

Giuseppe Lorini, Olimpia G. Loddo, Simone Pinna

Artefatti regolativi digitali, intelligenza artificiale e diritti dell'utente*

Abstract: Digital Regulatory Artifacts, Artificial Intelligence, and User Rights. The debate on the impact of artificial intelligence on personal freedom and fundamental rights has intensified in recent years, particularly with the rise of generative AI, which has brought both opportunities and challenges to the fore. However, as highlighted in the field of technological and digital regulation, digital artifacts not based on generative AI can also exert a profound influence on individual freedom. By examining the functioning of certain software tools designed to regulate online interactions and by distinguishing between deontic and adeontic (non-deontic) regulatory artifacts, it becomes possible to identify critical issues that do not necessarily depend on the adoption of intelligent technologies. After an introduction, this chapter is structured in three parts: the first briefly reconstructs the debate on the impact of artificial intelligence on user rights; the second introduces the distinction between deontic and adeontic digital regulatory artifacts and analyzes its usefulness in assessing the compatibility of regulatory tools with user rights. Finally, the third part examines two cases of digital regulatory tools belonging to these two different categories of artifacts, which, while not employing AI, nonetheless shape freedom of expression in different ways.

Keywords: Artificial Intelligence, Digital Regulation, Regulatory Artifacts, User Rights, Content Moderation.

* Il contributo è frutto di un lavoro congiunto e condiviso; i parr. 1 e 4 sono stati redatti da Giuseppe Lorini, il par. 2 da Simone Pinna e il par. 3 da Olimpia G. Loddo. Tutti gli autori hanno discusso e approvato i contenuti, assumendone congiuntamente la responsabilità scientifica.

Contributo realizzato nell'ambito del Piano Nazionale di Ripresa e Resilienza (PNRR), Missione 4, Componente 2, Investimento 1.1, Avviso n. 104 pubblicato il 02/02/2022 dal Ministero dell'Università e della Ricerca (MUR), finanziato dall'Unione Europea – NextGenerationEU – Titolo del progetto: Normative artifacts and normative drawings: investigating non-linguistic regulation (NAND) – CUP: F53D23003500006 – Decreto di concessione n. 968 adottato il 30/06/2023 dal Ministero dell'Università e della Ricerca (MUR).

SOMMARIO: 1. Introduzione – 2. Riflessione sull'IA e i suoi effetti perversi – 2.1. Note storiche sull'IA – 2.2. Il dibattito sul rapporto tra IA e libertà individuale – 3. Due strategie di regolazione digitale che non ricorrono necessariamente all'uso dell'IA – 3.1. Artefatti regolativi digitali *deontici* – 3.2. Artefatti regolativi digitali *adeontici* – 3.3. Strumenti deontici e adeontici di moderazione digitale dei contenuti – 4. Conclusioni: Tutta colpa dell'IA generativa? Effetti perversi di artefatti regolativi digitali senza l'ausilio dell'IA.

1. Introduzione

Il dibattito sull'impatto dell'intelligenza artificiale sulla libertà personale e sui diritti fondamentali si è intensificato negli ultimi anni, a partire dalla diffusione dell'intelligenza artificiale generativa¹. Nel corso di tale dibattito sono state messe in evidenza sia le potenzialità sia le criticità relative all'impatto che tali tecnologie possono avere sulla sfera giuridica individuale.

Alcuni autori pongono l'accento sulla necessità di un bilanciamento tra tutela di beni giuridici (es. protezione da *hate speech*, sicurezza) e rispetto delle libertà fondamentali². Altri insistono su trasparenza ed *explainability*³ come requisiti essenziali: in particolare emerge la necessità di rendere comprensibili le predizioni algoritmiche⁴; non a caso le *Ethics Guidelines for Trustworthy AI* (2019) hanno incluso la trasparenza tra i sette requisiti fondamentali che devono essere soddisfatti per garantire l'affidabilità dell'AI⁵. Un terzo filone sottolinea l'importanza

¹ L. Floridi, *The Ethics of Artificial Intelligence. Principles, Challenges, and Opportunities*, Oxford University Press, Oxford, 2023; S. Alegre, *Human Rights, Robot Wrongs*, Atlantic Books, London, 2024.

² L. Floridi, *Ibidem*.

³ Il termine si riferisce alla capacità di comprendere e descrivere il processo decisionale di un sistema di intelligenza artificiale in termini comprensibili per l'uomo, fornendo trasparenza su come e perché un modello è arrivato a un particolare risultato. Cfr. P.J. Phillips, C.A. Hahn, P.C. Fontana, D.A. Broniatowski, M.A. Przybocki, *Four Principles of Explainable Artificial Intelligence (NIST Interagency/Internal Report n. 8312)*, National Institute of Standards and Technology, Gaithersburg (MD), 2021.

⁴ T.Z. Zarsky, "Transparent Predictions", in *University of Illinois Law Review*, 4 (2013), pp. 1503-1570.

⁵ High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, European Commission, Brussels, 2019, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

della contestabilità e dei rimedi giuridici proposti dagli algoritmi: Hildebrandt⁶ e Yeung⁷ richiamano il diritto degli utenti a contestare decisioni algoritmiche e a disporre di canali di revisione umana. Secondo Hildebrandt, “[u]n divieto delle decisioni basate sulla profilazione dovrebbe mitigare l’erosione della presunzione di innocenza. [...] È imperativo che l’uso dell’IA venga testato e sottoposto a scrutinio da parte di poteri di controllo e che possa essere contestato in un tribunale”⁸. Infine, vi è un consenso crescente sulla necessità di prevenire fenomeni di discriminazione algoritmica. Ad esempio, si è mostrato come sistemi di gestione sanitaria possano incorporare *bias* razziali⁹; l’High-Level Expert Group on AI richiama espressamente nelle sue linee guida i principi di equità e non discriminazione¹⁰.

Tuttavia, come rilevato anche nell’ambito della regolazione tecnologica e digitale¹¹, artefatti digitali privi di componenti di intelligenza artificiale possono incidere profondamente sulla libertà individuale. Attraverso l’analisi del funzionamento di alcuni *software* destinati a regolare le interazioni online e tramite la distinzione tra artefatti regolativi deontici e adeontici (non-de-

⁶ M. Hildebrandt, “Law as Information in the Era of Data-Driven Agency”, in *The Modern Law Review*, 79 (2016), n. 1, pp. 1-30.

⁷ K. Yeung, “Algorithmic Regulation: A Critical Interrogation”, in *Regulation & Governance*, 12 (2018), pp. 505-523.

⁸ M. Hildebrandt, *Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology*, Edward Elgar Publishing, Cheltenham-Northampton, 2015, pp. 198-199, trad. nostra. Ecco il testo originale: “A prohibition of decisions based on profiling should mitigate the erosion of the presumption of innocence. [...] It is imperative that the usage of AI is tested and scrutinized by countervailing powers and made contestable in a court of law.”

⁹ Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, “Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations”, in *Science*, 366 (2019), n. 6464, pp. 447-453.

¹⁰ High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, cit.; Cfr. L. Floridi, J. Cowls, “A Unified Framework of Five Principles for AI in Society”, in *Harvard Data Science Review*, pubblicazione online ad accesso anticipato, 23 giugno 2019.

¹¹ M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.; V. Eubanks, *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin’s Press, New York, 2018; R. Brownsword, *Law 3.0. Rules, Regulation, and Technology*, Routledge, Abingdon, 2020.

ontici), è possibile evidenziare criticità che non dipendono necessariamente dall'adozione di tecnologie intelligenti¹².

Il presente contributo si articola in tre parti: nella prima parte viene ricostruito sinteticamente il dibattito sull'impatto dell'intelligenza artificiale sulle libertà dell'utente; nella seconda parte si presenta la distinzione tra artefatti regolativi digitali deontici e adeontici e si analizza l'utilità di questa distinzione per valutare la compatibilità degli strumenti regolativi con le libertà fondamentali dell'utente. Infine, si analizzano due casi di strumenti digitali regolativi appartenenti alle due diverse tipologie di artefatti che, pur non utilizzando tecnologie di intelligenza artificiale, incidono in modi diversi sulla libertà di espressione.

2. Riflessione sull'IA e i suoi effetti perversi

I recenti sistemi di IA generativa, come ChatGPT, Gemini o Claude, hanno reso evidente al grande pubblico la possibilità di interagire con macchine in grado di produrre testi, traduzioni e soluzioni a problemi complessi. Questo successo ha riportato al centro della scena l'idea di un'IA "forte" o intelligenza artificiale generale (AGI), ossia un sistema non più limitato a compiti specifici, ma capace di svolgere attività cognitive tipiche dell'essere umano, con possibili ricadute profonde sulla libertà personale, sulla distribuzione del potere e sulla stessa nozione di responsabilità.

2.1. Note storiche sull'IA

Pur con tratti inediti, l'attuale stagione tecnologica si inserisce in una storia lunga oltre settant'anni. Già Alan Turing si era chiesto se le macchine potessero pensare, aprendo la strada a una riflessione che ha accompagnato lo sviluppo dell'informatica e delle scienze cognitive¹³. Negli anni successivi, conferenze

¹² G. Lorini, S. Moroni, "Rule-Free Regulation: Exploring Regulation 'without Rules' and Apart from 'Deontic Categories'", in *Journal for the Theory of Social Behaviour* 52 (2022), n. 1, pp. 22-36; G. Lorini, S. Moroni, O.G. Loddo, "Regulatory Artifacts: Prescribing, Constituting, Steering", in *International Journal for the Semiotics of Law – Revue Internationale de Sémiotique Juridique*, 36 (2022), n. 1, pp. 211-225.

¹³ Alan Turing aprì la discussione con la domanda "Le macchine possono pensare?" e propose il Test di Turing come criterio operativo, cfr. A.M.

e programmi di ricerca hanno dato avvio a un campo disciplinare in cui approcci diversi hanno convissuto e si sono alternati, attraversando fasi di entusiasmo e di crisi¹⁴.

Dopo le prime delusioni, l'IA ha conosciuto nuove fasi di rilancio: dai sistemi esperti degli anni Ottanta, al *machine learning*, fino al *deep learning* che, a partire dagli anni 2010, ha segnato un punto di svolta con risultati nettamente superiori ai metodi tradizionali. Una svolta ulteriore è giunta nel 2017 con l'introduzione dell'architettura *Transformer*, che ha reso possibile lo sviluppo dei moderni *Large Language Models*¹⁵. La loro diffusione al grande pubblico a partire dal 2022 ha trasformato la percezione sociale dell'IA, alimentando entusiasmi e timori legati all'impatto che questi strumenti possono avere sulle forme di comunicazione, sulla conoscenza e sulle libertà fondamentali. In questa prospettiva, la nuova IA generativa non è un fenomeno del tutto nuovo, ma il risultato di un lungo percorso che ripro-

Turing, "Computing Machinery and Intelligence", in *Mind*, 59 (1950), n. 236, pp. 433-460.

¹⁴ La Conferenza di Dartmouth del 1956 segnò la nascita ufficiale dell'intelligenza artificiale, cfr. J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, Dartmouth College, Hanover (NH), 1956. Negli anni successivi si svilupparono approcci differenti – logico-simbolici, statistici e ispirati a modelli biologici – che conobbero alterne fortune. Seguirono infatti i cosiddetti "inverni dell'IA", dovuti a limiti teorici e tecnologici, cui fecero da contraltare le successive rinascite, dapprima con i sistemi esperti e, più recentemente, con il successo del *machine learning* e del *deep learning*; cfr. N.J. Nilsson, *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge University Press, Cambridge, 2010, e M. Mitchell, *Artificial Intelligence. A Guide for Thinking Humans*, Farrar, Straus and Giroux, New York, 2019.

¹⁵ Un momento decisivo fu la vittoria del *software* AlexNet a ImageNet nel 2012, che segnò la definitiva affermazione dei metodi di *deep learning*. Più recentemente, l'introduzione dell'architettura *Transformer* ha reso possibile lo sviluppo dei moderni *Large Language Models*, cfr. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention Is All You Need", in *Advances in Neural Information Processing Systems*, 30 (2017), pp. 5998-6008. Per un'introduzione non tecnica al funzionamento di questi sistemi, cfr. N. Cristianini, *Macchina Sapiens. L'IA e il futuro della nostra specie*, il Mulino, Bologna, 2023.

pone questioni ricorrenti: trasparenza, responsabilità e impatto sociale¹⁶.

2.2. Il dibattito sul rapporto tra IA e libertà individuale

Il dibattito contemporaneo sul rapporto tra intelligenza artificiale e libertà individuale si è articolato lungo diverse linee interpretative.

Da un lato, numerosi autori sottolineano i rischi di compressione delle libertà derivanti dall'uso pervasivo dell'IA. In particolare, l'accento è stato posto sul ruolo delle tecnologie di sorveglianza e sul potenziale di controllo sociale attraverso sistemi di tracciamento e profilazione su larga scala¹⁷. In questa prospettiva, l'IA non solo limita la *privacy*, ma incide direttamente sull'autonomia e sulla possibilità di esercitare in modo effettivo la libertà di scelta. Secondo Zuboff, in effetti, si tratta di aspetti strettamente connessi, perché “[e]sercitare il proprio diritto alla *privacy* produce una scelta [...]”. La *privacy* consente di decidere dove ci si vuole posizionare nello spettro tra segretezza e trasparenza in ogni situazione¹⁸. Analogamente, Eubanks¹⁹ ha mostrato come l'automazione del welfare possa sfociare in forme di *ghettizzazione digitale*, penalizzando soprattutto i gruppi più vulnerabili. Yeung ha descritto questo fenomeno come una “regolazione algoritmica” capace di “espandere, rafforzare e redistribuire potere, autorità e risorse, e [di] incidere sulle opportunità di vita fondamentali degli individui”²⁰.

¹⁶ M. Hildebrandt, “Law as Information in the Era of Data-Driven Agency”, cit.; R. Brownsword, *op. cit.*

¹⁷ M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.; S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Profile Books, London, 2019.

¹⁸ S. Zuboff, *op. cit.*, pp. 82-83, trad. nostra. Ecco il testo originale: “Exercising one’s right to privacy produces choice [...]. Privacy rights thus confer decision rights; privacy enables a decision as to where one wants to be on the spectrum between secrecy and transparency in each situation.”

¹⁹ V. Eubanks, *op. cit.*

²⁰ K. Yeung, *op. cit.*, p. 519, trad. nostra. Ecco il testo originale: “expand, reinforce and redistribute power, authority and resources, and [to] affect the critical life opportunities of individuals.”

Altri autori, per contro, evidenziano le *opportunità* offerte dallo sviluppo dell'IA. Floridi propone la distinzione tra *ethics by design* e *pro-ethical design*:

Entrambe sono prospettive liberali, ma l'*ethics by design* può assumere tratti lievemente paternalistici, nella misura in cui tende a favorire il compimento del “giusto” tipo di scelte, azioni, processi o interazioni da parte degli agenti coinvolti. Il *pro-ethical design*, invece, non implica necessariamente un orientamento paternalistico, poiché privilegia piuttosto la promozione della *riflessione critica* degli agenti stessi sulle proprie scelte, azioni e processi²¹.

D'altro lato, Gabriel suggerisce di valutare l'IA in termini di giustizia sociale, domandandosi se essa aumenti o riduca le capacità effettive degli individui. Come fa notare l'autore,

algoritmi di valutazione del rischio determinano sempre più spesso l'idoneità di una persona alla libertà su cauzione o alla libertà condizionale; la tecnologia di riconoscimento facciale è stata utilizzata per potenziare le capacità della polizia; e i sistemi di intelligenza artificiale indirizzano l'impiego delle attività di polizia mediante analisi predittiva. Nel contesto della mobilità economica e dell'accesso a servizi pubblici essenziali come il welfare, l'uso di strumenti algoritmici è parimenti influente, stabilendo chi ha diritto al sostegno sociale, chi può accedere all'edilizia popolare e quali famiglie vengono prese in carico dai servizi sociali²².

²¹ L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford, 2014, p. 190, trad. nostra. Ecco il testo originale: “Both are liberal, but *ethics by design* may be mildly paternalistic, insofar as it privileges the facilitation of the *right* kind of choices, actions, process, or interactions on behalf of the agents involved. Whereas *pro-ethical design* does not have to be paternalistic, insofar as it privileges the facilitation of *reflection* by the agents involved on their choices, actions, or process.”

²² I. Gabriel, “Toward a Theory of Justice for Artificial Intelligence”, in *Daedalus*, 151 (2022), n. 2, pp. 218-31, p. 220, trad. nostra. Ecco il testo originale: “risk-assessment algorithms increasingly determine a person's eligibility for bail or parole, facial recognition technology has been used to augment police capabilities, and AI systems direct the allocation of policing resources using predictive analytics. In the context of economic mobility and access to key public services such as welfare provision, the use of algorithmic tools is

In questa prospettiva, l'IA può contribuire ad ampliare l'accesso alle informazioni e ai servizi, favorendo inclusione e partecipazione, ma può anche avere effetti di cristallizzazione delle ingiustizie simili a quelli denunciati da Eubanks²³.

Un ulteriore filone di ricerca mette in luce il rischio di *concentrazione del potere economico e politico legato allo sviluppo dell'IA*. McQuillan interpreta le tecnologie di intelligenza artificiale come strumenti di “fascismo computazionale”, capaci di rafforzare assetti oligarchici²⁴, mentre Coeckelbergh insiste sui pericoli di dipendenza e manipolazione insiti nelle “tecnologie del sé”²⁵.

Infine, la questione della *responsabilità e della soggettività giuridica dei sistemi di IA* è oggetto di un acceso dibattito. Birhane, van Dijk e Pasquale hanno criticato l'idea di attribuire diritti agli algoritmi, sostenendo che “[d]a un punto di vista giuridico, la migliore analogia con i diritti dei robot non sono i diritti umani, bensì i diritti delle società, un concetto altamente controverso il cui effetto più importante è stato quello di minare i diritti dei lavoratori, dei consumatori e degli elettori”²⁶. Beckers e Teubner, in maniera più costruttiva, hanno proposto di ideare modelli innovativi di responsabilità collettiva per far fronte a una situazione in cui “[g]li esseri umani e gli algoritmi sembrano evolversi in un'entità simbiotica più grande della somma delle sue parti”²⁷.

similarly influential, determining who is eligible for welfare support, who has access to public housing, and which families are engaged by child services.”

²³ V. Eubanks, *op. cit.*

²⁴ D. McQuillan, *Resisting AI: An Anti-fascist Approach to Artificial Intelligence*, Bristol University Press, Bristol, 2022.

²⁵ M. Coeckelbergh, *Self-Improvement: Technologies of the Soul in the Age of Artificial Intelligence*, Columbia University Press, New York, 2022.

²⁶ A. Birhane, J. van Dijk, F. Pasquale, “Debunking Robot Rights Metaphysically, Ethically, and Legally”, in *arXiv preprint*, 2024, p. 1., trad. nostra. Ecco il testo originale: “From a legal perspective, the best analogy to robot rights is not human rights but corporate rights, a highly controversial concept whose most important effect has been the undermining of worker, consumer, and voter rights.”

²⁷ A. Beckers, G. Teubner, “Human–Algorithm Hybrids as (Quasi-)Organizations? On the Accountability of Digital Collective Actors”, in *Journal of Law and Society*, 50 (2023), n. 1, pp. 100-119, p. 112, trad. nostra. Ecco

In sintesi, il dibattito oscilla tra visioni che considerano l'IA prevalentemente come una minaccia alle libertà e altre che la interpretano come strumento di emancipazione. In posizione intermedia si collocano gli approcci che insistono sulla necessità di garantire criteri di *accountability*, trasparenza, contestabilità e non discriminazione, in modo da conciliare l'uso dell'IA con la tutela dei diritti fondamentali²⁸.

In questo contributo sosteniamo che il dibattito appena richiamato non riguarda, o non dovrebbe riguardare, soltanto l'IA – intesa come l'insieme di tecnologie che utilizzano tecniche di *deep learning* e di IA generativa – ma tocca più in generale le modalità attraverso cui gli artefatti digitali possono influenzare la libertà. Per questo, nella prossima sezione passeremo ad analizzare due strategie di regolazione digitale che operano anche senza l'uso dell'IA.

3. Due strategie di regolazione digitale che non ricorrono necessariamente all'uso dell'IA

Alcuni autori hanno sostenuto che le macchine non governeranno mai il mondo²⁹. Tuttavia, le tecnologie digitali stanno già rimodellando il modo in cui sono esercitate *governance* e regolazione. La nostra analisi si concentrerà ora sui meccanismi che consentono agli artefatti digitali di svolgere la loro funzione regolativa. In questo senso introduciamo ora la categoria di Artefatto Regolativo Digitale (ARD).

In letteratura emerge una distinzione tra il modo in cui le norme giuridiche svolgono solitamente la loro funzione regolativa e una nuova tendenza, che si manifesta soprattutto nel mondo digitale, consistente nell'opporre ostacoli cognitivi e materiali che scoraggiano le violazioni o addirittura le rendono impossibili.

il testo originale: “Humans and algorithms seem to evolve into a symbiotic entity greater than the sum of its parts.” Cfr. anche A. Beckers, G. Teubner. *Three Liability Regimes for Artificial Intelligence: Algorithmic Actants, Hybrids, Crowds*, Hart Publishing, Oxford, 2021.

²⁸ L. Floridi, J. Cows, “A Unified Framework of Five Principles for AI in Society”, cit.; High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, cit.

²⁹ J. Landgrebe, B. Smith, *Why Machines Will Never Rule the World. Artificial Intelligence without Fear*, Routledge, London-New York, 2022.

Interessante è la distinzione operata da Lessig tra due tipi di protezione. La prima è la protezione tradizionale della legge: definisce i confini dell'azione, vieta l'intrusione e punisce i trasgressori. La seconda è la protezione tramite barriera: un dispositivo tecnologico (una linea di codice) che, tra le altre cose, impedisce accessi indesiderati³⁰. Sulla stessa linea, Brownsword distingue tra libertà normativa e libertà pratica³¹. Altri studiosi hanno sottolineato come alcune forme di regolazione del comportamento basate su architetture di controllo, incidendo sulla libertà pratica, sollevino importanti questioni di legittimità e responsabilità rispetto alla regolazione tradizionale basata sulle norme³².

Per i nostri scopi, è interessante riflettere su come i dispositivi tecnologici in generale, e gli ARD in particolare, possano contribuire a entrambe le forme di protezione identificate da Lessig e possano incidere su entrambi i tipi di libertà distinti da Brownsword.

Alcuni ARD operano imponendo norme, possiamo quindi chiamarli artefatti regolativi digitali deontici (dARD). Altri ARD operano invece condizionando i comportamenti degli individui attraverso l'architettura. Questi artefatti non impongono regole ma sfruttano tendenze psicofisiche per spingere ad agire secondo modelli predefiniti. Tali dispositivi possono essere definiti artefatti regolativi digitali adeontici (-dARD). Nella sezione seguente viene fornita un'analisi più dettagliata della distinzione tra dARD e -dARD.

3.1. Artefatti regolativi digitali *deontici*

Con artefatti regolativi digitali deontici (dARD) si intendono ARD che svolgono la loro funzione regolativa attraverso regole. Ad esempio, i dARD possono indicare ciò che è consentito o proibito all'interno delle piattaforme digitali. Gli avvisi di età su piattaforme di streaming o giochi online informano gli utenti dei requisiti anagrafici per l'accesso, lasciando però agli individui la libertà di ignorare l'avviso e proseguire. Allo stesso modo,

³⁰ L. Lessig, *Code: Version 2.0*, Basic Books, New York, 2006.

³¹ R. Brownsword, *op. cit.*

³² K. Yeung, "Algorithmic Regulation: A Critical Interrogation", cit.; M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.

le piattaforme di *e-commerce* possono mostrare informazioni su norme fiscali o restrizioni alle importazioni. Tali messaggi guidano i comportamenti, ma gli acquirenti restano liberi di conformarsi o meno, seppure a rischio di sanzioni. In questo modo, i dARD condividono la caratteristica essenziale delle regole descritta da Schauer: guidano il comportamento pur rimanendo violabili³³. La loro forza regolativa non risiede nella coercizione, ma nell'offrire agli individui una scelta normativa: conformarsi o rischiare la sanzione.

Nella sfera digitale, dunque, i dARD svolgono un ruolo centrale nella comunicazione delle regole e nella trasmissione dei contenuti normativi, pur preservando un certo grado di flessibilità nell'osservanza. Non esercitano una "funzione di costrizione", cioè non impongono comportamenti specifici con la forza (come cancellare un post), né attivano automaticamente eventi specifici (come assegnare, cancellare o bloccare un profilo). Invece, segnalano l'esistenza di norme, lasciando agli utenti la libertà di decidere se conformarsi. In questa veste, incarnano una modalità di regolazione basata sulla consapevolezza e sulla scelta informata piuttosto che sulla coercizione diretta.

In generale, un dARD svolge un ruolo di comunicazione normativa. Non costringe l'utente a compiere o astenersi da un'azione, ma fornisce informazioni normative, modellando il comportamento attraverso la comunicazione. Questa funzione è stata discussa anche nella letteratura sulle architetture informative legali, che analizza come gli artefatti digitali incorporino la normatività nelle interfacce utente e medino il rapporto tra il sistema giuridico e i suoi soggetti³⁴.

3.2. Artefatti regolativi digitali *adeontici*

Nel quadro della sociologia del diritto digitale, l'analisi degli artefatti regolativi si è ampliata fino a includere una vasta gamma di strumenti non normativi. Nel caso degli artefatti regolativi digitali *adeontici* (-dARD) è il *design* stesso a imporre azioni specifiche attraverso processi automatizzati. In altre parole, i

³³ F.F. Schauer, *The Force of Law*, Harvard University Press, Cambridge (MA), 2015.

³⁴ M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.

-dARD non necessitano di strutture concettuali deontiche, anche se sono intenzionalmente ideati per influenzare e modificare il comportamento umano³⁵. Questi artefatti condizionano il comportamento degli utenti sia online sia “*onlife*” senza fare ricorso a norme. Hildebrandt ha sottolineato come gli ARD siano sempre più affiancati da forme di regolazione che non solo plasmano, ma anche costringono i comportamenti³⁶.

Possiamo menzionare diversi esempi. I sistemi avanzati di assistenza alla guida, ad esempio, traducono le informazioni normative in istruzioni non per gli utenti ma per il sistema di controllo elettronico, garantendo così il rispetto automatico delle regole. Un altro esempio sono i sistemi di gestione delle identità e degli accessi, che, incorporati nel codice delle piattaforme, concedono o negano l'accesso a sezioni specifiche in base a regole predefinite. Nelle piattaforme di *home banking*, ad esempio, gli utenti possono accedere solo ai propri conti autenticandosi. Nei *social network* i sistemi di moderazione algoritmica applicano le linee guida della comunità tramite meccanismi adeontici che possono rimuovere automaticamente contenuti o chiudere account³⁷.

Cohen mostra come architetture e artefatti non si limitino a vincolare o abilitare comportamenti, ma contribuiscano a plasmare attivamente gli orizzonti di scelta e di pratica entro cui gli individui operano³⁸. I -dARD incarnano precisamente queste forme di “regolazione intesa a influenzare i comportamenti delle persone senza ricorrere a norme, e senza alterare direttamente

³⁵ Si tratta di artefatti che svolgono una funzione regolativa senza ricorrere a strumenti deontici. Come sottolineano Lorini e Moroni: “Identifying (and distinguishing between) the set of normative non-regulatory interventions and the set of regulatory non-normative interventions helps to show that the regulatory function and the normative function can be performed separately.” G. Lorini, S. Moroni, “Rules without Regulation and Regulation without Rules”, in *Journal for the Theory of Social Behaviour*, 54 (2024), n. 2, pp. 216-234, p. 223.

³⁶ M. Hildebrandt, *Smart Technologies and the End(s) of Law*, cit.

³⁷ *Ibidem*.

³⁸ J.E. Cohen, *Configuring the Networked Self. Law, Code, and the Play of Everyday Practice*, Yale University Press, New Haven-London, 2012.

‘l’ambiente normativo’³⁹. La distinzione centrale tra artefatti deontici e adeontici risiede dunque nei loro modelli d’azione: i dARD si basano su norme esplicite, mentre i -dARD impongono comportamenti tramite il *design*. Questi vincoli possono essere fisici – in linea con il concetto di “*forcing function*” di Norman⁴⁰ – oppure psicologici, dove assomigliano a certi tipi di *nudges* identificati da Thaler e Sunstein⁴¹.

È utile un confronto tra il concetto di -dARD e la posizione di Floridi, che distingue tra *ethics by design* e *pro-ethical design*. Nel primo approccio, gli individui possono rinunciare a una preferenza preimpostata e sceglierne un’altra. Al contrario, le strategie basate sul *pro-ethical design* richiedono che gli individui prendano una decisione esplicita su una questione prima di procedere, senza orientare la scelta verso un’opzione predefinita. In questo modo, il *pro-ethical design* garantisce una scelta imparziale e informata, rispettando l’autonomia individuale nei processi decisionali⁴². Entrambi gli approcci possono essere implementati mediante i -dARD. Nell’*ethics by design* la regolazione adeontica risulta più incisiva, poiché alla *forcing function* che impone la manifestazione di una scelta si aggiunge un

³⁹ G. Lorini, S. Moroni, “Rule-Free Regulation: Exploring Regulation ‘without Rules’ and Apart from ‘Deontic Categories’”, cit., p. 23, trad. nostra. Ecco il testo originale: “regulation intended to influence people’s behavior without recourse to rules/norms, and without directly altering the ‘normative environment’.”

⁴⁰ Norman definisce una *forcing function* come un vincolo fisico “tale che un errore in una fase impedisce che avvenga la successiva”. Ad esempio, una *lock-in forcing function* può impedire a un utente di chiudere un programma prima di salvare le modifiche o di inserire un indirizzo e-mail valido. Cfr. D.A. Norman, *The Design of Everyday Things*, Revised and Expanded, Basic Books, New York, 2013.

⁴¹ Un *nudge* è un intervento di architettura delle scelte che, modificando il contesto decisionale senza imporre vincoli o incentivi significativi, orienta il comportamento degli individui in modo prevedibile. Secondo Thaler e Sunstein, le opzioni predefinite che spingono il cliente verso determinate scelte sono una forma di *nudge*. Esse sfruttano la tendenza psicologica delle persone a non modificare le impostazioni di default. Cfr. R.H. Thaler, C.R. Sunstein, *Nudge. Improving Decisions about Health, Wealth, and Happiness*, Yale University Press, New Haven, 2008.

⁴² L. Floridi, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, cit.

meccanismo di *nudging* che orienta il contenuto della decisione tramite l'impostazione di un *default*. Il *pro-ethical design*, al contrario, conserva la *forcing function* procedurale ma rinuncia consapevolmente a ogni orientamento della scelta. Artefatti che sfruttano tali meccanismi sono “oggetti che non sono intrinsecamente deontici”⁴³. Come osservano Lorini e Moroni, “essi permettono una regolazione senza regole e possono incoraggiare o scoraggiare determinati comportamenti senza dover ricorrere a regole e sanzioni formali”⁴⁴. Questa forma di regolazione, che opera senza riferimento a quadri deontologici, può nondimeno avere una portata etica.

Nella prossima sezione vedremo come la distinzione tra dARD e -dARD emerga chiaramente nei casi concreti di moderazione digitale, dove strumenti diversi incidono in maniera differente sulla libertà di espressione.

3.3. Strumenti deontici e adeontici di moderazione digitale dei contenuti

Quando parliamo di moderazione dei contenuti, non stiamo solo discutendo di *policy* o di algoritmi: stiamo osservando architetture che, in modi diversi, toccano la libertà di espressione. La distinzione tra artefatti regolativi digitali deontici (dARD) e adeontici (-dARD) ci aiuta a vedere ciò che spesso resta invisibile: come una piattaforma orienta o blocca le nostre azioni prima ancora che intervengano norme, sanzioni o giudizi.

Un dARD agisce tramite regole esplicite: rende saliente una norma, una raccomandazione, uno standard di condotta. Possiamo ascoltarlo, oppure no. È violabile per definizione. Un -dARD, invece, lavora sul *design*: crea barriere, instrada i flussi, inserisce attriti o cancella opzioni. Non chiede consenso: lo im-

⁴³ G. Lorini, S. Moroni, O.G. Loddo, “Deontic Artifacts. Investigating the Normativity of Objects”, in *Philosophical Explorations*, 24 (2021), n. 2, pp. 185-203, p. 190, trad. nostra. Ecco il testo originale: “objects that are not inherently deontic in nature.”

⁴⁴ G. Lorini, S. Moroni, “Rule-Free Regulation: Exploring Regulation ‘without Rules’ and Apart from ‘Deontic Categories’”, cit., trad. nostra. Ecco il testo originale: “can encourage or discourage behaviors without the need to rely on formal rules and sanctions.”

pone. In termini di libertà, i dARD incidono sulla *libertà normativa* (prescrivono cosa deve o non deve essere fatto), i -dARD incidono sulla *libertà pratica/ontica* (decidono cosa puoi fare, qui e ora).

Ad esempio, su YouTube, quando si sta per pubblicare un commento potenzialmente offensivo, compare il messaggio: “*Keep comments respectful.*” Questo suggerisce di riformulare il messaggio, magari moderandone i toni; tuttavia il messaggio può essere pubblicato comunque, anche senza variazioni⁴⁵. Qui l'artefatto è deontico: comunica una regola di convivenza, invita alla riflessione. È una forma di moderazione che non chiude possibilità: sposta l'attenzione, sollecita l'autocontrollo, ricorda la norma.

Se prendiamo in considerazione un *software* di moderazione come Coral-Moderation Dashboard, si configura uno scenario diverso poiché, in questo caso, i commenti non diventano pubblici finché un moderatore umano non li approva. Ci troviamo davanti ad un chiaro esempio di -dARD. Anche se la regola esiste, la funzione dell'artefatto non determina la sua imposizione e comunicazione. L'artefatto si limita a bloccare la comunicazione in attesa di una decisione del moderatore. In questo modo la libertà normativa non è rilevante dal momento che l'artefatto incide direttamente sulla libertà pratica (far arrivare quel messaggio a un determinato pubblico). Si tratta di un cambio di regime materiale della comunicazione: da flusso immediato a flusso condizionato.

I due modelli richiamati nell'esempio presentano rischi e garanzie molto diversi. Il *prompt* di YouTube rischia i falsi positivi ma con un costo relativamente basso in termini di libertà pratica: l'utente può pubblicare ugualmente. Coral, al contrario, può migliorare la qualità del dibattito e ridurre lo *spam*, ma può comportare ritardi, opacità in relazione ai criteri adottati dal moderatore e alle tempistiche e possibili *bias* individuali e istitu-

⁴⁵ Il *prompt* di YouTube può essere implementato anche senza modelli di machine learning (regola statica lato client/server o semplici euristiche). La presenza o l'assenza di ML non muta la natura deontica dell'artefatto, che comunica una norma senza bloccare la pubblicazione. Questo comunque mostra la rilevanza di questa distinzione anche per l'indagine sulla funzione regolativa di strumenti digitali che impiegano ML.

zionali, conferendo un potere di censura che, se non governato, può diventare invisibile e pervasivo. Occorre però prendere in considerazione il fatto che le piattaforme non sono tutte uguali, così come non lo sono i contesti entro i quali operano gli utenti. In *thread* ad alto rischio (sensibili a pratiche di disinformazione, incitamento all'odio, o che richiedono una particolare tutela per minori), una misura -dARD può essere necessaria e proporzionata, purché progettata in maniera rigorosa. Altrove, un "filtro deontico" come chiedere una conferma esplicita prima di pubblicare un contenuto che, agli occhi dell'IA, violi la *policy* della piattaforma, potrebbe offrire un ottimo compromesso al fine di garantire la qualità del dibattito senza soffocare la libertà di espressione.

4. Conclusioni: Tutta colpa dell'IA generativa? Effetti perversi di artefatti regolativi digitali senza l'ausilio dell'IA

In questo contributo, abbiamo visto in che senso i principi discussi nel dibattito sull'etica dell'IA (proporzionalità, trasparenza, contestabilità, non discriminazione) risultino rilevanti anche per valutare strumenti digitali che non integrano componenti di intelligenza artificiale (intesa come quella tecnologia che integra elementi di IA generativa e *deep learning*). Infatti, anche interfacce semplici, filtri automatici o barriere di *design* possono limitare la libertà di espressione o l'autonomia dell'utente, senza che vi sia alcuna "intelligenza" in gioco.

L'etica dell'IA e la regolazione degli artefatti digitali si muovono in un terreno ibrido, in cui principi giuridici consolidati (art. 21 Cost., art. 10 CEDU, art. 19 ICCPR) dialogano con i principi dell'*AI ethics* (*accountability, fairness, human oversight, transparency*). Su questo sfondo normativo, l'analisi proposta pone l'accento sul contesto in cui viene utilizzata una determinata tecnologia. Le questioni emergenti in tale dibattito non riguardano soltanto l'IA. Anche artefatti regolativi digitali più semplici, come meccanismi di pre-moderazione o *forcing functions* di *design*, possono compromettere la libertà pratica dell'utente. In questi casi, non vi è alcun modello predittivo da spiegare: il vincolo è inscritto nell'architettura stessa. Negli

esempi analizzati abbiamo riscontrato un dato apparentemente paradossale: è possibile che la libertà di espressione risulti maggiormente garantita da un moderatore automatico che da uno umano.

Per far fronte ai possibili effetti perversi derivanti dall'uso di artefatti regolativi digitali è dunque necessario conoscere a fondo anche le strategie di regolazione impiegate oltre agli ambiti di utilizzo. Sulla base di queste conoscenze preliminari, e non di principi etici dati *a priori*, è possibile intervenire per regolare l'utilizzo delle tecnologie e mitigarne i possibili effetti lesivi delle libertà personali.

La sfida, dunque, non è limitata alla fase decisionale dell'IA, ma investe il più ampio spettro degli ARD, deontici e adeontici. È proprio a questo livello che occorre chiedersi quali garanzie accompagnino l'uso di strumenti regolativi digitali e come possano essere rese compatibili con i diritti fondamentali degli utenti.

ANTONIO INCAMPO è Professore ordinario di Filosofia del diritto all'Università degli Studi di Bari Aldo Moro. È stato Direttore del “Dipartimento per lo studio del Diritto penale, del Diritto processuale penale e della Filosofia del diritto” e Presidente dei Corsi di laurea della Seconda Facoltà di Giurisprudenza nella stessa Università. Autore di numerosi studi di filosofia della validità giuridica, di semiotica e logica del linguaggio normativo, dirige per i tipi di Cacucci la collana editoriale “Unità del sapere giuridico. Quaderni di scienze penalistiche e filosofico-giuridiche”. È stato insignito nel 1998 dalla Presidenza della Repubblica d'Austria della “Gran Decorazione d'Onore” [*Großes Ehrenzeichen*] per meriti culturali

NICOLA TRIGGIANI è Professore ordinario di Diritto processuale penale all'Università degli Studi di Bari Aldo Moro e avvocato. È stato Direttore del “Dipartimento per lo studio del Diritto penale, del Diritto processuale penale e della Filosofia del diritto” e Coordinatore del Dottorato di ricerca in “Sistema penale e processo” nella stessa Università, nonché Giudice onorario presso il Tribunale per i minorenni di Bari. Autore di monografie e altri studi soprattutto in tema di investigazioni, prove e rapporti tra *media* e giustizia penale, è componente del comitato di direzione della rivista di classe A “Processo penale e Giustizia” e codirettore per i tipi di Cacucci della collana editoriale “Giustizia penale della post-modernità”.

ISBN 979-12-5965-606-3



9 791259 656063

La leggera colomba [...] potrebbe immaginare di volare più agevolmente
in uno spazio privo d'aria

Immanuel Kant, *Critica della ragion pura* (1787)

copia omaggio



GIUDIZIO PENALE E INTELLIGENZA ARTIFICIALE

a cura di A. Incampo e N. Triggiani

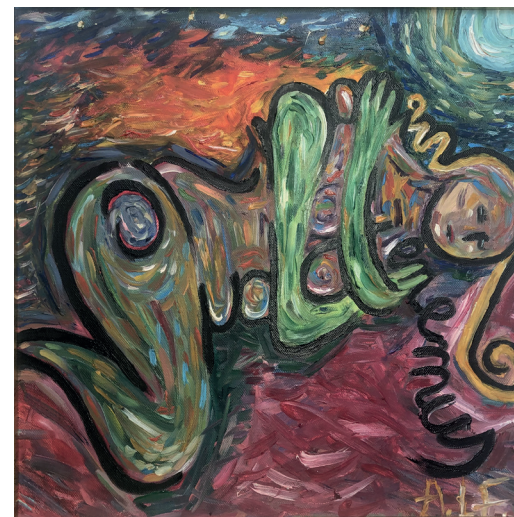
24

Unità del sapere giuridico

Quaderni di scienze penalistiche e filosofico-giuridiche

Collana diretta da

Antonio Incampo, Vito Mormando, Adolfo Scalfati



Giudizio penale e intelligenza artificiale

Una riflessione sistemica

a cura di Antonio Incampo e Nicola Triggiani

CACUCCI  EDITORE
BARI

Il volume esplora l’impatto dell’intelligenza artificiale sul diritto, mettendo in luce le profonde trasformazioni ontologiche, epistemologiche e normative che i sistemi esperti introducono nella razionalità giuridica contemporanea. Attraverso l’originale intreccio interdisciplinare di contributi teorici, tecnici e applicativi, l’opera riflette criticamente sulle nuove sfide poste dall’automazione dei processi decisionali e sui rischi connessi alle forme predittive di governance penale e all’uso della logica algoritmica nei contesti giudiziari e amministrativi. Ne emerge il ritratto di un diritto in transizione, chiamato a confrontarsi con un linguaggio “muto e opaco”, fondato sul calcolo più che sul significato, e con l’esigenza di preservare la centralità dell’essere umano in un ecosistema digitale dominato dalla potenza computazionale e dall’interconnessione globale.

The volume examines the impact of artificial intelligence on law, highlighting the profound ontological, epistemological and normative transformations that expert systems introduce into contemporary legal rationality. Through an original interdisciplinary interplay of theoretical, technical, and applied contributions, the work critically addresses the new challenges posed by the automation of decision-making processes, the risks associated with predictive forms of criminal governance, and the use of algorithmic logic in judicial and administrative contexts. What emerges is a portrait of a law in transition, called upon to confront a “mute and opaque” language grounded in calculation rather than meaning, and the need to preserve the centrality of the human being within a digital ecosystem dominated by computational power and global interconnection.

Annalisa la Forgia, *Non nomina nuda tenemus*
olio su tela, 37 x 38 cm (2012)
collezione privata

Giudizio penale e intelligenza artificiale
Una riflessione sistemica

a cura di
Antonio Incampo e Nicola Triggiani

CACUCCI  EDITORE
BARI

© 2026 Cacucci Editore – Bari

Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220

<http://www.cacuccieditore.it> e-mail: info@cacucci.it

Ai sensi della legge sui diritti d'Autore e del codice civile è vietata la riproduzione di questo libro o di parte di esso con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilms, registrazioni o altro, senza il consenso dell'autore e dell'editore.