

media LAWS

Anticipazioni

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

Maria Romana Allegri

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

Abstract

Questo scritto tenta, in primo luogo, di circoscrivere il perimetro del concetto di trasparenza algoritmica che, come si vedrà, viene in qualche modo a coincidere con quello di spiegabilità dell'algoritmo. Vengono poi evidenziate le difficoltà della traduzione delle regole tecnologiche dal linguaggio naturale a quello delle macchine e prospettati alcuni esempi di tentativi di definizione di indicatori misurabili di trasparenza algoritmica. Successivamente, il saggio rileva alcune incongruenze in tema di trasparenza algoritmica nella recente *digital regulation* dell'Unione europea (regolamento sulla protezione dei dati, regolamento sui servizi digitali, regolamento sull'intelligenza artificiale). L'utilizzo di algoritmi nel processo decisionale delle pubbliche amministrazioni è stato oggetto di alcune interessanti pronunce del Consiglio di Stato e della Corte di cassazione, non prive tuttavia di incertezze. In conclusione, si formulano alcune osservazioni relative ai principali risultati di questo studio e alla possibilità di considerare la trasparenza algoritmica un principio di rango costituzionale.

This paper attempts, first of all, to circumscribe the perimeter of the concept of algorithmic transparency which, as will be seen, somehow coincides with that of algorithmic explainability. The difficulties of translating technological rules from natural language to that of machines are then highlighted and some examples of attempts to define measurable indicators of algorithmic transparency are presented. Subsequently, this essay underlines some inconsistencies regarding algorithmic transparency in the recent digital regulation of the European Union (data protection regulation, digital services regulation, artificial intelligence regulation). The use of algorithms in the decision-making process of public bodies has been the theme of some interesting rulings from the Council of State and the Court of Cassation, albeit with some uncertainty. In conclusion, some comments are expressed regarding the main results of this work and the possibility of considering algorithmic transparency a principle of constitutional rank.

Sommario

1. Introduzione. – 2. Lo sfuggente concetto di trasparenza algoritmica. – 3. Il lessico della trasparenza algoritmica nella *digital regulation* dell'UE: nodi irrisolti. – 4. L'utilizzo di algoritmi da parte delle pubbliche amministrazioni: qualche indicazione giurisprudenziale e molte incertezze. – 5. Riflessioni conclusive.

Keywords

Intelligenza artificiale – algoritmi – *Machine learning* – trasparenza – pubblica amministrazione

1. Introduzione

I sistemi di intelligenza artificiale, indipendentemente dalle loro varie e non sempre univoche definizioni, stanno diventando rapidamente una parte integrante della nostra vita quotidiana, tanto da poter parlare di “macchine sociali”¹, di “costrutti sociotecnici”² o ancora di “istituzioni socio-digitali”³ in riferimento all’interconnessione di elementi umani e tecnologici. Il recente Regolamento europeo sull’intelligenza artificiale⁴ li definisce come sistemi automatizzati progettati per funzionare con diversi livelli di autonomia e adattabilità, che ricevono *input* e generano *output* – quali previsioni, contenuti, raccomandazioni o decisioni – che possono influenzare ambienti fisici o virtuali» (art. 3, c. 1).

L’associazione uomo-macchina – cioè, la stretta interazione fra esseri umani e sistemi informatici – è ormai un elemento strutturale nel nostro ecosistema: gli algoritmi definiscono e limitano l’orizzonte epistemico degli individui, determinando le modalità di connessione fra loro e con il mondo esterno, le opportunità di conoscenza e le possibilità di scelta, fino al punto da interferire con la libertà morale e con la costruzione e rappresentazione dell’identità personale⁵. Ciò è amplificato anche dal fatto che oggi gli elementi in base ai quali le persone compiono le proprie scelte sono sempre più frequentemente forniti da sistemi tecnologici che, attraverso varie forme di *nudging*⁶ più o meno esplicito, possono influenzare o persino determinare i comportamenti

¹ A. Simoncini, *Il linguaggio dell’intelligenza artificiale e la tutela costituzionale dei diritti*, in *Rivista AIC*, 2, 2023, spec. 4. Si veda anche N. Cristianini – T. Scartamburlo, *On social machines for algorithmic regulation*, in *AI & Society*, 35, 2020, 645 ss.

² C. Peroli – C.R. Pereira de Lima, *Democrazia come Value-Sensitive Design: un approccio sociotecnico allo sviluppo delle tecnologie basate su principi democratici*, in *Rivista italiana di informatica e diritto*, 1, 2021, 149 ss.

³ A. Beckers – G. Teubner, *Three Liability Regimes for Artificial Intelligence*, Oxford, 2021. Si veda anche O.G. Loddo, *L’agire sociale ai tempi dell’intelligenza artificiale. Il concetto di “istituzione sociale-digitale*, in *L’Icoerco*, 1, 2024, 359 ss.

⁴ Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull’intelligenza artificiale.

⁵ I. De Vivo, *Il sé allo specchio dell’algoritmo. Libertà epistemica e identità individuale*, in A. Sterpa (a cura di), *L’ordine giuridico dell’algoritmo*, Napoli, 2024, 51 ss., spec. 54.

⁶ Il *nudging*, traducibile in italiano come “spinta gentile”, consiste nell’attitudine espressa da poteri pubblici e privati a indirizzare il comportamento degli individui mediante pratiche persuasive più o meno esplicite, senza ricorrere a comandi, divieti o incentivi economici. Si tratta di una forma di “paternalismo debole”, che non vincola le scelte individuali ma si limita ad influenzarle, nel presupposto che gli obiettivi verso cui il *nudging* tende non possono che essere condivisi e che il loro mancato raggiungimento dipende essenzialmente da pregiudizi o preconcetti che è opportuno demolire. Tuttavia, posto che la mancanza di costrizioni legale non implica necessariamente l’assenza di interferenze nella sfera di libertà individuale, il *nudging* può essere considerato incompatibile con il disegno costituzionale secondo cui lo Stato ha il compito di assicurare i presupposti per il godimento delle libertà, non quello di orientare le scelte dei singoli. Si vedano in proposito, *ex multis*, Q. Camerlengo, *Costituzione e coscienza sociale: il contributo della teoria del “nudge”*, in *Lo Stato*, 21, 2023, 11 ss.; I. Carter, *Il disagio liberale di fronte al nudging*, in *Il Politico*, 1, 2023, 133 ss.; A. Gragnani, *Nudging e libertà costituzionale*, in *Dirittifondamentali.it*, 1, 2021, 498 ss.; M. Miravalle, *Gli orizzonti della teoria del “nudging” sulla normatività: verso un diritto senza sanzioni?*, in *BioLaw Journal*, 1, 2020, 441 ss.; C.R. Sunstein, *Effetto nudge: La politica del paternalismo libertario*, Milano, 2015; R.H. Thaler – C. R. Sunstein, *Nudge: The Final Edition*, London, 2021; A. Zito, *La nudge regulation nella teoria giuridica dell’agire amministrativo. Presupposti e limiti del suo utilizzo da parte delle pubbliche amministrazioni*, Napoli, 2021.

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

individuali e collettivi⁷.

Stiamo dunque assistendo a una progressiva perdita di centralità dell'elemento antropologico, data la progressiva delega totale o parziale di funzioni cognitive, fin qui considerate esclusivo appannaggio degli esseri umani, a sistemi tecnologici: la tecnologia non è più solo uno strumento per svolgere più efficientemente alcuni compiti, ma diviene soggetto agente che elabora autonomamente le informazioni e assume decisioni che hanno un impatto nella sfera giuridica individuale⁸. Come è stato efficacemente notato, «siamo organismi informazionali (inforg), reciprocamente connessi e parte di un ambiente informazionale (l'infosfera), che condividiamo con altri agenti informazionali, naturali e artificiali, che processano informazioni in modo logico e autonomo»⁹.

Non era mai accaduto finora che la tecnologia acquistasse autonomia rispetto agli esseri umani e diventasse capace di dar forma a un ecosistema in grado di svilupparsi, crescere e modificarsi autonomamente, superando gli umani limiti fisici e cognitivi¹⁰. Oggi, invece, la tecnologia digitale sta comportando una vera e propria ibridazione tra soggetti e oggetti: da una parte gli oggetti tecnologici divengono sempre più autonomi nelle decisioni e nelle azioni; dall'altra i soggetti umani delegano sempre più decisioni e azioni a tali strumenti, così che l'algoritmo può considerarsi «non tanto strumento neutrale soggetto al nostro arbitrio, ma quanto un portatore autonomo di una credibilità che sconfina, a tratti, in una veridicità di stampo divino»¹¹. Ciò comporta il rischio concreto che le decisioni algoritmiche, per il solo fatto di essere prodotte da sistemi tecnologicamente avanzati, che spesso rendono più semplici vari aspetti della vita umana, vengano considerate acriticamente come l'ordine naturale delle cose¹².

Tutto ciò mette in crisi i principi fondamentali del diritto costituzionale, incardinato sulla centralità della persona umana e sulla garanzia dei diritti individuali, se non del diritto *tout court*, inteso come tecnica di controllo del comportamento umano, veicolata attraverso enunciati linguistici. Nel campo dell'intelligenza artificiale, infatti, il diritto si confronta con altri sistemi di regolamentazione (in particolare, di matrice tecnologica) talvolta in sinergia e talaltra venendone sopraffatto. Non è escluso allora che, nel tempo, i tradizionali metodi e strumenti giuridici perdano gradualmente efficacia, a meno che il diritto non trovi il modo di affrancarsi da una logica semplicemente reattiva rispetto agli effetti della diffusione dell'IA, riuscendo invece ad anticipare la soluzione dei problemi che possono presentarsi nel prossimo futuro, mostrandosi sensibile alle nuove tendenze e ai nuovi rischi e disponibile all'interazione e all'integrazione con diversi sistemi di regole¹³.

⁷ A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal*, 1, 2019, 63 ss., spec. 69-70.

⁸ A. Simoncini, *Il linguaggio dell'intelligenza artificiale e la tutela costituzionale dei diritti*, cit., 8.

⁹ L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano, 2017, 106.

¹⁰ A. Sterpa e al., *L'ordine giuridico dell'algoritmo*, cit., 1124.

¹¹ P. Zellini, *La dittatura del calcolo*, Milano, 2018, 186.

¹² A. Sterpa – I. De Vivo – C. Capasso, *L'ordine giuridico dell'algoritmo: la funzione regolatrice del diritto e la funzione ordinatrice dell'algoritmo*, in *Nuovi Autoritarismi e Democrazie*, 2, 2023, 1120 ss., spec. 1152.

¹³ G. Mobilio, *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica*, in *BioLaw*

È stato efficacemente messo in luce come l'evoluzione del *digital environment* sia uno dei fattori alla base del crescente *judicial activism*, nello sforzo di supplire all'inerzia della politica e del legislatore: «*Political inertia (which is not always forced as sometimes power is delegated to courts with a view to avoiding difficult choices) has fostered judicial imagination within the digital era, along with the resulting use of metaphors and frames to adapt legal systems to the peculiarities of the digital realm*»¹⁴. Analogamente, è stata evidenziata la progressiva affermazione dei “poteri digitali privati”, per contrastare la quale sono state prospettate sostanzialmente due soluzioni: l'applicazione orizzontale dei diritti fondamentali nei confronti dei poteri privati e, contemporaneamente, l'identificazione di nuovi diritti adeguati a contrastare le sfide poste dagli algoritmi¹⁵. La sfida del cosiddetto “costituzionalismo digitale” è proprio quella di arginare l'affermazione del potere illimitato degli attori privati, attraverso strumenti giuridici di livello nazionale, ma soprattutto sovranazionale¹⁶. Ci si chiede, allora, se nel complesso degli atti normativi prodotti dall'Unione europea per regolare l'utilizzo di algoritmi e di sistemi di IA non possano essere rintracciati principi di rango costituzionale o para-costituzionale e se fra questi non possa essere annoverato il principio della trasparenza algoritmica.

In qualche modo, gli algoritmi svolgono una funzione regolativa simile a quella del diritto: reagendo al caos delle infinite possibilità, ordinano e razionalizzano una molteplicità di elementi, rendendoli fruibili per la mente umana; dunque, in parole povere, entrambi i sistemi (quello giuridico e quello algoritmico) assumono complessità e producono semplicità, anche se gli algoritmi lo fanno con velocità e capacità di elaborazione enormemente superiori rispetto all'armamentario giuridico¹⁷. Diversamente dagli algoritmi, però, il diritto è uno strumento ordinamentale flessibile, assiologicamente orientato, che lascia spazio all'innovazione, al discernimento e all'interpretazione, che permette il verificarsi di comportamenti e scelte umane in contrasto con i principi giuridici, che nei sistemi democratici è legittimato dalla volontà popolare¹⁸. Le regole giuridiche sono formulate in termini generali e astratti in modo da potersi adattare ad una varietà di casi, mantenere efficacia coercitiva nel tempo ed estendersi anche a nuove situazioni non previste al momento della loro originaria produzione. Per questo il diritto è espresso in un linguaggio intrinsecamente malleabile e per qualche verso ambiguo, che può produrre effetti solo ricorrendo all'interpretazione e al giudizio dell'uomo. Invece, le regole tecniche sono formulate in un linguaggio formalizzato e preciso, in base a categorie predefinite e scelte predeterminate¹⁹. Ora, posto che solo il diritto può far sì che la regolamentazione dei diversi fenomeni sia orientata alla preservazione della dignità e della libertà della persona e del pluralismo sociale,

Journal, 2, 2020, 401 ss., spec. 404-405.

¹⁴ O. Pollicino, *Judicial Protection of Fundamental Rights on the Internet*, Oxford, 2021, 12-13.

¹⁵ Ivi, 200.

¹⁶ G. De Gregorio, *Digital Constitutionalism in Europe*, Cambridge, 2022.

¹⁷ A. Sterpa e al., *L'ordine giuridico dell'algoritmo: la funzione regolatrice del diritto e la funzione ordinatrice dell'algoritmo*, cit., 1126.

¹⁸ *Ibid.*

¹⁹ E. M. Lombardi, *Norma e algoritmo: alcune considerazioni sul nuovo ordine tecnologico*, in *Giustizia Civile.com*, 7, 2020, spec. 10.

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

occorre difendere la primazia dell'ordine giuridico, cioè dei valori democratico-liberali del costituzionalismo²⁰. Tuttavia, tradurre le regole giuridiche in codice informatico, in modo da consentirne l'applicazione automatica, richiede che esse vengano trascritte nel linguaggio formale delle macchine, cosa che comporta inevitabilmente una riduzione del loro grado di generalità e astrattezza e, conseguentemente, della loro flessibilità e adattabilità²¹.

È chiaro che qualsiasi sistema tecnologico deriva originariamente da un'attività creativa umana, che quindi è alla base anche della progettazione delle funzionalità algoritmiche. Tuttavia, le dinamiche attraverso cui gli algoritmi “imparano” dai dati, basate su logiche inferenziali e calcoli statistico-probabilistici, spesso non sono comprensibili per le persone umane, nemmeno per gli stessi programmatori degli algoritmi: non possono essere spiegate prima dell'assunzione della decisione algoritmica né possono essere ricavate a posteriori ripercorrendo in senso contrario la catena dei processi, poiché la maggior parte delle decisioni algoritmiche, a differenza di quelle umane, sono basate sul calcolo delle probabilità e non su procedimenti logico-deduttivi in cui il giudizio è il frutto della ponderazione dei diversi elementi in senso non solo utilitaristico-funzionale, ma anche etico. Dunque, persino la pubblicità di informazioni sulle varie fasi dei processi algoritmici può non essere sufficiente ad assicurarne la trasparenza. Eppure, la questione della trasparenza algoritmica come possibile antidoto alla “tecnocrazia” o “algocrazia”²² ricorre costantemente nella recente produzione normativa e giurisprudenziale in tema di trattamento dei dati, servizi digitali, decisioni automatizzate e intelligenza artificiale, nella convinzione che le persone debbano essere poste in condizione di comprendere la logica sottostante alle decisioni algoritmiche, per poter tutelare i propri diritti. Occorre però evidenziare l'ambiguità connessa al concetto di trasparenza algoritmica e la vaghezza dei relativi standard, dovute alla difficoltà di tradurre i processi algoritmici nel linguaggio comune e all'assenza di parametri precisi per valutare l'adeguatezza delle “spiegazioni”.

Nelle pagine che seguono, si cercherà in primo luogo di circoscrivere il perimetro del concetto di trasparenza algoritmica che, come si vedrà, viene in qualche modo a coincidere con quello di spiegabilità dell'algoritmo. Verranno poi evidenziate le difficoltà della traduzione delle regole tecnologiche dal linguaggio naturale a quello delle macchine e verranno prospettati alcuni esempi di tentativi di definizione di indicatori misurabili di trasparenza algoritmica. Successivamente, si rileveranno alcune incongruenze in tema di trasparenza algoritmica nella recente *digital regulation* dell'Unione europea (regolamento sulla protezione dei dati, regolamento sui servizi digitali, regolamento sull'intelligenza artificiale). L'utilizzo di algoritmi nel processo decisionale delle pubbliche amministrazioni è stato oggetto di alcune interessanti pronunce del Consiglio di Stato

²⁰ A. Sterpa e al., *L'ordine giuridico dell'algoritmo: la funzione regolatrice del diritto e la funzione ordinatrice dell'algoritmo* cit., 1128.

²¹ E.M. Lombardi, *Norma e algoritmo*, cit., 11.

²² B. Chomanski, *Legitimacy and automated decisions: the moral limits of algocracy*, in *Ethics and Information Technology*, 3, 2022, 1 ss.; J. Danaher, *The Threat of Algocracy: Reality, Resistance and Accommodation*, in *Philosophy and Technology*, 3, 2016, 245 ss.; P. Marrone, *Democrazia epistemica, epistocrazia, algocrazia: alcuni problemi*, in *Paradigmi*, 2, 2022, 307 ss.; F. Zambonelli, *Algocrazia. Il governo degli algoritmi e dell'intelligenza artificiale*, Trieste, 2023.

e della Corte di cassazione, che verranno prese in considerazione, evidenziandone le incertezze. In conclusione, si formuleranno alcune osservazioni relative ai principali risultati di questo studio e alla possibilità di considerare la trasparenza algoritmica un principio di rango costituzionale.

2. Lo sfuggente concetto di trasparenza algoritmica

L'algoritmo è una procedura sistematica che produce risposte a interrogativi o soluzioni a problemi attraverso un numero finito di passaggi²³; ciò significa che l'algoritmo ha natura procedurale, consistendo in una sequenza di *input*, trasformazione e *output*. Mentre gli algoritmi più semplici, che possiamo definire *rule-based*, si basano su istruzioni predefinite, seguono una logica consequenziale a partire da premesse date e non sono in grado di elaborare nuove informazioni o gestire problemi imprevedibili, i più evoluti sistemi di *machine-learning* (MLS) sono invece strutturati per migliorare costantemente in base all'esperienza: possono valutare le correlazioni esistenti fra una moltitudine di variabili e risultati per arrivare a creare modelli predittivi in grado di stimare le probabilità di accadimento di eventi o comportamenti futuri. Tanto i MLS *supervised* – cioè i modelli di apprendimento costruiti sulla base di dati di addestramento etichettati, in cui gli esempi di input sono associati a output noti, in modo che l'algoritmo possa apprendere le relazioni tra *input* e *output*²⁴ – sia quelli *unsupervised* – cioè i modelli non basati su dati etichettati, ma lasciati liberi di analizzare i dati in modo autonomo per identificare *pattern* e relazioni²⁵ – sono progettati da creatori umani e, di conseguenza, non sono strumenti tecnologici neutrali, ma riflettono inevitabilmente le preferenze, le priorità e i pregiudizi che appartengono consciamente o inconsciamente ai loro progettisti; per questa loro caratteristica, i MLS possono essere strumenti di propagazione e perpetuazione di *bias* cognitivi, ingiustizie, disuguaglianze e distorsioni²⁶.

Ciò considerato, il concetto di trasparenza algoritmica (AT), che può essere definito come «*disclosure of information about algorithms to enable monitoring, checking, criticism, or intervention by interested parties*»²⁷, è inevitabilmente legato al funzionamento interno dell'algoritmo. Bisogna tuttavia distinguere fra una nozione di AT come “azione” – cioè il fatto di fornire informazioni sull'algoritmo ai suoi utilizzatori o alle terze parti interessate, tenendo presente che il “grado di disvelamento” dipende dall'attitudine dei proprietari o degli sviluppatori dell'algoritmo stesso in termini economici, etici e tecnici – e come “percezione” – ovvero l'impatto delle informazioni concernenti l'algoritmo sui destinatari, che è influenzato dalla metodologia usata nella spiegazione nonché dalle caratteristiche e dalla pregressa esperienza dei destinatari: può esistere, infatti, una di-

²³ *Algorithm*, in *britannica.com*.

²⁴ *Cosa è l'apprendimento supervisionato?*, in *ibm.com*.

²⁵ *Cos'è l'unsupervised learning?*, in *ibm.com*.

²⁶ N. Kossow – S. Windwehr – M. Jenkins, *Algorithmic Transparency and Accountability*, 2021, in *knowledgehub.transparency.org*, 5 febbraio 2021, 5 ss.

²⁷ N. Diakopoulos – M. Koliska, *Algorithmic Transparency in The News Media*, in *Digital Journalism*, 5, 2017, 809 ss., spec. 811.

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

screpanza fra la comprensione reale del funzionamento dell'algoritmo e quella semplicemente percepita, anche se gli studiosi divergono su quali siano i fattori più rilevanti alla base di tale discrepanza²⁸.

L'intrinseca opacità algoritmica e la crescente affermazione dei sistemi di intelligenza artificiale portano a focalizzare l'attenzione sul concetto di "spiegabilità algoritmica" (xAT)²⁹ come antidoto al *black box problem*³⁰, cioè al fatto che la realizzazione in concreto della trasparenza algoritmica si rivela estremamente difficoltosa in presenza di sistemi algoritmici di *machine learning* che, come le "scatole nere", non permettono ai programmatori, amministratori o utilizzatori di comprenderne i processi di funzionamento e il peso specifico dei diversi parametri elaborati, tanto da rendere impossibile capire le ragioni che hanno condotto il sistema a produrre un determinato risultato. In altre parole, «*a black box predictor is a data-mining and machine-learning obscure model, whose internals are either unknown to the observer or they are known but uninterpretable by humans*»³¹. Quindi, quando si sostiene che il rispetto del principio di trasparenza «si traduce nella garanzia di tracciabilità dei dati e dei procedimenti

che portano i sistemi intelligenti all'adozione della decisione finale»³² e che «tutte le informazioni relative al metodo di raccolta dei dati, alla loro classificazione, al tipo di algoritmo impiegato e i risultati ottenuti devono essere completamente documentate, cosicché tutti coloro che vi abbiano diritto o interesse possano comprendere le modalità di impiego dei dati esaminati, seguirne il flusso e verificare che la decisione finale sia stata elaborata e adottata correttamente»³³, bisogna aver ben presente il fatto che la tracciabilità dei dati è ostacolata dalla produzione e dall'utilizzo di dati inferiti (cioè di dati ottenuti attraverso processi di elaborazione secondaria di dati di partenza) e la comprensibilità dei procedimenti seguiti è impossibile nel caso dei sistemi *black box*.

Se esaminata dal punto di vista della progettazione dell'algoritmo, l'opacità può essere determinata intenzionalmente, allo scopo di prevenire o impedire infiltrazioni abusive nei sistemi o proteggere i diritti di proprietà intellettuale e i segreti industriali, ma può anche essere semplicemente una conseguenza, non voluta ma spesso inevitabile, della complessità tecnologica del sistema³⁴. Se considerata, invece, con riguardo ai destinatari delle decisioni algoritmiche o agli utilizzatori dei sistemi algoritmici, l'opacità può essere di tipo relazionale – quando è connessa alle diverse possibilità per diverse categorie di persone di avere accesso alle attività di elaborazione di dati che le riguardano

²⁸ T. Bitzer – M. Wiener – W. A. Cram, *Algorithmic Transparency: Concepts, Antecedents, and Consequences. A Review and Research Framework*, in *Communications of the Association for Information Systems*, 52, 2023, 293 ss., spec. 304-306 e 313-316.

²⁹ La sigla corrisponde a *algorithmic explainability*.

³⁰ T. Bitzer e al., *Algorithmic Transparency*, cit., 297.

³¹ R. Guidotti e al., *A Survey of Methods for Explaining Black Box Models*, in *ACM Computing Surveys*, 5, 2019, 1 ss., spec. 5.

³² M. Fasan, *I principi costituzionali nella disciplina dell'Intelligenza Artificiale. Nuove prospettive interpretative*, in *DPCE Online*, 1, 2022, 181 ss., spec. 191.

³³ *Ibid.*

³⁴ M. Grochowsky – A. Jablonowska – F. Lagioia – G. Sartor, *Algorithmic Transparency and Explainability for EU Consumer Protection: Unwrapping the Regulatory Premises*, in *Critical Analysis of Law*, 1, 2021, 43 ss., spec. 48.

– o di tipo tecnologico, se riferita al livello di comprensibilità del funzionamento del sistema da parte degli esperti che hanno pieno accesso ad esso³⁵. Occorre allora avere ben presente quale tipo di opacità caratterizza il sistema considerato e, in relazione ad essa, quale tipo di trasparenza occorre raggiungere³⁶. Nel caso di sistemi che presentano un’opacità di tipo relazionale, occorre realizzare la cosiddetta *exoteric transparency*, in modo da permettere alle persone che interagiscono con essi di ricevere e comprendere le informazioni “significative” rispetto ai loro interessi, acquisendo consapevolezza di quale impatto il sistema algoritmo ha avuto o potrà avere nella loro specifica situazione. La *exoteric transparency* non è legata solo ad aspetti tecnologici, ma dipende in larga misura da fattori comunicativi e dialettici finalizzati a far sì che le informazioni siano opportunamente adeguate alle caratteristiche dei loro destinatari³⁷ e, pertanto, tende a coincidere con la nozione di spiegabilità. Invece, neutralizzando l’opacità di tipo tecnologico – cioè, permettendo agli esperti di arrivare a conoscere i meccanismi di funzionamento interno del sistema – si ottiene la cosiddetta *esoteric transparency*, che può essere raggiunta tramite soluzioni tecniche (ad esempio, le metodologie di *reverse engineering* o di *transparent box design*)³⁸, certamente utili per gli “addetti ai lavori”, ma non per gli utenti comuni.

È stato giustamente osservato che uno dei fattori che rendono difficile la realizzazione in concreto della xAT è quello linguistico: «Attualmente, durante la fase di *input*, il programmatore deve spiegare in linguaggio naturale semplificato il risultato desiderato, elaborare il progetto utilizzando tale linguaggio e successivamente tradurlo in codice. Durante la fase di *output*, invece, è necessario tradurre il codice generato dalla macchina per ricostruire il processo logico e affrontare questioni di responsabilità e controllo umano *ex post*»³⁹. A ciò si aggiunge il fatto che alcune lingue – l’inglese in particolare – si prestano meglio di altre tanto nella programmazione algoritmica quanto nella traduzione del linguaggio algoritmico in quello naturale. Ne deriva che, poiché i sistemi di IA dispongono di un maggior numero di dati originariamente prodotti in lingua inglese rispetto alle altre lingue e li elaborano ed interpretano tenendo conto prevalentemente dei costrutti logico-sintattici della lingua inglese, possono produrre risultati viziati da *bias* di matrice linguistica.

Analogamente, va evidenziata la scarsa flessibilità del linguaggio algoritmico, che può essere modificato solo tramite processi intenzionali di sviluppo e programmazione, rispetto alle lingue naturali. Infatti, «il linguaggio algoritmico è basato su una struttura logica e sintattica chiusa, fondata sulla matematica, mentre le lingue naturali sono intrinsecamente complesse, ricche di sfumature, ambiguità, contestualizzazioni e contaminazioni [...] in grado di esprimere una vasta gamma di concetti complessi e astratti (si pensi alla metafora) che possono essere molto sfidanti da rappresentare con precisione nel linguaggio algoritmico: la comunicazione di emozioni, sentimenti, intuizioni

³⁵ *Ibid.*

³⁶ M. Grochowsky e al., *Algorithmic Transparency and Explainability*, cit., 54.

³⁷ *Ivi*, 55-57.

³⁸ Alcune di queste soluzioni tecniche sono descritte in R. Guidotti e al., *A Survey of Methods*, cit.

³⁹ G. Caravaggion – M. Orofino, *Lingua e costituzione: l’irrompere dei linguaggi algoritmici*, in *Rivista AIC*, 4, 2023, 155 ss., spec. 181.

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

e concetti filosofici richiede una ricchezza e una flessibilità che il linguaggio algoritmico non è in grado, nel suo stato attuale, di fornire. Inoltre, le lingue naturali si evolvono nel tempo, si adattano piuttosto rapidamente ai cambiamenti culturali e si contestualizzano all'interno delle comunità linguistiche. Il linguaggio algoritmico, al contrario, è soggetto a standard e regole specifiche che possono essere modificati solo attraverso processi di sviluppo e aggiornamento»⁴⁰.

Nonostante queste oggettive difficoltà, sono stati compiuti alcuni interessanti tentativi di definire alcuni standard di trasparenza algoritmica, intesa soprattutto come spiegabilità (xAT), attraverso i quali si mira a raggiungere un più elevato livello di *fairness* (giustizia, equità, correttezza) nell'utilizzo degli algoritmi. Fra questi, i sette *Principles for Algorithmic Transparency and Accountability*, elaborati nel 2017 dalla *Association for Computing Machinery* (ACM)⁴¹: si tratta sostanzialmente di obiettivi che gli sviluppatori di algoritmi e gli organismi pubblici e privati che se ne servono dovrebbero tener presente tanto nella progettazione quanto nell'utilizzo dei *software* algoritmici, che però non sono corredati da indicazioni su come questi obiettivi possano essere realizzati in concreto. Occorre poi menzionare le *Algo Rules*, un elenco di nove principi elaborati e pubblicati nel 2019 dal *think tank* tedesco Bertelsmann Stiftung⁴²: anche in questo caso, i principi indicati non sono accompagnati dalla definizione di accorgimenti da utilizzare in pratica per poterli realizzare; in particolare, poi, la regola n. 7 – relativa alla necessità di poter sempre mantenere il controllo umano su ogni fase del processo algoritmico, astenendosi dall'utilizzare i sistemi di *machine learning* la cui complessità lo impedisce di fatto – appare di fatto un po' ingenua, considerando la velocità con cui il progresso tecnologico raggiunge frontiere sempre nuove, inimmaginabili fino a poco tempo pri-

⁴⁰ Ivi, 183.

⁴¹ Statement on Algorithmic Transparency and Accountability by ACM U.S. Public Policy Council, approved January 12, 2017 ACM Europe Policy Committee, approved May 25, 2017. Questi principi consistono in: 1) consapevolezza (*awareness*) da parte degli utenti sulle modalità di *data processing* e il livello di automazione delle decisioni; 2) accesso e rimedio (*access* e *redress*), cioè la possibilità di individuare e correggere le decisioni errate; 3) responsabilità (*accountability*) da parte dei produttori e degli sviluppatori dei *software* algoritmici per le conseguenze delle decisioni algoritmiche; 4) spiegazione (*explanation*) della logica dell'algoritmo nel linguaggio umano; 5) consapevolezza della provenienza e dell'attendibilità dei dati su cui l'algoritmo si basa (*data provenance*); 6) verificabilità (*auditability*) dei processi seguiti nello sviluppo e nell'utilizzo degli algoritmi, in modo che possano eventualmente essere sottoposti a revisione; 7) accertamento che i sistemi automatizzati funzionino come previsto (*validation* e *testing*).

⁴² Bertelsmann Stiftung (ed.), *Algo.Rules Rules for the Design of Algorithmic Systems*, 7 March 2019. I nove principi consistono in: 1) rafforzamento della competenza degli sviluppatori e degli utilizzatori di sistemi algoritmici relativamente alle loro funzioni tecniche e ai loro effetti potenziali; 2) chiara definizione delle responsabilità associate a ciascun ruolo nella catena di progettazione e utilizzazione algoritmica; 3) chiara definizione *ex ante* e pubblicità delle finalità di ciascun sistema algoritmico, nonché periodiche valutazioni del suo effettivo impatto (*impact assesment*); 4) garanzia della capacità di resistenza del sistema algoritmico rispetto ad attacchi informatici, accessi non autorizzati e manipolazioni, da comprovare mediante appositi test; 5) *labelling*, cioè possibilità per i destinatari finali di poter chiaramente identificare la matrice algoritmica di determinate decisioni o risultati; 6) pubblicità delle informazioni concernenti l'algoritmo in termini facilmente comprensibili, in modo che i destinatari finali possano capire gli effetti diretti e indiretti del sistema algoritmico; 7) possibilità di mantenere sempre il controllo umano sull'intero processo algoritmico e non utilizzo di sistemi di *machine learning* per i quali il controllo umano sia impossibile; 8) monitoraggio costante sui risultati e sugli effetti prodotti dall'algoritmo e pronto intervento correttivo ove necessario; 9) predisposizione di meccanismi di ricorso contro le decisioni algoritmiche, efficaci e di facile utilizzo.

ma. Qualche passo in avanti è stato compiuto sempre dalla Bertelsmann Stiftung nel 2020, attraverso l'elaborazione di un documento intitolato *From Principles to Practice. An Interdisciplinary Framework to Operationalise AI Ethics*⁴³, con il quale si è tentato di pervenire alla definizione di indicatori osservabili, che rendano in qualche modo misurabile il raggiungimento di standard etici (*transparency, accountability, privacy, justice, reliability, environmental sustainability*) da rispettare nella progettazione e nell'utilizzo di algoritmi. In riferimento alla questione della trasparenza, ad esempio, il documento contiene alcune domande riguardanti l'origine dei *data set* e l'adeguatezza della *disclosure*⁴⁴, mentre altri interrogativi riguardano l'affidabilità dei modelli algoritmici utilizzati⁴⁵ oppure l'interpretabilità e comprensibilità dei modelli algoritmici per i *target groups* di destinatari⁴⁶. Va segnalata, da ultimo, l'attività dell'ISO⁴⁷, che a luglio 2024 ha predisposto una serie di standard tecnici in materia di trasparenza di sistemi di IA⁴⁸, la cui pubblicazione è prevista per novembre 2025.

3. Il lessico della trasparenza algoritmica nella *digital regulation* dell'UE: nodi irrisolti

La questione dell'implementazione pratica del principio della trasparenza algoritmica, *mantra* ricorrente nella recente produzione legislativa dell'UE in ambito digitale (a cominciare dal regolamento sulla protezione dei dati personali, per arrivare ai più recenti regolamenti sui servizi digitali e sull'intelligenza artificiale), è un problema non ancora risolto. È evidente, infatti, la tendenza ad una certa vaghezza e indeterminatezza lessicale nel dettato normativo, che nasconde l'incertezza di fondo sull'esatto significato da attribuire al concetto di AI. Tale incertezza, come si vedrà anche nel paragrafo seguente, rende difficile valutare la liceità e la correttezza dei provvedimenti amministrativi basati sull'utilizzo di algoritmi. Preliminarmente, quindi, è utile esaminare nel dettaglio le norme vigenti sulla trasparenza algoritmica, che costituiscono – e, nel caso di quelle relative ai sistemi di IA, costituiranno in futuro – la base sui cui si fonda la giurisprudenza che verrà più avanti considerata.

La trasparenza come principio applicabile al trattamento dei dati personali è menziona-

⁴³ Bertelsmann Stiftung (ed.), *From principles to practice: How can we make AI ethics measurable?*, 2 April 2020.

⁴⁴ *Is the data's origin documented? Is it plausible for each purpose, which data is being used? Are the training data set's characteristics documented and disclosed? Are the corresponding data sheets comprehensive?*

⁴⁵ *Has the model in question been tested and used before? Is it possible to inspect the model so far that potential weaknesses can be discovered? Taking into account efficiency and accuracy, has the simplest and most intelligible model been used?*

⁴⁶ *Are the modes of interpretability target-group-specific and have been developed with the target groups? Who has access to information about data sets and the algorithm/model used? Is the operating principle comprehensible and interpretable? Are the modes of interpretability in their target-group-specific form intelligible for the target groups? Are the hyperparameters (parameters of learning methods) accessible? Has a mediating authority been established to settle and regulate transparency conflicts?*

⁴⁷ L'ISO è un'organizzazione internazionale non governativa indipendente, che raggruppa gli organismi di standardizzazione di 164 paesi del mondo ed elabora standard internazionali che contribuiscono a migliorare la qualità e la sicurezza di beni e servizi e, non da ultimo, a facilitarne lo scambio.

⁴⁸ [ISO/IEC DIS 12792:2024](#).

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

ta, insieme alla liceità e alla correttezza, nell'art. 5, c. 1, lett. a) del GDPR⁴⁹. Ulteriori riferimenti, soprattutto in relazione alla trasparenza come spiegabilità, si trovano nell'art. 12, c. 1, a proposito delle informazioni sul trattamento dei dati da fornire all'interessato «in forma concisa, trasparente, intelligibile e facilmente accessibile, con un linguaggio semplice e chiaro, in particolare nel caso di informazioni destinate specificamente ai minori», nonché negli artt. 13 (c. 2, lett. f), 14 (c. 2, lett. g) e 15 (c. 1, lett. h), che attribuiscono all'interessato il diritto di ricevere «informazioni significative sulla logica utilizzata» nel trattamento dei dati. Eppure, come si è visto nei paragrafi precedenti, i modelli algoritmici di tipo *black box* possono produrre risultati viziati da pregiudizi e da errori logici non identificabili e non spiegabili, impedendo inoltre di individuare i soggetti su cui dovrebbe ricadere la responsabilità di risultati distorti o pregiudizievoli. Allora, anche in un'ottica di legittimità dell'azione amministrativa, occorre trovare una modalità di applicazione pratica del principio contenuto nell'art. 14, c. 2, lett. g), del GDPR, relativo al diritto dell'interessato di ottenere, in caso processi decisionali automatizzati, almeno le «informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato»: il dettato normativo, infatti, presuppone la consapevolezza di quali informazioni sono da considerarsi “significative”, nonostante finora la ricerca in quest'ambito riferita ai sistemi di IA abbia prodotto risultati frammentari e limitati⁵⁰.

Il dibattito sull'esistenza e sulla portata di un diritto alla spiegazione degli algoritmi e sul ruolo che dovrebbe (o potrebbe) spettare agli operatori umani nell'esercizio di questo diritto è stimolato anche da una certa ambiguità nella formulazione dell'art. 22 GDPR, che sancisce il diritto a non essere sottoposti a decisioni esclusivamente automatizzate e, comunque, ad essere messi in condizione di poter contestare tale decisione⁵¹. Mentre il c. 3 dell'art. 22 assicura all'interessato, nel caso di processi automatizzati, «almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione», i c. 1 e 2 del medesimo articolo sono finalizzati ad evitare che le macchine prendano il sopravvento sulla mente umana, ponendo alcune limitazioni all'utilizzo di algoritmi⁵²: l'interessato ha il diritto di non essere sottoposto a decisioni basate unicamente sul trattamento automatizzato dei dati, salvo che non sia necessario per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento o sia autorizzato dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento o si basi sul consenso esplicito dell'interessato.

Occorre però chiarire che l'ambito di applicazione degli artt. 15 e 22 del GDPR è diverso da quello degli artt. 13 e 14: questi ultimi si riferiscono al momento che precede l'inizio delle operazioni di *data processing* (trasparenza *ex ante*), mentre gli artt. 15 e 22

⁴⁹ *General Data Protection Regulation* (regolamento (UE) 2016/679 del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati).

⁵⁰ J. Lu e al., *Good Explanation for Algorithmic Transparency*, 2019, in *papers.ssrn.com*, 11 novembre 2019.

⁵¹ E. Spiller, *Il diritto di comprendere, il dovere di spiegare. Explainability e intelligenza artificiale costituzionalmente orientata*, in *BioLaw Journal*, 2, 2021, 419 ss., spec. 423.

⁵² G. De Minico, *Fundamental Rights, European Digital Regulation and Algorithmic Challenge*, in questa *Rivista*, 1, 2021, 9 ss., spec. 28.

alla fase in cui il trattamento è in corso, durante la quale l'interessato potrebbe avere necessità di ricevere ulteriori informazioni, come ad esempio quelle relative ai dati inferiti (generati attraverso l'elaborazione algoritmica) o al funzionamento degli algoritmi utilizzati nel trattamento⁵³. Dunque, il diritto dell'interessato di ottenere spiegazioni *ex post* concernenti una decisione algoritmica, che sarebbero invece necessarie per poterla eventualmente contestare, non è previsto da nessuna disposizione del GDPR, ma è relegato solo nel Considerando n. 71, che non è una norma giuridicamente vincolante: in assenza di tali spiegazioni, l'interessato non avrebbe la possibilità di esercitare un controllo sull'utilizzo dell'algoritmo. Nel caso di decisioni algoritmiche assunte da poteri pubblici, ci troveremmo davanti a una versione "modernizzata" e tecnologicamente avanzata di oppressione dei diritti e delle libertà individuali, analoga a quelle di cui è costellata la nostra storia⁵⁴.

Dunque, il problema del GDPR consiste proprio nel fatto di non prevedere diritti e obblighi in relazione a un livello minimo di requisiti di spiegabilità delle tecnologie algoritmiche⁵⁵: le disposizioni del GDPR sono focalizzate sui diritti dell'interessato (e sui connessi doveri del titolare del trattamento) e non contengono obblighi di *algorithmic disclosure* nei confronti di terzi interessati o della collettività in generale, che sarebbero invece necessarie per guidare l'azione amministrativa e, più in generale, quella di qualsiasi ente anche privato che assume decisioni algoritmiche. Del resto, il GDPR è un regolamento risalente al 2016, quindi piuttosto datato, se si considera la velocità del progresso delle tecnologie digitali e della loro diffusione.

Più recentemente il regolamento europeo sui servizi digitali (DSA)⁵⁶, in vigore dal 17 febbraio 2024, ha compiuto qualche progresso in termini di requisiti di trasparenza anche nei confronti di autorità pubbliche, revisori esperti indipendenti, ricercatori affiliati a organismi di ricerca, singoli utenti e "pubblico" in generale (in quest'ultimo caso, principalmente nella forma di *transparency reports* da pubblicare *ex post*, come indicato agli artt. 15 e 24). È particolarmente interessante l'art. 27, che postula la trasparenza *ex ante* relativa ai sistemi di raccomandazione, i cui «principali parametri utilizzati», nonché qualunque opzione a disposizione dei destinatari del servizio che consente loro di modificare o influenzare tali parametri principali» devono essere specificati «in un linguaggio chiaro e intellegibile». Non è affatto chiaro, tuttavia, in cosa consistano tali parametri, salvo il fatto che, ai sensi del c. 2, essi devono comprendere, come elementi minimi, i criteri più significativi per determinare le informazioni suggerite al destinatario del servizio e le ragioni per l'importanza relativa dei parametri stessi. Il Considerando n. 70 non fornisce ulteriori elementi interpretativi. Si evince, in ogni caso, il *favor* per la trasparenza *by design*, come elemento costitutivo della progettazione tecnologica. Un altro elemento importante in termini di *transparency by design* è costituito dal divie-

⁵³ C. Colapietro, *Gli algoritmi tra trasparenza e protezione dei dati personali*, in *Federalismi.it*, 5, 2023, 151 ss., spec.160.

⁵⁴ G. De Minico, *Fundamental Rights*, cit., 31.

⁵⁵ E. Spiller, *Il diritto di comprendere*, cit., 425.

⁵⁶ Regolamento (UE) 2022/2065 del 19 ottobre 2022, relativo a un mercato unico dei servizi digitali, noto come *Digital Services Act* (DSA). Sul regolamento in generale si veda, *ex multis*, M. Orofino, *Il Digital Service Act tra continuità (solo apparente) e innovazione*, in F. Pizzetti (a cura di), *La regolazione europea della società digitale*, Torino, 2024, 83 ss.

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

to, espresso nell'art. 25 DSA, di *dark patterns*, ovvero di interfacce *online* progettate e realizzate in modo da «ingannare o manipolare i destinatari dei loro servizi o da materialmente falsare o compromettere altrimenti la capacità dei destinatari dei loro servizi di prendere decisioni libere e informate». Si tratta però di una prescrizione da implementare, in concreto, sulla base di orientamenti che la Commissione europea dovrà produrre. Inoltre, dalla lettura del Considerando n. 67 si desume che questo tipo di trasparenza non ha molto a che fare con quella algoritmica, quanto piuttosto con la struttura, la progettazione e le funzionalità delle interfacce, che non devono essere ingannevoli né programmate per indirizzare le scelte individuali (*nudging*).

Infine, secondo l'art. 26 DSA, c. 1, lett. d), le piattaforme digitali che presentano pubblicità nelle loro interfacce devono fornire «informazioni rilevanti direttamente e facilmente accessibili dalla pubblicità relative ai parametri utilizzati per determinare il destinatario al quale viene presentata la pubblicità e, laddove applicabile, alle modalità di modifica di detti parametri». Nemmeno qui viene esplicitata la natura di tali parametri, ma il Considerando n. 68 precisa che i destinatari dei servizi devono ottenere «spiegazioni rilevanti sulla logica seguita» nella presentazione dei messaggi pubblicitari e sui principali criteri di profilazione utilizzati, nonché informazioni su come poter modificare tali criteri. Inoltre, il Considerando n. 69 mette in guardia sui rischi della profilazione (*targeting*) con finalità pubblicitarie, per via dei quali il c. 3 dell'art. 26 vieta di utilizzare per la profilazione le categorie speciali di dati personali di cui all'articolo 9, c. 1, del GDPR.

Sia il GDPR sia il DSA si basano su un approccio *risk-based* all'utilizzo di strumenti tecnologici⁵⁷. Il medesimo approccio è seguito dal recente regolamento europeo sull'intelligenza artificiale (*AI Act*)⁵⁸, che si propone di istituire un quadro giuridico uniforme per quanto riguarda lo sviluppo, l'immissione sul mercato, la messa in servizio e l'uso di sistemi di intelligenza artificiale nell'UE, in conformità ai valori dell'Unione, promuovendo la diffusione di un'intelligenza artificiale antropocentrica e affidabile, e 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 (Considerando n. 1 e art. 1)⁵⁹. Dunque, data la necessità di garantire la protezione dei dati personali e i diritti di proprietà intellettuale e industriale, eventuali obblighi di *algorithmic disclosure* previsti *ex lege* dovranno consentire «di sottoporre al requisito della trasparenza non l'intero algoritmo, ma solo le informazioni necessarie per conoscere e comprendere i motivi (quindi il procedimento “logico” utilizzato) sottesi alla decisione stessa»⁶⁰. In aggiunta a ciò, nel caso di sistemi di IA ad alto rischio⁶¹ potranno essere

⁵⁷ E. Longo, *La disciplina del “rischio digitale”* in F. Pizzetti (a cura di), *La regolazione europea*, cit., 53 ss.

⁵⁸ Regolamento (UE) 2024/1689 del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale.

⁵⁹ Sul regolamento in generale, *ex multis*: C. Casonato – B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 3, 2021, 415 ss.

⁶⁰ C. Colapietro, *Gli algoritmi*, cit., 172.

⁶¹ Gli artt. da 6 a 49 del regolamento, contenuti nel Capo III, si riferiscono ai sistemi di IA ad alto rischio, cioè quelli elencati nell'Allegato III, che ai sensi dell'art. 6 presentano rischi significativi di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, anche nel senso di influenzare i processi decisionali; sono sempre considerati ad alto rischio i sistemi che comportano la

previsti ulteriori obblighi di trasparenza in favore di pubbliche autorità e terze parti indipendenti coinvolte nel processo di valutazione tanto dei rischi quanto della conformità degli strumenti tecnologici ai requisiti previsti dalla normativa⁶².

Il Considerando n. 27 dell'AI Act richiama esplicitamente gli Orientamenti etici per un'IA affidabile, elaborati dal Gruppo di esperti ad alto livello sull'intelligenza artificiale (HLEG) istituito dalla Commissione europea nel 2018 e pubblicati nel 2019⁶³. In tali orientamenti il Gruppo ha elaborato sette principi etici non vincolanti (intervento e sorveglianza umani, robustezza tecnica e sicurezza, vita privata e *governance* dei dati, trasparenza, diversità, non discriminazione ed equità, benessere sociale e ambientale e responsabilità), volti a garantire che l'IA si sviluppi in modo affidabile e antropocentrico. In particolare, il requisito della trasparenza è declinato dal punto di vista della tracciabilità («i set di dati e i processi che determinano la decisione del sistema di IA, compresi quelli di raccolta ed etichettatura dei dati, come pure gli algoritmi utilizzati, dovrebbero essere documentati secondo i migliori standard»), della spiegabilità⁶⁴ e della comunicazione (i sistemi di IA devono essere identificabili come tali, deve essere preferita l'interazione umana a quella con il sistema e devono essere comunicate agli operatori del settore dell'IA o agli utenti finali le capacità e le limitazioni del sistema). Tuttavia, soprattutto in relazione al requisito della spiegabilità, non è affatto chiaro in che modo i processi tecnologici debbano e possano essere spiegati in modo che siano comprensibili per i destinatari delle decisioni e soprattutto in base a quali parametri si possa valutare l'adeguatezza della spiegazione fornita agli standard richiesti. Questo, come si è più volte evidenziato, rende difficile valutare la legittimità dell'azione amministrativa in caso di utilizzo di IA. Eppure, soprattutto nel caso di sistemi di AI ad alto rischio, i Considerando n. 71 e 72 evidenziano che disporre di informazioni comprensibili sulle modalità di e di funzionamento di tali sistemi è essenziale per consentirne la tracciabilità, verificarne la conformità ai requisiti normativi e svolgere il monitoraggio

profilazione delle persone. Ai sensi dell'art. 7, spetta alla Commissione europea il compito di modificare e integrare progressivamente l'Allegato III. Invece, gli artt. 50-56, contenuti nel capo IV e nel Capo V del regolamento, si riferiscono a sistemi di AI diversi da quelli ad altro rischio, ma comunque non privi di un certo livello di rischio.

⁶² A tale proposito, il Considerando n. 68 prevede che «ai fini dello sviluppo e della valutazione di sistemi di IA ad alto rischio, è opportuno concedere ad alcuni soggetti, come fornitori, organismi notificati e altre entità pertinenti, quali i poli europei dell'innovazione digitale, gli impianti di prova e sperimentazione e i ricercatori, l'accesso a set di dati di elevata qualità e la possibilità di utilizzarli nell'ambito dei settori di attività di tali attori soggetti al presente regolamento».

⁶³ Orientamenti etici per un'IA affidabile.

⁶⁴ Secondo il par. 77 del documento (a p. 20), «la spiegabilità attiene alla capacità di spiegare sia i processi tecnici di un sistema di IA che le relative decisioni umane (ad esempio i settori di applicazione di un sistema di IA). Affinché un sistema di IA possa essere tecnicamente spiegabile gli esseri umani devono poter capire e tenere traccia delle decisioni prese dal sistema stesso. Potrebbe inoltre essere necessario trovare un compromesso tra il miglioramento della spiegabilità di un sistema (sacrificando la precisione) e l'aumento della precisione (a scapito della spiegabilità). Se un sistema di IA influisce considerevolmente sulla vita delle persone, dovrebbe sempre essere possibile richiedere una spiegazione adeguata del processo decisionale del sistema. Tale spiegazione dovrebbe essere tempestiva e adeguata alle competenze del portatore di interesse in questione (un non esperto, un'autorità di regolamentazione o un ricercatore). Dovrebbero inoltre essere disponibili indicazioni sul grado in cui un sistema di IA influenza e plasma il processo decisionale organizzativo, sulle scelte progettuali del sistema e sulla logica alla base della sua distribuzione (garantendo così la trasparenza del modello di business)».

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

successivo all'immissione sul mercato. Quindi, i sistemi di IA ad alto rischio dovrebbero essere accompagnati da informazioni adeguate sotto forma di istruzioni per l'uso rivolte ai *deployer*⁶⁵ – soggetti privati, ma oggi spesso anche pubblici – concernenti le caratteristiche, le capacità e i limiti delle prestazioni del sistema, i rischi possibili e le pertinenti misure di sorveglianza umana, anche attraverso esempi illustrativi. Per questo «i fornitori dovrebbero garantire che tutta la documentazione, comprese le istruzioni per l'uso, contenga informazioni significative, complete, accessibili e comprensibili, tenendo conto delle esigenze e delle conoscenze prevedibili dei *deployer* destinatari. Le istruzioni per l'uso dovrebbero essere messe a disposizione in una lingua che possa essere compresa facilmente dai *deployer* destinatari, secondo quanto stabilito dallo Stato membro interessato». Ora, a prescindere dal fatto che i Considerando non hanno efficacia normativa, ma solo valore interpretativo, nemmeno la loro verbosità chiarisce adeguatamente in che modo sia in concreto possibile descrivere e spiegare parametri tecnici nel linguaggio naturale, quale livello di chiarezza e precisione nella spiegazione sia necessario e come sia possibile valutarne il raggiungimento.

Venendo al dettato normativo dell'AI Act, l'art. 11 prevede che i sistemi di intelligenza artificiale ad alto rischio debbano essere corredati di una documentazione tecnica da fornire alle autorità nazionali competenti e agli organismi notificati, che deve contenere almeno gli elementi che figurano nell'Allegato IV al regolamento stesso. Si tratta di informazioni, però, non destinate ai *deployer* o agli utilizzatori finali di tali sistemi, ma agli organi nazionali di regolamentazione e controllo. Effettivamente l'Allegato IV è piuttosto preciso e specifico e include, fra le varie informazioni da fornire, anche una descrizione dettagliata degli elementi del sistema di IA e del processo relativo al suo sviluppo, della quale vengono puntualmente indicati i vari aspetti. Permane tuttavia la perplessità sull'effettiva possibilità di tradurre in modo efficace e davvero comprensibile le regole tecniche dal linguaggio computazionale al linguaggio naturale, in modo da renderle comprensibili a persone non necessariamente esperte del linguaggio matematico-informatico.

Per quanto riguarda, invece, l'utilizzo di sistemi algoritmici da parte di pubbliche amministrazioni, rileva particolarmente l'art. 13, concernente le istruzioni per l'uso rivolte ai *deployer*; queste ultime devono contenere «informazioni concise, complete, corrette e chiare che siano pertinenti, accessibili e comprensibili», concernenti vari aspetti del sistema di IA, fra cui: «il livello di accuratezza che ci si può attendere, comprese le metriche, di robustezza e cibersecurity»; «le capacità e caratteristiche tecniche del sistema di IA ad alto rischio connesse alla fornitura di informazioni pertinenti per spiegarne l'output»; «le specifiche per i dati di input o qualsiasi altra informazione pertinente in termini di set di dati di addestramento, convalida e prova»; «informazioni che consentano ai *deployer* di interpretare l'output del sistema di IA ad alto rischio e di usarlo in modo opportuno»; «una descrizione dei meccanismi inclusi nel sistema di IA ad alto rischio che consente ai *deployer* di raccogliere, conservare e interpretare correttamente i log». Effettivamente, questo elenco di informazioni che devono essere contenute nelle

⁶⁵ Secondo l'art. 3, c. 4, questo termine, per il quale non è stata prevista una traduzione in lingua italiana, si riferisce a «una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un 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».

istruzioni per l'uso rappresenta il risultato più avanzato raggiunto finora nello sforzo di dotare il principio di trasparenza algoritmica di indicazioni operative quanto più possibili esaurienti e puntuali. Tuttavia, a parte il fatto che questo “decalogo” si riferisce solo a sistemi di IA considerati ad alto rischio e non a qualsiasi sistema di IA, non si può non evidenziare una certa difficoltà di comprensione di almeno alcune fra varie specifiche richieste nonché, come già sottolineato più volte, una certa ambiguità di fondo: quale livello di accuratezza e precisione è richiesto nelle spiegazioni affinché esse possano essere davvero considerate pienamente comprensibili per i loro destinatari?

Peraltro, è davvero necessario che i *deployer* – e particolarmente i soggetti pubblici che rivestono questo ruolo – acquisiscano un elevato livello di conoscenza del funzionamento del sistema di IA ad alto rischio, visto che l'art. 86, c. 1, attribuisce loro una sorta di ruolo di “mediatori” fra i progettisti e i proprietari dei sistemi algoritmici e i destinatari finali delle decisioni. Questi ultimi, infatti, hanno diritto di ricevere proprio dai *deployer* «spiegazioni chiare e significative sul ruolo del sistema di IA nella procedura decisionale e sui principali elementi della decisione adottata», qualora tale decisione abbia un impatto negativo sulla loro salute e sicurezza o sull'esercizio dei loro diritti fondamentali. Anche qui, però, non è affatto chiaro quale livello di dettaglio e di “tecnicismo” debbano avere queste informazioni trasmesse dai *deployer* ai destinatari finali. Comunque, l'art. 96 del regolamento attribuisce alla Commissione europea il compito di elaborare, senza però stabilire alcuna scadenza, orientamenti sull'attuazione pratica delle disposizioni del regolamento.

Le perplessità fin qui espresse assumono particolare rilievo nel caso dei sistemi di IA destinati a interagire direttamente con le persone fisiche (art. 50, c. 1), i sistemi per finalità generali che generano contenuti testuali o audiovisivi (art. 50, c. 2), i sistemi di riconoscimento delle emozioni o di categorizzazione biometrica (art. 50, c. 3) e quelli che generano *deep fakes* manipolando contenuti testuali o audiovisivi (art. 50, c. 4). Si tratta di sistemi non sempre e non necessariamente qualificabili ad alto rischio, ma che possono esserlo, per lo meno nel caso dei sistemi di identificazione e categorizzazione biometrica e quelli di riconoscimento delle emozioni, elencati nel par. 1 dell'Allegato III. Il Considerando n. 29 è particolarmente esauriente nel presentare tutti i possibili rischi derivanti dalle tecniche di manipolazione basate sull'IA, che «possono essere utilizzate per persuadere le persone ad adottare comportamenti indesiderati o per indurle con l'inganno a prendere decisioni in modo da sovvertirne e pregiudicarne l'autonomia, il processo decisionale e la libera scelta»; ciò è particolarmente vero nel caso di sistemi di IA che impiegano componenti subliminali, che vanno al di là della percezione umana, oppure nel caso delle interfacce *brain-computer* o ancora in quello della realtà virtuale. Occorrerebbe quindi vietare l'utilizzo di questi sistemi nei casi in cui possano verosimilmente provocare danni significativi alle persone. Analogamente, il Considerando n. 44 esprime preoccupazione per i sistemi di IA volti a identificare o inferire emozioni o intenzioni a partire da dati biometrici, non solo per via della loro scarsa attendibilità, ma anche perché potrebbero essere causa di trattamenti discriminatori.

Eppure, nonostante la prolissità dei Considerando, l'art. 50 si limita a prevedere che occorre informare le persone fisiche del fatto che stanno interagendo con un sistema di IA, «a meno che ciò non risulti evidente dal punto di vista di una persona fisica

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

ragionevolmente informata, attenta e avveduta, tenendo conto delle circostanze e del contesto di utilizzo»; inoltre, nel caso di IA generativa, occorre far sì che gli *output* dei sistemi siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente, soprattutto nel caso di produzione di *deepfakes*; infine, nel caso di sistemi di riconoscimento delle emozioni o di categorizzazione biometrica, occorre che le persone fisiche che vi sono esposte vengano informate in merito al loro funzionamento. Tutte queste informazioni devono essere «fornite alle persone fisiche interessate in maniera chiara e distinguibile al più tardi al momento della prima interazione o esposizione» (art. 50, c. 5).

Colpisce davvero la stringatezza del dettato normativo, concentrato sul generico obbligo di fornire informazioni, in confronto alla complessità dei possibili rischi evidenziati nei Considerando, senza considerare l'indeterminatezza del parametro relativo al punto di vista della persona ragionevolmente informata e avveduta e di quello relativo alla chiarezza e comprensibilità delle informazioni. Ci si chiede se le caratteristiche delle informazioni da fornire agli utenti di questi sistemi di IA particolarmente rischiosi debbano essere analoghe a quelle che l'Allegato IV e l'art. 13 del regolamento del regolamento hanno previsto rispettivamente per le autorità competenti e per i *deployer* oppure se debbano essere diverse, considerando che dovranno rivolgersi a persone comuni, probabilmente prive di conoscenze specifiche. In ogni caso, il regolamento sui sistemi di IA non lo specifica e tale omissione rappresenta certamente un *vulnus* in relazione alla garanzia dei diritti e delle libertà fondamentali degli individui.

Gli articoli da 51 a 56 del regolamento sulla AI riguardano i modelli di IA per finalità generali⁶⁶, definiti dall'art. 3, c. 63, come «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 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». Questi modelli – che tendenzialmente non sono inclusi fra quelli ad alto rischio, ma che talvolta possono comportare rischi sistemici nei casi di elevata capacità di impatto, secondo i criteri indicati nell'Allegato XIII – devono essere corredati da una documentazione tecnica (secondo le specifiche indicate nell'Allegato XI) e da informazioni sulla trasparenza, menzionate nell'art. 53, par. 1, lett. b e meglio specificate nell'Allegato XII. Vale anche in questo caso tutto ciò che è stato già scritto sopra a proposito dei requisiti indicati nell'Allegato IV: da un lato, si apprezza lo sforzo di declinare il principio della trasparenza algoritmica in termini pratico-operativi; dall'altro, però, non si può fare a meno di evidenziare l'ambiguità di alcune espressioni che figurano nell'Allegato.

Non va sottaciuto, infine, che l'art. 14 del regolamento sulla IA prevede il principio di sorveglianza umana per tutti i sistemi di IA ad altro rischio (*human in the loop* o HITL)⁶⁷, rinforzando così quanto già previsto dall'art. 22, c. 3, del GDPR. Ciò significa che

⁶⁶ I. Trigueiro e al., *General Purpose Artificial Intelligence Systems (GPAIS): Properties, definition, taxonomy, societal implications and responsible governance*, in *Information Fusion*, 103, 2024, 1 ss.

⁶⁷ B. Marchetti, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, in *BioLaw*

il sorvegliante umano – nel caso di sistemi di IA utilizzata dalle PA, un funzionario amministrativo o un tecnico esterno? – deve «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», nonché essere capace di «interpretare correttamente l'output del sistema di IA ad alto rischio», in modo da poter decidere di non utilizzare il sistema, intervenire sul suo funzionamento, modificarne il risultato o anche disattivarlo. Ora, a parte il fatto che appare piuttosto ingenuo immaginare che una persona fisica possa avere in ogni momento il pieno controllo di processi algoritmici prodotti da sistemi dotati di capacità di archiviazione, ricerca, elaborazione e calcolo infinitamente superiori a quelle umane, fino al punto da poter interferire con il funzionamento della macchina, è evidente che i descritti compiti di sorveglianza presuppongono a monte la piena realizzazione della xAT a beneficio del supervisore: ma che tipo di informazioni dovrebbero essere fornite a quest'ultimo per farlo in condizione di svolgere efficacemente i propri compiti?⁶⁸

Da ultimo, va evidenziato che il recente regolamento europeo sull'intelligenza artificiale non distingue fra enti pubblici e privati, ma solo fra fornitori e utilizzatori (*deployers*) di sistemi di intelligenza artificiale, senza tenere conto del fatto che, invece, gli enti pubblici dovrebbero essere tenuti a garantire un livello di trasparenza algoritmica tendenzialmente più elevato di quello richiesto ai privati, ai fini della sindacabilità del provvedimento amministrativo⁶⁹. In particolare, la normativa di livello unionale non obbliga i soggetti pubblici che utilizzano algoritmi decisionali a dare pubblicità a questa scelta⁷⁰, a meno che – come prescrive il GDPR – il sistema non preveda il trattamento di dati personali e quindi sia d'obbligo informare gli interessati delle modalità e finalità del *data processing*. Sarebbe invece utile vincolare per legge le pubbliche amministrazioni a dare pubblicità al fatto di utilizzare algoritmi per lo svolgimento di talune attività rientranti nei loro compiti, anche se a quel punto ci si troverebbe davanti all'ostacolo di come poter assicurare un efficace livello di spiegabilità algoritmica.

In conclusione, da questa rassegna della recente produzione normativa dell'UE sulla trasparenza algoritmica emerge una certa sudditanza delle regole giuridiche nei confronti di quelle tecniche⁷¹. Dovrebbe invece avvenire il contrario: la norma giuridica

Journal, 2, 2021, 367 ss.

⁶⁸ Inoltre, nel caso in cui sia la pubblica amministrazione a servirsi di sistemi algoritmici, ci si chiede se il compito di sorveglianza debba essere svolto da un funzionario amministrativo – che potrebbe non avere adeguate competenze tecniche – o da uno sviluppatore del sistema, che però sarebbe estraneo alla pubblica amministrazione (B. Marchetti, *La garanzia ...*, cit., 377).

⁶⁹ A. Papa, *Intelligenza artificiale*, cit., 110.

⁷⁰ Ivi, 112.

⁷¹ In effetti, gli articoli contenuti nella quinta sezione del regolamento UE n. 1689/2024 sui sistemi di IA richiamano il precedente regolamento UE n. 1025/2012 sulla normazione europea, modificato dal regolamento n. 2022/2480 del 14 dicembre 2022. Secondo questi regolamenti le cosiddette “norme armonizzate” (trad. it. di *harmonized standards*) – cioè le specifiche tecniche emanate da organismi di normazione internazionali (ISO, IEC), nazionali (per l'Italia, UNI e CEI), ma soprattutto europei (CEN, CENELEC o ETSI), su richiesta della Commissione europea, in base ad orientamenti generali elaborati da quest'ultima – stabiliscono i requisiti essenziali per l'immissione sul mercato europeo di determinati prodotti, volti soprattutto alla protezione della salute e della sicurezza dei consumatori, e

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

dovrebbe precedere – e non seguire faticosamente – le specifiche tecniche dei sistemi, dettando le linee guida per la loro disciplina. Ora, è pur vero che una tecnica normativa che prevede ampio ricorso a concetti indeterminati e interpretabili implica una certa flessibilità applicativa, che potrebbe favorire l'adattamento ai rapidi sviluppi tecnologici, oltre a lasciare spazio alla regolamentazione ad opera delle diverse autorità nazionali competenti; tuttavia, «in mancanza di pratiche applicative comuni, il rischio è che tale indeterminazione generi anche incertezza, aprendo la possibilità di applicazioni difformi della disciplina all'interno del territorio dell'Unione e un conseguente aumento del contenzioso giudiziario»⁷².

4. L'utilizzo di algoritmi da parte delle pubbliche amministrazioni: qualche indicazione giurisprudenziale e molte incertezze

La giurisprudenza italiana in anni recenti ha affrontato la questione della trasparenza algoritmica, cercando di definire – per la verità senza giungere a risultati soddisfacenti – quale livello di comprensibilità del funzionamento dell'algoritmo sia da considerarsi adeguato. Le sentenze qui esaminate⁷³ riguardano l'utilizzo di tecnologie algoritmiche nei processi decisionali delle pubbliche amministrazioni che, come è noto, sono tenute ad assicurare l'imparzialità loro agire (art. 97, c. 2, Cost.), nonché la trasparenza «intesa come accessibilità totale delle informazioni concernenti l'organizzazione e l'attività delle pubbliche amministrazioni, allo scopo di favorire forme diffuse di controllo sul

quindi rappresentano un elemento chiave per garantire la libera circolazione dei prodotti all'interno del mercato dell'Unione europea. Si tratta di un sistema di partenariato pubblico-privato nel quale, sotto la guida della Commissione europea, organismi tecnici di natura privata elaborano standard tecnici che conferiscono ai produttori la presunzione di conformità ai requisiti della legislazione. L'adozione delle norme armonizzate non è obbligatoria per le imprese; tuttavia, i prodotti che le rispettano mostrano di possedere un determinato livello di qualità, sicurezza e affidabilità e si presumono conformi alla legislazione armonizzata dell'Unione europea, così che, in caso di controversie, l'onere di provare la mancata conformità del prodotto ricade sull'autorità competente e non sull'impresa. I regolamenti UE sulla normazione richiedono, inoltre, che le autorità pubbliche si adeguino a tutte le specifiche tecniche pertinenti per l'acquisto di hardware, software e servizi di tecnologia dell'informazione. Dunque, poiché anche i sistemi di IA sono prodotti commerciali, il regolamento UE sull'IA (quinta sezione) presuppone, salvo prova contraria, che i sistemi di IA ad alto rischio o i modelli di IA per finalità generali conformi alle norme armonizzate – che gli organismi europei di normazione dovranno produrre su mandato della Commissione europea – siano conformi al regolamento. Si veda in proposito, per approfondimento, C. Marengi, *La proposta di Regolamento UE sull'intelligenza artificiale e la regolazione privata: spunti critici in tema di norme tecniche armonizzate*, in *Diritto comunitario e degli scambi internazionali*, 3-4, 2021, 563 ss.

⁷² C. Casonato-B. Marchetti, *Prime osservazioni*, cit., 422-423.

⁷³ A. Corrado, *La trasparenza necessaria per infondere fiducia in una amministrazione algoritmica e antropocentrica*, in *Federalismi.it*, 5, 2023, 175 ss.; L. Grimaldi, *Costituzionalismo "post-umano" alla prova della decisione pubblica algoritmica*, in *Federalismi.it*, 34, 2022, 75 ss.; G. Lo Sapia, *La trasparenza sul banco di prova dei modelli algoritmici*, in *Federalismi.it*, 11, 2021, 239 ss.; P. Otranto, *Riflessioni in tema di decisioni amministrative, intelligenza artificiale e legalità*, in *Federalismi.it*, 7, 2021, 187 ss.; A. Papa, *Intelligenza Artificiale e decisioni pubbliche tra tecnica, politica e tutela dei diritti*, in *Federalismi.it*, 22, 2022, 101 ss.; N. Rangone, *Intelligenza artificiale e pubbliche amministrazioni: affrontare i numerosi rischi per trarne tutti i vantaggi*, in *BioLaw Journal*, 2, 2022, 473 ss.; A. Rocca, *L'impatto dell'Intelligenza Artificiale nella Pubblica Amministrazione*, in *ratioiuris.it*, 18 marzo 2024; A. Valsecchi, *Algoritmo, discrezionalità amministrativa e discrezionalità del giudice*, in *iusinjinere.it*, 14 settembre 2020.

perseguimento delle funzioni istituzionali e sull'utilizzo delle risorse pubbliche» (art. 1, c. 1, d.lgs. 33/2013). A scanso di equivoci, va precisato che i casi qui di seguito prospettati si concentrano sulla nozione giurisprudenziale di algoritmo – che lo stesso Consiglio di Stato (sez. III, 25 novembre 2021, n. 7891) ha definito «una sequenza finita di istruzioni, ben definite e non ambigue, così da poter essere eseguite meccanicamente e tali da produrre un determinato risultato» – e non di intelligenza artificiale, che, secondo la medesima pronuncia del Consiglio di Stato, comporta una situazione in cui «l'algoritmo contempla meccanismi di *machine learning* e crea un sistema che non si limita solo ad applicare le regole *software* e i parametri preimpostati (come fa invece l'algoritmo “tradizionale”) ma, al contrario, elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico»⁷⁴. Il problema, però, è che fra le due categorie – algoritmo e intelligenza artificiale – non vi è una dicotomia, ma piuttosto una osmosi, posto che i sistemi di intelligenza artificiale funzionano grazie agli algoritmi e che gli algoritmi possono avere complessità variabile, rendendo praticamente impossibile una netta distinzione fra quelli che la sentenza in esame definisce “tradizionali” e quelli basati, invece, sul *machine learning*.

Proprio in considerazione dei diversi livelli di complessità algoritmica, occorre chiedersi, in primo luogo, se la tanto invocata trasparenza possa realizzarsi attraverso l'ostensione del codice-sorgente dell'algoritmo⁷⁵, che potrebbe eventualmente essere realizzata anche semplicemente vincolando le amministrazioni pubbliche a servirsi solo di *software* di tipo *open source*. Nemmeno così, però, si risolverebbe di per sé il problema della trasparenza algoritmica, non solo perché il codice è difficilmente comprensibile per i non addetti a lavori, ma anche perché non necessariamente riuscirebbe a spiegare le ragioni alla base di una decisione assunta da algoritmi di *machine learning*, progettati per “imparare” dai dati ed evolversi automaticamente di conseguenza⁷⁶.

La seconda e più delicata questione da considerare è quella linguistica. Si è già detto di come il linguaggio naturale – e specificamente il linguaggio giuridico, che è intriso di definizioni tecniche specifiche, estranee al linguaggio comune – non siano adeguati alla trasposizione delle regole algoritmiche. Infatti, «la declinazione della trasparenza come “diritto alla spiegazione”, e quindi una interpretazione delle norme che tiene conto dell'evoluzione tecnologica, è invero imposta dalla caratteristica, sopra già evidenziata, che accomuna tutti i sistemi algoritmici, ovvero la loro “opacità linguistica”, dietro il quale si nasconde il rischio non solo della non comprensibilità, ma anche della non corrispondenza tra ciò che prevede la regola giuridica e ciò che è stato “trasposto” nel linguaggio macchina», che è uno dei 2500 linguaggi di programmazione attualmente disponibili⁷⁷.

⁷⁴ N. Cappellazzo, *Algoritmi, automazione e meccanismi di intelligenza artificiale: la classificazione proposta dal Consiglio di Stato*, in *Federalismi.it*, 23 marzo 2022, 1 ss.; C. Filicetti, *Sulla definizione di algoritmo (nota a Consiglio di Stato, Sezione Terza, 25 novembre 2021, n. 7891)*, in *giustizjainsieme.it*, 8 febbraio 2023.

⁷⁵ Così ha prescritto il Tar Lazio (sez. III bis, 14 febbraio 2017, n. 3769) in una sentenza relativa all'algoritmo utilizzato per formare le assegnazioni alle sedi di servizio degli insegnanti.

⁷⁶ A. Corrado, *La trasparenza necessaria*, cit., 210-214; L. Grimaldi, *Costituzionalismo “post-umano”*, cit., 81-82; A. Simoncini, *Il linguaggio*, cit., 27 ss.

⁷⁷ G. Lo Sapio, *La trasparenza*, cit., 247.

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

Eppure, secondo il Consiglio di Stato (sez. IV, 8 aprile 2019, n. 2270)⁷⁸ il principio della trasparenza algoritmica implica la «piena conoscibilità di una regola espressa in un linguaggio differente da quello giuridico». Quindi occorre che «la “formula tecnica”, che di fatto rappresenta l'algoritmo, sia corredata da spiegazioni che la traducano nella “regola giuridica” ad essa sottesa e che la rendano leggibile e comprensibile, sia per i cittadini che per il giudice». In particolare, questa «conoscibilità dell'algoritmo deve essere garantita in tutti gli aspetti: dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti». Solo così, infatti, il giudice potrà «sindacare la stessa logicità e ragionevolezza della decisione amministrativa robotizzata, ovvero della “regola” che governa l'algoritmo». In sintesi, questo ed altri arresti della giurisprudenza amministrativa attraggono l'algoritmo all'interno del concetto di “atto amministrativo informatico” che, in ossequio al principio di trasparenza dell'azione amministrativa, deve essere accessibile da parte dei soggetti interessati⁷⁹. Il problema di fondo di questa decisione è che essa esplicitamente muove dall'erroneo presupposto che, citando la sentenza, «la regola tecnica che governa ciascun algoritmo resta pur sempre una regola amministrativa generale, costruita dall'uomo e non dalla macchina, per essere poi (solo) applicata da quest'ultima»; invece oggi i sistemi di intelligenza artificiale non si limitano ad applicare regole precostituite, ma riescono a produrre regole nuove sulla base dell'esperienza acquisita tramite di processi di elaborazione dei dati.

I medesimi principi sono stati espressi da Cons. Stato, sez. IV, 13 dicembre 2019, n. 8472⁸⁰, secondo cui occorre sempre garantire «gli elementi di minima garanzia per ogni ipotesi di utilizzo di algoritmi in sede decisoria pubblica: a) la piena conoscibilità a monte del modulo utilizzato e dei criteri applicati; b) l'imputabilità della decisione all'organo titolare del potere, il quale deve poter svolgere la necessaria verifica di logicità e legittimità della scelta e degli esiti affidati all'algoritmo». La conoscibilità e spiegabilità dell'algoritmo – declinata in modo analogo rispetto alla sentenza n. 2270 – consisterebbe, dunque, nel verificare se i criteri, le presunzioni e i risultati dell'algoritmo corrispondano ai requisiti e ai propositi prestabiliti per legge o per provvedimento amministrativo, in modo da poter esercitare il controllo sulle regole di funzionamento dell'algoritmo, tenute anch'esse al rispetto dei principi di imparzialità e buon andamento dell'amministrazione. In questo caso, la trasparenza algoritmica assume una “dimensione rinforzata a geometria variabile”, in quanto viene diversamente modulata a seconda dei destinatari: un livello di trasparenza maggiore rivolto al funzionario pubblico responsabile del procedimento decisionale o al giudice nel caso di *judicial review* dell'atto amministrativo; un livello più limitato rivolto alla persona su cui ricadono gli effetti della decisione amministrativa⁸¹.

⁷⁸ L. Grimaldi, *Costituzionalismo “post-umano”*, cit., 87-88; B. Marchetti, *La garanzia*, cit., 374-376.

⁷⁹ Corrado, *La trasparenza necessaria*, cit., 191. Sulle diverse opinioni dottrinali circa la qualificazione giuridica del *software* algoritmico si veda la ricostruzione di D. Diaco, *Brevi riflessioni sulla natura giuridica del software (a partire da TAR Lazio, sez. III-bis, n. 8384/2023)*, in *giustiziasieme.it*, 26 luglio 2023.

⁸⁰ G. Lo Sapio, *La trasparenza*, cit., 245 ss.

⁸¹ L. Grimaldi, *Costituzionalismo “post-umano”*, cit., 87.

Più o meno in linea con il Consiglio di Stato, la Corte di cassazione (I sez. civ., ord. 10 ottobre 2023, n. 28358) ha ritenuto soddisfatti i requisiti di trasparenza e conoscibilità dell'algoritmo (nella fattispecie, un algoritmo di *rating* reputazionale) nel momento in cui l'utente è posto in condizione di comprendere lo «schema esecutivo dell'algoritmo», cioè la sequenza dei passaggi necessari a raggiungere un risultato, partendo da determinati dati: tali passaggi devono essere «elementari, univoci, di numero finito, operabili in un tempo finito e con risultato unico» e devono essere spiegati all'utente «in modo non ambiguo e in maniera dettagliata» utilizzando i termini della lingua comune e non certo il linguaggio matematico, che l'utente non è tenuto a comprendere. Ma il problema permane: è davvero possibile descrivere senza ambiguità e con un sufficiente livello di dettaglio dei procedimenti matematici in linguaggio comune? E quali indicatori consentirebbero di valutare, in caso di controversie, se tale spiegazione è adeguata? Ad esempio, proprio nel caso di specie il giudice di primo grado (tribunale di Roma) aveva ritenuto che lo standard minimo di trasparenza algoritmica non fosse soddisfatto, poiché la sua descrizione si limitava all'incidenza dei dati presi in considerazione, senza spiegare le modalità con cui si giungeva al risultato finale (cioè, il “peso specifico” dei singoli fattori considerati e i meccanismi di interazione fra loro). Tale obiezione, per quanto non condivisa dalla Cassazione, è indicativa di un problema insolubile: la spiegazione “a parole” dell'algoritmo produce inevitabilmente un certo grado di oscurità e ambiguità.

Del resto, quando vengono utilizzati algoritmi nel processo decisionale pubblico «si finisce per affidare (spesso silenziosamente e quasi surrettiziamente) a sistemisti e progettisti di *software*, anche estranei all'ente, un ruolo rilevante nell'organizzazione amministrativa e nell'azione amministrativa retta, *in parte qua*, da regole tecniche non accessibili (o comunque non intelleggibili). Ne emerge l'immagine di un'“Amministrazione invisibile”, spettatrice delle decisioni ad essa stessa imputabili e di un “livello normativo clandestino” celato dall'algoritmo»⁸². Si sente spesso affermare, quindi, che il principio di legalità sostanziale dell'azione amministrativa imporrebbe di stabilire per legge i principi e le regole da osservare nella definizione, progettazione e implementazione del sistema algoritmico che conduce alla decisione amministrativa⁸³. Occorrerebbero norme «che regolino il rapporto (di controllo) dell'uomo nei confronti della macchina e non il contrario, affinché non si verifichi una vera e propria inversione della dialettica servo-padrone nei rapporti uomo-macchina»⁸⁴, incluse norme tecniche, «volte a disciplinare la progettazione degli algoritmi da parte dell'uomo che, se necessario, si spingano

coraggiosamente a prevedere limiti oltre i quali ingegneri, tecnici e informatici non si potranno spingere»⁸⁵. Queste norme dovrebbero avere «un alto livello di malleabilità e adattabilità, che consenta ai singoli di sperimentare un'ampia gamma di versioni e adattamenti della medesima regola, e in un'attuazione *ex ante* di regole tecniche»⁸⁶. Si

⁸² P. Otranto, *Riflessioni*, cit., 199.

⁸³ P. Otranto, *Riflessioni*, cit., 199 e 202.

⁸⁴ C. Colapietro, *Gli algoritmi*, cit., 173.

⁸⁵ *Ibid.*

⁸⁶ E. M. Lombardi, *Norma e algoritmo: alcune considerazioni sul nuovo ordine tecnologico*, in *Giustizia Civile.com*,

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

tratta però di un obiettivo difficilissimo da realizzare, se si considera che il legislatore non dispone normalmente delle competenze tecniche indispensabili a tal fine e che, in ogni caso, la legislazione non riesce a tenere il passo della rapidissima evoluzione tecnologica. Nello stesso tempo, però, «il tentativo di dare una definizione giuridica unica e assoluta a fenomeni malleabili, liquidi, in continua trasformazione come gli algoritmi potrebbe risolversi in un nulla di fatto in quanto la tecnologia non si cura delle partizioni giuridiche e sistematiche: essa semplicemente si diffonde»⁸⁷. Occorrerebbe, allora, trovare il modo di pervenire a una «*governance ibrida*, basata sulla compartecipazione tra *expertise*, tecnica e giuridica, relazione illuminata dal principio di trasparenza, in funzione di partecipazione, di controllo democratico e di legittimità del *decision making process*»⁸⁸.

La terza questione riguarda la definizione degli standard cui le amministrazioni pubbliche dovrebbero conformarsi nei casi in cui si servono di algoritmi. Il *Codice dell'amministrazione digitale*⁸⁹, in effetti, non offre molte indicazioni utili in tal senso. L'art. 12, si limita a stabilire al c. 1 che «le pubbliche amministrazioni, nell'organizzare autonomamente la propria attività, utilizzano le tecnologie dell'informazione e della comunicazione per la realizzazione degli obiettivi di efficienza, efficacia, economicità, imparzialità, trasparenza, semplificazione e partecipazione nel rispetto dei principi di uguaglianza e di non discriminazione, nonché per l'effettivo riconoscimento dei diritti dei cittadini e delle imprese ...» e al c. 2 che esse «utilizzano, nei rapporti interni, in quelli con altre amministrazioni e con i privati, le tecnologie dell'informazione e della comunicazione, garantendo l'interoperabilità dei sistemi e l'integrazione dei processi di servizio fra le diverse amministrazioni». L'art. 13-bis, c. 2, richiama il Codice di condotta tecnologica, che dovrebbe disciplinare «le modalità di progettazione, sviluppo e implementazione dei progetti, sistemi e servizi digitali delle amministrazioni pubbliche»; tuttavia, al momento in cui si scrive il Codice non risulta essere stato ancora approvato. Il c. 3 del medesimo articolo chiarisce che le amministrazioni pubbliche devono individuare una persona responsabile per la transizione digitale e possono servirsi di esperti «nello sviluppo e nella gestione di processi complessi di trasformazione tecnologica e progetti di trasformazione digitale». Sulla corretta attuazione di tali disposizioni sorveglia l'Agenzia per l'Italia digitale (AgID), istituita dall'art. 14-bis, incaricata anche di emanare «Linee guida contenenti regole, standard e guide tecniche». In particolare, secondo l'art. 71, l'AgID adotta Linee guida contenenti le regole tecniche e di indirizzo per l'attuazione del Codice. Le Linee guida pubblicate finora e disponibili sul sito dell'AgID⁹⁰, però, non riguardano la questione della trasparenza algoritmica. Il «Piano triennale per l'informatica nella Pubblica Amministrazione», adottato da AgID a dicembre 2023⁹¹, prevede l'obiettivo dell'aumento della consapevolezza della pubblica amministrazione nell'adozione delle tecnologie di intelligenza artificiale (obiettivo 5.4), per il quale

7, 2020, 10.

⁸⁷ N. Cappellazzo, *Algoritmi*, cit., 5.

⁸⁸ L. Grimaldi, *Costituzionalismo "post-umano"*, cit., 95.

⁸⁹ D.lgs.82/2005 e successive modifiche.

⁹⁰ [AGID, Linee Guida sull'accessibilità degli strumenti informatici – PA.](#)

⁹¹ [AGID, Piano triennale per l'informatica nella Pubblica Amministrazione, 2024-2026.](#)

dovranno essere adottate apposite Linee guida entro il 2024. Da ultimo, la “Strategia Italiana per l’Intelligenza Artificiale 2024-2026”⁹², elaborata da un gruppo di esperti e pubblicata il 22 luglio 2024, contiene solo un generico richiamo alla necessità di definire «un registro di dataset e modelli, che siano costruiti secondo principi di trasparenza e *fairness*» (p. 15). Dunque, per il momento non sembra che la normativa di settore sia sufficientemente progredita.

Sul piano operativo forse una soluzione, per quanto parziale, al problema di come realizzare la trasparenza algoritmica potrebbe consistere nella definizione, per via normativa, di indicatori e parametri il più possibile condivisi, in base ai quali misurare il livello di trasparenza algoritmica. Qualche indicazione in tal senso potrebbe forse provenire, in un prossimo futuro, dal Centro europeo per la trasparenza algoritmica (ECAT), inaugurato il 18 aprile 2023, che riunisce ricercatori ed esperti nell’ambito dell’informatica, dell’analisi dei dati, dell’intelligenza artificiale e delle scienze giuridiche e sociali, al fine di analizzare il funzionamento dei *software* delle grandi piattaforme e motori di ricerca. Sarebbe però opportuno, come è stato giustamente osservato⁹³, che venisse istituita a livello nazionale una specifica Autorità indipendente, con funzioni di garanzia nell’utilizzo dei sistemi di IA, così come opportuno potrebbe essere l’utilizzo, da parte delle pubbliche amministrazioni, solo di *software* algoritmici *open source*, in modo che il codice-sorgente sia per principio disponibile per gli interessati.

Una fonte di ispirazione potrebbe essere rappresentata dalle *Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration*, predisposte dallo European Law Institute (ELI) con sede a Vienna⁹⁴: si tratta di un interessante tentativo di integrare la legislazione europea sull’IA nel contesto specifico della pubblica amministrazione, al fine di stabilire alcune solide garanzie per accrescere la fiducia dei cittadini nell’uso della tecnologia da parte dei decisori pubblici. L’idea centrale alla base delle *Model Rules* è una procedura di valutazione d’impatto, volta a valutare la gravità dei rischi dei diversi sistemi di IA. La regola n. 6, in particolare, prescrive il contenuto del rapporto di valutazione d’impatto, che deve contenere diversi elementi riferiti alla trasparenza algoritmica, fra cui la natura e le caratteristiche tecniche del sistema, la selezione dei dati di formazione, convalida e test, il contesto in cui viene utilizzato il sistema, l’interrelazione del sistema con altri sistemi digitali, una valutazione delle prestazioni, dell’efficacia e dell’efficienza del sistema per quanto riguarda agli obiettivi pubblici previsti, una valutazione dell’impatto specifico e sistemico del sistema sui diritti o interessi fondamentali o altri diritti individuali, nonché sulla democrazia, sul benessere sociale e ambientale.

Un’altra esperienza significativa portata avanti nel Regno Unito è l’*Algorithmic Transparency Recording Standard Hub*⁹⁵, che ha l’obiettivo di aiutare le organizzazioni del settore pubblico a fornire informazioni chiare sugli strumenti algoritmici che utilizzano e sul

⁹² Dipartimento per la trasformazione digitale, *Publicato il documento completo della Strategia Italiana per l’Intelligenza Artificiale 2024-2026*, 22 luglio 2024

⁹³ A. Papa, *Intelligenza artificiale*, cit., 112.

⁹⁴ European Law Institute, *ELI Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration*.

⁹⁵ *Algorithmic Transparency Recording Standard Hub*, in *gov.uk*, 5 January 2023.

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

motivo per cui li utilizzano. A tal fine, è stato progettato un modello standardizzato per la registrazione e la condivisione delle informazioni⁹⁶, da compilare con le informazioni specifiche richieste, accompagnato da una guida⁹⁷ per supportare le amministrazioni pubbliche nella compilazione dei rapporti sulla trasparenza algoritmica. Sarebbe forse opportuno sviluppare qualcosa di simile a livello di Unione europea, raggiungendo non solo agli enti pubblici, ma anche a quelli privati.

5. Riflessioni conclusive

È ormai un dato di fatto che tutti noi siamo parte di un ecosistema “tecno-antropologico”, se non addirittura “antropo-tecnocratico”, nel quale siamo contemporaneamente progettisti-produttori di sistemi tecnologici e destinatari passivi e spesso inconsapevoli della tecnoregolazione. Ciò considerato, nonostante tutte le difficoltà definitorie e applicative esaminate fin qui, risolvere la questione della trasparenza e della comprensibilità dei *software* algoritmici, alla base dei sistemi di IA, è una priorità ineludibile, se davvero si vuole mantenere la persona umana al centro dei processi decisionali, in ossequio e all'imperativo kantiano della persona come fine e mai come mezzo⁹⁸. Del resto, «il principio democratico e il relativo circuito politico- rappresentativo richiedono di rendere noti al cittadino gli elementi tecnici su cui si è fondata la decisione»⁹⁹ e quindi «la trasparenza del *decision making process*, da estendersi alle logiche artificiali che hanno informato la decisione, costituisce un baluardo indispensabile per la legittimità della decisione guidata dall'algoritmo, poiché consente di discernere l'apporto della tecnica dalla valutazione della politica, compensando quel *vulnus* di democraticità che si realizza quando il processo decisionale è delegato, anche solo in parte, alla macchina»¹⁰⁰. Si tratta di affermazioni indiscutibili in linea di principio ma, come si è evidenziato nelle pagine scritte fin qua, accompagnate da una gran quantità di questioni irrisolte, che si proverà ora a riassumere.

In primo luogo, si è cercato di spiegare come la nozione di trasparenza algoritmica (AT) sia intrinsecamente vaga e ambigua, perché ingloba al suo interno tanto l'idea dell'accessibilità del codice sorgente – che non solo è protetto dal segreto industriale, ma che in ogni caso non chiarisce i procedimenti mediante i quali l'algoritmo produce l'output finale – quanto quella della spiegabilità dei processi algoritmici (xAT). Quest'ultima, peraltro, deve misurarsi con le difficoltà della “traduzione” di procedimenti informatici in enunciati linguistici nonché con il variabile livello di competenza e comprensione delle diverse categorie di interlocutori: esperti tecnici, *deployer* (privati o pubblici) che si servono di algoritmi nell'esercizio delle loro attività o destinatari finali delle decisioni. In secondo luogo, si è evidenziato come, al momento attuale, la spiegabilità algoritmica risenta della mancanza di standard certi e misurabili attraverso cui poter valutare la sua

⁹⁶ *Algorithmic Transparency Recording Standard*, in *gov.uk*, 29 November 2021.

⁹⁷ *Guidance for organisations using the Algorithmic Transparency Recording Standard*, in *gov.uk*, 5 January 2023.

⁹⁸ C. Colapietro, *Gli algoritmi*, cit., 173.

⁹⁹ L. Grimaldi, *Costituzionalismo “post-umano”*, cit., 85.

¹⁰⁰ *Ibid.*

adeguatezza. L'elevato livello di opacità dei sistemi algoritmici – dovuto in parte anche all'esigenza dei loro produttori di proteggere il segreto industriale – fa che le decisioni, le scelte e le azioni così realizzate siano sostanzialmente “inspiegabili”, dal momento che né i destinatari finali né di *deployer* e in certa misura nemmeno i programmatori stessi dei sistemi sono in grado di comprendere le ragioni che hanno dato luogo ad un determinato risultato finale¹⁰¹. Però, l'utilizzo di algoritmi da parte delle pubbliche amministrazioni, e quindi la delega del potere decisionale alle macchine, non può essere un alibi per la reintroduzione di nuove forme di “oscurità autoritativa”¹⁰², che può impattare negativamente sull'effettività delle garanzie procedurali e processuali¹⁰³. In un modo o nell'altro, quindi, occorre che i destinatari finali ricevano una esauriente spiegazione dei processi che hanno condotto all'adozione della decisione automatizzata e delle ragioni in base alle quali la pubblica amministrazione ha scelto di accettare o di rifiutare il risultato prodotto dal sistema. Questo, in effetti, è il nodo problematico intorno al quale ruota la giurisprudenza relativa all'utilizzo di algoritmi da parte delle pubbliche amministrazioni, esaminata nel par. 4: i reiterati richiami alla “conoscibilità” dell'algoritmo mediante spiegazioni che traducano le formule tecniche in regole giuridiche si scontrano, infatti, con l'assenza di indicatori standardizzati tali da consentirne la realizzazione in concreto.

Sono stati comunque avviati recentemente alcuni interessanti tentativi per porre rimedio a tale mancanza. Fra questi, sono state richiamate nelle pagine precedenti alcune esperienze tedesche (le *Algo Rules* elaborate nel 2019 dalla Bertelsmann Stiftung), austriache (le *Model Rules* per valutare l'impatto dei processi decisionali algoritmici da parte delle pubbliche amministrazioni, definite dallo European Law Institute), inglesi (l'*Algorithmic Transparency Recording Standard Hub*) e di livello internazionale (il documento, non ancora pubblico, sulla *Transparency taxonomy of AI systems*, recentemente prodotto dall'ISO), mentre l'Italia sembra ancora un po' indietro in questo campo. Si è inoltre messo in luce l'approccio utilizzato dal nuovo regolamento europeo sui sistemi di intelligenza artificiale, che è integrato da allegati che contengono indicazioni piuttosto precise sul tipo e sulla qualità delle informazioni da fornire alle diverse categorie di destinatari (organismi nazionali di supervisione, *deployer* dei sistemi di IA e utilizzatori finali) e che, nella quinta sezione, demanda alla Commissione europea il compito di interagire con gli organismi europei di normazione (CEN, CENELEC, ETSI), al fine di pervenire alla definizione di norme tecniche armonizzate cui adeguare i sistemi di IA. Va sottolineato, in terzo luogo, che una preconditione per legittimare l'utilizzo dei sistemi algoritmici nei processi decisionali pubblici è l'aspettativa ragionevole – tuttavia non comprovata – che le macchine “intelligenti” siano in grado di differenziare senza discriminare¹⁰⁴. Altrimenti, la presunta neutralità di tali sistemi potrebbe nascondere pratiche discriminatorie causate da *dataset* viziati da pregiudizi, impedendo in tal modo di identificare, contrastare e sanzionare le discriminazioni perpetrate su base tecnologica. Allora, nel caso di sistemi di IA generativa che producono e processano dati inferiti,

¹⁰¹ M. Fasan, *I principi costituzionali*, cit. 186.

¹⁰² G. De Minico, *Fundamental Rights*, cit., 31-32.

¹⁰³ N. Rangone, *Intelligenza artificiale*, cit. 483.

¹⁰⁴ A. Papa, *Intelligenza artificiale*, cit., 107.

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

bisognerebbe definire dei parametri da applicare per garantire un elevato livello di *fairness* dei dati di partenza e dei procedimenti di elaborazione, tenendo tuttavia presente la necessità di preservare la privacy dei dati personali.

Non a caso, l'aspetto più interessante dei documenti sopra richiamati relativi agli standard di trasparenza è quello di individuare un metodo operativo per la realizzazione pratica della *fairness* algoritmica, alla quale la questione della trasparenza è strettamente connessa. La *fairness* algoritmica può essere intesa come “equità negativa, ossia come assenza o eliminazione di *bias* che producono o acuiscono forme di discriminazione, ma anche come “equità positiva”, fondata sul riconoscimento dell'eguaglianza e del valore morale delle persone, che implica la giustizia distributiva (cioè, l'equa distribuzione di risorse e opportunità), il diritto alla giustificazione (cioè, la tutela dell'uguale diritto di ogni persona di richiedere una giustificazione per il trattamento decisionale algoritmico a cui è sottoposta) e l'uguaglianza di relazione (cioè, il rispetto delle persone come individui particolari anche nel caso di profilazione algoritmica)¹⁰⁵. Tuttavia, l'idea di equità che viene tenuta per lo più presente nella programmazione della maggior parte degli algoritmi attuali sembra essere affine a quella su cui si basano le democrazie liberali, incentrate sulla difesa dei diritti e delle libertà fondamentali dal potere coercitivo, mentre trovano spazio molto minore le concezioni della giustizia di tipo distributivo¹⁰⁶ o della democrazia ugualitaria¹⁰⁷, fondata sulla realizzazione di condizioni di uguaglianza di fatto attraverso la riduzione delle disparità socio-economiche¹⁰⁸. Al di là di sporadici esempi, infatti, alla dimensione della xAT manca una riflessione condivisa ed eventualmente sindacabile, delle differenti visioni di mondo implicite al concetto di *fairness*¹⁰⁹: solo attraverso un substrato valoriale condiviso è possibile programmare algoritmi il più possibile *unbiased*, che contribuiscano a ridurre le discriminazioni e le disparità di trattamento presenti nella società. La trasparenza algoritmica, dunque, può contribuire a rendere palese a quale concezione di *fairness*, l'algoritmo si ispira.

In conclusione, resta da chiedersi se vi sia qualche appiglio per ancorare la trasparenza algoritmica a principi di rango costituzionale, questione che richiederebbe certamente un approfondimento ben maggiore del contenuto di queste poche righe conclusive. Pur nella consapevolezza dei rischi attinenti alle interpretazioni evolutive della Carta costituzionale¹¹⁰, non si può non considerare che, se la nostra Costituzione non si adeguerà alle sfide della contemporaneità – come certamente è avvenuto nel momento

¹⁰⁵ B. Giovanola – S. Tiribelli, *Equità e decisioni algoritmiche*, in *Teoria*, 2, 2022, 117 ss.

¹⁰⁶ J. Rawls, *A theory of Justice*, Cambridge, 1971

¹⁰⁷ N. Bobbio, *Eguaglianza e libertà*, Torino, 1995.

¹⁰⁸ A. Santangelo, *Equità degli algoritmi e democrazia*, in *Scientific Journal on Digital Cultures*, 2, 2020, 21 ss.

¹⁰⁹ A. Sterpa e al. *L'ordine giuridico*, cit., 1141.

¹¹⁰ Solo qualche cenno bibliografico *ex multis* sulla questione dell'interpretazione della Costituzione: M. Betzu, *Interpretazione e sovra-interpretazione dei diritti costituzionali nel cyberspazio*, in *Rivista Aic*, 4, 2012, 1 ss.; E. Di Salvatore, *Interpretazione e nozioni della Costituzione*, in *Costituzionalismo.it*, 1, 2023, 1 ss.; F. Modugno, *Interpretazione costituzionale e interpretazione per valori*, in *Costituzionalismo.it*, 3, 2005, 1 ss.; R. Nania, *Su alcune questioni di metodo nell'insegnamento del diritto pubblico*, in *Nomos*, 2, 2014, 1 ss.; F. Petrillo, *L'interpretazione della Costituzione tra positivismo giuridico della modernità e stato di sicurezza*, in *Società e diritti*, 2, 2016, 138, 163; S. M. Pisacane, *Interpretazione costituzionalmente orientata e diritto vivente*, In *LUISS working paper*, 5, 2017; G. Pino, *Interpretazione costituzionale e teorie della Costituzione*, Modena, 2019; G. Pitruzzella, *L'interpretazione conforme e i limiti alla discrezionalità del giudice nell'interpretazione della legge*, in *Federalismi.it*, 3, 2021, 160 ss.

storico del suo originario concepimento – sarà progressivamente relegata alla marginalità.

Dunque, se lo sviluppo della IA impone di ripensare la persona umana nella sua interezza, di ricollocarla in una diversa dimensione caratterizzata dall'inevitabile relazione con le "macchine sociali"¹¹¹, può essere utile rileggere l'art. 2 Cost. alla luce del fatto che gli ecosistemi digitali, in cui gli elementi naturali e quelli tecnologici sono inscindibilmente interconnessi, potrebbero essere considerati oggi come formazioni sociali in cui l'uomo svolge la sua personalità, in cui dunque devono essere garantiti i diritti fondamentali dell'individuo e deve essere richiesto l'adempimento dei doveri inderogabili di solidarietà economica, politica e sociale, gravanti anche sui gestori delle infrastrutture tecnologiche su cui tali sistemi si basano. Peraltro, poiché l'art. 2 Cost. tutela l'individuo all'interno delle formazioni sociali, ma anche le formazioni sociali di per sé, si potrebbe azzardare di fondare sull'art. 2 Cost. anche la garanzia del fatto che i sistemi sociali algoritmici possano continuare a esistere e a svilupparsi. Inoltre, in riferimento all'art. 3 Cost., l'opacità algoritmica rischia sempre più frequentemente di caratterizzarsi come un ostacolo che limita di fatto la libertà e l'eguaglianza dei cittadini, impedendo il pieno sviluppo della persona umana e l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese: infatti, essa può generare disparità, incompatibili con il principio di uguaglianza sostanziale, fra chi possiede gli strumenti cognitivi necessari a comprendere i linguaggi algoritmici e chi non ne dispone; può far sì che molte persone vengano escluse non solo dalla comprensione dei processi decisionali, ma anche dall'accessibilità ai circuiti di *decision making*, producendo un *vulnus* sotto l'aspetto della loro piena partecipazione attiva alla vita democratica; può esporre le persone a lesioni della loro dignità personale o della loro *privacy*, senza che esse ne siano consapevoli e siano poste in condizioni di avvalersi efficacemente degli strumenti di tutela. Come sostenuto in più occasioni dal Consiglio di Stato, l'opacità algoritmica è contraria ai principi di imparzialità e buon andamento della pubblica amministrazione (art. 97 Cost.), senza contare che l'assenza o la carenza della motivazione della decisione amministrativa causata dall'opacità algoritmica compromette le garanzie processuali, qualora gli interessati vogliano agire in giudizio per la difesa dei propri diritti e interessi legittimi (art. 24 Cost. e art. 6 Cedu). I produttori, gli sviluppatori, i proprietari e persino i *deployer* privati di algoritmi e sistemi di intelligenza artificiale agiscono secondo il principio di libertà di iniziativa economica, di cui all'art. 41 Cost.: poiché quest'ultima «non può svolgersi in contrasto con l'utilità sociale o in modo da recare danno alla salute, all'ambiente, alla sicurezza, alla libertà, alla dignità umana», non è irragionevole ritenere l'opacità algoritmica un rischio da questo punto di vista (almeno in relazione alla libertà e alla dignità umana, ma probabilmente anche per la sicurezza). Infine, portando il ragionamento all'estremo, anche l'art. 9 Cost. nella sua più recente formulazione riferita alla tutela degli ecosistemi¹¹², potrebbe prestarsi ad essere applicato ai sistemi di interazione uomo-macchina, che di fatto sono – sempre più saranno in futuro – i moderni ecosistemi in cui si svolge gran parte della nostra vita (il *digital environment*). Pur trattandosi evidentemente di una suggestione assoluta-

¹¹¹ A. Simoncini, *Il linguaggio*, cit., 4.

¹¹² L. cost. 1/2022 che ha modificato gli artt. 9 e 41 Cost.

L'ambiguo principio (anche costituzionale?) della trasparenza algoritmica fra tecnologia, diritto e linguaggio

mente estranea alle intenzioni degli estensori della legge di revisione costituzionale, in prospettiva futura questa lettura potrebbe essere in qualche modo favorita dal richiamo alla ricerca scientifica e tecnica contenuto nel primo comma dell'art. 9 Cost. e a quello del secondo comma riferito alle future generazioni, che ancora più di quelle attuali saranno plasmate dalla tecnologia. Poiché la tutela di cui all'art. 9 si riferisce agli ecosistemi in sé – quindi a tutte le loro componenti e non a quella umana in particolare – la norma potrebbe prestarsi non solo alla protezione dei diritti individuali all'interno dell'ecosistema digitale, ma anche alla protezione dello sviluppo tecnologico, essenziale alla sussistenza dell'ecosistema stesso.