

# Guida etica all'intelligenza artificiale



Nel prossimo futuro le macchine intelligenti assumeranno un ruolo sempre più significativo nell'esecuzione dei nostri incarichi, con molte attività che richiederanno la collaborazione tra esseri umani e intelligenza artificiale (IA). In particolare, l'IA solleva sfide etiche da due prospettive diverse: una comune a tutte le tecnologie, che riguarda i suoi potenziali benefici e svantaggi, e un'altra derivante dall'acquisizione da parte delle macchine di specifiche capacità cognitive, come il giudizio, il processo decisionale e l'iniziativa, tutte con implicazioni etiche. Questo insieme di fattori solleva preoccupazioni significative riguardo all'impatto degli agenti intelligenti su individui, società e imprese, nonché interrogativi sulle possibili conseguenze.

ETICA//INTELLIGENZA ARTIFICIALE (IA)//ETICA PER DESIGN//ETICA PER APPRENDIMENTO//SISTEMI IBRIDI



## SALVATORE VICARI

è Professore ordinario di Economia e Gestione dell'Innovazione presso l'Università Bocconi di Milano e presidente del Comitato Community e Social Engagement.

La discussione sull'etica nell'intelligenza artificiale abbraccia diverse prospettive, tra cui quella filosofica, religiosa e sociale. In questa sede ci concentreremo in particolare sulle conseguenze, positive e negative, legate all'impatto dell'IA su individui, società e imprese. Eviteremo di affrontare la questione relativa allo status morale derivante dall'intelligenza intrinseca di queste tecnologie, e se in un futuro non troppo lontano possano essere riconosciuti loro dei diritti per proteggerne l'"individualità cognitiva". Parimenti, eviteremo di addentrarci nel dibattito circa la vera essenza dell'intelligenza nell'ambito dell'intelligenza artificiale. Nonostante questo dilemma filosofico coinvolga già numerosi esperti in tutto il mondo, di seguito forniremo alcuni chiarimenti rilevanti

ai fini della nostra discussione.

È importante riconoscere che l'intelligenza artificiale si distingue profondamente dall'intelligenza umana sotto due aspetti principali. In primo luogo, l'intelligenza umana si caratterizza per la sua intrinseca non finalità, esistendo indipendentemente da specifici obiettivi. L'intelligenza artificiale, d'altra parte, assume significato e valore attraverso la finalità a cui è orientata, operando come strumento o un mezzo verso uno scopo. Questa distinzione sottolinea che l'intelligenza umana non dovrebbe mai essere considerata solo come un mezzo per raggiungere un fine, mentre l'intelligenza artificiale trova la sua essenza nell'utilità.

In secondo luogo, l'autocoscienza, che rappresenta un tratto distintivo della coscienza umana, pur potendo essere simulata fino a un certo punto nelle reti neurali avanzate, non raggiunge mai la complessità o la profondità dell'autocoscienza umana. L'autocoscienza umana, intrecciata a caratteristiche come l'amore, la ricerca di significato, la spiritualità e il senso del divino, distingue in modo univoco l'esperienza umana da quella di qualsiasi altra entità o oggetto. La consapevolezza di sé, così come è vissuta dagli esseri umani, incorpora una profondità che trascende la mera funzionalità.

## ETICA E INTELLIGENZA ARTIFICIALE

L'intelligenza artificiale solleva questioni etiche da due prospettive principali: la prima riguarda l'impiego generale di questa tecnologia, mentre la seconda è legata alle sue funzioni cognitive avanzate, come il processo decisionale e l'abilità di avviare azioni autonome, entrambe con rilevanti implicazioni etiche. Quando si parla di IA ed etica, è importante operare una distinzione tra due ambiti: l'etica per l'IA e il comportamento etico dei sistemi di IA. *L'etica per l'IA* si riferisce ai doveri e alle responsabilità morali degli individui coinvolti nella progettazione, nello sviluppo, nell'impiego o nell'interazione con i sistemi di IA, una sfera che

non è esclusiva dell'IA, ma si estende all'uso di qualsiasi tecnologia. Invece, il *comportamento etico* dei sistemi di IA introduce una dimensione etica nuova e senza precedenti nella storia dell'umanità.

Questo articolo si concentrerà principalmente su quest'ultimo aspetto – il comportamento etico dei sistemi di IA – riconoscendolo come una frontiera che pone sfide uniche e richiede un'attenta considerazione nel contesto di un panorama tecnologico in continua evoluzione.

## ETICA PER L'INTELLIGENZA ARTIFICIALE

La tecnologia rappresenta un'arma a doppio taglio: se da una parte offre opportunità di progresso e soluzioni a una molteplicità di problemi, dall'altra porta in sé il rischio di essere utilizzata per scopi distruttivi, arrivando persino a costituire una minaccia per l'esistenza stessa dell'umanità. Questo paradosso etico nasce dalla natura ambivalente della tecnologia, capace di essere impiegata per il bene o per il male a seconda delle intenzioni di chi la sviluppa e di chi la utilizza. In essenza, la tecnologia è uno strumento moralmente neutro: le sue implicazioni etiche sono determinate da chi la progetta, la implementa e la adopera. La responsabilità morale ricade, quindi, sugli individui che maneggiano questi strumenti, non sulla tecnologia in sé, un concetto che si riflette nell'esempio del coltello, utilizzabile tanto in cucina quanto come arma, o dell'energia nucleare, sfruttabile per generare elettricità o causare distruzioni di massa.

I dilemmi etici associati all'intelligenza artificiale sono di fondamentale importanza a causa delle sue capacità senza precedenti. Da un lato, i sistemi di IA offrono prestazioni di straordinaria precisione ed efficienza, rivoluzionando il modo in cui svolgiamo molte attività. Dall'altro, queste capacità avanzate nascondono intrinseche vulnerabilità che possono essere sfruttate per scopi non etici. Ad esempio, l'IA può essere impiegata per mettere in atto campagne di phishing altamente mirate, sviluppare sofisticate tecniche di stalking



e generare *deepfake* – contenuti sintetici, siano essi audio, video o testuali artificiali, concepiti per imitare soggetti reali con l'intento di ingannare chi li fruisce.

Isaac Asimov, celebre autore di fantascienza, concepì un futuro in cui i robot sarebbero stati governati da una serie di “leggi” integrate nella loro programmazione, impedendo loro di nuocere agli esseri umani. Tuttavia, poiché l'IA continua a essere sviluppata senza adeguate salvaguardie contro i danni potenziali, questa visione rimane lontana. Non solo i sistemi di IA vengono programmati senza l'integrazione di chiari confini etici, ma vi è anche un'inquietante tendenza a progettare robot per scopi militari, sollevando profonde preoccupazioni sull'avvento di armi autonome in grado di porre fine a vite umane senza una supervisione diretta. Queste considerazioni etiche si estendono oltre l'IA stessa, toccando le azioni delle persone coinvolte nel suo sviluppo e nella sua implementazione: gli imprenditori devono affrontare i dilemmi morali legati all'uso di robot industriali che potrebbero causare la perdita di posti di lavoro o ostacolare future opportunità di impiego, così come le autorità di regolamentazione hanno il compito di stabilire un quadro etico per lo sviluppo e l'uso dei sistemi di intelligenza artificiale, in modo da garantire che la loro applicazione sia a favore e non a scapito dell'umanità.

Pur presentando potenziali rischi di uso improprio, l'intelligenza artificiale offre la promessa di promuovere risultati positivi, anche in termini di comportamento etico. Ad esempio, l'intelligenza artificiale può svolgere un ruolo fondamentale nell'individuazione e nella prevenzione di attività illecite nei mercati finanziari, come il riciclaggio di denaro, grazie alla sua capacità di analizzare grandi volumi di dati e riconoscere schemi sospetti con precisione superiore a quella umana. Nel campo della medicina, l'IA può migliorare l'accuratezza diagnostica, personalizzare i piani di trattamento e rilevare i primi segni di malattia.

Oltre a questi ambiti, l'intelligenza artificiale offre la possibilità di rafforzare l'etica individuale. Attraverso l'analisi di dati comportamentali, può identificare tendenze che preludono a comportamenti non etici e attivare misure preventive. Ad esempio, monitorando il comportamento alla guida, l'IA può rilevare abitudini pericolose e intervenire per correggere o evitare manovre rischiose, aumentando così la sicurezza stradale. Nel contesto aziendale, l'IA ha il potenziale di arricchire i programmi di formazione, guidando i dipendenti verso la pratica di principi etici solidi nelle attività professionali.

## INTELLIGENZA ARTIFICIALE ETICA

Una preoccupazione emergente nel campo della tecnologia riguarda quella che potremmo definire l'IA etica. Questo concetto non si limita a considerare il comportamento degli operatori umani, ma si estende anche al comportamento delle macchine intelligenti stesse.

Al fulcro di questo dilemma etico risiede un elemento di particolare importanza: l'autonomia degli agenti intelligenti. Una volta che un sistema di IA è stato addestrato a svolgere un compito specifico, gli esseri umani abbandonano il controllo sui processi decisionali e sui comportamenti, dando vita a quello che può essere descritto come uno scenario di “scatola nera”. Le macchine autonome

trascendono le intenzioni dei loro creatori e utenti e possiedono capacità cognitive simili a quelle degli esseri umani. L'autonomia, insieme ad altre funzioni cognitive, sta emergendo come punto focale del discorso etico, rendendo necessario un esame più approfondito delle sue implicazioni<sup>1</sup>.

Al di là delle considerazioni etiche tipicamente associate alla tecnologia, un dilemma etico più profondo riguarda la capacità dell'intelligenza artificiale di decidere e agire in modo autonomo, con la conseguente possibilità di fare delle scelte moralmente discutibili. Ciò sottolinea l'urgenza di affrontare le questioni irrisolte sulle implicazioni etiche dell'intelligenza artificiale, motivo per cui le preoccupazioni espresse da figure di spicco dell'industria hi-tech, come Bill Gates (Microsoft), Elon Musk (X.AI), Sundar Pichai (Google), Geoffrey Hinton (Google) e i numerosi firmatari che chiedono una moratoria sullo sviluppo dell'intelligenza artificiale, meritano una seria considerazione. Essi sostengono la necessità di istituire protocolli di sicurezza universali, progettati, implementati e convalidati con l'aiuto di esperti imparziali.

Un esempio emblematico dei rischi derivanti dall'autonomia dell'IA è rappresentato dal caso delle auto a guida autonoma, che evidenzia l'urgente necessità di mitigare questo specifico rischio etico e la complessità di implementare misure di sicurezza adeguate. Nel 2018 un veicolo sviluppato da Uber e dotato di un sistema di guida autonoma basato sull'intelligenza artificiale è stato coinvolto in un incidente mortale con un pedone, evidenziando i potenziali rischi per la sicurezza umana associati alle decisioni autonome delle macchine. Un altro esempio rilevante riguarda il sistema giudiziario, dove l'intelligenza artificiale è stata utilizzata per assistere i giudici nelle decisioni sulla cauzione e nella previsione dei tassi di recidiva. È plausibile ipotizzare che in futuro tali responsabilità possano essere completamente trasferite a questi sistemi, con conseguenze etiche di rilievo.

Le preoccupazioni etiche vanno oltre la mera

considerazione dei danni potenziali e si estendono a interrogativi morali, profondamente influenzati da contesti religiosi e socioculturali. Questi dilemmi sono intrinsecamente soggettivi e variano a seconda dei valori individuali. È altresì cruciale distinguere tra divieto e permesso, delineando non solo ciò che è vietato alle macchine, ma anche ciò che è loro consentito.

Consideriamo, ad esempio, lo scenario del prelievo di organi da una persona sana per aiutare un individuo malato. Sebbene di norma la società scoraggi l'auto inflizione di danni al proprio corpo, accetta l'atto altruistico della donazione di organi. Al centro del nostro quadro etico risiede il principio dell'autonomia degli individui nel sottoporsi a danni autoinflitti per perseguire un obiettivo moralmente superiore, come quello di aiutare gli altri. Tuttavia, dare alle macchine la capacità di stabilire obiettivi solleva la questione del concedere loro la libertà di arrecare danni per un bene "superiore".

Queste sfide possono essere affrontate in due modi principali: dando all'intelligenza artificiale la capacità di allineare il proprio comportamento con le norme e i valori etici di una determinata società, rispecchiando di fatto il ragionamento etico umano in quel contesto; oppure, in alternativa, delegando la risoluzione di questi dilemmi etici a un'autorità superiore sancita dalla società per questo specifico scopo, seguendo il modello *bottom-up* o il modello *top-down*.

Un dilemma etico più profondo riguarda la capacità dell'intelligenza artificiale di decidere e agire autonomamente, con la conseguente possibilità di fare scelte moralmente discutibili.

1. Si veda Grando e Vicari (2021) per ulteriori letture sulle capacità cognitive dell'IA.

## I DUE APPROCCI ALL'ETICA DEGLI AGENTI INTELLIGENTI

Gli agenti intelligenti introducono un livello di disimpegno da parte degli utenti, che spesso non comprendono pienamente il modello cognitivo della macchina, nemmeno durante l'iniziale fase di addestramento. Alcune sfide, infatti, si possono risolvere semplicemente grazie ai processi di apprendimento predisposti nei sistemi. Ad esempio, nel contesto del pilota automatico di un'automobile, l'addestramento può consentire il rispetto delle leggi sul traffico e delle norme precauzionali di base, come la riduzione della velocità su strade bagnate. Queste linee guida consentono alle macchine intelligenti di allineare le loro decisioni con le regole stabilite dalla comunità, mitigando così il potenziale comportamento scorretto (Etzioni e Etzioni, 2017).

Tuttavia, non tutte le decisioni si risolvono con la stessa semplicità; alcune richiedono un'analisi etica più approfondita. Si consideri lo scenario di un'auto con guida autonoma coinvolta in un incidente inevitabile, in cui diventa fondamentale determinare il male minore: un classico esempio di *trolley problem*, o problema del carrello ferroviario (Foot, 1967). Ad esempio, l'auto autonoma potrebbe trovarsi di fronte a un ostacolo improvviso e non essere in grado di frenare in tempo, dovendo scegliere tra mantenere la sua traiettoria, rischiando di colpire i pedoni o deviare nel traffico in arrivo, mettendo così a rischio altre persone alla guida, o addirittura scontrarsi con una barriera stradale, correndo il rischio di danneggiare i passeggeri.

Esistono due approcci principali per affrontare dilemmi etici di questo tipo: il primo, noto come etica per design, consiste nell'incorporare una serie di regole rigide, simili alle "leggi di Asimov", nella fase di progettazione del veicolo, che impongono l'adesione a linee guida etiche predefinite; l'approccio alternativo, invece, consiste nel consentire alla macchina di generare internamente principi etici attraverso processi di apprendimento, un paradigma noto come etica per apprendimento.

## ETICA PER DESIGN

Il primo approccio, noto come etica per design, prevede l'integrazione delle regole etiche direttamente nella progettazione della macchina, incorporandole nell'architettura e nella funzionalità del sistema. Questo approccio si contrappone all'etica per apprendimento, in cui la macchina sviluppa la propria bussola morale attraverso l'apprendimento esperienziale e le interazioni con l'ambiente. Il concetto di fondo è che i principi morali o le teorie etiche possono servire come linee guida per scegliere le azioni più appropriate. Attingendo da discipline quali religione, filosofia morale e psicologia, i ricercatori hanno identificato diversi "fondamenti morali" per fornire un quadro di base che può essere applicato in situazioni diverse (Graham et al., 2013). Questi fondamenti includono cinque dicotomie di base, integrate da una sesta: cura/danno (aiutare o recare danno agli altri), correttezza/tradimento (mantenere relazioni equilibrate), lealtà/tradimento (essere fedeli o tradire), autorità/sovversione (rispettare o sfidare l'autorità), santità/degradazione (mantenere l'integrità o compromettere la morale), libertà/oppressione (concedere autonomia o esercitare oppressione).

Sebbene questo approccio sembri semplice in teoria, la sua implementazione pratica si rivela estremamente complessa. I principi morali basilari spesso si confondono in scenari complessi in cui non tutte le eventualità sono prevedibili e non tutte le probabilità o utilità sono facilmente calcolabili. Inoltre, possono sorgere conflitti tra valori etici, come ad esempio nella scelta di salvare una vita a scapito di un'altra in situazioni di vita o morte. Va notato che questi dilemmi etici riguardano non solo il comportamento degli agenti artificiali, ma anche le azioni umane. In tali circostanze, è preferibile avere regole imperfette o limitate piuttosto che operare senza alcuna regola. Queste regole possono costituire un punto di partenza su cui gruppi di esperti, norme sociali, considerazioni religiose e sensibilità individuali possono contribuire collettivamente alla formulazione di principi etici generali.

Per affrontare le sfide poste dall'approccio *top-*

*down*, vengono comunemente proposti due quadri etici, ovvero quello consequenzialista e quello deontologico. Il primo approccio si concentra sulla valutazione delle conseguenze positive e negative di un'azione, con l'obiettivo di massimizzare i risultati positivi e minimizzare gli effetti negativi. Tuttavia, la valutazione dei pro e dei contro può essere soggettiva e difficile da quantificare con precisione, motivo per cui l'approccio deontologico propone invece l'aderenza ai principi e alle regole che governano le azioni, garantendo la coerenza con i principi etici e il rispetto dei diritti fondamentali. Seguendo questo quadro etico, quindi, quando ci si trova di fronte ad azioni contrastanti, si invoca un principio di ordine superiore per risolvere il dilemma.

Una valutazione completa di tutte le potenziali conseguenze è complessa e non sempre fattibile. Alcune situazioni, infatti, non possono essere adeguatamente rappresentate in termini astratti e richiedono decisioni specifiche al contesto. Occasionalmente, le complessità etiche possono essere prive di soluzioni definitive. Ad esempio, il valore della vita trascende l'aspetto individuale e comprende una serie di fattori contestuali che influenzano la nostra valutazione. Prospettive diverse, dettate dalla religione, dalla cultura e dalle convinzioni personali, possono condurre a interpretazioni e soluzioni differenti.

## ETICA PER APPRENDIMENTO

Un approccio distintivo al processo decisionale etico è l'etica per apprendimento, che consente agli agenti intelligenti di sviluppare un'agenzia morale, permettendo loro di distinguere tra giusto e sbagliato e di prendere autonomamente decisioni moralmente valide. Questo approccio promuove la creazione di sistemi intrinsecamente etici, in cui i principi etici sono incorporati nel quadro operativo di base e guidano i processi decisionali. I sostenitori di questo approccio vedono il processo di sviluppo dei principi etici negli agenti intelligenti come il processo di apprendimento di un bambino. Immagmano sistemi di apprendimento in cui le reti neurali ricevono

rinforzi positivi e negativi e acquisiscono principi etici attraverso l'apprendimento esperienziale. Ciò comporta l'esposizione a diversi scenari e la ricezione di *feedback*, sia positivi sia negativi, per modellare la comprensione etica. L'attuazione di questo approccio si confronta tuttavia con sfide significative, specialmente a causa del lungo processo di apprendimento coinvolto e della limitata rappresentatività delle situazioni reali ricreate in laboratorio. Lo sviluppo di tali procedure non solo richiederebbe uno sforzo considerevole ma anche un investimento significativo di risorse, con risultati altamente incerti. D'altra parte, le reti neurali offrono una prospettiva alternativa, essendo in grado di apprendere direttamente dai dati del mondo reale, superando così i vincoli delle situazioni simulate in laboratorio. Queste reti possono analizzare vasti insiemi di dati relativi alle azioni umane e ai giudizi etici associati, i quali emergono dall'interazione collettiva nella società umana. Benché le reti neurali non assimilino direttamente concetti morali astratti, esse apprendono osservando il comportamento umano e identificando i modelli di comportamento considerati eticamente accettabili o inaccettabili.

Pertanto, l'etica mediante apprendimento si fonda sull'osservazione del comportamento umano piuttosto che sulla rigida codifica di regole morali. Questo permette alle macchine di esaminare ampi set di dati relativi a una vasta gamma di azioni umane in contesti diversi, riconoscendo la varietà dei giudizi etici umani influenzati da fattori come la religione, il background culturale, l'esperienza individuale e le norme sociali. Diverse comunità culturali e religiose possono adottare quadri etici differenti, e le reti neurali sono in grado di individuare e adeguarsi a tali codici e norme prevalenti in gruppi sociali distinti. In società caratterizzate da tradizioni culturali, credenze etiche e affiliazioni religiose eterogenee, le macchine manifesterebbero modelli di comportamento articolati, capaci di riflettere la diversità etica intrinseca alle società umane con cui interagiscono.

Una sfida notevole a questo approccio, tuttavia, è l'ingente quantità di dati richiesta dall'intelligenza artificiale per sviluppare una solida comprensione del comportamento etico. Non è garantito che il

sistema incontri una varietà completa di situazioni necessarie per sviluppare pienamente la propria bussola morale e navigare con successo dilemmi etici complessi. Inoltre, il processo decisionale individuale tende spesso a privilegiare l'interesse personale. Da uno studio di Bonnefon, Azim e Iyad (2016) emerge che la maggioranza degli intervistati ha espresso preferenza per l'acquisto di un'auto autonoma programmata per dare priorità alla salvezza dei passeggeri rispetto alla vita di un maggior numero di pedoni. Un ostacolo significativo associato all'approccio *bottom-up*, quindi, è la probabilità che le imperfezioni e i pregiudizi insiti nel comportamento umano si riflettano nelle macchine che apprendono da tale comportamento. Ad esempio, i sistemi di IA progettati per la selezione del personale apprendendo dal comportamento umano potrebbero interiorizzare e perpetuare pregiudizi, soprattutto quelli legati al genere e alla razza, purtroppo diffusi nelle società umane.

## UNA POSSIBILE SOLUZIONE: I SISTEMI IBRIDI

Come si è visto, sia l'etica per design sia quella per apprendimento presentano limitazioni significative. Un approccio alternativo potrebbe essere quello di delegare l'autorità decisionale al giudizio umano, specialmente in scenari troppo complessi per una regolamentazione *ex ante* efficace o per una risoluzione autonoma. Questa strategia potrebbe tuttavia non essere sempre attuabile, poiché la concessione un'eccessiva autonomia decisionale ai sistemi di intelligenza artificiale potrebbe compromettere la possibilità di esercitare una supervisione umana, generando difficoltà nella gestione dei dilemmi etici che possono emergere tra la decisione e la sua implementazione. I limiti degli approcci esistenti sottolineano l'importanza di adottare una strategia più sofisticata. Infatti, nonostante gli agenti intelligenti siano capaci di prendere decisioni autonomamente, ciò non implica che siano in grado di farlo senza un quadro di riferimento. La combinazione degli approcci di etica per design e etica per apprendimento

consente di stabilire confini e linee guida coerenti con i principi etici predefiniti. L'autonomia non presuppone una libertà assoluta dall'intervento umano; è possibile stabilire limiti chiari entro cui i sistemi intelligenti devono operare. Una volta definiti, il sistema può determinare autonomamente metodi e risposte appropriate alle situazioni contingenti. Per superare il divario tra i due approcci, è quindi necessario adottare un sistema ibrido, integrando entrambi i metodi in un'unica cornice. Le linee guida etiche dovrebbero essere esplicitamente delineate, soprattutto per le aree in cui l'apprendimento diretto è difficile, mentre i sistemi di intelligenza artificiale possono acquisire autonomamente una comprensione morale dal comportamento umano e dalle norme sociali. Anziché escludersi a vicenda, questi approcci dovrebbero essere efficacemente integrati per garantire che le macchine esercitino l'autorità decisionale basandosi sull'apprendimento guidato dai dati, all'interno di un quadro normativo sociale stabilito.

### MANAGERIAL IMPACT FACTOR

- **Stabilire linee guida etiche precise:** sviluppare e formalizzare linee guida etiche che delineino in modo chiaro regole e limiti imprescindibili entro cui l'IA può operare. Queste direttive dovranno essere basate sia su principi etici ben definiti sia sull'apprendimento continuo del comportamento umano e delle normative sociali;
- **Sviluppare capacità di apprendimento adattivo:** investire in tecnologie e metodologie che consentano all'IA di apprendere dalle interazioni del mondo reale e di adattare le sue operazioni all'interno dei quadri etici stabiliti;
- **Revisione e aggiornamento continui:** rivedere e aggiornare periodicamente le linee guida etiche e gli algoritmi di apprendimento per riflettere i cambiamenti nelle norme sociali, nei requisiti normativi e nei progressi tecnologici.
- **Implementare meccanismi di supervisione umana:** stabilire protocolli per consentire un intervento umano immediato, in particolare nei processi decisionali complessi, per garantire che le decisioni siano coerenti con le norme etiche e sociali;

RIFERIMENTI  
BIBLIOGRAFICI

- Allen, C., Smit, I., Wallach, W. (2005). "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches." *Ethics and Information Technology*, 7, pp. 149-155. [doi](#).
- Benanti, P. (2018). *Le macchine sapienti. Intelligenze artificiali e decisioni umane*, Marietti 1820, Bologna.
- Bonnefon, J.F., Shari, A., Rahwan, I. (2016). "The Social Dilemma of Autonomous Vehicles." *Science*, 352(6293), pp. 1573-1576. [doi](#).
- Bostrom, N., Yudkowsky, E. (2018). "The Ethics of Artificial Intelligence." *Artificial intelligence safety and security*, Chapman and Hall/CRC, pp. 57-69. [doi](#).
- Commissione europea, Direzione generale delle Reti di comunicazione, dei contenuti e delle tecnologie (2019). *Orientamenti etici per un'IA affidabile*. Ufficio delle pubblicazioni, [doi](#).
- Conitzer, V. et al. (2017). "Moral Decision Making Frameworks for Artificial Intelligence." *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). [doi](#).
- Etzioni, A. (2018). "Incorporating Ethics into Artificial Intelligence (with Oren Etzioni)" in *Happiness is the Wrong Metric*. Library of Public Policy and Public Administration, vol. 11. Springer, Cham. [doi](#).
- Foot, P. (1967). "The Problem of Abortion and the Doctrine of the Double Effect." Included in Foot, 1977/2002 *Virtues and Vices and Other Essays in Moral Philosophy*. *Oxford Review*, n. 5.
- Graham, J., et al. (2013). "Moral Foundations theory: The Pragmatic Validity of Moral Pluralism." in *Advances in Experimental Social Psychology*, 47, pp. 55-130. [papers.ssrn.com](#).
- Grando, A., Vicari, S. (2021). *Gestione della tecnologia, dell'innovazione e delle operations*, Egea, Milano.
- Iyer, R., et al. (2012). "Understanding Libertarian Morality: The Psychological Dispositions of Self-Identified Libertarians." *PLoS ONE*, 7(8), articolo e42366. [doi](#).
- Kreutzer, R.T., Sirrenberg, M. (2020). "Outlook." In: *Understanding Artificial Intelligence. Management for Professionals*. Springer, Cham. [doi](#).
- Schiaffonati, V. (2022). "Guglielmo Tamburrini, Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiali, Carocci editore, Roma, 2020, pp. 151." *APhEx*, 25, pp. 1-12. [hdl.handle.net/10077/34250](#).
- Searle, J.R. (1980). "Minds, Brains, and Programs." in *Behavioral and brain sciences*, volume 3, numero 3, settembre 1980, pp. 417-424. [doi](#).
- Vicari, S. (2023). "Intelligenza artificiale ed etica" in Penco L. e Profumo G. (a cura di), *Strategie e Governo dell'Impresa. Scritti in onore di Pietro Genco*, G. Giappichelli Editore, Torino.
- Zanichelli, M. (2021). "L'intelligenza artificiale e la persona: tra dilemmi etici e necessità di regolazione giuridica" in *Teoria e Critica della Regolazione Sociale/ Theory and Criticism of Social Regulation*, 2(23), pp. 141-160. [doi](#).