPE
volpe.alessandro1@hsr.it

AFFILIAZIONE
Università Vita-Salute San Raffaele e European
Centre for Social Ethics

SOMMARIO

Il tema dell'autonomia di fronte ai sistemi automatizzati e informati da intelligenza artificiale (IA) è al centro di molti dibattiti da diversi anni, soprattutto dal punto di vista della libertà personale e della privacy. Una dimensione non sufficientemente indagata dell'autonomia è, tuttavia, quella *pubblica*, vale a dire quella direttamente legata ai meccanismi di regolamentazione e di direzione collettiva delle "macchine". A tal fine, si farà principale riferimento alle questioni etiche legate ai veicoli a guida autonoma. La tesi che si intende suggerire e difendere in questo contributo è che l'automazione e l'emergere dell'IA non solo non mettano a rischio l'autonomia umana *tout court*, bensì possano potenziare la sua dimensione pubblica (e, quindi, democratica), a patto che gli spazi di decisione collettiva non vengano "colonizzati" dalla stessa autonomia operativa delle macchine. Sfondo teorico di questa tesi è l'etica del discorso e un approccio all'etica dell'IA come teoria critica.

PAROLE CHIAVE

Autonomia
Automazione
Intelligenza artificiale
Etica del discorso

ABSTRACT

The topic of autonomy in the face of automated systems informed by artificial intelligence (AI) has been at the center of many debates for several years, especially from the point of view of personal freedom and privacy. However, a dimension of autonomy that has not been sufficiently investigated is the public one, i.e. the one directly linked to the mechanisms of regulation and collective management of the "machines". To this end, main reference will be made to the ethical issues related to self-driving vehicles. The thesis that we intend to suggest and defend in this contribution is that automation and the emergence of AI not only do not put human autonomy tout court at risk, but can enhance its public (and, therefore, democratic) dimension, provided that the collective decision spaces are not "colonized" by the operational autonomy of the machines themselves. The theoretical background of this thesis is discourse ethics and an approach to AI ethics as a critical theory.

KEYWORDS

Autonomy
Automation
Artificial intelligence
Ethics of discourse

DOI: 10.53267/20230105



1. AUTONOMIE NELLA TRASFORMAZIONE TECNOLOGICA IN ATTO

Il dibattito sull'autonomia personale di fronte alle macchine, all'intelligenza artificiale (IA) e ai sistemi robotizzati è ormai piuttosto avanzato e sfaccettato¹. In questione c'è tanto il margine di libertà umana nell'uso effettivo di sistemi automatizzati informati da IA, quanto la salvaguardia di spazi di autonomia privata *nonostante* l'impiego ormai capillare di questi sistemi: si pensi alla privacy, al consenso informato o alla difesa di libertà fondamentali, in particolare nei casi di profilazione e sorveglianza. Ciò ha reso il valore dell'autonomia, accanto a diversi altri principi, centrale in diverse iniziative, linee guida e regolamenti sull'intelligenza artificiale². Tuttavia, sul piano dell'analisi del suo effettivo restringimento o ampliamento, la questione dell'autonomia di fronte all'automazione va innanzitutto ricompresa entro uno schema che la specifica in una varietà di forme rispondenti alle diverse capacità umane e artificiali, inserite nel quadro di una trasformazione che unisce sempre più l'intelligenza artificiale a un nuovo paradigma industriale e sociale. Tali capacità sono messe effettivamente in opera, favorite o inibite, dalle trasformazioni in corso, soprattutto in una condizione nella quale la vita sociale è contraddistinta da attività mediate dall'automazione e dall'IA: dai trasporti agli acquisti, dal mondo del lavoro alla vita affettiva. Tra i casi di automazione più discussi, soprattutto per i dilemmi morali che comporta, è senza dubbio esemplare quello dei veicoli a guida autonoma³. L'ambito dei trasporti automatizzati e intelligenti offre efficacemente l'idea di un utilizzo di sistemi ad intelligenza artificiale nel quale gli individui, nella fattispecie i passeggeri, si trovano ad essere letteralmente *guidati*, presumibilmente all'interno di una viabilità stradale quasi integralmente adattata a questo genere di veicoli. Com'è ormai noto, i dilemmi morali relativi a questo tipo di veicoli sorgono soprattutto nell'ambito delle cosiddette collisioni inevitabili, situazioni nelle quali la 'scelta' di deviazione è delegata completamente alle macchine. Questi casi-limite esemplificano una condizione pratica nella quale l'azione diretta umana viene ad eclissarsi. La tradizionale responsabilità morale affidata alla sfera personale viene ad essere sostituita dalla responsabilità funzionale del congegno informato da IA, il quale sarebbe in grado di valutare in maniera precisa non solo i riferimenti spaziali e temporali della situazione d'impatto, ma anche l'identità

degli eventuali pedoni coinvolti. Sulla base di questi dati, il veicolo può dunque 'agire' in un senso o in altro, adattando la sua operazione, ad esempio, a modelli di tipo deontologico, consequenzialistico o secondo modelli etici integrati. La leva del famoso dilemma del carrello (*trolley problem*) è pertanto non più un congegno discrezionalmente manipolabile dall'agente morale, bensì un meccanismo interno di *input* e *output* che la macchina possiede e che fa sì che quest'ultima prenda una determinata 'decisione' di tipo etico. Questo aspetto illumina tanto l'assenza di controllo umano in situazioni eticamente dirimenti quanto il suo spostamento all'ambito della regolazione etica delle macchine. Come osserva Guglielmo Tamburrini: «È un errore categoriale ascrivere lode, biasimo e responsabilità morali alle macchine attuali che godono solo di autonomia operativa. La riflessione etica sulle attuali macchine autonome è esclusivamente incentrata sull'uomo come agente morale, su ciò che è bene o giusto che l'uomo faccia avvalendosi dell'autonomia operativa delle macchine, sulle responsabilità e sui giudizi di lode o di biasimo morale che riguardano coloro che sono coinvolti nelle attività di impulso politico ed economico, di ideazione, progettazione, realizzazione, diffusione, gestione e uso di queste macchine»⁴. Naturalmente, la priorità della valutazione morale umana sulle macchine si estende a molti altri casi di automazione legata all'intelligenza artificiale: algoritmi che decidono delle carriere lavorative dei dipendenti d'azienda (assunzione, avanzamento o licenziamento), profilazione in rete su larga scala, fino all'uso di armi autonome in guerra. La perdita di autonomia personale nell'uso di questi sistemi comporta però simultaneamente un guadagno, certo non scontato e inevitabile, in una sfera ulteriore di autonomia in senso allargato, *più orientato alla regolazione che all'azione*. Pertanto, continua Tamburrini⁵, si determina una continuità necessaria tra sfera etica e sfera politica, precisamente nei termini di 'politiche etiche', che, come si vedrà in seguito, coinvolgono tanto i decisori pubblici quanto, in prima istanza, tecnici, sviluppatori e ingegneri. In che modo, tuttavia, il rapporto umani-macchine intelligenti comporta una riconfigurazione dell'autonomia? Qui è utile, come anticipato, ricomprendere il concetto di autonomia entro uno schema che ne definisce gli ambiti in corrispondenza delle differenti capacità umane. In primo luogo, vi è una dimensione peculiarmente umana e individuale dell'autonomia, che riguarda tanto le scelte quanto

l'effettivo controllo sugli oggetti: l'*autonomia personale* è la capacità individuale di deliberare e determinare una scelta *in actu* in base a ragioni, credenze e contesto, in un'accezione anche più ampia di quella kantiana, mentre l'*autonomia d'utilizzo* è invece la capacità di controllo e utilizzo diretto di un certo tipo di macchina o congegno. A queste forme si aggiunge, tuttavia, una forma di autonomia non individuale, bensì collettiva e mediata spesso da procedure, norme o codici, ovvero l'*autonomia pubblica* o *sociale*. Per autonomia pubblica si deve intendere in generale la capacità degli individui di partecipare ai processi decisionali e di influenzare il contenuto delle norme a cui si è soggetti⁶. Queste forme variegata di autonomia umana si distinguono dall'*autonomia operativa* dei sistemi automatizzati, ovvero dalla capacità di una macchina o di un robot di compiere delle azioni senza il diretto intervento umano, a cui si è aggiunto con l'IA un perfezionamento rispetto a funzioni percettive, cognitive, di coordinamento sensomotorio e di interazione con altri agenti intelligenti, così come funzioni di auto-apprendimento su base prevalentemente statistica. Nel caso dei veicoli a guida autonoma, ad esempio, questa autonomia operativa si può differenziare secondo diversi livelli: veicoli ad autonomia crescente (dotati di frenata assistita, mantenimento della corsia di marcia e della distanza di sicurezza, e così via) e veicoli *completamente* autonomi, nei quali non è previsto alcun intervento umano nella guida. In parallelo, resta sempre valida l'ipotesi secondo cui, nel contesto di una piena automazione, possa approfondirsi l'*autonomia performativa* umana, ossia la capacità degli individui di determinare delle scelte di vita, di sviluppare facoltà, bisogni o desideri senza particolari interferenze, legate *in primis* al lavoro⁷. Come già previsto e salutato da diversi autori del Novecento come Herbert Marcuse⁸, ma anche più di recente⁹ – gli avanzamenti tecnologici e in particolare l'automazione avrebbero come effetto quello di estendere la sfera del tempo libero, riducendo non solo gli sforzi fisici ma anche il lavoro sociale necessario. Un dato effettivamente riscontrabile nel quotidiano, soprattutto nella liberazione dei tempi di vita, pur incorrendo in patologie interne alle forme di accelerazione¹⁰. Al di là dell'ottimismo di quest'ultima ipotesi, la tesi che si intende sostenere è che le trasformazioni tecnologiche in atto stiano mettendo in questione un particolare spettro di autonomia umana, in particolare quella *personale* e quella *d'utiliz-*

zo, mentre comportano un sempre maggiore coinvolgimento attivo della sfera *pubblica* dell'autonomia¹¹. Pertanto, l'argomento che si intende difendere è che l'automazione e l'emergere dell'IA non mettano a rischio l'autonomia umana *tout court*, bensì il rapporto positivo che lega l'autonomia personale all'autonomia d'utilizzo, vale a dire il ragionamento e la deliberazione morale individuali, da un lato, e l'effettivo controllo e utilizzo di un determinato oggetto o funzione algoritmica, dall'altro. L'autonomia davanti al combinato tra automazione e IA non si trova né in rischio né in crisi, bensì si troverebbe a subire una riconfigurazione che la slega dall'azione individuale e personale. Questo schema, va detto, si adatta soprattutto alle condizioni di intelligenza artificiale debole o ristretta (*narrow AI*), mentre resta agnostico rispetto alle possibilità che un'intelligenza artificiale forte o generale possa elaborare contenuti di coscienza e pertanto dotarsi di un vero e proprio ragionamento morale indipendente¹².

2. AUTONOMIA PUBBLICA SULL'AUTONOMIA OPERATIVA

In un ipotetico confronto tra ingegneri, tecnici e dirigenti d'azienda sulle modalità di intervento sulla macchina – ad esempio, nella correzione dell'algoritmo in casi di collisioni inevitabili, oppure nei casi di algoritmo discriminatorio – questi soggetti difficilmente si troveranno «nella condizione di poter risolvere unilateralmente» e «dirimere da soli le questioni morali»¹³. In altre parole, è piuttosto inverosimile immaginare che un'azienda o un singolo comparto tecnologico possano prendere delle decisioni di direzione morale delle macchine senza doversi inevitabilmente interfacciare con i decisori pubblici o con le norme che regolano il mercato. La portata delle questioni morali nonché l'effettiva applicazione di un determinato modello etico potrebbero costituire una responsabilità fin troppo elevata e ben superiore all'ambizione 'socialmente responsabile' di un'azienda o di una singola area di ricerca e sviluppo. Un'impossibilità, dunque, di deliberazione unilaterale riferita tanto agli utenti quanto ai costruttori. L'esito è un possibile slittamento procedurale dall'etica individuale all'ambito delle politiche etiche, ovvero, dalle scelte dei singoli alle scelte pubbliche. In tal senso, l'autonomia pubblica si presenta come un'estensione collettiva della valutazione morale individuale, anche se ciò può indurre a ritenerla una dimensione diversa dall'autonomia morale personale soltanto in termini quantitativi, per la presenza di

più soggetti coinvolti in una deliberazione, come se indicasse un insieme di valutazioni individuali. L'autonomia pubblica non fa riferimento però a un mero aggregato di individui che prendono decisioni individuali, bensì una sfera in cui le ragioni di ciascuno devono sapersi offrire anche in una modalità intersoggettiva e secondo un'ottica deliberativa. Sono, infatti, lo scambio di ragioni, di informazioni, di pretese di validità ad essere al centro dell'autonomia pubblica. Aspetto, quest'ultimo, sottolineato e valorizzato esemplarmente dai principali autori dell'etica del discorso¹⁴ così come della democrazia deliberativa¹⁵. Più precisamente, Jürgen Habermas ha qualificato il passaggio dall'autonomia morale di tipo individuale a quella pubblica e intersoggettiva come il transito da un'impostazione monologica di ragione pratica a una *dialogica*. Un'estensione che non sostituisce la sfera personale né la legittimità dell'autonomia personale come criteri normativi, bensì li ricomprende e in ultima istanza li promuove all'interno di un contesto di reciprocità e interazione. In particolare, Habermas ha contraddistinto eticamente questa sfera di interazione in un principio che, richiamandosi alla formula dell'universalizzazione kantiana, deve sapersi tradurre in tutti i contesti di deliberazione morale, secondo cui «ogni norma valida deve soddisfare la condizione che le conseguenze e gli effetti secondari derivanti [...] possano venir accettate da tutti gli interessati». Tale principio può ben adattarsi all'interno dell'ambito di regolazione collettiva dell'automazione informata da IA.

I primi approcci discorsivi e deliberativi ebbero come impulso a suo tempo, tra gli anni Settanta e Ottanta del secolo scorso, l'intensificarsi e il moltiplicarsi delle dinamiche di comunicazione nelle società avanzate, ma l'odierna immersione in sistemi automatizzati, per di più dotati di intelligenza artificiale, può conferire loro ulteriori elementi di adattabilità in un contesto in cui la regolazione sociale delle macchine prende il sopravvento sull'uso personale. Inoltre, l'enfasi delle teorie discorsive è posta precisamente sugli organi intermediari delle decisioni collettive, dalla sfera pubblica alla sfera istituzionale: comitati cittadini, associazioni, commissioni, fino agli organi rappresentativi. Si può fare riferimento qui, come casi felici consultazione etica di esperti ed anticipazione legislativa nel campo della regolamentazione dei veicoli a guida autonoma, alla Commissione etica sulla guida autonoma del Ministero dei Trasporti tedesco del 2017¹⁷ e a quella della Commissione europea del 2020¹⁸. In particolare, i con-

tenuti del rapporto della commissione tedesca mostrano come questo tipo di consultazione conduca anche ad esiti creativi sul piano delle teorie normative impiegate, con un adattamento reciproco tra conclusioni di tipo deontologico e consequenzialistico. L'articolo 9 del rapporto della Commissione tedesca difende, infatti, un approccio etico integrato che include il principio di non discriminazione («in caso di collisione ogni distinzione su caratteristiche individuali è severamente proibita»), il divieto di coinvolgere innocenti («in caso di collisione le parti coinvolte non devono sacrificare altre parti non coinvolte»), prescrivendo una programmazione generale volta a ridurre i danni alle persone e aumentare la sicurezza dei veicoli. Ragionevolmente, la commissione ha preso le mosse dal principio di non discriminazione – presente in tutte le principali costituzioni moderne europee –, così come dall'idea che altri soggetti non coinvolti nella dinamica di collisione siano sacrificati in virtù di un calcolo utilitaristico. L'aspetto consequenzialistico è però conservato dall'indicazione, rivolta alle case automobilistiche nonché ai gestori dell'infrastruttura autostradale, di tutelare e promuovere la sicurezza dei veicoli e della viabilità, per ridurre il più possibile circostanze di pericolo e di collisione.

3. PATOLOGIE DELL'AUTONOMIA PUBBLICA IN RAPPORTO ALL'INTELLIGENZA ARTIFICIALE

L'urgenza sempre maggiore di spazi di decisione democratica e collettiva sulle trasformazioni tecnologiche non determina di per sé che questi spazi godano di buona salute. Al contrario, il funzionamento dei processi di decisione collettiva in senso democratico può essere impedito o minacciato da quelle stesse tecnologie che in linea di principio dovrebbero essere governate e regolamentate. In altri termini, se l'autonomia personale è spesso compromessa nell'utilizzo dei sistemi dotati di IA, se non del tutto inibita, lo spazio dell'autonomia pubblica può rimanere intatto, a patto di a) estendere competenze e formazione etica a tutti i livelli; b) rafforzare il carattere deliberativo dei processi di decisione; c) evitare l'introduzione dei sistemi a IA nello stesso processo decisionale collettivo o la sua sostituzione con sistemi a IA.

In primo luogo, la formazione filosofica ed etica delle professioni va promossa innanzitutto in ragione della salvaguardia e potenziamento dell'autonomia pubblica *in generale*, oltre all'importanza che può assume-

re sul piano dell'allargamento culturale e dello sviluppo delle *soft-skills* utili al mercato del lavoro (importanza giustificata, fino ad adesso, prevalentemente dal lato dell'economia). In tal senso, la formazione etica dei quadri aziendali e dei reparti specializzati dev'essere un'iniziativa tanto favorita dai privati, per ragioni di vantaggio competitivo, quanto dal pubblico, Stato centrale ed enti locali, per favorire un'autonomia democratica sulle trasformazioni tecnologiche¹⁹. Ciò si lega tanto a obblighi di tipo etico, legati alla priorità dell'autonomia personale quanto a ragioni di carattere costituzionale (si vedano gli artt. 2 e 3 della Costituzione italiana).

La proposta enfatizza il ruolo dei costruttori e produttori diretti nel riconoscere le questioni di natura etica, mettendole in rapporto con le normative vigenti, nonché di anticipare e prevederle sulla base delle proprie conoscenze tecniche, come vere e proprie 'vedette etiche'²⁰. La specificità dell'ingegneria è infatti quella di avere come obiettivo il perseguimento di fini (realizzazione di prodotti e processi, progettazione e organizzazione di attività complesse) che hanno una rilevanza sociale riconosciuta. Nella fattispecie, per via delle loro competenze tecniche, solo i produttori e gli ingegneri specializzati possono intravedere e dunque anticipare possibili problemi etici riguardanti l'impiego di intelligenza artificiale e, più in generale, dei sistemi ad automazione. Come suggerisce a tal proposito Tamburrini, «[...] grazie alla loro capacità di reinterpretazione e di anticipazione creativa, lo scienziato e l'ingegnere si trovano spesso nella posizione di vedetta etica tecnologica, capace di allertare le parti interessate o, più in generale, l'opinione pubblica sulle questioni etiche che nascono alle frontiere della ricerca scientifica e tecnologica»²¹. La combinazione tra competenze etiche e tecniche permetterebbe anche una maggiore efficienza nell'individuazione di questioni moralmente rilevanti, giacché molto spesso è proprio l'assenza di competenze tecniche e dunque predittive da parte del teorico (il filosofo morale o politico, così come il decisore pubblico) a determinare un affanno nella comprensione di determinati fenomeni o una rincorsa tardiva nell'affrontarli.

In secondo luogo, rafforzare il carattere deliberativo dei processi decisionali significa dotarli di una forma inclusiva e fondata su pratiche di discorso, informazione e scambio di ragioni ad ogni livello di consultazione e legislazione²². Con ciò si intende anche proporre una visione normativa e multilivello che comprende tanto l'attività

legislativa quanto quella informale delle società civile e delle aziende, più nello specifico dei reparti specializzati in ricerca e sviluppo. Se così non fosse, la funzione regolatrice avrebbe come unica dimensione quella della decisione presa dall'alto, *top-down*, da apparati tecnocratici o autoritari. L'autonomia pubblica sfumerebbe in un'autoreferenzialità tipica dei sistemi chiusi e di dominio, il che conduce anche a una riflessione sul rapporto tra tecnologia e sistemi istituzionali, se non proprio sugli aspetti geopolitici²³. L'aspetto multilivello è peraltro giustificato alla luce delle peculiarità della rivoluzione industriale in atto, diversa sul piano sociologico e di composizione tecnica del lavoro rispetto a quelle precedenti. A differenza della stagione dell'operaio-massa, di tipo fordista e tayloristico, i livelli di influenza diretta sulle caratteristiche del prodotto e dei processi da parte del produttore diretto si sono moltiplicati, e così anche potenzialmente i livelli di decisione rispetto alla qualità e il controllo.

In terzo luogo, sarebbe in vario modo contraddittorio sostituire l'autonomia pubblica con l'autonomia operativa delle macchine. Non c'è necessità di richiamarsi a concezioni di tipo kantiano per sostenere che ci si troverebbe davanti a un'eteronomia della volontà, giacché si tratterebbe di alienare la propria valutazione morale a un 'terzo' artificiale, il che ci porrebbe al di fuori dello stesso terreno morale per condurci a una forma di fideismo di tipo tecnologico. Rientrano in questo problema anche le notevoli pressioni che le aziende tecnologiche, in virtù del loro peso economico, possono esercitare verso istituzioni pubbliche, in modo da indirizzarle verso una regolamentazione più lasca o vere e proprie omissioni di controllo²⁴. I casi recenti di utilizzo di ChatGPT nell'elaborazione delle relazioni in aule parlamentari sono in questo senso da supervisionare²⁵. Ciò non vuol dire che l'utilizzo di IA debba essere del tutto evitato all'interno dei processi decisionali, se non come supporto, ad esempio, nell'elaborazione dello stato dell'arte o di dati utili nella valutazione e sviluppo di un determinato dossier. Più in generale, 'chiedere' alle macchine di autoregolarsi significherebbe pretendere dall'autonomia operativa di svolgere un'operazione emulativa al pari delle altre, e non di 'riflettere' e quindi 'uscire fuori' da sé. Inoltre, dal punto di vista prudenziale, un'IA come supporto acritico alle decisioni pubbliche porterebbe a casi paradossali prossimi al '*paperclip maximizer*', un esperimento mentale ideato da Nick Bostrom²⁶ che aiuta a rendere l'idea dei rischi correlati al raggiungimento

di determinati fini strumentali da parte dell'IA, come, ad esempio, l'uso di tutte le risorse e mezzi per ottenerli, finanche la medesima esistenza umana.

4. CONCLUSIONI: ETICA DELL'IA COME DIFESA CRITICA DEGLI SPAZI DI AUTONOMIA

In conclusione, le criticità che riguardano il corretto funzionamento degli organi regolatori in campo tecnologico emergono precisamente quando le diverse sfere di razionalità in gioco tendono a confondersi o assumere pretese di 'colonizzazione sistemica'²⁷. L'urgenza è pertanto quella di indicare delle clausole interne che preservino lo statuto di autonomia tra la sfera dell'azione umana e quella artificiale, in un'ottica di complementarità e non di semplice contrapposizione. Affermare che l'automazione e l'intelligenza artificiale siano al 'nostro servizio' vuol dire anche aspettarsi che i sistemi intelligenti non siano 'come noi', bensì esattamente diversi e migliori nello svolgere determinate funzioni operative. Per tornare all'esempio della guida autonoma, il punto essenziale non è quello di recuperare margini di controllo umano sul veicolo, riportando 'la mano sul volante': verrebbe meno il senso dell'innovazione tecnologica in questione, con anche i suoi potenziali vantaggi in termini di sicurezza a lungo termine. Il punto è tuttavia quello di far *agire bene* la macchina, come noi vorremmo, in uno scambio continuo di giustificazioni. Non è dunque del tutto errato avanzare l'ipotesi che l'etica dell'intelligenza artificiale coincida in parte o del tutto con una teoria critica dell'IA²⁸, se con quest'ultima si intende una riflessione interdisciplinare orientata all'emancipazione, e nello specifico, alla difesa e alla promozione di un'autonomia umana residua o potenziale in rapporto alle tecnologie, tenendo anche conto delle relazioni di potere in gioco.

NOTE

1. Lawless William Frere et al., *Autonomy and Artificial Intelligence: A Threat or Savior?*, Berlin: Springer, 2017.

2. A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, 1 (2019): 389–399. doi: <https://doi.org/10.1038/s42256-019-0088-2>

3. Si vedano Guglielmo Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, (Roma: Carocci, 2020); Fabio Fossa, *Ethics of Driving Automation: Artificial Agency and Human Values*, Cham: Springer, 2023.

4. Guglielmo Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, p. 64.

5. Ibid, cap. 2.

6. Si vedano, a tale proposito, Jürgen Habermas, *Fatti e norme*, Roma-Bari: Laterza, 2012; Christopher Zurn, "Private and Public Autonomy", in Amy Allen e Edoardo Mendieta (a cura di), *The Cambridge Habermas Lexicon*, Cambridge: Cambridge University Press, 2012, pp. 348-351.

7. Sfera dell'autonomia che può coincidere con la nozione di libertà negativa, come sfera di non-interferenza, cfr. Isaiah Berlin, "Two Concepts of Liberty", in Id., *Four Essays on Liberty*, London: Oxford University Press, 1969, pp. 118-72.

8. Herbert Marcuse, *L'uomo a una dimensione. L'ideologia della società industriale avanzata*, Torino: Einaudi, 1967.

9. Luciano Floridi et al., "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", *Minds & Machines* 28, (2018): 689-707. <https://doi.org/10.1007/s11023-018-9482-5>. Così anche da una prospettiva accelerazionista di stampo neo-marxista, N. Srnicek, A. Williams, *Inventare il futuro*, Nero Edizioni, Roma, 2018.

10. Hartmut Rosa, *Accelerazione e alienazione. Per una teoria critica del tempo nella tarda modernità*, Torino: Einaudi, 2015.

11. Questa tesi, ribadiamo, si applica a un contesto di automazione e intelligenza artificiale, non all'ambito più generale della rivoluzione digitale, nel quale le responsabilità morali individuali possono restare intatte e risultare persino più cogenti, se si pensa all'uso individuale dell'uso della rete e soprattutto dei media sociali.

12. Questo assunto è puramente di carattere metodologico, per sottolineare il carattere limitato della presente analisi a un ambito di IA orientato a compiti specifici.

13. Guglielmo Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, cit., p. 31.
14. Jürgen Habermas, *Fatti e norme*, (Roma-Bari: Laterza, 2012); Rainer Forst, *The Right to Justification: Elements of a Constructivist Theory of Justice*, New York: Columbia University Press, 2012.
15. Joshua Cohen, "Deliberation and Democratic Legitimacy," in *The Good Polity: Normative Analysis of the State*, a cura di Alan Hamlin e Philip Pettit, Oxford, UK: Basil Blackwell, 1989, pp. 17-34; Robert A. Dahl, *La democrazia e i suoi critici*, Roma: Editori Riuniti, 1990.
16. Jürgen Habermas, *Etica del discorso*, Roma-Bari: Laterza, 1990, p. 71.
17. Federal Ministry of Transport and Digital Infrastructure, "Report, Ethics Commission Automated and Connected Driving (extract)", 2017. https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission-automated-and-connected-driving.pdf?__blob=publication-file
18. Horizon 2020 Commission Expert Group to advise on specific ethical issues raised by driverless mobility (E03659), "Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility". Publication Office of the European Union: Luxembourg, 2020. <https://op.europa.eu/en/publication-detail/-/publication/89624e2c-f98c-11ea-b44f-01aa75ed71a1/language-en>
19. Si può fare qui l'esempio delle attività di consulenza su IA per aziende della società italiana Ammagamma, composta da team multidisciplinari di ingegneri, matematici, storici filosofi e designer. Il suo approccio di consulenza aziendale si ispira a una 'concezione umanistica dell'intelligenza artificiale'. Casi studio e di successo sono consultabili all'indirizzo www.ammagamma.com
20. Guglielmo Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, cit., p. 77.
21. Ibid.
22. Questa idea di autonomia pubblica sull'automazione informata da IA può essere anche sensibile a un adattamento non occidentale del concetto di autonomia, in termini di autonomia relazionale, cfr. Sábèlo Mhlambi e Simona Tiribelli, "Decolonizing AI Ethics: Relational Autonomy as a Means to Counter AI Harms", *Topoi* 42, 3 (2023): 867-880. Il carattere elastico e fondato su un'interazione costante che contraddistingue l'autonomia pubblica apre spazi di interpretazione creativa delle differenti comunità nell'adottare e introdurre politiche etiche dell'IA.
23. Si veda, a tale proposito, Alessandro Aresu, *Le potenze del capitalismo politico. Stati Uniti e Cina*, Milano: La Nave di Teseo, 2012.
24. Si vedano le pressioni di OpenIA sull'Unione europea (UE), rispetto alla nuova regolamentazione UE in tema IA: <https://time.com/6288245/openai-eu-lobbying-ai-act/>
25. Mi riferisco a vicende accadute negli ultimi mesi nelle sedi parlamentari di Italia, Danimarca, Gran Bretagna e Giappone, prevalentemente a scopo provocatorio o sperimentale. A titolo di esempio si veda David Deans, "ChatGPT: "Welsh politician uses AI chatbot to write speech", *BBC News*: <https://www.bbc.com/news/uk-wales-politics-65976541>;
26. Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford: Oxford University Press, 2013.
27. Mi riferisco qui all'analisi critica che Habermas compieva rispetto all'intrusione dei sistemi di tipo strumentale nel mondo della vita, cfr. Jürgen Habermas, *Teoria dell'agire comunicativo* (2 voll), Bologna: Il Mulino, 1985.
28. Rosalie Waelen, "Why AI Ethics Is a Critical Theory", *Philosophy and Technology* 35, 9 (2022). doi: <https://doi.org/10.1007/s13347-022-00507-5>