ISSN 1826-3534

federalismi.it
RIVISTA DI DIRITTO PUBBLICO ITALIANO, COMPARATO, EUROPEO

OSSERVATORIO TRASPARENZA
10 LUGLIO 2024

L'*Artificial Intelligence Act* e la prova di resistenza per la legalità algoritmica

di Germana Lo Sapio
Consigliere T.A.R. Campania

L'*Artificial Intelligence Act* e la prova di resistenza per la legalità algoritmica*

di Germana Lo Sapio
Consigliere T.A.R. Campania

Abstract [It]: L'AI Act è la prima normativa trasversale che si occupa della tecnologia general purpose di Intelligenza Artificiale (IA). Poiché la pubblica amministrazione, oltre che potenziale regolatore, è anche guardando alla classificazione dei sistemi di IA ad alto rischio tra i principali utilizzatori dei sistemi, l'articolo si pone l'obiettivo di verificare la tenuta dei principi affermati con riguardo alla cd. legalità algoritmica nel nuovo quadro regolatorio.

Title: The Artificial Intelligence Act and the test of endurance for algorithmic legality
Abstract [En]: The AI Act represents the inaugural comprehensive regulatory framework addressing the general-purpose Artificial Intelligence (AI) technologies. Since the public administration, besides being a potential regulator, is also, considering the classification of high-risk AI systems, among the main users of the systems. This paper seeks to assess the regulatory integrity of principles pertaining to what is referred to as algorithmic legality within the novel legislative context.

Parole chiave: Intelligenza Artificiale, Tecnologie generali, Trasparenza, Sorveglianza Umana, Modelli a scopo generale, ChatGPT, Large Language Models, Intelligenza Generale Generativa
Keywords: Artificial Intelligence, General Purpose Technologies, Transparency, Human Oversight, General Purpose Models, ChatGPT, Large Language Models, Generative Artificial Intelligence

1. Un Regolamento trasversale per l'Era dell'Intelligenza Artificiale e l'irruzione dell'IA Generativa

L'AI Act[1] approvato al Parlamento dell'Unione Europea, dopo oltre due anni di percorso legislativo, ha comunque un primato. È la prima iniziativa normativa globale che affronta in modo trasversale una tecnologia "*general purpose*" e in rapidissima evoluzione come quella che va sotto il nome di "Intelligenza Artificiale" (IA), passato così dal gergo quotidiano, cui era entrato solo pochi mesi fa, al linguaggio

* Articolo sottoposto a referaggio.
[1] Nel testo si usa l'acronimo IA per indicare Intelligenza Artificiale. Il testo cui si fa riferimento nell'articolo è quello, in versione italiana, che reca la data 13 giugno 2024, PE-CONS 24/1/24 REV 1, Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i Regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le Direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 d'ora in poi indicato anche come *AI Act* o *Regolamento*. Al momento della redazione, si prevede la pubblicazione Gazzetta Ufficiale dell'Unione europea nella data del 12 luglio 2024. Il Regolamento sarà pienamente applicabile solo a partire da agosto 2026, anche se alcune specifiche disposizioni saranno efficaci già in periodi antecedenti, come quelle sulle "pratiche vietata" e quelle sui sistemi di IA con finalità generali.

normativo. Per *general purpose technologies* (GPTs) ci si riferisce alla locuzione, elaborata dagli economisti[2], per descrivere l'impatto dirompente di invenzioni tecnologiche, come la macchina a vapore, l'elettricità, Internet, che sono il perno intorno a cui si sono compiute le Rivoluzioni Industriali. Per quanto affine, nel richiamo comune al "general purpose", questa locuzione economica va tenuta distinta da quella, che si colloca tutta dentro il mondo variegato dell'IA, di "*General Purpose Models*" (GPMs), la quale invece, grazie alla definizione ora contenuta nel Regolamento, ha valenza oramai giuridica[3]. Si tratta, in particolare, di due nozioni che prendono in considerazione profili e contesti diversi. Con la prima (GPTs), ci si riferisce a tecnologie che hanno un impatto vasto e vario su molti settori dell'economia, della società, della cultura, del lavoro, capaci pertanto di provocare trasformazioni significative nella produttività e nelle strutture economiche e sociali. Si sviluppano lungo ondate caotiche, a tratti disordinate, non prevedibili; anzi, in effetti, mai previste nella portata rivoluzionaria, come è facile comprendere, ricordando a metà degli anni '90 la comparsa sulla scena di Internet. Ma si contraddistinguono per una parabola peculiare: tanto sono dirompenti all'inizio, facendo compiere veri e propri salti al ritmo della storia e del progresso umano, tanto diventano silenziose con il tempo, a mano a mano che entrano a far parte della vita quotidiana, diventando accessibili, disponibili a costi sempre più bassi, necessarie.. L'IA rientra a pieno titolo nelle GPTs poiché ha già mostrato di incidere profondamente in tutti i settori economici, rivelando anche, in quanto tecnologia digitale, la sua integrabilità con altre tecnologie della cd. IV Rivoluzione industriale[4] (bio-tecnologie, Internet of Things, Blockchain, Realtà virtuale ed aumentata, Quantum Computing); procede con spinte di accelerazione, basti pensare all'irruzione di ChatGPT a novembre 2022; promette benefici nella medicina, nelle scoperte scientifiche, nella ottimizzazione e personalizzazione dei servizi pubblici e porta con sé rischi, per alcuni di carattere catastrofico per la specie

[2] Cfr. T.F. BRESNAHAN E M. TRAJTENBERG "*General purpose technologies Engines of growth?*", in *Journal of econometrics*, 1995, 65. Sul ciclo evolutivo, cfr. anche M. SULEYMAN, M BHASKAR, *The Coming Wave*, *Crown*, United States, 2023, p. 27-30 «*the irony of general-purpose technologies is that, before long, they become invisible and we take them for granted. Language, agricolture, writing –each was a general-purpose technologies at the center of an early wave. These three waves formed the foundation of civilization as we know it. Now we take them for granted (…) General-purpose technologies become waves when they diffuse widely. Without an epic and near-uncontrolled global diffusion, it'not a wave. It's a historical curiosity (…) proliferation is catalized by two forces: demand and the resulting cost decreases, each of wich divres technology to become even better and cheaper*». La rapidità della diffusione dell'AI è una delle novità più dirompenti ed è principalmente dovuta al fatto che essa si fonda su infrastrutture digitali (dati e potenze computazionali utilizzabili attraverso la Rete, anche in luoghi lontani da dove sono collocati i Super-Computer), cosicché non è comparabile a quella delle più recenti tecnologie *general purpose*, come l'elettricità.

[3] Art. 3 Definizioni, paragrafo 1, n. 63: *"modello di IA per finalità generali": un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato.*

[4] K. SCHWAB, *Governare la quarta rivoluzione industriale,* Franco Angeli, 2019, p. 23: "*l'espressione "quarta rivoluzione industriale" descrive l'insieme delle trasformazioni in atto nei sistemi che ci circondano, che spesso molti di noi danno per scontati. Sebbene a coloro i quali vivono piccoli, seppure significativi cambiamenti nella vita di tutti i giorni possa sembrare un fenomeno irrilevante, la quarta rivoluzione industriale segna un nuovo capitolo nello sviluppo umano, la cui importanza è pari a quelle delle guerre mondiali e il cui avanzamento come in passato è favorito dalla crescente disponibilità ed interazione tra innovazioni tecnologie straordinarie".*

dell'Homo Sapiens, di cui però nessuno riesce a predire i tempi. Intorno alla forza dirompente dell'IA sorgono questioni metodologiche nuove ed emerge l'urgente "*bisogno collettivo di tornare a pensare filosoficamente gli orizzonti socio-tecnologici emergenti*"[5], di trovare punti fermi per orientarsi in questa re-ingegnerizzazione della realtà[6]. Con i *General Purpose Model*, invece, nel contesto delle diverse branche dell'IA, ci si riferisce modelli di IA[7] che possono eseguire una vasta varietà di compiti in diversi domini e che sono solitamente addestrati su enormi quantità di dati con metodi diversi. La "generalità" riguarda pertanto i compiti che tali modelli possono svolgere sulla base del medesimo addestramento e quindi la loro attitudine ad essere impiegati nei più disparati domini, compreso quello legale. La sottocategoria di GPM sotto l'attenzione negli ultimi mesi, e dagli effetti dirompenti ancora da esplorare nel campo giuridico, sono i modelli linguistici (*Large Language Models*) [8], capaci di elaborare e generare testi in linguaggio naturale (sottocategoria entro cui rientra anche GPT – Generative Pre-trained Transformer - che è il modello base su cui si fonda nota applicazione ChatGPT, nelle sue più recenti versioni però già multimodale, avendo l'abilità di integrare linguaggio naturale, video, immagini, suoni). Così, almeno per tutto il 2023, e proprio per effetto dello tsunami ChatGPT, il dibattito politico-normativo si è incentrato sui GPM riconducibili all'area della IA cd. Generativa[9]; modelli che, basandosi su strutture di calcolo di

[5] C. Accoto*, Il mondo ex machina. Cinque brevi lezioni di filosofia dell'automazione,* Egea, 2019, p. 7.

[6] A. PAJNO, *La costruzione dell'infosfera e le conseguenze sul diritto*, in (a cura di) A. PAJNO, F. DONATI, A. PERRUCCI, *Intelligenza artificiale e diritto: una rivoluzione?*, Il Mulino, 2022, p. 19.

[7] Il legislatore europeo sottolinea anche nel Recital n. 97 la distinzione tra i "modelli di IA" e i "sistemi di IA", destinati però lungo la catena di valore dell'IA ad integrarsi: così "*sebbene i modelli di IA siano componenti essenziali dei sistemi di IA, essi non costituiscono di per sé sistemi di IA. I modelli di IA necessitano dell'aggiunta di altri componenti, ad esempio un'interfaccia utente, per diventare sistemi di IA. I modelli di IA sono generalmente integrati nei sistemi di IA e ne fanno parte*".

[8] Per una panoramica generale dei *Large Language Models*, cfr. OCSE, *AI Language Models. Technological, socio-economic and policy considerations*, Digital Economy papers, n. 352, 2023. Sui profili regolatori, cfr. P. HACKER, A. ENGEL, M. MAUER, *Regulating ChatGPT and other Large Generative AI Models*, 12 marzo 2023, disponibile nel sito arxiv; C. NOVELLI, F. CASOLARI, P. HACKER, G. SPEDICATO, L. FLORIDI, *Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity,* 14 gennaio 2024, *disponibile nel sito SSRN;* G. FASANO*, Le 'informazioni sintetizzate' generate dai large language models e le esigenze di tutela del diritto all'informazione: valori costituzionali e nuove regole, in www.Dirittifondamentali.it, 6 febbraio 2024* che sottolinea come di fatto i *Large Language Model* «*sono in grado di generare una narrativa testuale per mezzo di un'attività che non è meramente riproduttiva di fonti già esistenti, sono talmente pervasivi, consentendo a tutti di poter attingere informazioni da qualsiasi luogo e in qualsiasi momento, generalisti perché non hanno limiti di materie su cui pronunciarsi, persuasivi per via del loro stile narrativo calibrato sul profilo della persona e corrispondente alle sue aspettative di conoscenza, dall'aspetto confidenziale considerato il contesto in cui viene veicolata l'informazione sintetica, da risultare molto convincenti e seduttivi, in grado di inculcare un forte affidamento sui contenuti trasmessi*» p. 113*;* A. MALASCHINI, *ChatGPT e simili: questioni giuridiche ed implicazioni sociali,* in *www.Consultaonline.it., 6 luglio 2023;* da ultimo, S. DA EMPOLI, *L'economia di ChatGPT. Tra false paure e veri rischi,* Egea, Milano *2023*. Con riguardo invece all'impatto dei *Large Language Model* nella attività decisionale della pubblica amministrazione*, cfr.* G. GARULLO, *Large Language Models for Transparent and Intelligible AI-Assisted Public Decision-Making,* 19 settembre 2023, in www.ceridap.eu.; G. LO SAPIO, *Le nuove frontiere dei sistemi di elaborazione del linguaggio naturale tra ChatGPT e dintorni*, in *Atti del convegno su "intelligenza artificiale, diritti, giustizia e pubblica amministrazione"*, Palazzo Spada, 18 maggio 2023, www.giustizia-amministrativa.it.

[9] CEPEJ Working group on Cyberjustice and Artificial Intelligence (CEPEJ-GT-CYBERJUST), *Use of Generative Artificial Intelligence (AI) by judicial professionals in a work-related context,* 12 febbraio 2024 "*Generative Artificial Intelligence (AI) are software systems that communicate in natural language, able to give answers to relatively complex questions and can create content (provide a text, picture, or sound) following a formulated question or instructions (prompt). These tools include OpenAI ChatGPT, Copilot, Gemini, and Bard, all of which are developing rapidly*". Nel Regolamento non c'è una definizione di IA Generativa. Ma ad essa fanno

complesse (per la maggior parte dei casi costituite da "*reti neurali artificiali*") [10], addestrati ed alimentati da enormi quantità di dati e gestiti con le capacità computazionali di Super Computer, hanno l'abilità generativa, essendo capaci di produrre contenuti nuovi in testi linguistici, musica, immagini, video[11] rispetto ai dati di addestramento. L'ascesa repentina sotto i riflettori anche mediatici mondiali dell'IA Generativa ha fortemente inciso sul già tormentato percorso legislativo dell'*AI Act*. Ma il legislatore eurounitario non ha perso di vista l'obiettivo della *leadership* normativa e, aggiustando in corsa la propria rotta, ha disciplinato i fenomeni fino ad allora noti soprattutto agli addetti ai lavori, dei modelli *general purpose*[12], cui ora è dedicato l'intero Capo V (articoli 51 e ss) e dei *deep fake* (oggetto, tra l'altro, del Recital 134 e dell'articolo 3 n. 60[13]) e soprattutto, acquisita la consapevolezza dei rischi collettivi che la nuova dimensione diffusiva dell'IA Generativa comporta, ha dichiaratamente preso in carico anche la salvaguarda della democrazia e dello Stato di diritto (oltre che dell'ambiente), che ora compaiono tra gli scopi dell'articolo 1, accanto alla triade originaria "salute, sicurezza e diritti fondamentali"[14].

---

riferimento diverse disposizioni, almeno tutte quelle che riguardano i sistemi di IA che "*possono generare*" contenuti, testi, output, al punto da non rendere facile "*distinguere dai contenuti autentici e generati da esseri umani. L'ampia disponibilità e l'aumento delle capacità di tali sistemi hanno un impatto significativo sull'integrità e sulla fiducia nell'ecosistema dell'informazione, aumentando i nuovi rischi di cattiva informazione e manipolazione su vasta scala, frode, impersonificazione e inganno dei consumatori*" (Considerando 133).

[10] Il nuovo linguaggio che l'IA fa emergere, contaminando quello giuridico, è ricco di locuzioni evocative di similitudini con il comportamento umano oramai entrate anche nella comunicazione divulgativa. Ad esempio, una descrizione efficace di Deep learning e reti neurali artificiali è quella offerta da M. MITCHELL, *L'intelligenza Artificiale, una guida per essere umani pensanti,* I Maverick, 2022, p. 60 "*Deep learning indica semplicemente i metodi volti all'addestramento delle deep neural network, ossia reti neurali con più di uno strato nascosto. (…) Gli stati nascosti sono gli strati di una rete compresi tra l'input e l'output. La profondità di una rete è il suo numero di strati nascosti: una rete "poco profonda" (…) ha un solo strato nascosto; una rete "profonda" me ha più di uno. E' opportuno evidenziare questa definizione: il profondo (deep) del deep learning non si riferisce alla complessità delle cose imparate, ma alla profondità degli strati della rete addestrata*". Ancora più semplice la descrizione di R. CUCCHIARA, L'Intelligenzza non è artificiale, Mondadori, 2021, p. 60 "*Gli algoritmi di machine learning sono tanti (…), ma quelli che ora funzionano meglio sono quelli basati su reti neurali artificiali, più o meno profonde. Quando in machine learning si iniziano ad usare le reti neurali artificiali, tutto cambia. L'algoritmo si generalizza, rimane sempre lo stesso, si basa su neuroni, appunto, e tutto dipende dai dati. Con lo stesso modello neurale si può riconoscere una musica, una voce, la strada per un'auto autonoma in corsa*".

[11] *L'Artificial Intelligence Index Report* del 2024 dell'Università di Stanford conferma che nel 2023 i maggiori investimenti privati (in un solo anno pari ad otto volte quelli dell'anno precedente) e le maggiori performance si sono avute proprio nell'ambito dell'IA cd. Generativa. Secondo il Report, nonostante un generale declino degli investimenti privati nel settore dell'IA, quelli per la *Generative IA* sono aumentati fino ad oltre 25 miliardi di dollari nel mondo. Solo in Italia, nel 2023 la crescita del giro di affari per questa branca di IA è stata pari al 52%. È però anche cresciuta non solo l'attenzione dei regolatori, alle prese con le nuove sfide, ma anche la diffusa consapevolezza circa i rischi e l'impatto di questa tecnologia per le professioni, specie quelle intellettuali, come riportato nel capitolo del Report dedicato.

[12] Accanto ad essi, in una fase del percorso parlamentare, anche i *Foundation Model*, dai quali però sono sostanzialmente indistinguibili; cfr. nota 55.

[13] "*deep fake*": *un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona*"

[14] Articolo 1 paragrafo 1: "*Lo scopo del presente regolamento è migliorare il funzionamento del mercato interno e promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, contro gli effetti nocivi dei sistemi di IA nell'Unione, e promuovendo l'innovazione*". Nella proposta originaria della Commissione del 21 aprile 2021 erano democrazia e Stato di diritto erano citati solo con riguardo alle pratiche vietate di manipolazione, sfruttamento e controllo sociale e con riferimento ai sistemi destinati all'amministrazione della giustizia e ai processi democratici.

Questo articolo esplora alcuni profili chiave dell'AI Act, ma con lo sguardo rivolto al diritto amministrativo nazionale, nell'ambito del quale è in via di formazione quasi un "diritto speciale" dell'amministrare per algoritmi: orientamenti giurisprudenziali, elaborazioni scientifiche, disposizioni settoriali, che ruotano intorno a principi chiave come quelli della trasparenza algoritmica, della supervisione umana e del divieto di non discriminazione. A tale scopo, premessa una ricognizione dei canoni della legalità algoritmica (paragrafo 2) e degli scopi, espressi e taciti, del Regolamento (paragrafo 3), l'articolo si sofferma sulla nuova definizione normativa di IA (paragrafo 4) che, enfatizzando il concetto chiave di "autonomia", solleva delicate questioni in relazione alla necessaria "*supervisione umana*" (Human In The Loop); si sofferma, poi, sull'impatto che l'irruzione della IA generativa ha avuto rispetto all'approccio originario *risk-based* del Regolamento (paragrafo 5); e, infine, prende in considerazione la dimensione specifica che il principio di trasparenza assume nell'AI Act (paragrafo 6), al fine di verificare se esso è allineato con quello elaborato nell'alveo della legalità algoritmica (paragrafo 7).

2. L'AI Act e la legalità algoritmica: la sfida è appena cominciata

Regolare tecnologie innovative, dirompenti e dall'evoluzione rapida ed imprevedibile è già una missione quasi impossibile per ogni legislatore. Dal punto di vista del diritto amministrativo, poi la cosa è ancora più complicata perché il terreno delle decisioni basate su algoritmi non era proprio inesplorato. Il legislatore eurounitario ha scelto però la strada della trasversalità: stesse regole per tutti, dentro il mercato europeo. L'AI Act è infatti destinato ad applicarsi a qualunque settore[15] e a qualunque soggetto coinvolto nel ciclo di vita dei sistemi di IA, dovunque abbia la sede, purché il sistema di IA sia immesso nel mercato unico europeo. Tra gli attori della complessa filiera[16]compare ovviamente anche la pubblica amministrazione in senso lato ("*organismi di diritto pubblico o gli operatori privati che forniscono servizi pubblici*" Recital 96 e articolo 27[17]) quale categoria speciale, e oltremodo onerata dai doveri di *compliance*, nella famiglia dei cd. *deployer*[18], almeno nell'ipotesi, che sarà frequente, di sistemi o modelli di AI acquistati sul

[15] La natura trasversale dell'IA e l'ambizione generale dell'AI Act non escludono ampi ambiti dal suo perimetro di applicazione. Anche a prescindere dalla espressa esclusione dei sistemi di IA impiegati per scopi militari, di difesa o di sicurezza nazionale, domìni che restano riservati alla sovranità degli Stati membri, una ulteriore esclusione si rinviene nel primo allegato (Allegato I, sezione B, cui rinvia l'articolo 2, paragrafo 2) e riguarda settori civili che sono già all'avanguardia nelle applicazioni di IA tra cui compaiono ad esempio, alcuni ambiti dell'aviazione civile e il mercato dei veicoli a motore.

[16] Ancora più complessa nel caso dei modelli con finalità generali: cfr. S. KUSPERT, N. MOES, C. DUNLOP, *The value chain of general-purpose AI*, 10 febbraio 2023, *Ada Lovelace Institute*.

[17] Articolo 27 paragrafo 1: "*Prima di utilizzare un sistema di IA ad alto rischio di cui all'articolo 6, paragrafo 2, ad eccezione dei sistemi di IA ad alto rischio destinati a essere usati nel settore elencati nell'allegato III, punto 2, i deployer che sono organismi di diritto pubblico o sono enti privati che forniscono servizi pubblici e i deployer di sistemi di IA ad alto rischio di cui all'allegato III, punto 5, lettere b) e c), effettuano una valutazione dell'impatto sui diritti fondamentali che l'uso di tale sistema può produrre*".

[18] Il *deployer,* secondo l'articolo 3 n. 4. è definito come «*persona fisica o giuridica, autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale*». Nella proposta originaria veniva utilizzato il termine molto più equivoco di "user" che però faceva

mercato. L'AI Act non è cioè pensato avendo in mente in particolare le applicazione dell'IA nel settore pubblico e quindi i principi cui deve conformarsi l'azione amministrativa, a prescindere dalle tecnologie utilizzate; non ha come scopo specifico l'abilitazione di una PA "*aumentata*" [19], ma quello di garantire, mediante la forza cogente un Regolamento ex art. 288 TFUE, certezza agli operatori economici e ai cittadini del mercato digitale, promuovendo uno sviluppo della tecnologia in linea con i valori unionali[20] in tutti i settori pubblici o privati che siano, tranne quelli espressamente esclusi. Senonché, l'AI Act non è una un'isola solitaria in mezzo al mare. Giunge nell'ordinamento nazionale, quando, sul tema delle decisioni della pubblica amministrazione fondate su algoritmi (decisioni *robotiche*, decisioni *automatizzate*, decisioni *algoritmiche*, secondo le diverse nomenclature; anche ADM, *Automated Decision Making*), orientamenti giurisprudenziali ed elaborazioni dottrinali hanno già compiuto notevoli passi avanti, conducendo alla elaborazione di una versione "in salsa digitale" del principio di legalità: la cd.. legalità algoritmica[21] che ruota intorno ai principi della trasparenza, della sorveglianza umana e del divieto di

pensare ai destinatari finali che, tramite le applicazioni anche di facile uso, si interfacciano con l'IA ormai in ogni aspetto della vita quotidiana.

[19] C. CUTILI, *Intelligenza artificiale & Pubblica amministrazione, Guida alle applicazioni dell'AI per il settore pubblico*, ISSRF, 2024, p. 116.

[20] Pertanto, nel caso i sistemi di IA siano sviluppati o applicati dalla pubblica amministrazione, vengono in gioco, oltre alla disciplina generale che si applica ai soggetti privati, anche alcune specifiche disposizioni: ad esempio, per i sistemi classificati "ad alto rischio" (Allegato III), sorge per l'amministrazione, anche se non ha il ruolo di "fornitore", l'obbligo della valutazione d'impatto sui diritti fondamentali e l'obbligo di registrazione nella banca dati pubblica. L'obbligo di registrazione è sancito ora dall'articolo 49 paragrafo 3 che prevede "*Prima di mettere in servizio o utilizzare un sistema di IA ad alto rischio elencato nell'allegato III, ad eccezione dei sistemi di IA ad alto rischio elencati nel punto 2 dell'allegato III, i deployer che sono autorità pubbliche, istituzioni, organi e organismi dell'Unione o persone che agiscono per loro conto si registrano, selezionano il sistema e ne registrano l'uso nella banca dati dell'UE di cui all'articolo 71*".Per i *deployer* privati la registrazione è invece su base volontaria.

[21] A. SIMONCINI, *Diritto costituzionale e decisioni algoritmiche*, in S. DORIGO (a cura di) *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Pisa, 2020, p. 37 e ss.; La letteratura in materia è già sterminata. In considerazione del contesto limitato di questo articolo, si suggeriscono, in particolare, tra gli studi più recenti, E. BELISARIO, G. CASSANO (a cura di), *Intelligenza artificiale per la pubblica amministrazione*, Pacini giuridica, 2023, p. 427, anche per la copiosa bibliografia ivi citata e A. DI MARTINO, *Tecnica e potere nell'amministrazione per algoritmi*, Editoriale scientifica, 2023, che si segnala per l'approfondimento scientifico e sistematico al tema. In attesa dell'istituzione degli enti preposti alla *governance* nell'ordinamento interno (Autorità nazionali), imposta dallo stesso AI Act che, come già accaduto per il diritto dell'ambiente con l'istituzione del Ministero preposto con la legge n. 349/1986, potrebbero fungere anche da centri di riferimento per coagulare in un "sistema" i principi giurisprudenziali, l'elaborazione scientifica e le norme giuridiche per ora solo settoriali, si può perfino ritenere che abbia preso l'abbrivio, nell'ambito del diritto pubblico, un settore scientifico autonomo quale quello dell'amministrazione algoritmica, ad alto contenuto di tecnicità, da indagare necessariamente con un approccio interdisciplinare e nel quale sono messe a dura prova le categorie generali tradizionali del diritto amministrativo. Sul cambiamento culturale che si impone sotto il profilo del necessario "dialogo tra conoscenze" cfr. MARIO R. SPASIANO, *Frammentazione delle conoscenze e unità del sapere*, in Diritto e Società, 2, 2022, p. 447 e ss.; Sulla *governance*, cfr. F. ALUIGI, A. GAROFALO, J. PIEMONTE, Governance dell'AI: la strategia Ue e il dilemma italiano, 13 marzo 2024, in agendadigitale.it. Per l'Unione Europea, cfr. la decisione della Commissione Europea del 14 gennaio 2024 per l'istituzione a decorrere dal 21 febbraio 2024 dell'AI Office. La Spagna è stata invece il primo Stato membro ad istituire l'Agenzia nazionale per la supervisione dell'IA, con il decreto regio n. 729/2023: cfr. decreto regio. Il disegno di legge in materia di intelligenza artificiale di cui alla nota precedente prevede l'istituzione di due autorità nazionali per l'intelligenza artificiale, l'Agenzia per l'Italia digitale (AgID) e l'Agenzia per la cybersicurezza nazionale (ACN). Deve in ogni caso sottolinearsi che secondo l'ultima versione dell'articolo 88 dell'AI Act, la Commissione ha il potere esclusivo di supervisionare l'attuazione delle regole previste per i cd. General Purpose Model (sulla cui definizione, cfr. infra paragrafo seguente), che sono il motore, almeno all'attualità, della recente evoluzione

discriminazione. Tanto è stata rapida l'evoluzione della letteratura scientifica e della giurisprudenza in materia, che i canoni della legalità algoritmica sono già diventati diritto positivo, mediante l'entrata in vigore, ormai un anno fa, dell' art. 30 del decreto legislativo 31 marzo 2023, n. 36 *"Codice dei contratti pubblici in attuazione dell'articolo 1 della legge 21 giugno 2022, n. 78, recante delega al Governo in materia di contratti pubblici"*); e, a breve, mediante l'articolo 6 del decreto legislativo recante semplificazione dei controlli sulle attività economiche in attuazione della delega al governo di cui all'articolo 27, comma 1, della legge 5 agosto 2022, n. 118[22] che ha una formulazione analoga[23]. La curva evolutiva della tutela di nuovi interessi, e dell'esigenza di mitigazione dei nuovi rischi associati all'IA, non è sotto questo profilo originale. Appartiene al DNA del diritto amministrativo la sua natura "giurisprudenziale" ovvero il fatto che, in prima istanza, il bisogno di tutela sia colmato nelle aule giudiziarie, prima che intervenga il legislatore, applicando i principi generali che regolano l'azione amministrativa, di derivazione costituzionale e sovranazionale, ed interpretando in senso evolutivo le norme vigenti. E tale natura, che si radica nella storia del diritto amministrativo, non è contraddetta dalla bulimica produzione legislativa contemporanea che ne esce perfino rafforzata: perché più il quadro normativo è frammentario, incoerente, instabile nel tempo, più la giurisprudenza è costretta ad assumere un ruolo essenziale di tutela delle situazioni giuridiche soggettive, specie dinnanzi a fenomeni nuovi, con sforzi interpretativi anche dei principi generali che a volte si pongono ai limiti dell'ossequio ai criteri ermeneutici indicati nelle preleggi. Nel mondo, solo in parte esplorato, delle decisioni algoritmiche, la storia si è pertanto ripetuta: "*di fronte alla rapidità del progresso la normativa può farsi attendere, ma la tutela no, e il giudice si trova non di rado ad essere l'istituzione di prima linea che deve far fronte al problema, dovendo comunque fornire una risposta, una soluzione giuridica, alla richiesta di tutela. Richiesta che arriva inaspettatamente, senza annunci o definizioni, tramite l'impugnazione di un provvedimento*

dell'AI generativa. La letteratura italiana si è anche interrogata sull'impatto dell'AI nella giustizia: cfr. G. CARLOTTI, *La giustizia predittiva e le fragole con la panna*, in Atti del convegno, cit. www.giustizia-amministrativa.it; N. DURANTE, *Brevi considerazioni in tema di amministrazione algoritmica e di giustizia predittiva*, in www.giustizia-amministrativa.it, M. BARBERIS, Giustizia predittiva: ausiliare e sostitutiva Un approccio evolutivo, in Milan Law Review, 2, 2022; Da ultimo, A. CIRIELLO, *Amministrazione della giustizia e intelligenza artificiale*, in E. BELISARIO, G. CASSANO (a cura di), *Intelligenza artificiale per la pubblica amministrazione*, *cit*., p. 427 e ss. Considerando la disposizione specifica dedicata al settore della giustizia nel disegno di legge governativo sull'IA approvato il 23 aprile 2024 dal Consiglio dei Ministri e ora in lettura al Senato (art. 14), che mira a imporre significative restrizioni sull'uso dell'intelligenza artificiale, è ragionevole prevedere che questo argomento scatenerà ancora di più intense discussioni

[22] Il testo, dopo una lunga gestazione, è stato approvato al Consiglio dei ministri del 3 luglio 2024, e di esso è stata data ampia risonanza mediatica; G. Trovati, *Imprese, stop per 10 mesi ai controlli su chi è in regola*, IlSole24ore, 3 luglio 2024.

[23] Ora, va precisato che la legalità algoritmica è stata studiata e poi fatta propria dalla giurisprudenza amministrativa in fattispecie che di "intelligenza artificiale" non avevano nulla (come nel noto caso della cd. "buona scuola"). Ad oggi, non pare che una decisione amministrativa, che nella fase istruttoria o decisionale si sia avvalsa di sistemi di IA, sia giunta alle porte dei T.A.R. e del Consiglio di Stato. Si trattava piuttosto di utilizzo, in procedimenti collettivi di mobilità, di utilizzo di "software" acquistati sulla base di gare pubbliche e fondati su algoritmi complessi (e come poi si è scoperto mal programmati e mal funzionanti), senza alcun livello di autonomia, (cd. "deterministici") , ma il legislatore nazionale, con la espressa menzione nelle disposizioni di rango primario sopra citate dell'IA ha già proiettato quei medesimi principi nel nuovo mondo dei sistemi "intelligenti", con una vis espansiva che va ben oltre gli specifici settori oggetto di disciplina.

interamente digitale, che va comunque esaminata, insieme a tutte le altre impugnazioni di provvedimenti umani. Se la vita si evolve, l'amministrazione si evolve con essa e la digitalizzazione pervade entrambe. Progressivamente, irreversibilmente. Tutto ciò avviene con una velocità tale che non si fa in tempo a legiferare ex ante. E il primo a doversene occupare è il giudice. In Italia, questo giudice è il giudice amministrativo"[24]. Per comprendere appieno quanto vi sia di distonia o meno tra alcuni principi affermati nell'AI Act e gli approdi della legalità algoritmica, appare però opportuno fare un passo indietro e chiedersi: "*perché l'AI Act*".

3. L'obiettivo di policy del Regolamento e l'auspicato *Brussel Effect*

Il Regolamento prende l'abbrivo dalla dichiarazione del suo scopo: «*migliorare il funzionamento del mercato interno e promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, contro gli effetti nocivi dei sistemi di intelligenza artificiale (sistemi di IA) nell'Unione nonché promuovere l'innovazione*» (articolo 1, paragrafo 1). Lo scopo è creare, attraverso le regole giuridiche, un punto di equilibrio tra sviluppo dell'innovazione tecnologia e tutela dei valori fondanti l'ordinamento giuridico europeo che i rischi associati ai sistemi e modelli di IA, specie di quelli più avanzati di *Deep Learning*, sollevano: autodeterminazione, privacy, salute, sicurezza, discriminazione e aumento delle disuguaglianze, partecipazione alla vita democratica, giusto processo, oscurità dei processi decisionali, disinformazione[25]. Al di là dello scopo dichiarato, ci sono però due obiettivi che le Istituzioni europee intendono perseguire con il Regolamento. Uno, più volte rappresentato nei diversi atti che hanno preceduto la proposta della Commissione dell'Unione europea del 2021, è ripreso nel *Recital* 3: evitare la frammentazione disciplinare tra i diversi Stati membri, alcuni dei quali, già anni fa, muovevano i primi passi per disciplinare questa tecnologia dirompente, in modo da

[24] L. CARBONE, *L'algoritmo e il suo giudice. Relazione al convegno "Digital administration – Daily efficiency and smart choices". Università degli studi Federico II. Napoli, 9-10 maggio 2022.,* in www.giustizia-amministrativa.it.

[25] Sulla assunzione di consapevolezza dei rischi associati all'IA per i valori della democrazia e dello Stato di diritto, come accennato nel paragrafo 1, è stata dirompente la svolta dell'IA Generativa, vista la sua enorme diffusione per effetto della facilità di uso di molte sue applicazioni (come ChatGPT, appunto) e il potenziale loro effetto manipolatorio sugli strumenti di comunicazione, mediante la generazione di contenuti fake o comunque mediante sofisticate tecniche di manipolazione. La questione ha solo ricevuto una più ampia risonanza per effetto della ubiquità di questi sistemi. Ma era già emersa, con riguardo alla profilazione degli utenti, nello scandalo *Facebook/Cambridge Analytica*, scoppiato nel 2018 ma riferito a fatti verificatisi prima, in cui 87 milioni di dati di utenti *Facebook* sono stati utilizzati, mediante l'app *Thisisyourdigitallife*, allo scopo di influenzare le manifestazioni di voto per l'elezione del Presidente degli Stati Uniti e il referendum sulla Brexit; cfr. SERBANESCU, *Why Does Artificial Intelligence Challenge Democracy? A Critical Analysis of the Nature of the Challenges Posed by AI-Enabled Manipulation*, in Retskraft – *Copenhagen Journal of Legal Studies*, 2021, 5, p. 105-128, consultabile sul sito ssrn. cfr. *Risoluzione del Parlamento europeo del 3 maggio 2022 sull'Intelligenza Artificiale nell'era digitale* (P9_TA(2022)0140, punto 97. F. PIZZETTI, *Intelligenza artificiale, quali rischi per la democrazia (e quali regole)*, 26 luglio 2018, in agenda digitale. C. SERBANESCU, *Why Does Artificial Intelligence Challenge Democracy? A Critical Analysis of the Nature of the Challenges Posed by AI-Enabled Manipulation*, in *Retskraft – Copenhagen Journal of Legal Studies*, 2021, 5, p. 105-128, in SSRN.

salvaguardare la "*certezza del diritto*" su cui dovrebbero fare affidamento sia il mercato che i cittadini[26]. L'altro è l'ambizione di ripetere il cd. *Brussel Effect*[27] ovvero affermare la leadership normativa "*by example*" dell'Unione Europea, sperimentata con il GDPR[28]. In realtà, entrambi gli obiettivi rischiano di non essere centrati. In primo luogo, i tempi lunghi di completa attuazione del Regolamento (da sei mesi per i divieti dei sistemi a rischio inaccettabile, a tre anni per i sistemi ad alto rischio, non *standalone,* ma inseriti come componenti in prodotti; cfr. articolo 113) appaiono incompatibili con la rapidità dell'evoluzione tecnologica, come la lezione dell'impatto della AI Generativa sullo stesso percorso normativo dell'AI dovrebbe aver insegnato. E poiché però, oltre all'aspettativa di tutela di fronte ai rischi dell'IA, un ruolo di primo piano assume anche l'esigenza di certezza delle regole, le pressioni sui governi nazionali delle organizzazioni sociali, degli operatori economici, degli opinion leader sono molto forti e gli Stati membri cominciano a giocare la carta parallela delle legislazioni nazionali, con conseguenti notevoli questioni interpretative di coordinamento e soluzione delle antinomie[29], demandate in ultima analisi agli operatori del diritto nella fase applicativa. In secondo luogo, il mito del *Brussels Effect* si confronta ora con un mondo radicalmente diverso da quello in cui sono nati e proliferati Internet e i social network. Nell'Era dell'Intelligenza Artificiale, competizione è la parola chiave. Non si tratta solo di una sfida tra i grandi *player*, ma anche tra i regolatori stessi (USA, Cina, UK, Canada, Giappone). In questo contesto, la

[26] In questa ottica, le regole armonizzate sull'IA servivano a supportare gli investimenti allo scopo di tentare di recuperare il tempo perduto rispetto a Cina e Stati Uniti, leader sempre più indiscussi nella geopolitica disegnata dall'incombere dell'AI. Nello stesso periodo della proposta dell'AI Act da parte della Commissione, si prevedeva infatti anche la revisione del Piano Coordinato sull'Intelligenza Artificiale, risalente al 2018 sottoscritto dagli Stati Membri e dell'Unione Europea per supportare il flusso di capitali, pubblici e privati, per rilanciare il Mercato Unico Digitale cavalcando l'onda (che è poi diventato uno tzunami) dell'Intelligenza Artificiale. Le iniziative delle Istituzioni europee in materia di IA cominciano a delinearsi nel 2017, con la convocazione del vertice europeo, sotto la presidenza estone; nel 2018, viene pubblicata la comunicazione della Commissione sulla strategia UE per una AI affidabile (L'intelligenza artificiale per l'Europa, COM(2018) 237 final).

[27] Mediante l'imposizione di standard regolatori assistiti da sanzioni anche molto severe che, formalmente, vigono solo nel territorio unionale, si vorrebbe ottenere l'effetto sostanziale di fornire a quegli standard una forza persuasiva extraterritoriale, facendo leva sul fatto che dovrebbe convenire alle imprese multinazionali adottare strategie di *compliance* uniformi, anche se i servizi e i prodotti sono offerti su mercati diversi. La locuzione entrata nella letteratura è stata coniata da A. BRADFORD, *The Brussels Effect: How the European Union Rules, World Oxford University Press*, 2020, p. 18: "*the EU has become the global regulatory hegemon unmatched by its geopolitical rivals. [This] challenges the critics' view that portrays the EU as a powerless global actor, and shows how such a criticism focuses on a narrow and outdated vision of what power means today (…) promulgates regulations that influence which products are built and how business is conducted, not just in Europe but everywhere in the world. In this way, the EU wields significant, unique, andhighly penetrating power to unilaterally transform global markets, be it through its ability to set the standards in competition policy, environmental protection, food safety, the protection of privacy, or the regulation of hate speech in social media*" cfr. brusselseffect. Una ricognizione accurata delle policy e delle specifiche legislazioni adottate nel 2023 nel mondo, e in particolare, negli Stati Uniti è riportata nell'*Artificial Intelligence Index di Stanford*, capitolo 7. Con specifico riferimento invece alle policy dedicate alla IA generativa, cfr. il report OECD *Initial policy considerations for generative artifiicial intelligence*, pubblicato a settembre 2023.

[28] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati).

[29] F. META, *Intelligenza artificiale, si delinea la strategia nazionale: un miliardo di investimenti*, 12 marzo 2024, in corrierecomunicazioni.; S.A. DI CAPRIGILIA, *Intelligenza artificiale: una sfida globale tra rischi, prospettive e responsabilità. Le soluzioni assunte dai governi unionali, statunitense e sinico. Uno studio comparato*, 17 aprile 2024, in www.Federalismi.it.

"sovranità digitale" è un campo di battaglia dove non solo si investono risorse pubbliche per accaparrarsi le risorse umane e materiali (basti pensare alla corsa al nuovo oro, le *Graphics Processing Unit* che sono necessarie per far "girare" le reti neurali artificiali, prodotte, insieme a pochi altri, da NVIDIA *Corporation,* che a giugno 2024 è diventata la seconda azienda al mondo per capitalizzazione), ma si delineano anche strategie normative, che ovviamente riflettono le diverse tradizioni giuridiche e culturali in cui vengono adottate[30].

4. Autonomia dell'IA e mito della sorveglianza umana

Sulle reali finalità dell'AI Act, si può anche non concordare. Invece è fuori discussione che non vi sia ad oggi una definizione unanime di *cosa* sia l'intelligenza artificiale, anche perché la locuzione, come si sa, è stata inventata a tavolino nel 1956 da John McCarthy, uno degli organizzatori della Conferenza di Dartmouth presso il Dartmouth College nel New Hampshire, il quale poi ha rivelato che ne aveva un'altra in mente, che molto probabilmente non avrebbe avuto lo stesso successo mediatico: "studio dell'automata"[31]. Fatto sta che, una volta imboccata la strada del Regolamento trasversale che ha ad oggetto un fenomeno (l'IA), è necessario innanzi tutto definirlo, cercando anche di trovare la quadra tra certezza e flessibilità, tenuto conto della rapidità con cui il fenomeno evolve. Sulla definizione di IA il percorso svolto dall'AI Act è stato davvero tortuoso. Dopo ampie discussioni, la definizione da ultimo accolta, molto diversa anche per impostazione da quella contenuta nella proposta originaria della Commissione del 21 aprile 2021, attesta che, non avendo l'AI confini geografici, anche il Regolamento dell'Unione Europea risente di interferenze globali. Le Istituzioni europee hanno abbracciato la definizione di AI che si rinveniva negli atti non vincolanti dell'OCSE fin dal 2019 (ovviamente anch'essa

[30] Come osservato, infatti, «*una condizione essenziale per il dispiegarsi del Brussel Effect (…) almeno nell'arena digitale delle ultime decadi, è stata l'inerzia regolamentare degli Stati Uniti*» (che non c'è per l'IA), S. DA EMPOLI, cit., pag. 127 Il riferimento è ovviamente ai due documenti adottati dalla Casa Bianca nel 2023: il *Blueprint for an AI Bill of Rights* e *l'AI Risk Managenement Framework*. Il fenomeno dell'oligopolio economico da parte delle grandi imprese che ora si contendono anche la fetta di mercato dell'AI generativa, si è consolidato proprio grazie alle maglie larghe della regolazione negli Stati Uniti, un «*regime regolatorio molto favorevole, caratterizzato da pochi vincoli e molti incentivi grazie alla scelta del governo americano, durante la presidenza Clinton, di non applicare ai "servizi di informazione" la disciplina già in essere per gli operatori telefonici. Con l'Administration's Telecommunication Act del 1996 si consentì alle start up di sviluppare la propria attività e i propri investimenti in assenza di vincoli sui prezzi, sull'accesso alla infrastruttura e ai servizi e sulla posizione e crescita nel mercato di riferimento. Quella scelta è stata una condizione decisiva per la creazione di servizi innovativi, la trasformazione di start up in piattaforme di enormi dimensioni globali e la crescita, ancora oggi in corso, di investimenti e del mercato*», L. Torchia, *Lo Stato digitale*, Il Mulino, 2023, p. 23. Per una ricognizione dello scenario geopolitico in cui si sono mosse le iniziative regolatorie in materia di IA, cfr. A. MALASCHINI, *Regolare l'intelligenza artificiale. Le risposte di Cina, Stati Uniti, Unione Europea, Regno unito, Russia e Italia*, in P. SEVERINO (a cura di), *Intelligenza Artificiale, Luiss University Press*, Roma, 2022, p. 104.

[31] "*Forse il più notevole, anche se trascurato, risultato della proposta di Darthmouth è l'improbabile e probabilmente involontaria capacità dell'espressione intelligenza artificiale di attrarre interesse e attenzione ben oltre la sua origine accademica. Non c'è nulla nella vita di John McCarthy a suggerire che covasse un segreto interesse o talento per coniare brillanti slogan pubblicitari, eppure la scelta di questo particolare moniker ha acceso un duraturo interesse da parte della stampa, del pubblico e dei media di intrattenimento – un risultato, questo, che spesso sfugge ai più navigati professionisti dell'advertising",* J. Kaplan, *Intelligenza Artificiale. Guida al prossimo futuro*, Luiss, 2016, p. 41.

soggetta a continue rivisitazioni[32]), dichiarando espressamente che non può prescindersi dalla "*convergenza internazionale*"[33]. Nella definizione ora contenuta nell'articolo 3 n. 1), l'accento è posto sul carattere più peculiare dei sistemi di IA, ovvero su ciò che rende dell'IA una tecnologia incomparabile con le altre tecnologie *general purpose* che hanno segnato la storia dell'Uomo: il loro variabile *"livello di autonomia*"[34]. Comprendere cosa si intenda per *autonomia,* e, anche, quale sia la sua differenza rispetto all'*automazione,* è il perno intorno a cui ruotano le questioni etiche e giuridiche che inondano il campo dell'IA. Un certo livello di "*autonomia*" significa, in sintesi, che i sistemi di IA riescono ad intraprendere azioni per raggiungere gli obiettivi assegnati, senza essere costantemente "guidati" dagli esseri umani, ivi compresi i programmatori[35]. Autonomia dai programmi iniziali; autonomia nella scelta della soluzione più efficiente

[32] L'ultima definizione concordata dai paesi OCSE è la seguente: «*An AI system is a machine-based system that explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical real or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment*». Sul sito istituzionale dell'OCSE sono rinvenibili le ultime modifiche e le spiegazioni del progressivo affinamento della definizione, cfr *OCSE*

[33] Articolo 3 paragrafo 1 n. 1) *"sistema di IA": un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali*; *Recital* 12: «*La nozione di "sistema di IA" di cui al presente regolamento dovrebbe essere definita in maniera chiara e dovrebbe essere strettamente allineata al lavoro delle organizzazioni internazionali che si occupano di IA al fine di garantire la certezza del diritto, agevolare la convergenza internazionale e un'ampia accettazione, prevedendo nel contempo la flessibilità necessaria per agevolare i rapidi sviluppi tecnologici in questo ambito*».

[34] «*Sistema di IA*": *un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali"* (art. 3, par. 1 lett.a). Già nel riferimento al "Sistema" si evoca la complessità strutturale di una applicazione di IA che è composta da almeno tre component: dati, potenza computazionale, algoritmi.

[35] R. CUCCHIARA, *L'intelligenza non è artificiale. La rivoluzione tecnologia che sta già cambiando il mondo*, Mondadori, 2021, p. 122, *«I sistemi che progettiamo stanno piano piano imparando ad apprendere dal mondo esterno, sono in grado di farsi proprie rappresentazioni interne di capire di ragionare e quindi di generare azioni. Sbagliano. E sono capaci di ricominciare sapendo come migliorare dai propri errori. Come facciamo noi dalla mattina alla sera. Come dovremmo fare se fossimo intelligenti»*. Questa "autonomia" è in effetti il salto di qualità dei sistemi di IA contemporanei ed ha sorpreso la stessa comunità scientifica, attraverso veri e propri *turning point* che hanno segnato la storia dell'IA, alcuni dei quali anche con forti risonanze mediatiche. Sintomatica è stata ad esempio la vittoria nel 2016 di AlphaGo di DeepMind alla gara internazionale del gioco del GO, soprattutto perché all'inizio la mossa vincente della macchina, inventata dal sistema che aveva imparato a giocare sulla base di milioni di partite e utilizzando le tecniche d Deep Learning, fu addirittura letta dagli esperti come un errore, lasciando di stucco il campione mondiale in carico Lee Sedol. E invece dimostrava che le capacità computazionali di individuare autonomamente la propria strategia di gioco avevano già raggiunto, oramai circa otto anni fa, livelli tali da confondersi con l'intuizione e la creatività La vittoria di AlphaGo ebbe un notevole impatto mediatico, poiché il gioco del GO è considerato nei paesi asiatici più una forma d'arte che un vero e proprio gioco da tavolo, con pedine e scacchiera e la partita fu seguita in diretta da migliaia di spettatori. La sconfitta del campione Lee Sedol sorprese anche il mondo scientifico perché fece comprendere che da allora in poi che lo sviluppo dell'IA avrebbe chiaramente preso la strada del Deep Learning. Il racconto dell'evento e della sua portata storica per il mondo scientifico si ritrovano in M. TEGMARK, *Vita 3.0. Esseri umani nell'era dell'intelligenza artificiale,* 2017, Raffaello Cortina, p. 122; in C. ACCOTO, *Il mondo ex machina*, 2019, p. 33, «*Su cinque gare consecutive, AlphaGo ne vincerà ben quattro e lo scoramento di Lee sarà più visibile sul suo volto. Ma la cosa maggiormente sorprendente non è questa. È piuttosto il modo con cui la macchina ha vinto: con creatività di gioco. Non solo con potenza di calcolo (…), ma anche e soprattutto con potenza di immaginazione. E di conseguenza anche di innovazione. La Macchina AlphaGo ha fatto una mossa strategica che nessuno dei milioni di giocatori umani di Go, in più di due millenni di partite, aveva immaginato: l'oramai celebre "mossa 37" derisa dai commentatori umani esperti durante la gara come un errore madornale della macchina, poi dimostratasi na straordinaria e, fino ad allora, inimmaginata mossa vincente*". Ancora più efficace il racconto riportato in C. METZ, *Costruire l'intelligenza. Google, Facebook, Musk e la sfida del futuro*, 2022, Mondadori, p. 177, «*la probabilità della mossa 37 era una contro diecimila. AlphaGo sapeva che questa non era una mossa che un giocatore professionista di Go avrebbe mai scelto, ma la fece comunque sulla*

rispetto all'obiettivo; autonomia anche nelle strategie vincenti per raggiungere il risultato prefissato dall'uomo, come quelle individuata dal sistema di IA di AlphaGo nella storica partita al gioco del GO del 2016, che consentì alla macchina di battere il campione mondiale in carica Lee Sedol; una strategia in quella partita talmente "autonoma", rispetto a tutte le partite giocate da umani esperti in questo gioco molto diffuso nei paesi asiatici, e sulle quali la macchina era stata addestrata, che gli osservatori avevano inizialmente ritenuto la mossa vincente (la mossa 37) come un errore. È l'autonomia, come emerge dalla definizione dell'articolo 3, la vera novità delle tecnologie ascrivibili all'area dell'IA; ciò che, unita alla potenza computazionale e alla capacità di riprodurre le informazioni su scale incomprensibili per la mente umana, fa temere che si possa giungere nel giro di pochi anni ad un *punto di non ritorno* nell'evoluzione dell'umanità. Il fatto che un sistema di IA sia dotato di un *certo livello di autonomia* non vuol dire, se si guarda alla decisione finale destinata ad impattare sui cittadini, che esso sia anche *automatico*. Autonomia e automaticità sono caratteri separati, che possono essere indipendenti l'uno dall'altro. È, ad esempio, "*automatico*" un sistema di distribuzione di bevande, poiché con l'inserimento della monetina, scatta il sensore e la bibita viene erogata, ma nessuno immaginerebbe che esso possa essere definito *intelligente* o dotato di *un certo livello di autonomia* (e sotto il profilo giuridico, già i sistemi automatici di forniture di cose hanno sollevato enormi questioni nell'ambito delle teorie civilistiche del negozio giuridico). E' invece *autonomo,* oltre che *automatico,* un sistema di IA che, ad esempio in campo finanziario sulla base di algoritmi di apprendimento, discrimini la classe dei debitori in base alla probabilità stimata della loro capacità di ripagare il debito, elaborando i dati in suo possesso riferiti ovviamente al passato e poi stabilisca, senza il filtro del funzionario preposto, se ammettere il richiedente al finanziamento, nel caso, mediante una applicazione web[36]. L'autonomia dei sistemi di IA, perno della sua definizione giuridica, si riflette però, come in un gioco di specchi, su uno dei principi chiave della cd. legalità algoritmica, quello della "non

base dei milioni di partite che aveva giocato contro se stesso, a cui non aveva partecipato alcun essere umano. Aveva compreso che, nonostante nessun essere umano avrebbe pensato a quella mossa, era quella giusta (…) Lo aveva scoperto da solo, attraverso un processo di introspezione».

[36] È, pertanto, *«opportuno ribadire che quando parliamo di intelligenza artificiale non si implica anche il suo impiego in un processo automatico che la usa, salvo i casi in cui al termine dell'esecuzione dell'algoritmo esso esegue anche delle funzioni programmate da umani che non hanno conoscenza o della ragione finale delle sue determinazioni, in quanto sono il risultato di un numero di variabili non trattabile da parte del nostro cervello»,* G.F. ITALIANO, E. PRATI, *Storia, tassonomia e sfide future dell'intelligenza artificiale*, , in (a cura di) P. SEVERINO, *Intelligenza artificiale* cit., p. 77. Pare opportuno evidenziare che la giurisprudenza amministrativa è già giunta a conclusioni analoghe. Con la sentenza della Terza Sezione del Consiglio di Stato del 25 novembre 2021, n. 7891, nell'ambito di una procedura di gara di fornitura di dispositivi medicali (pacemaker) con elevato grado di automazione, è stata infatti delineata la differenza – nel caso specifico rilevante ai fini della valutazione dell'offerta – tra i sistemi *intelligenti* e quelli solo "automatici", precisando che l'automazione "*tradizionale*" segue la logica *if-then*, mentre i primi prevedono algoritmi di machine learning e, invece di applicare regole predeterminate, rinvengono essi stessi le regole nella analisi dei dati, elaborando costantemente nuovi criteri di inferenza tra i dati. Sulla sentenza, cfr. N. CAPPELLAZZO, *Algoritmi, automazione e macchinismi di intelligenza artificiale: la classificazione proposta dal Consiglio di Stato*, in www.federalismi.it, 23 marzo 2022.; G. NATALE, *Quadro normativo vigente e questioni insolute in materia di dispositivi medici intelligenti*, in (a cura di) U. RUFFOLO, M. GABRIELLI, *Intelligenza artificiale, dispositivi medici e diritto*, Giappichelli, 2023, p. 179.

esclusività algoritmica"[37] o della sorveglianza umana[38], in tutte le sue plurime denominazioni, compresa quella presa a prestito dalle scienze informatiche del *Human-in-the-loop*. In effetti, negli orientamenti etici

[37] In termini giuridici, il tema è declinato come principio di non esclusività algoritmica o, con una locuzione che coglie il cuore del nuovo volto uomo-macchina, "riserva di umanità". Da ultimo, l'art. 30 del Codice dei contratti pubblici positivizza il principio della "non esclusività della decisione algoritmica" che era già stato affermato dalla giurisprudenza amministrativa con riferimento alle cd. decisioni robotiche (comma 3, lett. b). La disposizione costituisce la prima base normativa di rango primario della cd. riserva di umanità e rivela potenzialità espansive che vanno ben oltre il settore dei contratti pubblici. G. GALLONE, *Riserva di umanità e funzioni amministrative*, 2023, CEDAM, che con una ricostruzione scientifica pregevole delinea anche la fonte costituzionale, e le conseguenze in caso di violazione, del principio del necessario "*contributo umano*" preso sul serio nell'azione amministrativa algoritmica: artt. 28, 97, terzo comma, 52 secondo comma, e 98 Cost. Secondo l'Autore, l'appiglio su cui si è fondato il principio di conio giurisprudenziale, l'art. 22 del G.D.P.R. non è sufficientemente saldo, sia perché l'ottica del regolamento sulla protezione dei dati personali è individuale, «*sia perché esso intercetta il tema della gestione automatizzata dei dati unicamente in tale prospettiva e non in quella, assai differente, delle garanzie di una buona amministrazione*» e, comunque, delimita il principio con una articolata serie di deroghe. Dello stesso a., *Digitalizzazione, amministrazione e persona: per una "riserva di umanità" tra spunti codicistici di teoria giuridica dell'automazione*, in *Pubblica Amministrazione, Persona e Amministrazione*, 1, 2023, 329 e ss. P. BENANTI, *Human in the loop. Decisioni umane e intelligenze artificiali*, Mondadori, 2023; D.U. GALETTA *Human-stupidity-in-the-loop? Riflessioni (di un giurista) sulle potenzialità e i rischi dell'Intelligenza Artificiale*, in www.federalismi.it, 22 febbraio 2023.

[38] Ma tale carattere si infrange anche sul principio, declinato in senso rafforzato per i sistemi algoritmici utilizzati nell'azione amministrativa, della trasparenza, poiché l'autonomia elaborativa e propositiva dei sistemi di AI è ottenuta grazie ad algoritmi avanzati soprattutto di *Deep Learning*, che sono però caratterizzati dalla impossibilità di spiegare il processo di elaborazione seguito per giungere ai risultati (impossibilità anch'essa descritta con una metafora entrata nel linguaggio comune: Black Box). La metafora è stata resa celebre da F. PASQUALE, *The black box Society: the secret algorithms that control money and information*, Harvard University Press, 2016. Sulla dimensione articolata della opacità algoritmica, che nella tecnologia di intelligenza artificiale fondata sulle tecniche di *Machine Learning* e *Deep Learning* porta il nome metaforico della black box, cfr. V. PAPADOULI, *Transparency in Artificial Intelligence: A Legal Perspective* 30 maggio 2022, *Journal of Ethics and Legal Technologies*, 4, 25-40; A MASCOLO, *Gli algoritmi amministrativi: la sfida della comprensibilità*, in Giorn. Dir. amm., 2020, pag. 366 e ss.; G. LO SAPIO, *La Black box: l'esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, in www.federalismi.it, 2021, pag. 114 e ss.; G. LO SAPIO, *La trasparenza sul banco di prova dei modelli algoritmici*, in www.federalismi.it, 2021, pag. 239 e ss. Occorre considerare che i sistemi di IA di *Deep Learning*, non solo imparano costantemente dai dati, ma lo fanno in modo non comprensibile. La causa della impossibilità di guardare "dentro al cofano" della macchina computazionale è la medesima ragione della sua super-efficienza nel raggiungere gli obiettivi assegnati: la estrema complessità matematica, descrivibile in cifre, quanto a parametri utilizzati, che sfuggono anch'essi alla reale percezione umana della scala, che è in esponenziale aumento lungo i salti evolutivi dei sistemi. Per fornire in pochi secondi una risposta ad un "prompt" in linguaggio naturale, e in qualunque lingua, ad esempio GPT 3.5., il modello di base di ChatGPT al momento del suo lancio sul mercato a novembre 2022, era basato su 175 miliardi di parametri; la precedente versione aveva 1,5 miliardi; il numero di parametri utilizzati dalla attuale versione di GPT 4.0 è stata stimata in 100 trilioni; quest'ultimo dato però, a proposito di trasparenza, non è stato reso pubblico da OpenAI (I Large Language Model sono definiti per questo anche iperparametrici e da questo deriva il termine "large"). Tanto è nera la *Black Box* che anche sull'area di ricerca, ancora tutta da esplorare, della cd. Explainable AI (XAI) che si pone l'obiettivo di disvelare il meccanismo di funzionamento di un certo sistema di IA, vi sono opinioni discordanti, poiché secondo una parte della scienza si tratterebbe di costruire modelli paralleli che "approssimano" il comportamento del sistema di IA che si vorrebbe disvelare, senza alcuna garanzia che la spiegazione fornita per il procedimento seguito corrisponda effettivamente a ciò che è accaduto nei calcoli svolti in nanosecondi: cfr. C. RUDIN, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead,* in *Nature Machine Intelligence*, 1, 2019, p. 206-2015, disponibile nella rivista Nature «*Rather than trying to create models that are inherently interpretable, there has been a recent explosion of work on "Explainable ML," where a second (posthoc) model is created to explain the first black box model. This is problematic. Explanations are often not reliable, and can be misleading, as we discuss below. If we instead use models that are inherently interpretable, they provide their own explanations, which are faithful to what the model actually computes*». Di fronte a tale potenza tecnologica, disvelare il meccanismo di funzionamento che conduce ad elaborare gli input e a fornire una risposta per perseguire l'obiettivo assegnato è di fatto impossibile; "*benché sarebbe facile programmare un computer affinché stampi una lista delle addizioni e moltiplicazioni eseguite da una rete per un dato ingresso, la lista darebbe a noi umani un'idea pari a zero della strada seguita dalla rete per ricavare la risposta. Una lista di un miliardo di operazioni non è una spiegazione che un essere umano può capire*" M. MITCHELL, *L'intelligenza artificiale. Una guida per essere umani pensanti*, cit., p. 105.

per una AI affidabile dell'Unione Europea pubblicati l'8 aprile 2019, elaborati dal gruppo di esperti ad alto livello sull'intelligenza artificiale istituito nel 2018, il principio della sorveglianza umana viene declinato come: a) *Human-in-the-loop* ovvero la capacità di intervento dell'uomo in ogni ciclo decisionale del sistema; b) *Human-on-the-loop,* inteso come capacità di intervento umano durante la fase della progettazione e del monitoraggio; c) *Human-In-Command,* relativo alla capacità di supervisione dell'impatto complessivo, anche economico e sociale, dei sistemi di IA. Se però si passa dal livello dell'affermazione del principio a quello della sua effettiva applicazione in concreto nell'esercizio dell'azione amministrativa, la domanda che sorge è: come assicurare che i funzionari pubblici assumano e mantengano effettivamente il "comando" della decisione amministrativa che si avvalga dell'ausilio di algoritmi di IA i quali a loro volta hanno un "certo livello di autonomia"? Come fare in modo che la decisione sia assunta in ogni caso dall'essere umano "al comando" e non completamente delegata all'algoritmo di IA? Non c'è il rischio, vista l'autonomia dei sistemi di IA che il controllo si riveli, a ben guardare, una "illusione di controllo"? E, in ogni caso, come fare in modo che la sorveglianza sia efficace, ma anche svolta in maniera efficiente, evitando di perdere, con misure di verifiche lente e complesse, i risultati di efficienza che invece i sistemi di IA promettono? I risultati delle sparute ricerche condotte sul campo non sono incoraggianti[39], poiché, nelle sperimentazioni oggetto di analisi, la fase della validazione sostanziale delle proposte algoritmiche è proprio quella più trascurata. La difficoltà di realizzare, mediante regole efficaci, il principio della cd. non esclusività algoritmica dipende da diversi fattori. Ma qui preme evidenziarne uno che deriva da un'altra caratteristica dei sistemi di IA, non presa in considerazione nella definizione giuridica contenuta ora nell'articolo 3 dell'AI Act: la loro concreta persuasività, supportata dalla apparente oggettività meccanicistica del digitale e agevolata dalla facilità d'uso di molte applicazioni, già entrate nella vita quotidiana. L'efficacia dei sistemi di IA nel convincere, influenzare, modellare comportamenti umani, in alcuni contesti privati, è alla base di strategie innovative, ad esempio, di *marketing* digitale personalizzato, o degli algoritmi di raccomandazione sui *social media* che, mediante la profilazione degli utenti, selezionano i contenuti che hanno la maggior probabilità di "ingaggio", condivisione, commento e propagazione. Ma la potenza persuasiva è sperimentabile da chiunque si avvicini alle applicazioni dei *Large Language Models*: uno dei motivi del successo mondiale di ChatGPT sta infatti nella sua estrema facilità di utilizzo e nella capacità dei testi linguistici generati di sembrare attendibili ed accurati, vista la correttezza stilistica, la logicità delle argomentazioni, la precisione grammaticale; mentre in una elevata percentuale di casi non

[39] E. CHITI, B. MARCHETTI, N. RANGONE, *L'impiego di sistemi di intelligenza artificiale nelle pubbliche amministrazioni italiane: prove generali,* in *BioLaw Journal, Rivista di BioDiritto*, 2, 2022, p. 503 *«l'esigenza che sia assicurato un controllo umano sui vari passaggi del funzionamento della macchina è talore riconosciuta, ma si pone l'accento sulle difficoltà che tale controllo incontrerebbe (…). In ogni caso, l'esigenza di un controllo umano sul funzionamento della macchina non si traduce mai nella definizione di specifici requisiti, tanto meno di requisiti definiti da norme giuridiche».*

vi è alcuna accuratezza[40], anche quando i modelli sono stati "affinati" in domini specifici, come quello legale. Il principio etico-giuridico di *Human-In-The-Loop* è ora sancito nell'articolo 14 dell'IA Act, rubricato *Human Oversight*. La stessa collocazione della disposizione delimita però il suo campo di applicazione, poiché il principio non riguarda tutti i sistemi di IA, ma solo quelli ad alto rischio, così classificati nell'Allegato III (e sempreché essi restino classificati come tali anche dopo la valutazione del rischio in concreto che l'amministrazione deve svolgere ai sensi dell'articolo 27 citato). L'articolo 14 attua il principio con regole dettagliate, imponendo ai *provider* l'adozione di misure[41] già nella fase del *design* dei sistemi. Tra queste, due, in particolare, suggeriscono alcune considerazioni. In primo luogo, solo per una particolare categoria di sistemi ad alto rischio, ovvero per quelli di riconoscimento biometrico in tempo reale e in spazi aperti al pubblico dove eccezionalmente consentiti, si prevede che la sorveglianza debba essere affidata a *due persone fisiche* separatamente. In sostanza, il legislatore sospetta che, quando il sistema rischia di incidere sui diritti fondamentali dei cittadini come quelli del riconoscimento biometrico, il controllo affidato ad un solo essere umano non dia sufficiente garanzia; un po' come quando per l'aviazione, sui voli civili di una certa soglia, è stato stabilito che nella cabina di pilotaggio ci fossero almeno due piloti, il *pilot flying* e il *pilot monitoring*. Il significato di questa dettagliata disposizione va peraltro ben oltre la sua *ratio* ispirata ad esigenze di massima prudenza: conferma che quando si fanno stime sugli impatti organizzativi del fenomeno IA, anche rispetto ai livelli occupazionali, il dato normativo non è

[40] È l'effetto delle cd. "allucinazioni" o "confabulazioni", che è il rovescio della medaglia delle elevate performance dei sistemi generativi; ovvero la facilità con cui possono produrre errori macroscopici e fatti falsi, poiché il risultato è basato su stime della somiglianza semantica tra le parole senza alcuna verificabilità delle fonti (perché in realtà non ci sono fonti, ma frammenti di parole tokenizzate) *https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-queries*

[41] L'art. 14 prevede in particolare che la sorveglianza umana si articola in misure che devono essere integrate nel sistema fin dalla sua progettazione e quindi prima della immissione sul mercato, se però questo è "tecnicamente possibile"; o almeno individuate dal fornitore ma in modo da poter essere adottate dagli sviluppatori. In ogni caso, le persone fisiche cui è attribuita la sorveglianza devono poter comprendere "correttamente le capacità e i limiti pertinenti del sistema di IA ad alto rischio ed essere in grado di monitorarne debitamente il funzionamento, anche al fine di individuare e affrontare anomalie, disfunzioni e prestazioni inattese" (paragrafo 4 lett. a); "interpretare correttamente l'output del sistema di IA ad alto rischio, tenendo conto ad esempio degli strumenti e dei metodi di interpretazione disponibili; (paragrafo 4 lett. c); "decidere, in qualsiasi situazione particolare, di non usare il sistema di IA ad alto rischio o altrimenti di ignorare, annullare o ribaltare l'output del sistema di IA ad alto rischio"; (paragrafo 4 lettera d,); "intervenire sul funzionamento del sistema di IA ad alto rischio o interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga che consenta al sistema di arrestarsi in condizioni di sicurezza" (paragrafo 4 lett. e) Il potere di intervenire in concreto fino all'arresto del funzionamento della macchina è la traduzione in termini giuridici del principio eticamente riconosciuto della cd. meta-autonomia, ovvero la necessità che la delega da parte dell'uomo alla decisione (o alla proposta di decisione) del sistema computazionale non sia irrevocabile; che ci sia sempre la possibilità di rientrare nel pieno dominio della sfera decisionale. *«Quando adottiamo l'IA e il suo agire smart, cediamo volontariamente parte del nostro potere decisionale ad artefatti tecnologici (…) il rischio è che la crescita dell'autonomia artificiale possa minare il fiorire dell'autonomia umana (…) È chiaro dunque sia che l'autonomia umana debba essere promossa, sia che l'autonomia delle macchine debba essere limitata e resa intrinsecamente reversibile, qualora l'autonomia umana debba essere protetta o ristabilita (si pensi al caso di un pilota in grado di disattivare il pilota automatico e riprendere il pieno controllo dell'aereo). Ciò introduce una nozione che può essere definita come meta-autonomia, o modello della decisione di delega. Gli esseri umani dovrebbero mantenere il potere di decidere quali decisioni prendere, esercitando la libertà di scelta dove necessario e cedendola nei casi in cui ragioni di primaria importanza, come l'efficacia, possano prevalere sulla perdita di controllo. Ma qualsiasi delega dovrebbe rimanere in linea di principio reversibile, adottando come ultima garanzia il potere di decidere di decidere di nuovo»*, L. FLORIDI, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide, Raffaello Cortina editore, 2022,* pag. 99.

neutro potendo direttamente incidere sull'organizzazione, oltre che sul fabbisogno di competenze. L'altra osservazione deriva dall'espresso richiamo – unica disposizione in tal senso in tutto il testo - contenuto nel medesimo articolo 14 all'efficacia persuasiva dei sistemi di IA[42]. Il paragrafo 4, lettera b) stabilisce in particolare che le persone, preposte alla sorveglianza, debbano restare consapevoli «*della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio ("automation bias"), in particolare per i sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche*»[43]. Per *automation bias* si intende la tendenza da parte dell'essere umano, coinvolto nell'interazione con la macchina, ad affidarsi ai suoi *output*, fino a trascurare o ignorare altre informazioni che derivano da fonti diverse: una pigrizia cognitiva che ognuno, in esperienze diverse, ha avuto modo di sperimentare, ancorandosi ad esempio alle prime informazioni e/o intuizioni in un processo decisionale. Tener conto di questa tendenza e quindi del fatto che le norme si rivolgono, per la loro attuazione, non a robot, ma a persone in carne ed ossa con i loro *bias* cognitivi e comportamentali è già un punto di svolta, peraltro in linea con la spinta dell'Unione Europea a valutare, per la elaborazione di "*regole a prova di futuro*", anche le scienze comportamentali[44]. Senza l'adozione di un approccio globale che consideri tutti gli aspetti, tecnologici, organizzativi, culturali, che riguardano la complessa relazione tra uomo-macchina, *l'automation bias* rischia infatti di far restare sulla carta

[42] Nella Comunicazione della Commissione UE "Legiferare meglio: unire le forze per produrre leggi migliori" del 29 aprile 2021 si prende atto dell'effetto dirompente che i cambiamenti dell'era digitale hanno anche sugli approcci e sugli strumenti a disposizione di regolatori, e nel Better Regulation Toolbox del 25 novembre 2021 indica l'uso delle scienze comportamentali e la regulatory sandbox come metodi emergenti utilizzabili per elaborare regole future-proof.
Peraltro la necessità di evitare fallacie normative, elaborando regole che tengano conto di come esse siano poi effettivamente applicate alla luce dei bias cognitivi e comportamentali degli uomini era già stata affermata anche nel parere Consiglio di Stato, Sezione consultiva per gli Atti Normativi, 7 giugno 2017, sullo schema di decreto del Presidente del Consiglio dei Ministri recante «*Disciplina sull'analisi dell'impatto della regolamentazione, la verifica dell'impatto della regolamentazione e la consultazione*». Di qui il rilievo dei principi etici che poi dovrebbero tradursi in regole giuridiche; «*se infatti l'etica è quel rampo della filosofia che si occupa in generale del comportamento umano, appare evidente che essa non può non essere attenta alle manifestazioni del diritto, sia quando vede in esso una rappresentazione, in un determinato momento storico, dei valori ritenuti consoni allo sviluppo della personalità dell'uomo, sia quando il diritto si manifesta in una sorta di traduzione in regole giuridiche del comportamento etico dell'uomo o delle istituzioni; si pensi al principio di affidamento, a quello di buona fede, a quello di leale collaborazione»,* A. PAJNO, *La costruzione dell'infosfera e le conseguenze sul diritto*, in A. PAJNO, F. DONATI, A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, Il Mulino, 2022, p. 19. Vi è pertanto il rischio che se i principi etici riconosciuti in astratto sono tradotti in regole giuridiche ineffettive, l'impegno sbandierato di una IA antropocentrica diventi solo una manifestazione di cd. bluewhashing etico; cfr. L. FLORIDI, cit., 111.

[43] Si tratta di un *bias* cognitivo studiato in settori come l'aviazione civile, dove l'affidamento ai sistemi di pilotaggio automatico da parte dei piloti è stato indicato come una concausa di disastri aerei, come quello del AIr France 447 del 2009; B. HOFFMAN, *Automation Bias: What It Is And How To Overcome It, 10 marzo 2024, disponibile on line nella rivista Forbes*. J. P. BROWN, *The Effect of Automation on Human Factors in Aviation, The Journal of Instrumentation, Automation and System*, 2016. Ci sono molteplici fattori che contribuiscono all'*Automation Bias*: l'avarizia cognitiva che è la tendenza umana a compiere il minor sforzo possibile per prendere una decisione, cercare scorciatoie e risparmiare energie; ma anche la dispersione della responsabilità, maniera simile a quanto avviene quando gli umani collaborano con altri umani e nel lavoro di squadra si perdono di vista le responsabilità decisorie individuali (il cd. social loafing).

[44] "*Behavioural insights (BI) show that human beings are often not rational. They do not always base their decision on an analysis if all possible courses of actions. Policy initiative may fall if they expect rational behaviour by the public. By understanding how people really behave, we can make policies more effective", Commissione UE, Better Regulation Toolbox* del 25 novembre 2021, p. 598.

l'affermazione del *Human-In-The-Loop* e di far emergere, sotto l'egida della "supervisione umana" una nuova forma di burocrazia tecnologica, in cui la proposta algoritmica viene validata solo formalmente dal funzionario che assume la decisione, "catturato" dalla macchina, ma privo della concreta abilità di metterne in discussione il risultato. Occorrerà nella fase di attuazione ragionare su quali misure e come realizzarle in concreto, potendo esse riferirsi riguardare ad esempio l'interfaccia utente, sistemi di allarme specifici o piani formativi innovativi, fondati non solo sulla formazione teorica (ivi compresa l'AI Literacy), ma anche su simulazioni pratiche che riproducano il momento critico della decisione finale. Tutto però senza dimenticare che l'intensità dell'*automation bias* è direttamente proporzionale alla diffusione dell'innovazione tecnologica. Se la prudenza, la circospezione e l'attenzione della persona fisica che interagisce con la macchina possono essere elevate in una fase iniziale e sperimentale, come quella attuale almeno nell'ambito della pubblica amministrazione; è verosimile immaginare che la tendenza ad affidarsi alla macchina potrebbe assumere un peso ben più rilevante se, in un arco temporale che oggi non è ancora prevedibile, l'IA divenisse ausilio ordinario nella pratica quotidiana.

5. L'approccio risk-based di fronte all'irruzione della IA Generativa

Mentre lo scopo dichiarato dall'AI Act è quello di proteggere dai rischi associati all'IA la salute, la sicurezza e i diritti fondamentali, l'ambiente, la democrazia e lo Stato di diritto riconosciuti dall'ordinamento eurounitario, la base legale per l'adozione del Regolamento è individuata nell'art. 114 del TFUE che legittima l'Unione Europea ad adottare regole di armonizzazione per il funzionamento del mercato interno. Tale base legale è illuminante anche con riguardo all'impostazione generale: se il focus è il *mercato* dell'era digitale, si spiega perché il Regolamento abbia preso in considerazione i sistemi di IA secondo un approccio *risk-based,* analogo a quello seguito per la "*sicurezza dei prodotti pericolosi*" la cui libertà di movimento deve essere bilanciata con le esigenze di protezione dei valori dell'ordinamento eurounitario[45]. Sui livelli di rischio previsti nell'AI Act la dottrina si è a lungo soffermata[46]. In sintesi,

[45] Nella prospettiva del legislatore eurounitario, i rischi rispetto ai quali cautelarsi sono peraltro di natura valoriale, poiché «*le entità esposte a queste fonti di pericolo sono qualificate in maniera assiologica: esse corrispondono a valori e diritti fondamentali dell'EU come diritti umani fondamentali, la sicurezza e le procedure democratiche*».

[46] Per tutti, da ultimo, C. NOVELLI, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, 24 gennaio 2024, in www. federalismi.it. In realtà, a prescindere dalla classificazione dei diversi livelli di rischio, neanche vi è unanimità tra gli scienziati su quali siano effettivamente i rischi derivanti dalla diffusione dell'IA. Mentre su alcuni, come (a) il rischio dei bias discriminatori che si annidano nell'enorme mole di dati elaborati dai sistemi; b) il rischio di perdita del controllo umano e della responsabilità per gli eventuali danni, in considerazione del loro carattere di "autonomia; c) il rischio di non comprensibilità delle decisioni che hanno un impatto diretto sulla sfera giuridica dei destinatari, vista la "opacità" (*Black Box*) dei meccanismi di operatività di alcuni sistemi come quelli di *Deep Learning*"), vi è una certa condivisione; su altri invece si brancola nel buio, né vi sono oggi stime sui rischi futuri e prevedibili, fondate sugli accadimenti passati), perché in quanto tecnologia *general purpose* l'IA ha la abilità di combinarsi con altre diverse tecnologie o dar luogo ad applicazioni trasversali in un'onda di innovazione di cui non sono ancora percepibili gli effetti, ed è in ogni caso, in quanto fondata su dati e gestita attraverso una complessa catena di valore che vede molteplici agenti (fornitori, sviluppatori, utilizzatori per scopi professionali o meno) anche suscettibile di essere utilizzata per scopi malevoli.

invocando il principio di proporzionalità, nell'impianto originario del Regolamento sono individuati quattro diversi livelli di rischio[47], che, al gradino più alto di pericolosità, vede le "*pratiche di IA vietate*" nel mercato europeo (articolo 5); a quello immediatamente sotto, i sistemi ad alto rischio (capo 3, artt. 6- 49, ed Allegato III) oggetto del *corpus* più cospicuo di norme; infine, nella base della piramide, i sistemi a basso rischio o che richiedono solo obblighi minimi di informazione a favore degli utilizzatori finali, cittadini compresi. Dal punto di vista della classificazione, e guardando in primo luogo ai *provider* che intendono affacciarsi sul mercato europeo, rileva poco che tali sistemi siano utilizzati da soggetti privati o dalle autorità pubbliche. Eppure, se si scorre l'elenco dei sistemi ad alto rischio contenuti nell'allegato III, ai quali si rivolge il corpus più cospicuo di requisiti e obblighi e relative sanzioni, emerge che la maggior parte di essi riguardano proprio l'attività amministrativa[48].

Ma la prospettiva originaria risk-based del Regolamento è stata messo a dura prova per effetto della recente evoluzione dell'IA. Tutta la classificazione di rischiosità per livelli è operata infatti *ex ante*, guardando alla finalità "*prevista*" fin dalla fase della progettazione ("*intended purpose*") quella «*prevista dal fornitore, compresi il contesto e le condizioni d'uso specifici, come dettagliati nelle informazioni comunicate dal fornitore nelle istruzioni per l'uso, nel materiale promozionale o di vendita e nelle dichiarazioni, nonché nella documentazione tecnica*»

[47] Ad ogni livello di "pericolosità" corrispondono misure di salvaguardia di quei valori che vanno dal divieto assoluto di immissione nel mercato europeo di specifiche applicazioni oggetto di disciplina nel Titolo II (divieti, ma anche alcune eccezioni, come nelle ipotesi consentite di riconoscimento biometrico in tempo reale in luoghi aperti al pubblico) all'incentivo ad adottare codici di condotta volontari (come nel caso dei sistemi di IA per i videogiochi, di cui al Titolo IX). Sull'approccio risk-based, cfr. G. FINOCCHIARO, *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in *Diritto dell'Informazione e dell'Informatica*, 2, 2022, pag. 303; D. MESSINA, *La proposta di Regolamento europeo in materia di intelligenza artificiale: verso una "discutibile" tutela individuale di tipo consumer-centric nella società dominata dal pensiero artificiale*, in *Medialaws*, 2022, 2; C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale,* in *Biolaw Journal*, 2022, 415; G. MAZZINI, S. SCALZO, *The proposal for Artificial Intelligence Act: considerations around some key concepts*, 2022, disponibile nel sito SSRN, G. LO SAPIO, *Intelligenza artificiale: rischi, modelli regolatori, metafore,* in *www.federalismi.it*, 19 ottobre 2022.

[48] A prescindere dal riconoscimento biometrico, quando ammesso, sono considerati ambìti ad altro rischio quello delle infrastrutture critiche (traffico stradale, fornitura di acqua, gas, riscaldamento ed elettricità); istruzione e formazione professionale (accesso, l'ammissione o l'assegnazione a istituti educativi e di formazione professionale a tutti i livelli; valutazione dei risultati di apprendimento; valutazione del livello di istruzione appropriato per un individuo); occupazione, gestione dei lavoratori (dalla fase della selezione, a quella della risoluzione dei contratti, anche in questo caso, senza alcuna distinzione tra dipendenti di datori di lavoro privati e dipendenti di pubbliche amministrazioni); accesso ai servizi pubblici e privati essenziali (sanità, sicurezza, assegnazione di finanziamenti, servizi assicurativi); funzioni di polizia anche amministrativa; gestione della migrazione, asilo e controllo delle frontiere; infine, amministrazione della giustizia e processi democratici ("*Sistemi AI utilizzati nella ricerca e nell'interpretazione dei fatti e nell'applicazione della legge ai fatti concreti o utilizzati nella risoluzione alternativa delle controversie. Influenzare gli esiti delle elezioni e dei referendum o il comportamento di voto, escludendo i risultati che non interagiscono direttamente con le persone, come gli strumenti utilizzati per organizzare, ottimizzare e strutturare le campagne politiche*" Sull'amministrazione della giustizia, in particolare, l'allegato III deve necessariamente leggersi alla luce del Recital n. 61, per quanto questo sia privo di efficacia vincolante. Il Regolamento distingue infatti tra i sistemi diretti ad assistere le attività giurisdizionali come quella di ricerca e interpretazione delle norme tali da incidere sulla decisione del singolo caso, da applicazioni che invece dovrebbero relegarsi al piano dell'organizzazione (come ad esempio quelle volte alla anonimizzazione o pseudonimizzazione, o l'assegnazione e distribuzione delle risorse, o la comunicazione tra il personale.

(art. 3, paragrafo 1, punto 12)[49]. Così, per fare un esempio specifico che riguarda l'erogazione di un servizio pubblico essenziale come quello scolastico, un sistema di IA rientra nell'elenco di quelli ad alto rischio, se è stato progettato, creato, sviluppato ed applicato proprio allo scopo specifico di fornire un ausilio agli insegnanti nella programmazione didattica personalizzata, con una finalità che dovrebbe pertanto emergere in tutto suo il ciclo di vita[50]. In realtà, gli osservatori più accorti avevano già evidenziato l'irriducibilità dei sistemi di IA ai "*prodotti pericolosi*" o ai servizi *on-off*[51], poiché innanzi tutto, componendosi di algoritmi, dati e potenza computazionale, essi sono sistemi dinamici, passano di mano in mano, seguono una catena di valore complessa, e, quelli di *Machine Learning* in particolare, continuano ad imparare dai dati anche dopo la fase di addestramento; sicché i profili di rischio sono essi stessi dinamici, sono influenzati dai nuovi dati, dai nuovi usi, dai contesti specifici in cui i sistemi sono adattati (le allucinazioni di ChatGPT o sistemi analoghi hanno ben diverso valore se il sistema è utilizzato per creare una poesia nello stile di Ungaretti o per una pubblicazione scientifica in neuroscienze)[52]. In secondo luogo, lo stato dell'arte dell'IA è oramai caratterizzato dalla diffusione dei modelli di IA "*finalità generali*" (GPMs), cui si è già accennato, adattabili a diversi contesti specifici; modelli di base (i cd. *Foundation Model*) che già da tempo la scienza computazionale aveva approfondito, quanto a opportunità e rischi[53]. Dopo

[49] L'AI Act nell'ambizione di perseguire l'obiettivo della certezza del diritto, introduce però, almeno sul piano generale, una classificazione rigida degli ambiti di rischio, suscitando preoccupazioni di sottovalutazione o sovrastima dei rischi. È emblematico il posizionamento dei sistemi per videogiochi tra quelli a rischio minimo (per i quali si prevede solo la sollecitazione ad adottare codici di condotta), nonostante possano comportare rischi significativi come la dipendenza o la manipolazione del comportamento per promuovere acquisti tra i giocatori soprattutto se minorenni. È un pericolo di sottostima altamente probabile, se si considera che l'industria dei videogiochi è da sempre un terreno fertile per l'innovazione digitale e ora dei sistemi di Intelligenza Artificiale. D'altra parte, anche per alcuni ambiti a rischio elevato, potrebbero rivelarsi *ex post* delle sovra-stime, poiché nell'impianto normativo si è deciso di guardare ad un intero ambito e non alle diverse e plurime attività che in esso possono rinvenirsi, ciascuna con un diverso valore di rischio; per tale profilo, cfr. le acute osservazioni di C. NOVELLI, *L'artificial intelligence Act Europeo: alcune questioni di implementazione*, 2/2024 in www.federalismi.it.

[50] Alla regola però corrisponde anche in questo caso un'eccezione. Per superare l'eccessiva rigidità della qualificazione dei sistemi ad alto rischio basata su contesti e scopo specifici, è infatti ora previsto che solo per i sistemi stand alone (ovvero quelli non incorporati in prodotti) vi sia una sorta di prova del nove: un giudizio di significatività del rischio a valle. Il fornitore ha pertanto la facoltà – e l'onere – di dimostrare, fornendo una sintesi di informazioni e ragioni , basate sul "livello di severità, intensità, probabilità" o durata del rischio e sull'impatto del sistema rispetto ad individui a collettività o a particolari gruppi di persone, che nella fattispecie concreta, pur trattandosi di un sistema di IA che rientrerebbe nell'elenco dell'Allegato III, non vi sono però effettivi rischi per i valori oggetto di tutela, diritti fondamentali ed ecosistema compresi.

[51] L. EDWARDS, *Regulating AI in Europe: four problems and four solution*, disponibile nel sito dell'Adalovelaceinstitute.

[52] Un eclatante esempio di dinamicità dei sistemi di IA è fornito dal famoso caso Loomis considerato il *leading case* dei rischi di IA nella cd. giustizia predittiva. Il sistema di IA Compas, prodotto da Northpoint, il cui algoritmo era (e tuttora è) coperto da segreto industriale e quindi sottratto alla conoscibilità sia dei terzi che degli stessi giudici, era stato progettato e messo sul mercato all'origine non per stimare il rischio di recidiva (finalità poi utilizzata nel sistema giudiziario americano ed oggetto del caso giudiziario), ma per supportare i direttori delle carceri americani nella distribuzione della popolazione carceraria, sulla base di diversi fattori dei detenuti, tra i quali l'età e i loro precedenti.

[53] Secondo la definizione di Stanford, un *Foundation model* (modello di base) è qualsiasi modello addestrato su dati ampi (generalmente utilizzando l'auto-supervisione su larga scala) che può essere adattato (ad esempio, affinato) per una vasta gamma di compiti successivi. Gli esempi attuali includono BERT, GPT-4.o, CLIP. Dal punto di vista tecnologico, i modelli di base si basano su reti neurali profonde e sull'apprendimento auto-supervisionato e come tali non sono di per

un acceso dibattito in realtà non solo europeo, ma planetario, l'AI Act, come anticipato nel primo paragrafo, disciplina adesso anche i *Modelli con finalità generali*[54], prevedendo oltre, alla loro definizione di cui all'articolo 3 n. 63, anche una *summa divisio (*Capo V), a seconda che tali modelli presentino o meno un "*rischio sistemico*"[55] (Sezione 1 Capo V, articoli 51-52); solo per questi ultimi, prevede infatti la pubblicazione di un elenco in cui essi devono essere registrati ed obblighi per i fornitori analoghi a quelli previsti per i sistemi di AI classificati ad alto rischio (valutazione *ex ante*, misure dirette a mitigare i rischi anche per la cibersicurezza, notifica di eventuali incidenti). Per tutti gli altri modelli di IA con finalità generale (Sezione 2 del Capo V, articoli 53-54) i quali, di per sé considerati, cioè prima di essere integrati in sistemi di IA, non possono essere considerati "ad alto rischio" (perché non è neanche definita ancora la loro specifica finalità), la tutela della sicurezza, della salute e dei diritti fondamentali è affidata ad alcuni obblighi di trasparenza; ovvero ad una concezione di trasparenza che appare diversa dal significato che il principio di trasparenza ha nell'ordinamento interno, come *paradigma* dello Stato democratico di diritto[56]

sé una innovazione, esistendo da decenni. Tra gli studi più recenti, E. JONES, *Explainer: What is a foundation model?*, 17 luglio 2023, disponibile nel sito dell'Adalovelaceinstitute. Anche sui Foundation model, il 2023 ha rappresentato un momento di svolta. Secondo *l'Artificial Intelligence Index di Stanford*, nell'anno 2023 sono stati rilasciati 149 Foundation model, il doppio di quelli dell'anno precedente ed è indicativo che il 65,7% di essi siano *open source*. Tra gli ultimi Llama 3 di Meta AI reso disponibile a metà aprile 2024; cfr i dettagli sul sito di MetaAI.

[54] Precisando che essa «*dovrebbe essere chiaramente distinta dalla nozione di sistemi di IA per consentire la certezza del diritto*» (Recital 97) e che è tale un modello che sia caratterizzato da "*una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato*».

[55] Con il quale si intende *«un rischio specifico per le capacità di impatto elevato dei modelli di IA per finalità generali, avente un impatto significativo sul mercato dell'Unione a causa della sua portata o di effetti negativi effettivi o ragionevolmente prevedibili sulla salute pubblica, la sicurezza, i diritti fondamentali o la società nel suo complesso, che può propagarsi su larga scala lungo l'intera catena del valore»* (Recital 110 e ss.). Il rischio sistemico sussiste se ricorrono alcune condizioni specifiche, alternative tra loro: a) se il modello presenta capacità di impatto elevato valutate sulla base di strumenti tecnici e metodologie adeguati, compresi indicatori e parametri di riferimento (b) se vi è una decisione della Commissione, ex officio o a seguito di una segnalazione qualificata del gruppo di esperti scientifici, che ritiene sussistente comunque una capacità o un impatto equivalenti a quelli di cui alla lettera a) secondo criteri di cui all'allegato XIII. Nel paragrafo 2 dell'articolo 51, si specifica anche che vi è una presunzione di rischio sistemico, quando un modello di IA per finalità generali la quantità cumulativa di calcolo utilizzata per il suo addestramento misurata in operazioni in virgola mobile è superiore a 10^25 FLOPs,).

[56] E. CARLONI, *Il paradigma trasparenza. Amministrazioni, informazione, democrazia*, Il Mulino, 2022, p. 35. «*Il principio di trasparenza (…) costituisce anche un caposaldo del principio di buon funzionamento della pubblica amministrazione, quale "casa di vetro" improntata ad imparzialità, intesa non quale mera conoscibilità, garantita dalla pubblicità, ma anche come intelligibilità dei processi decisionali e assenza di corruzione*»; Cons. Stato, Ad. Plen. 2 aprile 2020, n. 10; cfr. L. GIAMPIETRO, *La P.A. nella "casa di vetro": i concorrenti "diritti di accesso" in materia di appalti pubblici*, in *Ambiente e sviluppo*, 2020, 8-9, 671; A.G. OROFINO, *La trasparenza oltre la crisi. Accesso, informatizzazione e controllo civico*, Cacucci Editore, 2020; A. CORRADO, *Conoscere per partecipare: la strada tracciata dalla trasparenza amministrativa*, Edizioni Scientifiche Italiane, 2019. Non è possibile ripercorrere in questa sede l'evoluzione della giurisprudenza amministrativa in tema di trasparenza *algoritmica*; si citano, anche per i plurimi riferimenti bibliografici, E. CARLONI, *I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo*, in Dir. Amm., 2020, pp. 277; A. CORRADO, *La trasparenza necessaria per infondere fiducia in una amministrazione algoritmica e antropocentrica*, 22 febbraio 2023, in www.federalismi.It.; della stessa a. anche *Discrezionalità algoritmica e sindacato del giudice amministrativo*, E. BELISARIO, G. CASSANO (a cura di), *Intelligenza artificiale per la pubblica amministrazione*, cit., p. 173 e ss.; da ultimo, S. FOÀ, *Intelligenza artificiale e cultura della trasparenza amministrativa. dalle "scatole nere" alla "casa di vetro"?*, in *Diritto amministrativo*, n. 3-2023, 515 e ss.;

e che, anche nell'ambito della cd. legalità algoritmica, è stato addirittura configurato in una sua "declinazione rafforzata".

6. Dalla informazione alla spiegabilità: il nuovo volto della trasparenza "mediata"

Sulla scia della dottrina costituzionalista che ha delineato i canoni della legalità algoritmica, la giurisprudenza amministrativa, tra i principi cui devono conformarsi le decisioni amministrative che siano basate su algoritmi nella fase istruttoria o decisoria, ha infatti attribuito alla trasparenza un ruolo di primo piano[57], logicamente antecedente al principio della necessaria riferibilità della decisione al funzionario pubblico. L'idea di fondo è che, per quanto mutino gli ausili tecnologici e gli strumenti di comunicazione ed informazione, condizione necessaria della democrazia è la piena conoscenza della modalità di esercizio del potere pubblico. Lo Stato di diritto democratico – valore protetto anche dai rischi di IA, secondo l'art. 1 dell'AI Act – impone che vi sia una relazione stretta tra democrazia, sovranità popolare ex art. 1 Cost., e trasparenza, mediante la quale sia garantito un reale e dinamico confronto tra governanti e governati non in astratto, ma in relazione alle singole decisioni; "*un'amministrazione, quindi è democratica se trasparente, e se cerca la propria legittimazione nel confronto costante con i cittadini, e non solo attraverso il veicolo elettorale o nel principio di legalità, che sulla trasparenza politica pone i suoi fondamenti dogmatici*"[58]. La trasparenza che consente la conoscibilità piena dei meccanismi decisionali nelle fattispecie concrete disvela la sua forza, non solo nella fase fisiologica dell'esercizio del potere, ma anche rispetto ad una tutela giurisdizionale piena ed effettiva contro tutti gli atti amministrativi, ai sensi degli artt. 24 e 113 Cost., poiché senza conoscibilità delle ragioni poste a base della decisione (e degli atti e documenti in cui si rinvengono quelle ragioni di fatto e di diritto) non vi è effettività del diritto di difesa. Senonché, «*come l'amministrazione è*

[57] Il legislatore ha previsto peraltro, sia pure nel settore dei contratti pubblici, che in caso di uso di procedure automatizzate nel ciclo di vita dei contratti pubblici, con il ricorso a soluzioni tecnologiche anche di intelligenza artificiale, le stazioni appaltanti e gli enti concedenti devono rendere disponibile al codice sorgente, alla relativa documentazione e comunque ad ogni elemento utile a comprenderne la logica di funzionamento (articolo 30 D.lgs. 26/2023); ma ha anche precisato che il risultato finale della messa a disposizione degli elementi della soluzione adottata (e quindi dei sistemi di IA) sia la "conoscibilità e comprensibilità" delle decisioni assunte mediante automazione a favore degli operatori economici. Deve però anche osservarsi che l'accesso alle piattaforme digitali e alle infrastrutture informatiche che siano coperte da "diritti di privativa industriale" è escluso, ai sensi dell'articolo 35 comma 4, lett. b). Il tema è destinato ad assumere un ruolo di primo piano nel prossimo futuro, incidendo anche queste specifiche disposizioni (oltre che i criteri indicati, in senso però non vincolante, nell'art. 68 CAD) sull'opzione "open source" o regime proprietario che l'amministrazione deve considerare nella scelta della soluzione tecnologica più adeguata all'obiettivo. Deve segnalarsi che nel panorama mondiale si registra peraltro, già da qualche anno ma con una accelerazione nel 2023, Da alcuni anni tuttavia, "*un'inversione di tendenza caratterizzata dalla progressiva diffusione dei c.d. open source software a scapito del paradigma del software proprietario. Sul punto, cfr. per "gli importanti riflessi" che la scelta dell'amministrazione a favore di codici sorgenti aperti, anche per quanto riguarda il rispetto dei principi di imparzialità e trasparenza e l'effettività delle garanzie partecipative*" cfr. il prezioso contributo di S. DEL GATTO, *I sistemi proprietari, l'open source e la pubblica amministrazione,* in *Giornale di diritto amministrativo*, 5, 2021, p. 571 e ss. Significativi oltremodo i dati riportati nell'AI Index di Stanford 2024: nel 2023 sono stati rilasciati un totale di 149 modelli di base (Foundation Model), più del doppio rispetto al 2022. Di questi nuovi modelli, il 65,7% era *open source*, rispetto al solo 44,4% nel 2022 e al 33,3% nel 2021.

[58] A.G. OROFINO, *La trasparenza oltre la crisi. Accesso, informatizzazione e controllo civico*, Cacucci Editore, 2020, p. 50.

soggetta ad una necessaria mutabilità, così la trasparenza che ne è snodo di collegamento con il tessuto sociale. Al cambiare dell'amministrazione, al mutare degli interessi che si rapportano con l'amministrazione, devono mutare anche la trasparenza e si suoi meccanismi che presidiano alla relazione tra individui e istituzioni nell'ordinamento democratico»[59]. Rispetto alla decisione algoritmica, la trasparenza diventa allora, come è stato osservato, allo stesso tempo flessibile e vulnerabile. Flessibile, perché per cavalcare l'onda in arrivo, la trasparenza innanzi tutto si rafforza, cosicché il giudice amministrativo afferma che la conoscibilità deve investire tutti gli elementi essenziali del sistema algoritmico utilizzato: «*dai suoi autori, al procedimento usato per la sua elaborazione, al meccanismo della decisione comprensivo delle priorità assegnate nella procedura di valutazione e decisionale dei dati selezionati come rilevanti*»[60], con una declinazione "*rafforzata*" perché diretta appunto a bilanciare l'opacità algoritmica[61]. Vulnerabile, perché in concreto, se riesce già difficile conoscere gli elementi di un sistema algoritmico tradizionale, per i sistemi di IA avanzati come quelli fondati sul Deep Learning, la conoscibilità effettiva sembra essere preclusa dalla stessa elevata complessità matematica che consente a quei sistemi di svolgere i compiti che, se fossero svolti da un essere umano, presupporrebbero intelligenza. Se questo è l'*humus* in cui il Regolamento è destinato ad attecchire, occorre allora chiedersi se in esso vi sia la medesima idea di trasparenza che è stata affermata dalla giurisprudenza e dal legislatore italiani, o piuttosto una sua versione "praticabile", un surrogato possibile, il massimo risultato ottenibile nel *trade-off* tra sviluppo tecnologico ed efficienza da esso promesso, da un lato, e i principi dello Stato di diritto democratico, dall'altro. Ad una prima lettura del testo del Regolamento pare che tutto il clamore anche mediatico intorno al principio di trasparenza si sia alla fine tradotto in una marcata asimmetria a favore di chi detiene la conoscenza (i *provider* di IA) e che ha il potere di filtrare *pro domo sua* ciò che deve essere reso trasparente e quindi conoscibile da parte sia dei cittadini che degli altri soggetti della filiera (tra cui la stessa amministrazione che utilizza i sistemi di IA). In sostanza, se la fiducia è l'elemento cardine che regge l'era digitale, qui si assiste ad una inversione logica della questione. Mentre nella prospettiva della tutela offerta sul piano giurisprudenziale e normativo interno, la trasparenza è il presupposto per riporre fiducia nelle decisioni algoritmiche; nel contesto dell'IA Act, la trasparenza elargita dai provider presuppone che vi sia a monte fiducia da parte degli utenti finali che essi effettivamente mettano a disposizione le informazioni utili per conoscere e comprendere i sistemi immessi sul mercato. Con l'effetto, sottolineato da autorevole dottrina, secondo cui parlare «*di trasparenza potrebbe quindi risultare fuorviante*", emergendo "*un ennesimo momento di tensione tra la funzione ordinatrice e di garanzia del diritto e il ruolo ancillare della tecnica, arricchita dalla oscurità intrinseca che riguarda non solo l'algoritmo, ma, più in generale, la traiettoria di sviluppo delle decisioni pubbliche che hanno*

[59] E. CARLONI, cit, p. 36.
[60] Cons. Stato, sez. IV, n. 8472/2019.
[61] Così, si diffonde l'opinione che, per le decisioni algoritmiche, gli strumenti giuridici tradizionali, come il diritto di accesso al codice sorgente, rischino di essere «un mero simulacro di natura formale», cfr. L. TORCHIA, *Lo Stato digitale. Una introduzione*, Bologna, 2022, p. 154.

traghettato i pubblici poteri in una black box society»[62]. In sostanza, l'idea di trasparenza che trapela dall'AI Act sembra ispirarsi, oltre che alle logiche proprietarie degli operatori privati che investono quantità ingenti di risorse nella ricerca e sviluppo dei sistemi e modelli di IA, una *ragion pratica*: non vi è possibilità di accesso diretto alle informazioni che sarebbero rilevanti per conoscerne il funzionamento in concreto, perché i sistemi e i modelli di IA restano di fatto così complessi da essere di fatto incomprensibili. Allora si opta pragmaticamente per una sorta di "narrazione" della trasparenza dall'alto verso il basso, sotto forma di obblighi di istruzioni per l'uso, comunicazioni, documenti di sintesi elaborati dagli stessi fornitori. In altri termini, visto che è impossibile avere contezza dei dati di addestramento, dei modelli, delle infrastrutture informatiche dei sistemi di IA, viene imposta al provider l'obbligazione di una comunicazione "*chiara e comprensibile*", secondo indeterminati standard di ragionevolezza che però lasciano ampi spazi di valutazione a chi deve "raccontare" e quindi selezionare le informazioni. Non è escluso che, almeno allo stato dell'arte dell'evoluzione tecnologica, questo possa essere l'unico punto di caduta tra garanzia e sviluppo tecnologico cui aspira il mercato, tra performance elevate e rispetto dei principi giuridici tradizionali[63]; ma i segnali della trasfigurazione della trasparenza in salsa digitale emergono in varie disposizioni dell'IA Act che di seguito si accennano. In primo luogo, va menzionato l'articolo 13 – relativo solo i sistemi di AI "ad alto rischio" - che già nella rubrica ("*Trasparenza e fornitura di informazione ai deployer*"), conferma il modello meramente "*informativo*" prescelto[64]. La disposizione citata prevede infatti che tali sistemi debbano essere progettati e sviluppati in modo da garantire che il loro funzionamento «*sia sufficientemente trasparente da consentire ai deployer di interpretare l'output del sistema e utilizzarlo adeguatamente. Sono garantiti un tipo e un livello di trasparenza adeguati*». Viene promessa una trasparenza *sufficiente* e *adeguata*, ma il livello di sufficienza ed adeguatezza rientrano nella piena discrezione e competenza tecnica dei fornitori, poiché, essendo "oscuro" il meccanismo di raggiungimento dell'output, ogni interpretazione del suo

[62] A. DI MARTINO, *Tecnica e potere nell'amministrazione per algoritmi*, cit., p. 194.

[63] Eppure al legislatore sovranazionale non sfugge l'oggetto che la trasparenza, intesa senso forte, dovrebbe disvelare a favore dei cittadini. Lo dimostra la cura con cui ha indicato le condizioni dei modelli rilasciati con licenza libera e open source, per i quali non si applicano molti degli obblighi informativi da parte dei provider. Sono tali quelli che infatti consentono, stavolta in forma diretta e non "mediata" da logiche comunicative, l'accesso, l'uso, la modifica e la distribuzione del modello e «i cui parametri, compresi i pesi, le informazioni sull'architettura del modello e le informazioni sull'uso del modello, sono resi pubblici». Se vi è questo elevato livello di trasparenza, gli utenti hanno accesso diretto a tutte le componenti del modello, così da consultarlo liberamente, utilizzarlo, modificarlo e ridistribuirlo e la sottrazione agli obblighi informativi si spiega perché la trasparenza è garantita alla fonte, eliminando il regime proprietario del modello. Così congegnato, per open source non si intende, pertanto solo "accessibilità al codice sorgente": essa è un elemento necessario ma non sufficiente, poiché è richiesta la possibilità di permettere modifiche e ridistribuzioni del modello a qualsiasi scopo, per le più diverse applicazioni.

[64] Nel Recital 72, si sottolinea che "*To address concerns related to opacity and complexity of certain AI systems and help deployers to fulfil their obligations under this Regulation, transparency should be required for high-risk AI systems before they are placed on the market or put it into service. High-risk AI systems should be designed in a manner to enable deployers to understand how the AI system works, evaluate its functionality, and comprehend its strengths and limitations. High-risk AI systems, should be accompanied by appropriate information in the form of instructions of use. Such information should include the characteristics, capabilities and limitations of performance of the AI system"*

significato è aperta, né vi sono criteri cui tale attività deve informarsi. In aggiunta, la norma richiede che i sistemi di IA ad alto rischio siano accompagnati da *istruzioni d'uso*, sia in formato digitale che non, che forniscano informazioni "*concise, complete, corrette e chiare che siano pertinenti, accessibili e comprensibili per i deployer",* anch'esse selezionate da chi le detiene, non potendosi escludere neanche che la divergenza possa, in concreto, dipendere da manipolazioni dirette a tutelare i segreti industriali di operatori economici che si muovono in un settore estremamente competitivo. La natura "*comunicativa*" della trasparenza è riproposta anche con riguardo ai sistemi di IA che, a prescindere dal livello di rischio, interagiscono direttamente con le persone fisiche. L'articolo 50 che apre il Capo IV dedicato agli obblighi di trasparenza per "certi sistemi" prevede infatti che «*i fornitori garantiscono che i sistemi di IA destinati a interagire direttamente con le persone fisiche sono progettati e sviluppati in modo tale che le persone fisiche interessate siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dal punto di vista di una persona fisica ragionevolmente informata, attenta e avveduta, tenendo conto delle circostanze e del contesto di utilizzo*» (cd. *bot-disclosure*); regola di trasparenza limitata a dichiarare che si interloquisce con un sistema e non con un essere umano; eccezione alla regola, se ciò è evidente, demandando all'interprete la verifica della ragionevole evidenza nel caso concreto. Il dovere informativo è poi previsto per l'IA Generativa, sulla quale si sono concentrate, come accennato, le discussioni politiche degli ultimi mesi. Per tali sistemi il paragrafo 2[65] del medesimo articolo 50 impone ai fornitori il "*watermarking*", una sorta di rivisitazione in chiave contemporanea della filigrana utilizzata nelle banconote o nelle marche da bollo, con lo scopo di evidenziare, in questo caso, non l'autenticità come nel caso della filigrana, ma la natura "artificiale" del contenuto generato: video, immagine, musica, testo linguistico. Anche in tale fattispecie è riservato ai fornitori un notevole spazio di valutazione tecnica, poiché la regola generale è delimitata da un ampio ventaglio di limitazioni (articolo 50, paragrafo 2, secondo periodo). I fornitori devono infatti garantire che le soluzioni tecniche adottate per la marcatura siano «*efficaci, interoperabili, solide e affidabili nella misura in cui ciò sia tecnicamente possibile, tenendo conto delle specificità e dei limiti dei vari tipi di contenuti, dei costi di attuazione e dello stato dell'arte generalmente riconosciuto, come eventualmente indicato nelle pertinenti norme tecniche*». In ogni caso, l'obbligo non si applica «*se i sistemi di IA svolgono una funzione di assistenza per l'editing standard o non modificano in modo sostanziale i dati di input forniti dal deployer o la rispettiva semantica, o se autorizzati dalla legge ad accertare, prevenire, indagare o perseguire reati*». Un'analoga disposizione (articolo 50 paragrafo 4) è prevista anche per i sistemi che generano *deep fake* con video, immagini, musica; o testi linguistici su questioni di pubblico interesse. In tal caso, però l'obbligo di informazione è traslato sul *deployer* - ivi compresa pertanto la pubblica amministrazione – che è tenuto ad informare il pubblico che il contenuto è stato generato o

[65] «*I fornitori di sistemi di IA, compresi i sistemi di IA per finalità generali, che generano contenuti audio, immagine, video o testuali sintetici, garantiscono che gli output del sistema di IA siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente*».

manipolato artificialmente, sempreché non ricorra una delle eccezioni previste dalla norma. Infine, la duttilità della trasparenza digitale emerge senza veli con riferimento proprio ai *Modelli con finalità generali* (GPMs) per i quali i *provider*, oltre che tenere traccia della documentazione, devono elaborare e mettere a disposizione del pubblico una «*sintesi sufficientemente dettagliata dei contenuti utilizzati per l'addestramento del modello di IA per finalità generali, secondo un modello fornito dall'ufficio per l'IA*» (articolo 53, paragrafo 1 lett. d). In attesa del modello che sarà elaborato dall'AI Office, appena istituito, e restando al dato normativo, resta il fatto che la sintesi – e quindi la selezione delle informazioni da rendere disponibili - è elaborata dal medesimo soggetto che è tenuto a conformarsi alla normativa; ed in ogni caso essa deve essere solo "sufficientemente" dettagliata. Questo articolato flusso comunicativo che parte dal *provider*, almeno per i sistemi ad alto rischio è destinato, a sfociare in un punto finale, che se colto in tutte le sue potenzialità, rischia di mettere in luce proprio l'astrattezza degli obblighi di comunicazione e pubblicazione disseminati nel testo. Per tali sistemi, infatti l'articolo 86 sancisce a chiare lettere il "*diritto alla spiegazione*" a favore dei singoli ovvero a favore di coloro su cui la decisione che si avvale di sistemi di IA incide. E la spiegabilità non riguarda qui il "sistema" o al "modello" di IA prodotto o utilizzato per quella categoria generale di compiti, attività, decisioni, ma riguarda proprio il caso concreto, come chiaramente si evince già dalla rubrica "*diritto alla spiegazione dei singoli processi decisionali*": tradotto in termini di diritto amministrativo, la spiegazione riguarda il ruolo del sistema di IA nel procedimento che ha condotto al provvedimento finale e deve essere una spiegazione "*chiara e significativa*" relativa ai "*principali elementi della decisione adottata*". Si tratta, dal punto di vista delle garanzie, di una delle più rilevanti innovazioni introdotte nella fase parlamentare del lungo percorso legislativo dell'AI Act (la norma non era prevista infatti nella proposta originaria della Commissione del 2021, nella quale i cittadini, o comunque "gli utenti finali" non comparivano proprio sulla scena). Ma se gli obblighi di trasparenza a monte non sono sufficienti a consentire tale spiegabilità, il sistema normativo rischia di entrare in un *loop* e, nel caso di sistemi ad alto rischio utilizzati dall'amministrazione, a danno anche della stessa amministrazione che si ritrova al centro tra la pretesa alla spiegazione del cittadino e il potere conoscitivo detenuto dal *provider* che ha fornito il sistema di IA. La sfida che si apre sarà in primo luogo pertanto quella di fare in modo che le due facce si parlino tra loro: che la trasparenza "comunicata" percorra senza troppi ostacoli tutta la catena di valore dei sistemi di IA, dai fornitori a chi li utilizza fino a chi ne subisce l'impatto per effetto di singole decisioni amministrative. In relazione il diritto alla spiegabilità, molte questioni però dovranno essere affrontate in fase applicativa (anche a prescindere dal problema irrisolto della Black Box[66]): il suo perimetro di

[66] La spiegabilità è peraltro invocata quale una delle auspicabili regole di opportuna cautela anche con riferimento all'area, dal perimetro ancora incerto, della cd. giustizia predittiva, ma con ficcante ricostruzione, essa non è sufficiente se isolata, dovendo integrarsi con altre misure che guardano al modello organizzativo dei sistemi di IA quali *«a)il controllo pubblico sugli algoritmi di giustizia predittiva che … dovrebbero essere open source o di proprietà pubblica o, quanto meno, certificati da un'autorità*

applicazione, il livello di "*chiarezza e significatività*" circa l'effettiva influenza del sistema rispetto alla decisione finale formalmente assunta da un essere umano, tenendo conto che la spiegazione deve essere comprensibile, ma anche pertinente; gli strumenti e le tecniche processuali con i quali il diritto deve essere garantito in caso di violazione; il modello di sindacato giurisdizionale sulla verifica della attendibilità della spiegazione e della sua adeguatezza, vista la acclarata opacità dei sistemi di IA e comunque l'incompetenza tecnica degli operatori giuridici; a monte, il peso del diritto alla spiegazione sulla opzione di acquisto del sistema da parte della pubblica amministrazione, tra sistemi *open source* e sistemi proprietari.

7. Conclusioni

L'AI Act rappresenta certamente un passo pionieristico nel tentativo di disciplinare una famiglia di tecnologie *disruptive* come quella di IA. Tuttavia, nel campo specifico del diritto amministrativo, si cala in una dimensione in cui l'esigenza di controllo umano e di trasparenza delle decisioni algoritmiche sono state già esplorate, con approdi che ora fanno i conti con la nuova stagione di diritto positivo. Le problematiche, che sono state da poco messe a fuoco dalla dottrina e dalla giurisprudenza, sono destinate a complicarsi con l'entrata in vigore delle nuove norme, sia in previsione della loro implementazione pratica che della loro interpretazione sistematica alla luce della peculiarità dell'esercizio del potere pubblico. La fase di implementazione dell'AI Act, che è funzionale al miraggio della "certezza del diritto", richiederà pertanto un particolare sforzo interpretativo, avendo in mente che *"l'inevitabile innalzamento del livello di tecnicità del procedimento non deve depotenziare la concreta contestabilità da parte dell'interessato della decisione robotica, non soltanto sotto il profilo della sua conformità ai parametri legali, ma anche in relazione alla sua credibilità razionale*"[67]. Solo all'esito del percorso appena cominciato, che dovrà compiersi in un dialogo intellettualmente onesto tra competenze diverse, giuridiche, economiche ed informatiche, potrà effettivamente verificarsi se la prova di resistenza cui sono ora sottoposti i canoni della cd. legalità algoritmica potrà dirsi superata.

pubblica, previa verifica sul loro funzionamento; b) il controllo pubblico sulla qualità dei dati, ossia sulle banche di dati di dottrina, giurisprudenza e di normativa utilizzate dagli algoritmi», G. CARLOTTI, *La giustizia predittiva e le fragole con la panna*, *cit*., p. 11.

[67] D. SIMEOLI, *L'automazione dell'azione amministrativa nel sistema delle tutele di diritto pubblico*, in in (a cura di) A. PAJNO, F. DONATI, A. PERRUCCI, cit., 2022, Vol. II, p. 638.