LA PROPOSTA DI REGOLAMENTO EUROPEO SULLA INTELLIGENZA ARTIFICIALE E LA GESTIONE DEI RISCHI: UNA BATTAGLIA CHE PUÓ ESSERE VINTA? *

(The proposal for a european regulation on artificial intelligence and risk management: a battle that can be won?)

ABSTRACT: *La proposta di Regolamento europeo sull'Intelligenza Artificiale, approvata dal Parlamento dell'Unione Europea nel giugno 2023, rappresenta una notevole sfida normativa: infatti, le tecnologie e gli algoritmi di IA presentano un'ampia varietà di profili e pongono diversi problemi sui diritti dell'individuo e sulla stessa tenuta democratica; allo stesso tempo, possono rappresentare un'opportunità per un significativo miglioramento della vita umana. Il saggio si concentra su un aspetto particolare della proposta, la gestione del rischio, che ne costituisce per certi versi l'elemento più qualificante. In particolare, si cerca di verificare se le nuove regole possano essere sufficientemente adeguate a rispondere alla necessità che gli aspetti più controversi dell'IA siano adeguatamente regolamentati e, se necessario, vietati.*

The proposal for a European Regulation on Artificial Intelligence, which the EU Parliament approved in June 2023, represents a notable regulatory challenge: in fact, AI technologies and algorithms present a wide variety of profiles and pose several problems about the rights of the individual and the democratic resilience itself; at the same time, they may represent an opportunity for a significant improvement in human life. The essay focuses on one particular aspect of the proposal, risk management, which is in some respects its most qualifying element. In particular, it tries to examine whether the new rules may be adequate enough to respond to the

* Contributo approvato dai revisori.

need for the most controversial aspects of AI to be adequately regulated and, where necessary, prohibited.

SOMMARIO: 1. L'intelligenza artificiale, tra informatica, sociologia, diritto – 2. L'AI Act: impianto generale – 3. L'approccio al rischio: i divieti – 4. I sistemi ad alto rischio – 5. I sistemi a medio e minimo rischio – 6. Conclusioni

1. Il termine "intelligenza artificiale" è da tempo entrato non solo nel linguaggio comune, ma anche nell'immaginario collettivo. Il motivo è legato al fascino (ambivalentemente unito a paura) che da sempre l'Uomo prova per una entità "aliena" che possa competere con il suo tratto specifico fondamentale, ossia la razionalità e la capacità di ragionamento astratto[1].

Tuttavia, definire cosa in concreto sia una "intelligenza artificiale" è tutt'altro che semplice[2]. Il rischio, infatti, è di cadere nella totale ambiguità. Si pensi solo che il computer, agli albori dell'informatica, veniva chiamato in Italia "cervello elettronico", con evidente richiamo al funzionamento del cervello umano. Tuttavia, all'epoca, il quesito sulla possibilità di configurare una "intelligenza artificiale" era stato già posto da Alan Turing, in uno dei suoi più geniali scritti[3], ove si poneva il problema "*can machines think?*"[4].

Dunque, in senso generale (ma efficace) l'intelligenza artificiale può definirsi come un programma informatico in grado di interagire con la realtà come una

[1] Per un esame, sia pure sommario, del fondamento del mito dell'intelligenza artificiale rinvio a G. LEMME, *La transizione giuridica. La crisi del diritto di fronte alla sfida tecnologica*, Torino, 2023, pp. 127 ss.

[2] Sottolinea l'estrema ambiguità, se non l'impossibilità, di definire l'intelligenza artificiale P. TRAVERSO, *Breve introduzione tecnica all'intelligenza artificiale*, in Dir. pubbl. comp. ed eur., 2022, p. 157

[3] A. TURING, *Computer machinery and intelligence*, in Mind, 1950; nota G. FINOCCHIARO, *La regolazione dell'intelligenza artificiale*, in Riv. trim. dir. pubb., 2022, 1087, come il termine "intelligenza" sia ambiguo, e tenda a creare un pregiudizio di antropomorfizzazione di ciò che, in realtà, non è una vera e propria intelligenza, ma una macchina che si comporta "come se" fosse intelligente.

[4] Sulla portata filosofica dello scritto di Turing v. G. LANDI, *Intelligenza artificiale come filosofia*, Trento, 2020, pp. 24 ss.

persona[5]. Oppure, se vogliamo, un programma informatico in grado di implementare alcune delle funzioni superiori del cervello umano.

Una definizione più completa è quella data dallo High Level Expert Group della Commissione Europea sulla IA[6]: "*Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions*".

Infine, va menzionata la definizione fornita nella proposta di Regolamento UE sull'intelligenza artificiale, che fa riferimento (art. 3, lett. a.) a "*un sistema automatizzato progettato per operare con livelli di autonomia variabili e che, per obiettivi espliciti o impliciti, può generare output quali previsioni, raccomandazioni o decisioni che influenzano gli ambienti fisici o virtuali*"[7]; si noterà, in questa definizione, l'accenno alla possibilità di influenzare la realtà in base ad un *output*, ma tale definizione è tutt'altro che chiara, in quanto ogni tecnologia è teoricamente in grado di influenzare l'ambiente esterno in base ai propri *output*.

Se però vogliamo passare ad una definizione meno tecnica, ma che ponga le basi

[5] D. DOBREV, *A definition of artificial intelligence*, 2012, https://arxiv.org/pdf/1210.1568.pdf; anche agli albori dell'informatica di consumo, che ha portato ad un significativo aumento della potenza di calcolo dei computer, la definizione era sostanzialmente identica: "*The term artificial intelligence denotes behavior of a machine which, if a human behaves in the same way, is considered intelligent*" (A. B. SIMMONS – S. G. CHAPPELL, *Artificial intelligence – Definition and practice*, in IEEE Journal of Oceanic Engineering, 1988, p. 14).

[6] Commissione Europea, Rapporto *AI Watch. Defining artificial intelligence*, 2020

[7] La formulazione è frutto dell'emendamento del Parlamento Europeo rispetto al testo base proposto dalla Commissione, che recitava "*un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono*".

per le argomentazioni di tipo economico/giuridico, oltre che sociologico, possiamo pensare alla AI come ad un programma in grado di:

a. elaborare, a partire da alcuni dati, un ragionamento di tipo deduttivo, sia a livello matematico, sia soprattutto a livello simbolico;

b. riuscire ad apprendere, ossia a modificare il tipo di risposta in base alla esperienza dei dati precedentemente acquisiti[8].

Questa definizione ha il vantaggio di chiarire la differenza essenziale tra il semplice programma per computer (che da risposta identica a identica domanda, indipendentemente dalle circostanze) e l'intelligenza artificiale, che viceversa adatta le risposte alla medesima domanda basandosi sull'analisi delle risposte precedenti; ad esempio, verificando se l'operatore abbia o no scelto di validare una risposta precedentemente suggerita.

Per fare un caso pratico, appartiene all'intelligenza artificiale il programma operativo di un aspirapolvere robot che verifichi che in un certo punto della casa si accumula polvere più facilmente, e pertanto lo privilegi nell'esecuzione della routine di pulizia; oppure il software di gestione di una *playlist* musicale che proponga per primi i motivi che in passato sono stati selezionati più spesso[9].

Va peraltro chiarito – anche agli scopi del presente studio – che il presupposto in base al quale opera normalmente l'intelligenza artificiale è la disponibilità di una grande mole di dati[10]. Ciò appare naturale, dal momento che l'autoapprendimento, anche nelle intelligenze biologiche, si forma attraverso l'interazione con l'ambiente

[8] v. M. GABBRIELLI, *Dalla logica al* deep learning*: una breve riflessione sulla intelligenza artificiale*, in AA.VV., *XXVI lezioni di diritto dell'intelligenza artificiale*, a c. U. Ruffolo, Torino, 2021, p. 27; sottolinea l'importanza dell'elemento esperienziale anche F. DONATI, *Intelligenza artificiale e giustizia*, in Rivista AIC, 2020, pag. 415, mentre, per un profilo informatico (ancorché di tipo esplicativo) rimando a G. F. ITALIANO, *Le sfide interdisciplinari dell'intelligenza artificiale*, in Anal. Giur. Economia, 2019, 1, pp. 10 ss.

[9] Sui processi di autoapprendimento delle macchine, è stimolante la lettura di P. DOMINGOS, *L'algoritmo definitivo*, Torino, 2016, in particolare a pp. 47 ss.

[10] D. IACOVELLI, M. FONTANA, *Nuove sfide della tecnologia e gestione dei rischi nella proposta di regolamento europeo sull'intelligenza artificiale: set di* training*, algoritmi e profilazione dei dati. Rilievi critici*, in Il diritto dell'economia, 2022, p. 109

circostante (ivi comprese le altre intelligenze). Dunque, lo sviluppo dell'intelligenza artificiale va necessariamente di pari passo con quello dell'economia dei *Big Data* e con applicazioni particolari come la *Internet of Things*[11].

La differenza fondamentale tra l'apprendimento di una IA e quella di un essere umano è dovuta, peraltro, al fatto che la prima deve partire necessariamente da un *input* dato dal secondo, che in un certo qual modo potremmo paragonare ad un atto creativo di significato quasi religioso: l'Uomo crea la macchina, la quale, peraltro, proprio per le caratteristiche conferitele nella creazione, è poi in grado di evolvere autonomamente il proprio apprendimento.

A questo punto, l'agire della IA (o meglio, il modo in cui accumula la raccolta di dati che la fa "evolvere") può essere totalmente opaco per lo stesso "creatore"[12], con accentuazione del senso di distacco tra la volontà umana e quella artificiale[13]. Tale mancanza di trasparenza nel modello di apprendimento della IA fa sì che non possa essere identificato, ad esempio, se una determinata risposta ad un problema derivi dai *bias* forniti all'atto della programmazione, ovvero sia frutto di una corretta ed oggettiva interazione con i dati. Non dimentichiamo, infatti, che la caratteristica desiderata del risultato delle risposte della IA è in generale (e con alcune rare eccezioni) proprio l'oggettività, ossia l'assenza di condizionamenti che limitino la funzionalità della risposta fornita.

Questi sia pur sommari rilievi pongono immediatamente in luce la vera criticità dell'intelligenza artificiale: gli aspetti etici connessi al suo utilizzo, specie in settori sensibili, come la regolazione o la giustizia[14]. Ad esempio, è stato dimostrato[15] che gli

[11] Su quest'ultima mi permetto di rinviare a G. LEMME, *op. cit.*, pp. 121 ss.
[12] D. IACOVELLI, M. FONTANA, *op. cit.*, p. 110, ove ulteriori riferimenti; v. anche C. CASONATO, *L'intelligenza artificiale ed il diritto pubblico comparato ed europeo*, in Dir. pubbl. comp ed eur., 2022, pp. 172 ss.; P. TRAVERSO, *op. cit.*, p. 164
[13] Stigmatizza l'alienazione derivante dal rapporto uomo/algoritmo B. ROMANO, *Algoritmi al potere. Calcolo giudizio pensiero*, Torino, 2018, specie a pp. 23 ss. e p. 104
[14] S. GUERRA, *L'intelligenza artificiale tra sperimentazione normative e limiti etici del mercato globale*, in Tigor – Rivista di scienze della comunicazione e di argomentazione giuridica, 2022, pp. 105 ss.; G. LEMME, *op. cit.*, pp. 145ss.

algoritmi di giustizia predittiva (tipica applicazione della IA in campo giudiziario) tendono ad avere una funzione conservativa, mentre in altri casi possono addirittura alimentare pregiudizi verso alcune categorie di popolazione[16].

Se dunque è indubbio che l'intelligenza artificiale possa costituire – se ben gestita – una straordinaria opportunità di progresso e di semplificazione di alcune *routines* della vita umana, è altrettanto vero che non bisogna sottostimare i rischi connessi al suo utilizzo. È appena il caso di aggiungere, in proposito, che i sistemi di intelligenza artificiale si prestano ad essere usati sia per scopi criminali[17], sia per favorire il mantenimento del potere da parte delle dittature (si pensi, a quest'ultimo proposito, all'utilizzo distorto dei sistemi di riconoscimento facciale). Nota giustamente parte della dottrina[18] come l'intelligenza artificiale tenda fondamentalmente a sovrapporre l' "essere" al "dover essere", finendo per frustrare le spinte innovative tipiche dell'intelligenza umana, tanto da potersi potenzialmente porre, in alcune sue applicazioni, in contrasto con la nostra Carta costituzionale, la cui funzione adattiva ed evolutiva è nota.

È in questo quadro, quantomai complesso ed in continuo mutamento, che l'Unione Europea ha elaborato la proposta di Regolamento sull'intelligenza artificiale (AI Act).

2. Di fronte alla sfida posta dall'intelligenza artificiale, che sta pervasivamente estendendosi a svariati settori di mercato, il legislatore europeo, da sempre in prima linea nella regolazione dell'economia digitale, non poteva rimanere inerte.

---

[15] A. CHRISTIN, A. ROSENBLAT, D. BOYD, *Courts and predictive algorithms*, in atti del convegno *Data & civili rights: A new era of policy and justice*, Washington, 27 ottobre 2015, http://www.law.nyu.edu/sites/default/files/upload_ documents/Angele%20Christin.pdf.; G. TERRACCIANO, *I.I.A. (intelligenza artificiale amministrativa) e sindacato giurisdizionale*, in Amministrativamente, 2022, 2.

[16] F. LAGIOIA, G. SARTOR, *Il sistema compas: algoritmi, previsioni, iniquità*, in AA.VV., *XXVI lezioni di diritto dell'intelligenza artificiale*, cit., p. 230 ss.

[17] K. J. HAYWARD, M. M. MAAS, *Artificial intelligence and crime: a primer for criminologists*, in Crime Media Culture, 2020

[18] T. E. FROSINI, *L'ordine giuridico del digitale*, in Giur cost., 2023, pp. 391 ss.

È dunque stata emanata dalla Commissione la proposta di "Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale".

La proposta, presentata il 21.4.2021, è stata rivista dal Consiglio il 25.11.2022 e quindi finalmente emendata dal Parlamento Europeo il 14.6.2023[19].

Nell'esaminare il testo, che a questo punto dovrebbe essere definitivo, occorre partire da alcune considerazioni sull'impianto e la filosofia generale.

Partiamo anzitutto dalla considerazione che la fonte primaria del Regolamento è l'art. 114 Trattato FUE e dunque esso deve soddisfare una serie di obiettivi principali: sicurezza, rispetto dei diritti fondamentali, assicurazione della certezza del diritto, facilitazione dello sviluppo del mercato unico[20]. Inoltre, si vuole con il Regolamento raggiungere un elevato livello di standardizzazione, necessario per raggiungere gli obiettivi fissati nel regolamento stesso[21].

Lo strumento del quale il Regolamento si avvale è quello della regolazione orizzontale, ossia di un sistema di definizioni ampio e di applicazioni non indirizzate a casi specifici, ma a tutti gli ambiti possibili di applicazione della intelligenza artificiale[22].

Il dato, sottolineato da tutti i commentatori, è che l'approccio del

[19] Per una analisi delle modifiche introdotte nei vari passaggi, rinvio a F. FEDORCZYK, *AI legislation in flux: tracking evolving modifications of the AI Act*, in Diritti comparati – Comparare i diritti fondamentali in Europa, 2023

[20] K. YORDANOVA, *The EU AI Act – Balancing human rights and innovation through regulatory sandboxes and standardization*, 2022, https://kuleuven.limo.libis.be/discovery/search?query=any,contains,LIRIAS3708651&tab=LIRIAS&search_scope=lirias_profile&vid=32KUL_KUL:Lirias&offset=0; S. GUERRA, *op. cit.*, p. 107

[21] M. EBERS, *Standardizing AI – The case of the European Commission's proposal for an Artificial Intelligence Act*, 2021, https://ssrn.com/abstract=3900378; A. TARTARO, *Regulating by standards: current progress and main challenges in the standardization of artificial intelligence in the support of the AI Act*, in European Journal of Privacy Law and Technologies, 2023; J. LAUX, S. WACHTER, B. MITTELSTADT, *Three pathways for standardization and ethical disclosure by default under the European Union Artificial Intelligence Act*, Oxford Internet Institute, 2023, https://universitypress.unisob.na.it/ojs/index.php/ejplt/article/view/1792

[22] G. FINOCCHIARO, *op. cit.*, p. 1093; B. TOWNSEND, *Decoding the proposed European Union Artificial Intelligence Act*, 2021, https://eprints.whiterose.ac.uk/178738/

Regolamento si focalizza in particolar modo sui rischi[23]. Ci focalizzeremo di seguito, pertanto, su questo aspetto; preme intanto sottolineare come l'idea sottesa a tale approccio sia comprensibile, nella misura in cui – dati per scontati i risvolti positivi della tecnologia – se ne vogliano più che altro regolare in senso repressivo gli aspetti negativi. Per fare un paragone forse azzardato, ma efficace, la velocità delle auto è presa in considerazione dal Codice della Strada per limitarla e sanzionare coloro che superino i limiti stabiliti, e non in sé.

L'altro aspetto dell'impostazione del Regolamento, per certi aspetti conseguenza di quello ora sottolineato, è il tentativo di rendere il più possibile trasparente la IA e favorire il controllo umano sui risultati[24]. In entrambi i casi, l'evidente scopo è quello della più efficace gestione del rischio, e di evitare una eccessiva autonomizzazione della IA. È noto, infatti, che uno dei maggiori problemi insiti nella tecnologia è quello del c.d. *black box effect,* per il quale il modo in cui l'algoritmo arriva ad un determinato risultato non è percepibile all'esperienza umana[25]. Ciò, chiaramente, ha come conseguenza l'affidamento cieco delle persone al responso algoritmico, che tuttavia potrebbe essere – già lo accennavamo – il risultato di pregiudizi.

Comprensibile, dunque, è l'intento del legislatore europeo di – parafraso un'espressione usata per le società – *piercing the algorithmic veil*, e consentire una verifica sia *ex ante* che *ex post* del responso dell'intelligenza artificiale.

---

[23] D. IACOVELLI, M. FONTANA, *op. cit.*, p. 118; C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla proposta di Regolamento europeo in materia di intelligenza artificiale*, in BioLaw Journal – Rivista di BioDiritto, 2021, p. 421; J. SCHUETT, *Risk managemnt in the Artificial Intelligence Act*, in Europan Journal of Risk Regulation, 2023; C. NOVELLI, F. CASOLARI, A. ROTOLO, M. TADDEO, L. FLORIDI, *Taking the AI risk seriously: a new assesment model for the AI Act*, in AI & Society, 2023

[24] C. PANIGUTTI et al, *The role of excplainable AI in the conext of the AI Act*, Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, 1144

[25] W. J. VAN ESCHENBACH, *Transparency and the black box problem: why we do not trust AI*, in Philosophy and Technology, 2021; F. PASQUALE, *The black box society: The secret algorithms that control money and information*, Cambridge-London, 2015;

3. Come si accennava, tra i punti focali e qualificanti del regolamento vi è senza dubbio l'approccio al rischio.

La struttura immaginata in proposito nel Regolamento è sostanzialmente piramidale: si distingue, infatti, tra sistemi a rischio inaccettabile, rischio alto, rischio basso, rischio minimo, con conseguenze diverse sulle prescrizioni adottate.

I punti particolarmente delicati sono soprattutto quelli volti a identificare i sistemi di IA vietati in quanto intrinsecamente pericolosi (art. 5) e regolamentare in forma specifica quelli che, pur non essendo ineluttabilmente pericolosi, possano presentare un alto profilo di rischio (art. 6).

Quanto ai primi, l'elenco è lungo e articolato. In sostanza e in sintesi, sono vietati tutti i sistemi che tendono naturalmente a discriminare tra gruppi di persone, o ad influenzarne i comportamenti in maniera pervasiva (come le tecniche subliminali). L'elenco è frutto di un compromesso[26], e non è immune da critiche[27]. Vi è peraltro da notare che la versione approvata dal Parlamento è notevolmente migliorativa rispetto all'originale formulazione della Commissione. Ad esempio, le tecniche subliminali sono ora definite in maniera più chiara, usando, tra l'altro, una terminologia per certi aspetti analoga a quella rinvenibile nel Codice del Consumo in merito alle pratiche commerciali ingannevoli[28].

Analogamente, la proibizione delle tecniche di sfruttamento specifico di vulnerabilità appare improntata al tentativo di evitare che la IA possa indirizzarsi

[26] K. YORDANOVA, *op. cit.*, sottolinea come le grandi imprese, che usano algoritmi di profilazione, spingessero per un approccio più liberista, mentre le associazioni di consumatori ed in generale i gruppi a sostegno dei diritti umani avrebbero voluto restrizioni ancora più severe.

[27] G. CONTISSA, F. GALLI, F. GODANO, G, SARTOR, *Il regolamento europeo sull'intelligenza artificiale. Analisi informatico – giuridica*, in I-Lex. Scienze Giuridiche, Scienze Cognitive e Intelligenza Artificiale, 2021; M. FRANKLIN, H. ASHTON, R. GORMAN, S. ARMSTRONG, *Missing mechanisms of manipulation in the EU AI Act*, The International FLAIRS Conference Proceedings, 2022.

[28] Si, veda, all'art. 5, par. 1, lett. a), il riferimento alle tecniche aventi "*lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la capacità della persona di prendere una decisione informata, inducendo pertanto la persona a prendere una decisione che non avrebbe altrimenti preso, in un modo che provochi o possa provocare a tale persona, a un'altra persona o a gruppo di persone un danno significativo*".

insidiosamente a categorie di soggetti la cui capacità di resistere alle tecniche comunicative dell'algoritmo siano in qualche modo menomate. Anche in questo caso, a fronte delle critiche dei primi commenti dottrinari[29] il Parlamento è intervenuto estendendo notevolmente l'ambito di applicazione. Ciò che si vuole evitare, per usare i termini della proposta, è la distorsione comportamentale, che porta le categorie vulnerabili di persone a ricevere un danno dalla manipolazione indotta dalla IA.

Gli esempi di applicazioni della IA in regimi dittatoriali hanno indotto il legislatore europeo a vietare esplicitamente applicazioni di *social scoring*, basate sull'inquadramento dei soggetti alla luce di comportamenti sociali o tratti della personalità (ad esempio, partecipazione a manifestazioni politiche). Anche in questo caso, l'intervento del Parlamento europeo ha rimosso uno dei profili di criticità prontamente evidenziati[30], ossia la limitazione del divieto al software usato dalle pubbliche autorità.

Infine, il Regolamento (anche su questo punto, modificato in senso garantista dal Parlamento) prevede il divieto di uso di mezzi di identificazione biometrica a distanza in tempo reale, se non in specifiche circostanze.

Per quanto, come si diceva, vi siano numerosi miglioramenti apportati dal Parlamento rispetto alla proposta originale della Commissione, ciò non ha impedito di rimuovere del tutto alcune perplessità di fondo della dottrina, in particolare sulla eccessiva vaghezza del Regolamento, che alla fine rischia di non reprimere adeguatamente pratiche di marketing *border line* che sono volte, comunque, ad influenzare i comportamenti delle persone[31]. Ad esempio, per le pratiche di neuromarketing si è scelto l'approccio della trasparenza (art. 52 del Regolamento), allo scopo di non impedire tecniche ormai collaudate e giudicate irrinunciabili dall'industria, ma che evidentemente sono fortemente distorsive dei comportamenti economici dei

[29] G. CONTISSA, F. GALLI, F. GODANO, G, SARTOR, *op. cit.*
[30] *Id.*
[31] S. ORLANDO, *Regole di immissione sul mercato e "pratiche di intelligenza artificiale" vietate nella proposta di Artificial Intelligence Act*, in Persona e mercato, 2022, pp. 356 ss.

soggetti.

Il risultato, paradossalmente, è che spingere all'estremo una interpretazione delle pratiche proibite ai sensi dell'art. 5 del Regolamento renderebbe automaticamente illegale qualsiasi comunicazione pubblicitaria basata sulla IA, proprio per l'attitudine di questa ad "insinuarsi" maggiormente nelle menti dei soggetti, influenzandone e distorcendone i comportamenti[32]. Ma, dal momento che questo non è evidentemente possibile, l'art. 5 dovrà interpretarsi in chiave restrittiva, con il rischio (è il caso di dirlo) di renderlo una norma di impatto estremamente limitato dal punto di vista pratico.

Questi rilievi critici non fanno in verità che confermare – e con questo per certi aspetti anticipo, in parte, quelle che saranno le mie considerazioni finali – come quello del legislatore europeo sia, in un certo senso, un lavoro di inseguimento della tecnologia, il cui substrato economico ne comporta l'estrema elusività rispetto alle possibilità di compiuto inquadramento regolatorio. In altri termini, risulta arduo circoscrivere gli aspetti resi palesemente inaccettabili dalla loro pervasività e dannosità, da quelli che ricadono in pur sofisticate e sottilmente manipolatorie tecniche avanzate di comunicazione pubblicitaria. Per fare un esempio tratto da quest'ultima: c'è davvero una differenza sostanziale tra la pubblicità subliminale in senso stretto (vietata) e le tecniche di *product placement*, come l'inserimento, all'interno di un film o di una serie TV, di un prodotto il cui marchio sia chiaramente visibile?

4. Inquadrati così i sistemi di intelligenza artificiale il cui uso è proibito, il Regolamento, come anticipavo, prende in esame, questa volta allo scopo di regolarne e limitarne l'uso, i sistemi ad alto rischio.

La definizione del sistema ad alto rischio, piuttosto articolata, è contenuta

[32] *Id.*, pp. 359 ss.. Lo stesso A., a p. 363, suggerisce di inserire tutti i sistemi di neuromarketing basato su AI tra quelli "ad alto rischio", di cui appresso diremo.

all'art. 6 del Regolamento, e prende in considerazione due fattispecie fondamentali: anzitutto, il fatto che un sistema sia soggetto a valutazione di conformità *ex ante* in quanto presenti criticità e rischi per la salute e la sicurezza (quest'ultimo requisito è stato introdotto dal Parlamento); secondariamente, il fatto che il sistema rientri in settori critici individuati all'Allegato III al Regolamento.

Anche in questo caso, ove si esaminino le fattispecie elencate nell'Allegato III, il quadro che emerge rischia di essere frammentario e poco efficace[33], anche se la modificabilità dell'elenco di cui all'Allegato III, demandata in via autonoma alla Commissione (art. 7) presenta una maggior possibilità di adeguamento continuo della normativa[34]. Tuttavia, come sottolineato[35] rispetto al modello seguito per un approccio *risk based*, ossia il GDPR, l'attribuzione al legislatore, e non al gestore, di stabilire le regole di *compliance* introduce inevitabili meccanismi di rigidità, incompatibili con la rapida evoluzione del quadro tecnologico.

Se un sistema di IA è classificato ad alto rischio, il gestore dovrà mantenere costantemente in essere un sistema articolato di gestione dei rischi (artt. 9 ss. del Regolamento). La complessità di tale sistema ha portato giustamente a concludere[36] che il Regolamento rischia di avere effetti di contrazione dei corretti meccanismi concorrenziali in un settore così strategico, in quanto solo le imprese di grandi dimensioni potranno sostenere gli oneri, economici e non, dei requisiti imposti dalla normativa.

Nei sistemi ad alto rischio deve essere inoltre garantita la trasparenza del funzionamento, intesa come "*funzionamento…sufficientemente trasparente da consentire agli utenti di interpretare l'output del sistema e utilizzarlo adeguatamente*". Mi sembra – e non sono il solo a rilevare questo aspetto[37] - che

[33] G. CONTISSA, F. GALLI, F. GODANO, G, SARTOR, *op. cit.*
[34] C. CASONATO, B. MARCHETTI, *op. cit.*, pp. 424 ss.
[35] G. FINOCCHIARO, *op. cit.*, pp. 1095 ss.
[36] *Id.*, pp. 1096 ss.
[37] G. CONTISSA, F. GALLI, F. GODANO, G, SARTOR, *op. cit.*

l'illusione di una trasparenza dell'*output* sia per l'appunto tale, considerato che solo un ristretto margine di professionisti sono presumibilmente in grado di comprendere adeguatamente i processi decisionali della IA.

Il fatto che il Parlamento, nel modificare l'art. 13 del Regolamento, abbia esplicitamente spiegato come i dati debbano essere comprensibili ai "*fornitori e utenti*", risolve il problema solo in parte; così come non mi appare decisiva la spiegazione, sempre introdotta dal Parlamento, su cosa debba intendersi per "trasparenza"[38]. Infatti, si presuppone comunque che l'utente abbia una tale conoscenza del sistema di IA, da spiegare le decisioni adottate da quest'ultima. In altri termini, si trascura del tutto la non completa eliminabilità del *black box effect* di cui abbiamo già trattato, e si aderisce alla illusione che l'utente medio sia in grado di comprendere i processi di formazione della "volontà" della IA[39]. Viceversa, i processi informatici, di per sé, sono sempre opachi per l'utente medio non professionale[40], il che, come si accennava, rischia di rendere del tutto teorico e di fatto inutilizzabile (se non in ambiti e casi estremamente ristretti) il ricorso al principio di trasparenza immaginato dal Regolamento.

Non a migliori risultati, a mio avviso, giungono le norme sulla sorveglianza umana, contenute nell'art. 14 del Regolamento per i sistemi ad alto rischio. Come giustamente notato[41] lo scopo della IA è di sostituire efficacemente l'uomo in una

[38] "*per trasparenza si intende pertanto che, al momento dell'immissione sul mercato del sistema di IA ad alto rischio, sono utilizzati tutti i mezzi tecnici disponibili conformemente allo stato dell'arte generalmente riconosciuto per garantire che l'output del sistema di IA sia interpretabile dal fornitore e dall'utente. L'utente è in grado di comprendere e utilizzare adeguatamente il sistema di IA avendo una conoscenza generale del funzionamento del sistema di IA e dei dati trattati, il che gli consente di spiegare le decisioni adottate dal sistema di IA alla persona interessata*" (art. 13, par. 1)

[39] Ho qualche dubbio che a risolvere tale problema sia sufficiente una mera comprensione della "*logica di fondo*" del funzionamento della IA, come auspicato da A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in BioLaw Journal – Rivista di BioDiritto, 2019, p. 76 e C. CASONATO, B. MARCHETTI, *op. cit.*, p. 427.

[40] P. ZUDDAS, *Brevi note sulla trasparenza algoritmica*, in Amm. in cammino, 2020, p. 11; R. COVELLI, *Lavoro e intelligenza artificiale: dalla trasparenza alla conoscibilità*, in Labour & Law Issues, 2023, pp. 93 ss.

[41] A. D. SELBST, *Negligence and AI's human users*, in Boston University Law Review, 2020, p. 1315; G. CONTISSA, F. GALLI, F. GODANO, G, SARTOR, *op. cit.*

serie di compiti che, da solo, avrebbe difficoltà a svolgere; appare dunque utopistico immaginare che, in tali azioni, l'uomo possa sostituire la macchina, o anche solo arrestarla tempestivamente (art. 14, par. 4, lett. e))[42].

Anche sotto questo profilo, dunque, il Regolamento sembra presentare evidenti punti di debolezza. È facile immaginare che non potrà attuarsi efficacemente una sorveglianza continua della IA per ogni giorno e per ogni ora, specie, ancora una volta, per le PMI e le start-up, destinate pertanto ad essere nuovamente marginalizzate nella gestione dei sistemi ad alto rischio. D'altro canto, anche nei casi in cui la sorveglianza possa essere realizzata, è difficile che l'intervento umano possa essere efficace e tempestivo nel prevenire i rischi[43].

Va peraltro segnalata l'opinione[44] (non del tutto priva di fondamento) di quanti sostengono che alcuni sistemi ad alto rischio, interferendo con i diritti fondamentali dell'individuo, dovrebbero essere ricompresi tra quelli vietati. Sotto questo profilo, il Regolamento sembra effettivamente frutto di un compromesso con i creatori e gestori di IA artificiale, cui si sono voluti concedere margini ampi per continuare ad

[42] Significativamente, il Parlamento ha a tal proposito aggiunto che l'arresto può avvenire "*tranne se l'interferenza umana aumenta i rischi o è suscettibile di incidere negativamente sulle prestazioni in considerazione dello stato dell'arte generalmente riconosciuto*".

[43] Più ottimista appare, al riguardo, C. CASONATO, *op. cit.*, pp. 176 ss., il quale ritiene che l'art. 14 sia caratterizzato da sufficiente realismo, ma deve poi riconoscere che difficilmente la persona fisica incaricata della sorveglianza del sistema di IA possieda competenze e "coraggio" necessari ad interferire con il funzionamento di questa, se non in casi estremi.
Sotto un profilo ancora più pregnante, M. FASAN, *I principi costituzionali nella disciplina dell'intelligenza artificiale. Nuove prospettive interpretative*, in Dir. pubbl. comp. eur., 2022, pp. 197 ss., fa assurgere il principio di sorveglianza umana alla stregua di cardine costituzionale: "*la sorveglianza della persona umana diventa, quindi, una garanzia costituzionale, affinché l'applicazione di questa tecnologia non si traduca in una violazione dei diritti e delle libertà che lo Stato costituzionale di diritto attribuisce agli esseri umani nella loro dimensione individuale e collettiva*". Anche l'Autrice si pone, però, il problema della effettività e dell'efficacia dell'intervento umano, in relazione alla concreta comprensibilità del meccanismo di azione della IA (e, dunque, della permanenza del *black box effect*).
Sulle interazioni tra intelligenza umana ed artificiale v. anche N. RANGONE, *Intelligenza artificiale e intelligenza umana a supporto di una buona amministrazione*, in AA.VV., *Governance of/through Big Data*, a c. G. Resta, V. Zeno-Zencovich, Roma, 2023, pp. 159 ss.

[44] N. SMUHA *et al.*, *How can the EU achieve legally trustworthy AI: a response to the Commission's proposal for an Artificial Intelligence Act*, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3899991

operare i loro sistemi.

5. Nella struttura piramidale che, come abbiamo visto, caratterizza il Regolamento, la massima attenzione è rivolta ai sistemi a rischio inaccettabile ed alto. Molto minor spazio, comprensibilmente, è stato dedicato ai sistemi a rischio medio e a rischio minimo, per i quali le previsioni sono estremamente limitate.

In sostanza, per quanto riguarda i primi, ci si limita a prescrivere che le persone fisiche che interagiscano con la IA ne siano consapevoli; che, cioè, siano messe in grado di sapere che non hanno di fronte un interlocutore umano. Questo, in particolare, per il c.d. *deepfake*, ossia, per usare l'espressione contenuta nelle modifiche approvate dal Parlamento, "*testi o contenuti audio o visivi che potrebbero apparire falsamente autentici o veritieri e che rappresentano persone che sembrano dire cose che non hanno detto o compiere atti che non hanno commesso*".

È proprio il realismo del *deepfake*, assieme alla attitudine a potersi prestare alla diffusione di false notizie, ad aver suggerito che, pur non essendo a stretto rigore un sistema ad alto rischio, vi sia l'esigenza di renderne trasparente la natura.
Meno problematica appare la fattispecie dei *chatbots*, essendo in questo caso spesso molto evidente la loro natura artificiale; mentre, nel caso della terza categoria presa in considerazione dall'art. 52 del Regolamento (sistemi di riconoscimento delle emozioni) questi appaiono certamente più neutrali, non essendo volti alla identificazione degli individui[45].

Per quanto riguarda i sistemi a minimo rischio, il Regolamento (che non li definisce, se non per sottrazione (*sistemi di IA diversi dai sistemi di IA ad altro rischio*, art. 69) si limita a prevedere l'adesione volontaria degli operatori a codici di condotta. L'approccio minimale è giustificato dall'impatto limitatissimo, se non nullo, di questi

[45] G. CONTISSA, F. GALLI, F. GODANO, G, SARTOR, *op. cit.*; sui dati biometrici, in generale, v. D. IACOVELLI, M. FONTANA, *op. cit.*, pp. 130 ss.

sistemi (essenzialmente, filtri antispam e videogiochi[46])

6. Il titolo di questo saggio pone una domanda; è questo, dunque, il momento di darvi risposta.

Tuttavia, tale risposta non può essere univoca. Dobbiamo, infatti, porci il problema se la proposta di Regolamento abbia tentato di risolvere un problema la cui soluzione era necessaria (in altri termini, fosse a propria volta necessaria la regolazione sul tema) e se le norme siano adeguate nel contenuto.

Sul primo punto, penso che non si possa che rispondere affermativamente. È stato giustamente notato, prima ancora che la proposta venisse formulata, che "*la mancanza totale di regole può aprire la strada a pesanti conflitti tra innovazione ed altri valori tutelati*"[47]. Effettivamente, se nessuno può dubitare che l'innovazione tecnologica vada incoraggiata, in quanto chiave essenziale per l'attuale modello di sviluppo, è altrettanto vero che la tecnologia pone numerosi rischi, che minacciano i diritti fondamentali dell'individuo[48] e pongono in discussione lo stesso fondamento del diritto come lo conosciamo[49].

L'Unione Europea, che aspira ad un ruolo centrale nella regolamentazione delle tecnologie (ruolo che, sul piano dell'innovazione normativa, certamente ha raggiunto) non poteva astenersi dal dettare regole per quella che potrebbe essere, per certi aspetti, la più innovativa ma anche la più pervasiva e "rivoluzionaria" delle tecnologie.

Venendo però al merito delle regole, non sfuggirà certamente a chi ha letto le pagine che precedono come l'approccio scelto (di cui ho voluto trattare solo

[46] L. EDWARDS, *The EU AI Act: a summary of its significance and scope*, 2022, https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf

[47] L. AMMANNATI, *Verso un diritto delle piattaforme digitali?*, in AA.VV., *Algoritmi, Big Data, piattaforme digitali*, a c. L. Ammannati, A. Canepa, G.L. Greco, U. Minneci, Torino, 2021, pp. 13 ss.

[48] G. GUIGLIA, *L'Intelligenza artificiale e la tutea dei diritti umani nella prospettiva del diritto europeo*, in AA.VV., *Diritto costituzionale e nuove tecnologie*, a c. G. Ferri, Napoli, 2022, pp. 267 ss.

[49] G. LEMME, *La transizione giuridica*, cit., pp. 205 ss.

l'elemento a mio avviso più qualificante, ossia la gestione del rischio) sia colmo di criticità, che ho cercato di mettere in evidenza caso per caso, ma che in generale riguardano tre aspetti.

Anzitutto, l'innovazione tecnologica nel campo della IA creerà presumibilmente sistemi sempre più complessi ed invasivi, per i quali anche i meccanismi di adeguamento del Regolamento, pur apparentemente snelli, rischiano di risultare inadeguati. Si può creare, in altri termini, uno iato temporale nel quale sistemi ad alto rischio o addirittura da vietare non siano specificamente regolamentati, e possano pertanto essere lecitamente adottati.

Il secondo aspetto è quello relativo al rischio che il Regolamento, ponendo regole di sorveglianza molto stringenti, escluda dal mercato tutti i soggetti (PMI, start-up) non in grado di sostenere i requisiti previsti da tali regole.

Infine, la stessa classificazione dei sistemi di IA tra vietati, ad alto rischio, a rischio medio desta alcuni dubbi, rischiando di essere in alcuni casi eccessivamente "tollerante" verso algoritmi particolarmente critici.

Come sempre, non rimarrà che vedere il Regolamento alla prova dei fatti, per fare un bilancio, questa volta più accurato, sul suo impatto e sulla sua attitudine a tutelare i diritti individuali, senza al contempo chiudere la porta a tecnologie il cui scopo ultimo dovrebbe essere, almeno teoricamente, quello di migliorare la qualità della vita umana.

Giuliano Lemme

Ordinario di Diritto dell'Economia

nell'Università di Modena e Reggio Emilia