

*Pasquale Grieco\**

## **Design for action: sviluppo tecnologico e metamorfosi delle categorie etiche**

### **Abstract**

#### **Design for action: technological development and metamorphosis of ethical categories**

This paper explores key conceptual intersections between subjectivity, technological mediation, and agency, focusing on the influence of information technologies, artificial intelligence, and digital systems on these foundational philosophical and ethical categories. The objective is to offer critical insights for rethinking these concepts in the context of the digital revolution. In light of the design turn, the study emphasizes the growing importance of embedding ethical values into artifacts and design processes, highlighting their implications for political and legislative decision-making at both national and supranational levels.

### **Keywords**

Design Ethics; Artificial Intelligence Ethics; Technology Ethics; Digital Design; Digital Ethics

## **1. HAL e i pappagalli: l'etica nella società dell'informazione**

Le domande paradigmatiche della *computer ethics*, di cui può essere considerata emblema quella di Daniel Dennett, “*When HAL kills, who’s to blame?*” (Dennett, 1997), un tempo relegate a una dimensione fantascientifica, affascinante ma essenzialmente astratta, stanno diventando sempre più concrete nella nostra quotidianità. Immaginiamo, ad esempio, di chiedere un consiglio medico online e di ricevere una risposta utile per il nostro benessere. È rilevante sapere se tale risposta provenga da un medico in carne e ossa (o meglio, di carbonio) o da un chatbot (digitale, di silicio) basato su un *Large Language Model* (LLM) open-

\* Università degli Studi di Bari Aldo Moro



source? La questione si complica ulteriormente, sia dal punto di vista etico che filosofico, nel caso in cui l'output risulti dannoso per la salute. Dove risiede la responsabilità morale se il suggerimento è fornito da un "pappagallo stocastico" (Bender et al., 2021), uno dei diffusi programmi di intelligenza artificiale che utilizzano il linguaggio naturale? Chi o cosa deve essere considerato responsabile in tali circostanze? A queste domande si aggiungono ulteriori interrogativi, riguardanti il ruolo della tecnologia digitale come medium dell'azione moralmente significativa.

Il modo in cui queste tecnologie vengono progettate e implementate ha una rilevanza etica? Le tradizionali categorie morali e filosofiche sono sufficienti per affrontare queste nuove tipologie di problemi? Rispondere a tali interrogativi significa confrontarsi direttamente con l'impatto etico delle tecnologie digitali e dell'informazione, in particolare delle *Information and Communication Technologies* (ICT), tra cui l'Intelligenza Artificiale (AI). Quest'ultima rappresenta, per diffusione, disponibilità e accessibilità, una potente e nuova "riserva di capacità di agire a portata di mano" (Floridi, 2022, p. 53). Spesso, le capacità di agire delle tecnologie dotate di algoritmi di *Machine Learning* (ML) e *Deep Learning* (DL) si manifestano attraverso caratteristiche quali *interattività*, *adattabilità* e *autonomia*<sup>1</sup>.

Queste tre caratteristiche costituiscono il livello minimo necessario per considerare un agente capace di produrre effetti morali, indipendentemente dalla sua natura umana, artificiale o ibrida. Infatti, se un agente è interattivo, adattabile e autonomo, può partecipare al *moral game*<sup>2</sup> (Allen

<sup>1</sup> Le tre caratteristiche rappresentano il Livello di Astrazione (LoA) minimale per considerare gli agenti artificiali come agenti morali. Un LoA si riferisce a un insieme di proprietà rilevanti (osservabili, cioè varianti tipizzate interpretate) di un sistema che vengono selezionate per descrivere, analizzare e comprendere tale sistema in un determinato contesto. Cfr Floridi, 2020.. In generale, sul concetto di Livello di Astrazione. Nello specifico, le tre caratteristiche di: Interattività (1), Autonomia (2), Adattabilità (3) significano: (1): che l'agente e il suo ambiente possono agire l'uno sull'altro. Esempi tipici sono l'immissione o l'emissione di un valore, o l'esecuzione simultanea di un'azione da parte dell'agente e del paziente, come la forza gravitazionale tra corpi; (2): che l'agente è in grado di cambiare stato senza rispondere direttamente all'interazione. risposta diretta all'interazione: può eseguire transizioni interne per cambiare il proprio stato. Quindi un agente deve avere almeno due stati. Questa proprietà conferisce all'agente un certo grado di complessità e di indipendenza dall'ambiente; (3): che le interazioni dell'agente possono modificare le regole di transizione con cui cambia stato. Questa proprietà garantisce che un agente possa essere visto, a un dato LoA, come un agente che apprende il proprio modo di operare in un modo che dipende criticamente dalla sua esperienza. Si noti che se le regole di transizione di un agente sono memorizzate come parte del suo stato interno, discernibile a questo LoA, l'adattabilità deriva dalle altre due condizioni. Cfr. Floridi, Sanders, 2004.

<sup>2</sup> Allen et al. hanno proposto la possibilità di un test per valutare la capacità morale delle macchine ispirato al test di Turing, *l'imitation game*. Il test morale proposto per le macchine consiste nel valutare se un agente artificiale possa superare un "*Moral Turing*

et al., 2000), anche se non è un essere umano, contrariamente a quanto previsto dalla tradizione etica della modernità filosofica.

L'apertura alla considerazione morale delle azioni di entità artificiali o ibride è di grande attualità, vista la crescente pervasività di tali tecnologie e il ruolo che rivestono nello spazio pubblico delle democrazie. Si pensi, ad esempio, agli effetti deleteri delle *fake news*, spesso propagate da *bot* sui social media, che influenzano l'orientamento del voto. Diventa quindi essenziale sviluppare una prospettiva etica sulle ICT che non solo incida sulla vita quotidiana degli individui, ma che contribuisca anche a costruire un quadro di riferimento, un *framework* morale, entro il quale gli organismi democratici (a livello nazionale, comunitario e globale) possano regolamentare la progettazione, l'implementazione e l'uso di tali tecnologie.

Il ruolo delle macchine e degli altri agenti non umani è cruciale per comprendere la società odierna, e ciò è particolarmente vero in un contesto in cui l'*informazione*<sup>3</sup> non è solo il bene economico per eccellenza delle società capitalistiche avanzate, né soltanto il concetto chiave per interpretarne le dinamiche socio-politiche e culturali, ma rappresenta anche il livello ontologico minimo comune a tutte le entità che vi operano (cfr. Floridi, 2009, pp. 25-57). L'*agency* di macchine, agenti artificiali, sistemi ibridi e quella mediata dalle tecnologie assume dunque un ruolo

*Test*" (MTT), in cui si verifica se l'agente è in grado di comportarsi in modo moralmente indistinguibile da un essere umano, per un giudice umano in situazioni etiche significative. Allen, et. al., 2000. La questione che emerge è quella se è possibile considerare accettabile che le macchine debbano essere in grado di mentire riguardo questioni che hanno rilevanza etica, riguardante cioè la loro capacità di considerare e produrre bene e male, e quindi eventuali danni ad altri enti, compreso l'essere umano. Questa idea che nella possibilità di mentire vi sia la base per la capacità degli agenti artificiali di sviluppare una moralità di tipo umano si ritrova anche nel summenzionato saggio di Dennett su HAL. Posto che le macchine per sviluppare una loro eticità debbano accedere a una intenzionalità di ordine superiore ("high-order intentionality"), il prezzo da pagare per questa intenzionalità superiore è la duplice capacità di ingannare: di ingannare gli altri e se stessi. "Another price we pay for higher-order intentionality is the opportunity for duplicity, which comes in two flavors: self-deception and other-deception". Dennett, 1997.

<sup>3</sup> L'informazione è un concetto che può essere inteso in diversi modi. Per *informazione* si può intendere l'informazione di Shannon, di natura quantitativa, che viene considerata indipendentemente dal contenuto e dalla dimensione semantica che essa può assumere. L'informazione di Shannon è definita in relazione alla misura dell'entropia e dell'incertezza di un messaggio e può essere espressa tramite il logaritmo medio dell'improbabilità del messaggio stesso, formalizzata come:  $H = -\sum p_j \log_2 p_j$  (H = quantità di informazione, l'entropia;  $\Sigma$  =somma di tutti i messaggi j;  $p_j$  = probabilità del messaggio j). C.E. Shannon, 1948, L'informazione può essere considerata anche nella sua dimensione semantica, intesa in senso "forte" o meno. Per il dibattito sulla teoria dell'informazione semantica in senso forte si può fare riferimento a Allo, 2010. Un'introduzione efficace, accessibile e non specialistica al concetto di informazione, che ne ricostruisce anche la storia, si trova in Gleick, 2012. Per una panoramica sul concetto di filosofia dell'informazione si veda Adriaans 2012/2020.

decisivo per il futuro delle nostre democrazie. Altrettanto pressante e importante è la questione della materializzazione dei valori all'interno delle tecnologie.

Una chiave di lettura per comprendere l'impatto e le conseguenze delle tecnologie digitali in campo etico è la metamorfosi che queste hanno prodotto in una delle categorie fondamentali della modernità: quella di Soggetto, insieme alle caratteristiche che lo definivano come l'unico detentore della moralità. L'idea che il digitale eserciti un potere di re-ontologizzazione e re-epistemologizzazione sul modo in cui percepiamo la realtà e noi stessi (Floridi, 2022, p. 39), e che il suo impatto si estenda alla sfera etica, sociale e politica – malgrado l'urgenza di una riflessione su queste dinamiche è oggi più pressante che mai – non è certo nuova. Norbert Wiener, il padre della cibernetica, prendendo in esame le istanze etiche che avrebbero dovuto condurre la nuova scienza, la *scienza della comunicazione e del controllo nella macchina e nell'animale* verso un *uso umano degli esseri umani*<sup>4</sup>, affermava che:

La società può essere compresa soltanto attraverso lo studio dei messaggi e dei mezzi di comunicazione relativi ad essi; e che nello sviluppo futuro di questi messaggi e mezzi di comunicazione, i messaggi fra l'uomo e le macchine, fra le macchine e l'uomo, e fra macchine e macchine sono destinati ad avere una parte sempre più importante (Wiener, 1963, pp. 23-24).

## 2. Soggetti frantumati: dall'Io All'I/O

L'impatto rivoluzionario del digitale rappresenta solo l'ultima fase della crisi del Soggetto che abbiamo ereditato dalla modernità filosofica. Le ICT e il pensiero informazionale e computazionale che le accompagnano evidenziano una frattura nelle sue strutture concettuali fondamentali. Il Soggetto moderno non è più monolitico e individuale, unico e identico a sé stesso, come la cartesiana *res cogitans*, ontologicamente ipostatizzata e separata dal mondo, né tantomeno come un "Io" trascendentale kantiano che osserva il mondo.

I germi di questa scissione sono già presenti nel cuore della filosofia moderna, e il digitale non fa che amplificarli. Hobbes, nel *Leviatano*, definisce la caratteristica essenziale del Soggetto, la Ragione, come un calcolo:

<sup>4</sup> Queste locuzioni non sono casuali. Sono infatti i titoli delle opere maggiori di Wiener, *Cybernetics, or control and communication in the animal and the machine* del 1948 e *The Human Use of Human Beings* del 1950.

Da tutto ciò, possiamo definire (vale a dire determinare) che cosa si vuol dire con questo vocabolo ragione, quando lo calcoliamo tra le facoltà della mente. Infatti, la ragione, in questo senso, non è che il calcolo (cioè l'addizione e la sottrazione) delle conseguenze dei nomi generali su cui c'è accordo per contrassegnare e significare i nostri pensieri; dico contrassegnarli, quando li calcoliamo per noi stessi e significarli, quando dimostriamo o proviamo i nostri calcoli agli altri uomini (Hobbes, 1651, I, Cap. V).

Ragione e pensiero, quindi, non sono altro che un'attività computazionale, un calcolo di simboli. Questa idea espressa da Hobbes nel *Leviatano*<sup>5</sup> si ritrova, tre secoli dopo, nella teorizzazione dell'Intelligenza Artificiale simbolica – quella che precede la rivoluzione subsimbolica e connessionista del *Deep Learning* – la *Good Old Fashioned Artificial Intelligence* (GOFAI). “Questa è basata su una particolare teoria dell'intelligenza e del pensiero: sostanzialmente l'idea di Hobbes che la ragione sia calcolo” (Haugeland, 1988, p. 107).

Ma chi, o quale istanza, opera queste operazioni? Rispondere a questa domanda comporta l'immediato incontro con un paradosso intrinseco alla concezione computazionale della mente, sia nella modernità filosofica e sia nella GOFAI: il paradosso della ragione meccanica, ovvero la presenza di un *homunculus* intelligente, non puramente meccanico, che manipola i simboli. Questo paradosso implica necessariamente una regressione infinita di *homuncoli* che manipolano simboli (Ivi, pp. 39-43). Non è possibile approfondire ulteriormente qui questo paradosso, ma è importante evidenziare come la teoria computazionale frantumi l'unità del Soggetto.

Il distacco dalla visione tradizionale di soggettività, identità e coscienza, considerate come entità uniche, semplici e completamente trasparenti, si amplifica quando viene preso in esame il paradigma subsimbolico dell'intelligenza artificiale e la correlata teoria della mente, il *Connessionismo*. Questa teoria si fonda su processi sub-cognitivi che operano al di sotto delle strutture logico-simboliche di alto livello, come il linguaggio e la dimensione semantica. I meccanismi che producono gli output delle *Artificial Neural Networks* (ANN), modello esplicativo della mente nelle teorie connessioniste e motore del successo degli algoritmi di *Machine Learning* e *Deep Learning*, sono per loro natura inconoscibili. Questo perché tali processi avvengono all'interno di una *black box* dove la convoluzione dei nodi si sviluppa su più livelli negli strati nascosti delle reti

<sup>5</sup> È da notare come la definizione di ragione come computazione si presenti per la prima volta nell'opera per eccellenza sul *potere*. Sembra quasi prefigurare la contemporaneità dove il potere computazionale (es. dei giganti del capitalismo digitale) è la base per le altre forme di potere (economico, sociale, politico e così via).

neurale. Pertanto, sia le tecnologie algoritmiche, come l'odierna *Intelligenza Artificiale*, sia la mente umana risultano essere delle scatole nere.

Marvin Minsky, uno dei pionieri dell'AI, parla di "Società della Mente" (Minsky, 1989). Stabilita l'identità tra cervello e mente, Minsky afferma che esiste una pluralità di agenti privi di mente, dalla cui interazione, che una natura "sociale", in un rapporto "talvolta di cooperazione, ma più spesso di conflitto" (Ivi, p. 20) deriva la mente stessa. Non esiste un singolo "trucco magico" che spieghi l'intelligenza (Ivi, p. 308). Di conseguenza, non vi è più un unico principio che possa spiegare e giustificare mente e coscienza; al contrario, sono la diversità e la frammentazione a caratterizzarle.

La disgregazione dell'unità del soggetto moderno "tradizionale" è visibile anche nell'analisi di questa categoria dal punto di vista informativo e nel rapporto con le ICT: il Sé è un Sistema Multi-Agente (MAS) costituito da diverse dimensioni (le "membrane": corporea, cognitiva e cosciente) che interagiscono e si influenzano reciprocamente in un processo poietico (Floridi, 2011). Lo scopo del Sé è semantizzare, ossia trovare schemi di senso nell'ambiente circostante. Il Soggetto si costruisce "narrativamente" nel rapporto con il proprio ambiente, e l'ambiente contemporaneo è sempre più pervaso dalle tecnologie digitali (Ivi). Queste ampliano in modo esponenziale la tipologia e la natura di questo processo autopoietico in relazione ad altri oggetti e al mondo, mettendo in risalto il ruolo decisivo che la mediazione operata dalle ICT riveste e l'importanza, altrettanto decisiva, di un loro design etico.

Prima di approfondire ulteriormente questi temi, è opportuno mettere in luce come, in due dei più celebri *loci* relativi all'Intelligenza Artificiale, si definisca l'intelligenza come azione. Il primo è "l'atto di nascita dell'IA" nel *workshop* di Dartmouth nell'estate del 1956, organizzato da John McCarthy, con la partecipazione di futuri giganti del pensiero computazionale, come Minsky, Rochester, Shannon, Simon e Newell. Il secondo è l'*Imitation Game*, il cosiddetto Test di Turing.

Nella celebre *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, in cui il termine Intelligenza Artificiale è usato per la prima volta, si stabilisce che: "per l'attuale scopo, il problema dell'intelligenza artificiale è considerato quello di far comportare una macchina in modi che sarebbero definiti intelligenti se fosse un essere umano a comportarsi così"<sup>6</sup>. Il primo vagito dell'IA associa quindi l'intelligenza a un comportamento.

Lo stesso vale per l'*Imitation Game* proposto da Alan Turing nel 1950. Il matematico inglese, dopo aver affermato che la domanda "Possono

<sup>6</sup> "For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving". McCarthy et al., 1955.

pensare le macchine?” (Turing, 1950) è insensata e andrebbe scartata, propone un gioco per stabilire se le macchine siano intelligenti. Ispirandosi a un popolare gioco di ruolo dell’epoca – in cui tre partecipanti, A (uomo), B (donna) e C (giudice), sono separati e il giudice deve determinare chi sia l’uomo e chi la donna attraverso semplici domande e risposte scritte – Turing riformula la domanda sul pensiero delle macchine in un’ipotesi scientifica verificabile: se una macchina sostituisse uno dei partecipanti e ingannasse il giudice, facendogli credere di essere un essere umano, allora alla macchina andrebbe attribuita l’intelligenza.

Anche in questo caso, l’intelligenza, o meglio, il giudizio sull’intelligenza, è legato all’azione, al comportamento. Nell’azione, le differenze tra uomo e macchina sfumano, non trovando più un ostacolo epistemologico nella concezione sostanziale e ipostatizzata della soggettività intelligente. Se l’intelligenza si dissolve nell’agire, e questo è mediato dagli artefatti tecnologici, è necessario interrogarsi su come questi artefatti modifichino l’*agency* morale.

L’analisi computazionale e informazionale della soggettività evidenzia come, più che un’irriducibile conflittualità tra “Uomo” e “Macchina” – entrambi concetti filosoficamente problematici – le ICT, il digitale e l’AI debbano essere compresi, con distinzioni nette che vanno sfumando, in relazione alla capacità di agire. È qui, nel legame con l’azione – che le tecnologie digitali amplificano, mediano e concretizzano – che va ricercata la dimensione etica delle nuove tecnologie. Queste hanno contribuito a modificare l’*agency* e la responsabilità morale tipica della modernità e pongono il problema di una loro implementazione eticamente sostenibile.

### **3. Trasformazioni dell’*agency* e della responsabilità: una genealogia sociotecnica**

Le riflessioni precedentemente espone sulla trasformazione, sia ontologica che epistemologica, della natura del soggetto, causata dalla rivoluzione digitale, sollevano un ulteriore tema. Se le ICT e l’AI rappresentano oggi una riserva della nostra capacità di agire, diventa essenziale comprendere l’impatto che tali tecnologie hanno avuto sul concetto di *agency* e come esse, già dalla loro progettazione, mediano e trasformano il nostro agire morale.

Se l’introduzione delle ICT ha messo in discussione la concezione del Soggetto ereditata dalla modernità filosofica, allora le categorie etiche tradizionali, che attribuivano al Soggetto “moderno” il monopolio della moralità, non sono più sufficienti per giudicare l’azione mediata dagli artefatti digitali. Questi artefatti, come abbiamo già osservato, possono non solo essere agenti autonomi, ma anche costituire il mezzo attraverso

cui si costruisce narrativamente la nostra identità. Di conseguenza, diventa imprescindibile considerare nuovi fattori che rendono insostenibile il modello tradizionale di relazione lineare tra utente e oggetto nella produzione di atti moralmente qualificabili. V. Flusser offre una sintesi efficace e originale della metamorfosi in corso.

Negli anni '90 del secolo scorso, Flusser osservava come fosse diventato estremamente arduo attribuire la responsabilità morale in una società in cui le tecnologie sono sempre più pervasive. Secondo il pensatore ceco-brasiliano, questo aumento di complessità era dovuto a tre fattori principali:

1) La perdita di credibilità delle autorità tradizionali (religiose, politiche, morali) nel definire norme sociali, particolarmente nel contesto di una produzione industriale complessa;

2) La crescente complessità della produzione industriale e del design, che comportano una quantità di informazioni tale da eccedere la capacità cognitiva individuale, richiedendo una cooperazione organizzativa nei processi di progettazione e implementazione. L'azione, spesso svolta da "team" che combinano componenti umane e artificiali, non può pertanto essere attribuita a un singolo autore: "Per questo motivo, nessuna persona può essere ritenuta responsabile di un prodotto";

3) In passato, la responsabilità morale ricadeva solo sull'utilizzatore di un prodotto. "Se una persona pugnava un'altra con un coltello, ne era l'unico responsabile, non il designer del coltello". In questo scenario, la produzione di coltelli era considerata un'attività pre-etica, priva di valutazioni morali. Tuttavia, oggi, con l'automazione dei processi produttivi, ci si potrebbe chiedere se, ad esempio, siano i robot o le macchine a essere responsabili dell'uso dei prodotti. Dal momento che questo è un assurdo, a chi deve essere attribuita la responsabilità morale?

Alla persona che ha costruito il robot, a quella che ha realizzato il coltello, oppure a chi ha programmato il funzionamento del robot? Si potrebbe forse attribuire la responsabilità a un errore di progettazione, di programmazione o di produzione? E perché non imputare la responsabilità morale all'intero settore industriale che ha prodotto il robot? O magari all'intero complesso industriale, e, in ultima analisi, al sistema di cui tale complesso fa parte? (Flusser, 2003, pp. 63-65).

Queste domande evidenziano come il quadro della responsabilità morale sia cambiato irreversibilmente. Questi problemi non sono contingenti o legati esclusivamente allo sviluppo industriale del XX secolo, ma sono strettamente legati al ruolo degli artefatti tecnologici nel nostro agire e al loro design. Un esempio particolarmente drammatico, utilizzato da Flusser per illustrare il ruolo della progettazione/produzione nell'*agency* morale, è emerso durante il processo di Norimberga. In una lettera scritta

da un industriale tedesco a un ufficiale nazista, l'industriale si scusava per aver progettato male le camere a gas che gli erano state commissionate: “invece di uccidere migliaia di persone per volta, ne venivano uccise solo centinaia” (Ivi, p. 66).

L'episodio è emblematico di come “a) non esistano più norme applicabili alla produzione industriale; b) non ci sia più un singolo autore di un crimine; e c) la responsabilità sia così diluita che di fatto ci troviamo in una situazione di totale irresponsabilità rispetto agli atti derivanti dalla produzione industriale” (Ibidem).

A questi tre punti possiamo aggiungere una considerazione ulteriore: il ruolo di un artefatto, come nell'esempio citato da Flusser, nel determinare un risultato moralmente giudicabile. Non è più possibile considerare la relazione tra oggetto e utente come lineare, senza tener conto del designer e del processo di progettazione. I designer, dunque, non possono eludere una riflessione etica sul loro lavoro, pena il totale dissolvimento della responsabilità morale individuale. L'episodio sopra menzionato mette drammaticamente in evidenza la funzione di mediazione degli artefatti tecnologici nella sfera morale.

#### 4. Mediazione tecnologica e metamorfosi fenomenologiche

Un contributo decisivo all'analisi del concetto di mediazione tecnologica, dal punto della *postfenomenologia* – sulla scia del concetto di *script* di Latour (1992) –, è stato quello di P. P. Verbeek. Il filosofo olandese sostiene che la mediazione tecnologica è costitutiva dell'azione degli utilizzatori, con le tecnologie che plasmano le azioni dell'utente oltre la sfera della mera funzionalità. Il design iscrive modalità d'uso specifiche nella tecnologia, con un carattere interazionale. “Le tecnologie hanno “intenzioni,” non sono strumenti neutrali ma giocano un ruolo attivo nella relazione tra gli esseri umani e il loro mondo” (Verbeek, 2006)<sup>7</sup>. Gli artefatti tecnologici non solo influenzano il modo in cui gli esseri umani agiscono, ma anche il modo in cui organizzano la loro vita sociale e ambientale.

Le riflessioni di Verbeek implicano una responsabilità condivisa tra utenti e designer per le conseguenze etiche delle azioni mediate dalla tecnologia. Responsabilità che si articola in tre modalità:

1. Responsabilità dei progettisti;
2. Responsabilità degli utilizzatori;
3. Consapevolezza etica condivisa.

<sup>7</sup> “Technologies have ‘intentions’, they are not neutral instruments but play an active role in the relationship between humans and their world”.

La mediazione tecnologica deve essere integrata pienamente nella riflessione etica. “Gli artefatti tecnologici non sono intermediari neutrali ma plasmano attivamente il modo in cui le persone sono nel mondo: le loro percezioni e azioni, l’esperienza e l’esistenza”<sup>8</sup> (Ivi). Occorre riconoscere che le tecnologie non sono semplici strumenti passivi, ma coinvolgono attivamente gli esseri umani e altri attori non umani in assemblaggi di azione che influenzano il nostro modo di vivere. Sia i progettisti che gli utenti devono essere consapevoli delle implicazioni etiche connesse alle tecnologie, adottando un approccio critico e olistico che consideri le complesse interazioni tra esseri umani, tecnologie e ambiente. In questo senso, attraverso la mediazione tecnologica, il design realizza materialmente la moralità, poiché offre risposte mediate alla domanda su come si debba agire, delineando la responsabilità morale specifica dei designer<sup>9</sup>. Il concetto di *agency* morale si evolve: non è più limitato al solo essere umano (o al Soggetto tradizionale), ma include anche l’oggetto, la sua non neutralità etica e la storia della sua progettazione. Queste istanze si comprendono meglio

L’ambiente informazionale, perennemente connesso e potenziato dalla globalizzazione, dal digitale e dalle opportunità offerte dalla mediazione tecnologica, dimostra chiaramente come le categorie etiche tradizionali non siano più sufficienti, ma necessitino di un ampliamento. Van den Hoven sottolinea come, insieme al mutamento del contesto sociale in cui si svolge l’azione morale, sia cambiata anche quella che S. Scheffer definisce la fenomenologia dell’*agency* ispirata al senso comune.

La concezione tradizionale della moralità prevedeva:

1. la priorità dei doveri negativi (evitare di danneggiare gli altri) rispetto ai doveri positivi di fare il bene o aiutare gli altri nel perseguimento dei loro obiettivi;
2. una divisione del lavoro morale, in cui gli obblighi verso le persone più vicine sono più forti rispetto a quelli verso individui lontani, sia nello spazio che nella società;

<sup>8</sup> “Technological artifacts are not neutral intermediaries but actively shape people’s being in the world: their perceptions and actions, experience, and existence”

<sup>9</sup> La responsabilità del designer è sempre più chiamata in causa, via via che la capacità persuasiva degli artefatti diviene maggiore a causa dello sviluppo tecnologico. Il testo chiave per comprendere le tecnologie della persuasione è Fogg, 2017. Fogg, specialmente grazie all’influenza del suo *Behaviour Design Lab* a Stanford, è “l’eminenza grigia” che sta dietro molte delle testimonianze di protagonisti delle industrie hi-tech contenute nel documentario di Netflix, *The Social Dilemma* del 2020. Per quanto riguarda la responsabilità morale dei designer in relazione ad azioni intenzionali o non-intenzionali ed effetti prevedibili o non prevedibili, si rimanda a Berdichevsky, Neuenschwander, 1999.

3. il primato dell'atto sull'omissione, con l'atto considerato sempre come autonomo, mentre l'omissione è vista come contestuale;
4. la preminenza degli effetti immediati delle nostre azioni rispetto a quelli lontani. Inoltre, la percezione della nostra influenza causale diminuisce proporzionalmente alla distanza spazio-temporale degli effetti. La nostra *agency* individuale, infine, viene percepita in modo attenuato quando agiamo in contesti sociali complessi o in collaborazione con altri (Van den Hoven, 2017).

Questa fenomenologia dell'*agency* non è più in grado di rispondere adeguatamente alle sfide del mondo contemporaneo. Come, nel paragrafo precedente le analisi di Flusser hanno mostrato, non è più possibile considerare una semplice relazione lineare tra agente/utilizzatore e oggetti. L'agente morale individuale, modellato da questa visione, oggi si trova subordinato a strutture globali che dirigono, interconnettono e influenzano profondamente la sua azione morale. Un esempio emblematico di ciò è la facilità con cui è possibile compiere azioni eticamente disastrose utilizzando semplicemente un computer connesso a Internet.

Proseguendo con l'analisi della tesi di Scheffer, Van den Hoven osserva che oggi “1) il contributo degli individui ai risultati complessivi è limitato; 2) il loro controllo è ridotto; 3) i processi in cui sono coinvolti sono pervasivi e inevitabili; e 4) è molto difficile ottenere informazioni accurate, affidabili e tempestive su tali processi e i loro effetti”<sup>10</sup> (Ivi). Questa situazione, come descritto da Scheffer, evidenzia chiaramente che la nostra concezione della moralità, così come espressa dalla fenomenologia del senso comune, necessita di essere ampliata. Tuttavia, per Van den Hoven, questo non è sufficiente a spiegare un fattore ulteriore:

Molte delle azioni che compiamo nel ventunesimo secolo non le compiamo direttamente l'uno verso l'altro, ma quasi sempre tramite la tecnologia e attraverso un milieu tecnologico che è stato progettato e che è emerso nel corso dei secoli a causa di innumerevoli contributi da parte di altri separati da noi nel tempo e nello spazio (Ivi).<sup>11</sup>

<sup>10</sup> “The subsumption of individual agents in complex global interconnected structures has a number of morally relevant features according to Scheffer: (1) the contribution of individuals to overall results is limited; (2) their control is limited; (3) the processes in which they are subsumed are pervasive and inescapable; and (4) it is very difficult to get accurate, reliable and timely information about these processes and their effects”

<sup>11</sup> “Much of what we do in the twenty-first century we do not do directly to each other, but almost always by way of technology and via a technological milieu that was designed, and that came about over the course of centuries as a result of myriad contributions by others separated from us by time and place”

Il Soggetto della contemporaneità, influenzato dall'impatto delle ICT, si costruisce narrativamente attraverso l'interazione con un ambiente sempre più permeato da artefatti tecnologici e informativi. Il ruolo del design nella considerazione morale implica che la riflessione etica debba essere accompagnata da una genealogia del design, capace di spiegare la stratificazione delle scelte progettuali: "se Nietzsche ha tracciato la genealogia della morale, la svolta progettuale ha tracciato la genealogia delle scelte progettuali" (Umbrello, 2023). La trasformazione dell'*agency* e del soggetto agente rende obsoleta la tradizionale relazione diretta tra agente e oggetto. La metamorfosi dell'ambiente tecnologico, unita alla pluralità di stakeholder coinvolti (designer, utenti, pazienti indiretti, agenti e pazienti artificiali o ibridi), riformula la fenomenologia morale, chiarendo il cambiamento qualitativo e quantitativo dell'*agency* rispetto alla moralità dei soggetti tradizionali.

Anche alla luce di queste considerazioni sul Soggetto, diventa allora evidente l'importanza cruciale della progettazione e della sua storia. Nella riflessione morale, è imprescindibile prendere in esame "1) le storie di progettazione dei sistemi e dell'ambiente tecnologico in cui si realizza l'*agency*; e 2) le scelte e il ruolo degli agenti di progettazione nella formazione delle architetture di scelta" (Ivi).

In qualunque ambiente informativo – sia esso quello operativo diretto dell'utente o quello integrato in un sistema di controllo o di utilizzo – le scelte di design risultano decisive. Un operatore nella sala di controllo del traffico aereo, un trader di borsa, un chirurgo in una moderna sala operatoria o un tecnico in una centrale nucleare hanno precise responsabilità e doveri morali. Tuttavia questi agenti "non possono svolgere la loro azione [...] senza un ambiente di lavoro tecnologico adeguatamente progettato. Se il software dei loro sistemi contiene bug, se i modelli utilizzati si basano su presupposti errati o se le informazioni sono imprecise, le decisioni e le azioni umane erediteranno questi difetti"<sup>12</sup> (Ivi).

In questo contesto, gli agenti, umani e non, interagiscono con artefatti e ambienti informativi generati da processi di progettazione e implementazione. Il *design turn* dell'etica diventa così il mezzo per dare forma ai valori in un'epoca in cui i confini tradizionali tra le categorie stabilite si fanno sempre più sfumati. La capacità di coniugare sviluppo tecnologico e dimensione etica emerge quindi come una sfida cruciale per le società

<sup>12</sup> "Think of an operator in the control room of a nuclear power plant, an air traffic control leader, a surgeon in a modern operating theatre, or a trader in a trading room. All of these professionals have duties and responsibilities, and they cannot possibly do what they are expected to do — and what everyone justifiably expects them to do — without a properly designed technological work environment. If the software of their systems contains bugs, the models used are based on incorrect assumptions, or the information is inaccurate, then human decisions and actions will inevitably inherit these flaws"

democratiche, chiamate a trovare un equilibrio tra innovazione e rispetto dei principi morali fondamentali.

## 5. Etica e design tecnologico

La rivoluzione digitale ha messo in luce come il nostro ambiente sia sempre più di natura informazionale, caratterizzato da livelli di complessità crescenti. Viviamo e agiamo all'interno dell'*Infosfera*<sup>13</sup>, un'epoca dominata dai *Big Data* e dalla proliferazione di *stakeholder*, che arriva a comprendere “specie non umane, superorganismi, la Terra stessa, i social robot”<sup>14</sup> (Friedman, Hendry, 2019, p. 27). In questo orizzonte, la sfera legislativa, sia a livello nazionale che sovranazionale, ha progressivamente posto al centro dell'attenzione la dimensione etica delle nuove tecnologie, con particolare riferimento alla loro progettazione e implementazione. Un chiaro esempio di questo approccio è rappresentato dal *Regolamento Generale sulla Protezione dei Dati* (GDPR) dell'Unione Europea, entrato in vigore nel 2018, che introduce concetti fondamentali relativi alla gestione e protezione dei dati personali. Tra le novità principali sancite dal GDPR, e recepite nelle legislazioni nazionali, emergono i concetti di *Privacy by Design* e *Privacy by Default*, definiti rispettivamente nell'art. 25 e nel considerando 78 del Regolamento.

Il principio di *Privacy by Design* sottolinea come le istanze etiche, come la protezione della privacy, debbano essere incorporate direttamente nei sistemi mediante il loro design. Ciò significa che la protezione dei dati personali deve essere integrata sin dalle prime fasi di progettazione di sistemi, processi, prodotti o servizi che trattano tali dati, rendendo la privacy una parte intrinseca dell'architettura tecnica e concettuale delle tecnologie stesse.

Il concetto di *Privacy by Default*, complementare a quello di *Privacy by Design*, stabilisce che i sistemi devono essere configurati in modo da garantire automaticamente il massimo livello di protezione dei dati personali, senza che l'utente debba intervenire per modificarne le impostazioni. In questo modo, le impostazioni predefinite minimizzano la raccolta e l'uso dei dati, rafforzando il rispetto dei diritti degli utenti.

<sup>13</sup> “L'infosfera indica l'intero ambiente informazionale costituito da tutti gli enti informazionali, le loro proprietà, interazioni, processi e reciproche relazioni”. Il concetto può essere esteso sino a comprendere l'intero reale. L. Floridi, 2017, p. 44.

<sup>14</sup> “Yet the technologies we design and build reach far beyond human beings to implicate other nonhuman entities. Here are four for consideration: nonhuman species, superorganisms, the Earth, and social robots”

Quanto sancito dal GDPR riflette la crescente consapevolezza normativa della necessità di affrontare le sfide etiche poste dall'uso delle tecnologie digitali e degli algoritmi, assicurando che istanze etiche fondamentali, come la privacy<sup>15</sup>, siano tutelate fin dalla fase di progettazione delle tecnologie stesse. In questo contesto, si può riconoscere una formale consacrazione del *design turn* nell'etica applicata, in cui il processo di progettazione diventa centrale per garantire il rispetto dei principi etici (Van den Hoven, 2017). Sebbene non priva di criticità in alcune metodologie, come il *Value Sensitive Design* (VSD)<sup>16</sup> attraverso cui tenta di concretizzarsi, la svolta progettuale dell'etica, ovvero l'incarnazione dei valori negli artefatti, può rappresentare una via verso un'innovazione sostenibile. Insieme ad approcci come l'*Artificial Intelligence for Social Good* (AI4SG)<sup>17</sup>, essa offre

<sup>15</sup> Per una panoramica sull'implementazione della privacy tecnologica, che sfata i miti sulla protezione dei dati attraverso algoritmi, si veda Hoepman, 2021. Per una trattazione snella ed efficace delle strategie di progettazione orientate alla privacy, si rimanda a Hoepman, 2022.

<sup>16</sup> Il VSD è un approccio teorico alla progettazione tecnologica che incorpora sistematicamente i valori umani nel processo di design, sviluppato da Batya Friedman e colleghi negli anni '90. Il VSD adotta una metodologia tripartita, basata su indagini concettuali, empiriche e tecniche, e riconosce l'importanza di valori morali come giustizia, benessere e dignità (Friedman et al., 2008; Friedman, Hendry, 2019). Tuttavia, tra le principali criticità emerge la mancanza di una chiara definizione di "valore", oltre al rischio di oscillare tra relativismo e universalismo. In particolare, viene imputata al VSD la tendenza all'ipostatizzazione valoriale, ovvero la tendenza a reificare un insieme ristretto di valori che potrebbe non adattarsi ai contesti locali, ostacolando la scoperta di nuovi valori emergenti. Il VSD tenderebbe, inoltre, alla "fallacia naturalistica", dove si confonde ciò che è con ciò che dovrebbe essere. Questo problema sottolinea la difficoltà di bilanciare l'universalità dei valori con la loro adattabilità ai diversi contesti socioculturali e progettuali, una questione ancora aperta nella metodologia del VSD. Per le critiche al VSD, si veda Albrechtslund, 2007; Le Dantec et al., 2009; Mander-Huits, 2011. Intorno al rapporto VSD-AI4SG si veda Umbrello, 2023. Gli oggetti, come evidenziato nell'analisi, non sono neutri; al contrario, essi operano mediando l'azione e esercitano una forma di influenza, più o meno diretta, che può manifestarsi sotto forma di spinta, di *nudging*: "Qualsiasi aspetto dell'architettura delle scelte che altera il comportamento delle persone in modo prevedibile senza vietare alcuna opzione o cambiare significativamente i loro incentivi economici". Thaler & Sunstein 2009, p.6, trad. mia. L'incarnazione di valori nella materialità degli oggetti, l'*ethics by design*, può essere vista come una forma di *nudging* ontologico, che viola l'autonomia decisionale del soggetto che non può non agire come l'oggetto impone. A questa forma in cui rientra il VSD L. Floridi oppone il *pro-ethical design*, una forma di *nudging* informazionale, modifica la natura delle informazioni a cui le persone sono esposte, influenzando così le decisioni senza alterare le opzioni fisiche disponibili, come più rispettoso dell'autonomia dell'agente. La forma dell'*ethics by design*, via scelta dagli organismi legislativi dell'Unione Europea per proteggere beni etici in relazione all'implementazione tecnologica, può trovare da queste riflessioni spunti per una progressiva riduzione di potenziali criticità etiche. Floridi, 2016.

<sup>17</sup> L'AI4SG è il "design, lo sviluppo e l'implementazione di sistemi di IA in modo da (i) prevenire, mitigare o risolvere i problemi che incidono negativamente sulla vita umana o/e sul benessere del mondo naturale e/o (ii) consentire sviluppi preferibili dal punto di vista sociale e/o sostenibili dal punto di vista ambientale.". Floridi, 2022, p. 227.

una prospettiva promettente per coniugare progresso tecnologico e rispetto dei principi etici.

L'aver preso in esame questi snodi critici dell'impatto dello sviluppo tecnologico delle ICT nell'orizzonte socio-tecnico del digitale, può essere utile per offrire spunti affinché moralità e valori trovino maggiore tutela nell'implementazione tecnologica e nella produzione legislativa, incarnandosi in quegli artefatti che influenzano profondamente le nostre azioni eticamente significative, permettendone un uso rispettoso delle prerogative umane.

## Bibliografia

- Albrechtslund, A.  
2007 *Ethics and technology design*. *Ethics and Information Technology*, 9, 63-72. <https://doi.org/10.1007/s10676-006-9129-8>
- Allen, C., Varner, G., Zinser, J.  
2000 *Prolegomena to any future artificial moral agent*, *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3), 251-261. DOI:10.1080/09528130050111428
- Allo, P. (guest editor)  
2010 *Special Issue: Luciano Floridi and the Philosophy of Information*, 41(3), *Metaphilosophy*, Wiley. <https://doi.org/10.1111/j.1467-9973.2010.01642.x>
- Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S.  
2021 *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. <https://doi.org/10.1145/3442188.3445922>
- Berdichevsky, D., & Neuenschwander, E.  
1999 *Toward an Ethics of Persuasive Technology*. *Communications of the ACM*, 42(5), pp. 51-58.
- Dennett, D.  
1997 *When Hal Kills, Who's to Blame? Computer Ethics*, in D. Stork (a cura di), *Hal's Legacy: 2001's Computer as Dream and Reality*, MIT Press, Cambridge, MA, pp. 351-365.
- Floridi, L.  
2008 *The method of levels of abstraction*, *Minds and Machines*, 18(3), 303-329. <https://doi.org/10.1007/s11023-008-9113-7>
- 2010 *Information Ethics*, in *The Cambridge Handbook of Information and Computer Ethics*, Edited by Luciano Floridi, Cambridge University Press, Cambridge.
- 2011 *The informational nature of personal identity*, *Minds and Machines*, 21(4), 549-566. DOI:10.1007/s11023-011-9259-6
- 2016 *Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration*, *Science and Engineering Ethics*, 22(6), 1669-1688. <https://doi.org/10.1007/s11023-011-9259-6>
- 2017 *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina, Milano.

- 2020 *Pensare l'infosfera, la filosofia come design concettuale*, Raffaello Cortina, Milano.
- 2022 *Etica dell'Intelligenza Artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina, Milano.
- Floridi, L., Sanders, J. W.
- 2004 *On the morality of artificial agents*, *Minds and Machines*, 14(3), 349-379. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>
- Flusser, V.
- 2003 *Filosofia del design*, Bruno Mondadori, Milano.
- Friedman, B., Hendry, D. G.
- 2019 *Value Sensitive Design: Shaping Technology with Moral Imagination*, MIT Press Cambridge MA.
- Friedman, B., Kahn, P. H. Jr., Borning, A.
- 2008 *Value Sensitive Design and Information Systems*, in K.E. Himma, H.T. Tavani (a cura di), *The Handbook of Information and Computer Ethics*, John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 69-101.
- Gleick, J.
- 2012 *L'informazione: Una storia. Una teoria. Un diluvio*, Feltrinelli Milano.
- Haugeland, J.
- 1988 *Intelligenza Artificiale. Il significato di un'idea*, Bollati Boringhieri Torino.
- Hobbes, T.
- 1651 *Leviathan, or The Matter, Forme and Power of a Commonwealth Ecclesiastical and Civil*; trad. it, *Il Leviatano*; BUR, Milano 2011.
- Hoepman, J.-H.
- 2021 *Privacy Is Hard and Seven Other Myths: Achieving Privacy through Careful Design*. Cambridge, MA: The MIT Press.
- 2022 *Privacy Design Strategies (The Little Blue Book)*. <https://creativecommons.org/licenses/by-nc/4.0/>.
- LATOURE, B.
- 1992 *Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts*, in W. E. Bijker, J. Law (a cura di), *Shaping Technology/Building Society: Studies in Sociotechnical Change*, MIT Press, Cambridge, MA, 225-258.
- Le Dantec, C. A., Poole, E. S., & Wyche, S. P.
- 2009 *Values as lived experience: Evolving value sensitive design in support of value discovery*. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1141-1150. New York, NY: ACM. <https://doi.org/10.1145/1518701.1518875>
- Manders-Huits, N.
- 2011 *What values in design? The challenge of incorporating moral values into design*. *Science and Engineering Ethics*, 17(2), 271-287. <https://doi.org/10.1007/s11948-010-9198-2>
- McCarthy, J., Minsky, M. L., Rochester, N., Shannon, C. E.
- 1955 *A proposal for the Dartmouth summer research project on artificial intelligence*, *AI Magazine*, 27(4).
- Minsky, M.
- 1989 *La società della mente*, Adelphi Milano.
- Shannon, C. E.
- 1948 *A Mathematical Theory of Communication*, *Bell System Technical Journal*, 27(3), 379-423.

- Thaler, R. H., Sunstein, C. R.  
2008 *Nudge: Improving decisions about health, wealth, and happiness*, Yale University Press, New Haven CT.
- Turing, A. M.  
1950 *Computing Machinery and Intelligence*, *Mind*. Vol. 59, No. 236 (Oct., 1950), pp. 433-460; trad. it. in *Intelligenza Meccanica*, Bollati Boringhieri, Torino 1994.
- Umbrello, S.  
2023 *Oggetti buoni. Per una tecnologia sensibile ai valori*, Fandango Roma.
- Van Den Hoven, J.  
2017 *The Design Turn in Applied Ethics*, in J. van den Hoven, S. Miller, T. Pogge (a cura di), *Designing in Ethics*, Cambridge University Press, Cambridge, 11-31. DOI:10.1017/9780511844317.002
- Verbeek, P. P.  
2006 *Materializing Morality: Design Ethics and Technological Mediation*, *Science, Technology, & Human Values*, 31(3), 361-380. DOI:10.1177/0162243905285847
- Wiener, N.  
1948 *Cybernetics; or control and communication in the animal and the machine*. John Wiley & Sons New York.
- 1950 *The Human Use of Human Beings: Cybernetics and Society*, Houghton Mifflin, Boston. trad. it. *Introduzione alla cibernetica – L'uso umano degli esseri umani*, Einaudi Torino, 1953