

DISCLOSURE, EXPLAINABILITY E HUMAN OVERSIGHT: NUOVI DIRITTI FONDAMENTALI NELL'ERA DELL'IA E DEI PRIVATE GOVERNORS

Andrei Mihai Pop*

SOMMARIO: 1. Introduzione – 2. La genesi dei diritti fondamentali – 3. *Human first* e l'IA – 4. Dai poteri pubblici ai *private governors*: lotta alla disinformazione come banco di prova per i nuovi diritti – 5. Conclusioni.

1. Introduzione

Alla luce dello sviluppo della tecnologia appare opportuno ripensare il catalogo dei diritti fondamentali. Invero, essi dovranno tutelare il singolo dalle ingerenze dei poteri pubblici, ma anche dall'ascesa di un potere algoritmico che proviene dal mondo privato – in particolare dai c.d. *private governors*.

Ogni rivoluzione ha necessitato di una elaborazione e previsione di nuovi diritti e norme, quindi appare coerente operare nello stesso modo anche per l'IA. Un primo elemento da comprendere è quello legato al quesito: cosa è un diritto fondamentale? La ragione di una simile domanda è semplice, ossia per sostenere la necessità di introdurre nuovi diritti appare opportuno comprendere l'oggetto della discussione.

Si segnala che un diritto fondamentale è composto da due termini: *in primis*, da diritto, ossia una situazione giuridica di vantaggio, tutelata dall'ordinamento, rispetto all'attore privato e pubblico; *in secundis*, da un aggettivo, ossia fondamentale, il quale richiama l'idea che di una certa utilità-vantaggio non si può fare a meno, poiché fondativa di un certo ordinamento. Allo scopo di distinguere la generica categoria di diritti da quella di diritti fondamentali occorre effettuare un certo sforzo, poiché, di per sé, questa seconda nozione non ha un contenuto predeterminato¹. Infatti, la fundamentalità dei diritti non è

* Dottorando di ricerca in Filosofia del diritto, Università degli Studi di Milano.

1. La Costituzione utilizza il termine diritto fondamentale una sola volta all'articolo 32, ma ciò non equivale alla inammissibilità di altri diritti.

un attributo che, in sé e per sé, assume un significato all'interno dell'universo giuridico essendo necessario ricercare le risposte all'esterno di esso. Dunque, un diritto verrà considerato come fondamentale in base anche a criteri esterni, quindi a dei giudizi di valore. Invero, osservando la migliore dottrina, si possono individuare due grandi filoni di pensiero nell'interpretazione della nozione di diritto fondamentale.

Una prima tesi formalistica afferma che i diritti sono fondamentali a seconda della fonte che li afferma². In tale visione, essi sarebbero quelli enunciati nelle Costituzioni e nelle varie leggi. Una seconda impostazione, adottando una prospettiva sostanziale, osserva i contenuti e le posizioni giuridiche oggetto di tutela per affermare se un diritto sia dotato dell'attributo della fundamentalità³. A oggi i più, preferendo la seconda tesi, si concentrano maggiormente sul contenuto degli stessi. Infatti, i diritti fondamentali andrebbero a proteggere uno zoccolo duro di libertà nei confronti di ogni manifestazione del potere.

A questo punto risulta opportuno comprendere se risulti necessario un intervento del legislatore, ovvero se esista già, all'interno della nostra Carta dei diritti, uno strumento idoneo all'accoglimento delle nuove istanze della *societas*. La risposta a tale interrogativo potrebbe avere ricadute immense sulla tutela dei diritti dei singoli, poiché accogliendo la prima tesi poc'anzi esposta non si potrebbe che dubitare della capacità del Legislatore di intervenire tempestivamente con una disciplina organica all'altezza delle sfide lanciate dall'IA⁴. Ovvero, allo stato, la via preferibile da percorrere sarebbe la seconda per il tramite dell'articolo 2 della Costituzione.

L'interpretazione di tale norma ha generato immensi scontri tra gli esperti: è una clausola chiusa o una aperta? Secondo l'impostazione chiusa⁵, i diritti fondamentali sarebbero solamente quelli riconosciuti dalla Costituzione, ovvero quelli potenzialmente riconoscibili per il tramite di una interpretazione estensiva. Invece, la prospettiva aperta⁶ osserva nella detta norma una finestra tramite la quale nuovi diritti potrebbero essere affermati. Tale questione non può essere definita come una astratta. Ovverosia, le sfide dell'IA difficilmente

2. Così N. Bobbio, *Presente e avvenire dei diritti dell'uomo*, in *L'età dei diritti*, Einaudi, 2014.

3. Si veda G. Pino, *Il costituzionalismo dei diritti. Struttura e limiti del costituzionalismo contemporaneo*, il Mulino, 2017.

4. L'*AI Act*, e i vari interventi nazionali, indubbiamente introducono nuovi diritti fondamentali, ma non riescono a enuclearli tutti. Inoltre, l'*iter* legislativo richiede delle specifiche tempistiche e lo stesso *AI Act* entrerà pienamente in vigore solamente gradualmente.

5. Si rimanda ad A. Pace, *Problematica delle libertà costituzionali. Parte generale*, Cedam, 2003.

6. Così A. Barbera, *Art. 2*, in *Commentario della Costituzione. Principi fondamentali*, Zanichelli, 1975.

potrebbero consentire una risoluzione dei futuri problemi semplicemente con l'estensione delle vecchie libertà e garanzie.

Come accennato, a seguito di lunghi confronti, la tesi della fattispecie aperta è prevalsa. La scelta di riconoscere la possibilità che l'evoluzione dei costumi e dei bisogni della società possa innovare il catalogo dei diritti fondamentali non deve sorprendere. Invero, uno dei timori di una costituzione scritta è quella di divenire "lettera morta".

2. La genesi dei diritti fondamentali

Alla luce della ricostruzione svolta, e assodato che l'interpretazione dell'articolo 2 della Costituzione come clausola aperta fornisce lo strumento per accogliere nuove istanze di tutela, occorre analizzare la genesi di tali nuovi diritti⁷. L'IA è un potere, anzi sarà "Il Potere" nel futuro prossimo, quindi occorrono nuovi interventi.

Analizzando la genesi di tale categoria, si possono svolgere alcune osservazioni. In genere un diritto segue una certa sequenza: *in primis*, esso nasce da una istanza sociale; *in secundis*, riceve lentamente espliciti segnali di riconoscimento da parte della giurisprudenza; successivamente, essa verrà positivizzata; infine, e a una certa distanza temporale, trova un accoglimento all'interno delle carte costituzionali. Di seguito si concentrerà l'attenzione sul quarto momento di tale processo con lo scopo di individuare quei nuovi diritti fondamentali che potrebbero, alla luce dell'IA, trovare accoglimento all'interno della Costituzione. Purtroppo, tale analisi non verrà svolta con uno sguardo panoramico, bensì calando i diritti all'interno della vita e dell'azione dei *private governors*, in modo da avere un contesto di riferimento. Il rapporto tra i *private governors* e i cittadini dovrà essere modificato e adattato all'impatto che l'IA avrà sia nel settore privato sia in quello pubblico.

Occorreranno nuovi protocolli, procedure e diritti a favore dei soggetti coinvolti, poiché le piattaforme digitali sono la *longa manus* di un potere algoritmico, oltre che spesso l'attore più vicino al territorio e agli utenti⁸. Invero, i *private governors* impattano, con la loro attività, sulle basilari libertà democratiche del nostro ordinamento, in quanto la manipolazione delle informazioni è pericolosa specialmente se il soggetto privato rimane irresponsabile.

7. Si veda A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *Rivista di Biodiritto*, n. 1/2019.

8. Innegabile la pervasività delle piattaforme digitali all'interno della quotidianità (lo *smart phone* è l'esempio per eccellenza).

A seguito di queste riflessioni iniziali, si può già intuire che la questione non è “se” l’IA verrà utilizzata dai *private governors*, bensì il “come” e con quali limiti: il cittadino avrà un controllo circa l’andamento del procedimento informativo, oppure diverrà passivo rispetto al calcolo della macchina? I *private governors* saranno un potere responsabile?

Per tale ragione, al fine di tutelare l’esistente, servirà una ricerca e una operazione di individuazione al fine di enucleare dal presente i diritti fondamentali futuri imprescindibili per una *societas* complessa come quella attuale.

Tra i più importanti candidati, pur non esaurendo essi la lista, vi sarebbero: il diritto a una *AI disclosure* (riconoscibilità-conoscibilità), ossia il conoscere la natura artificiale di un determinato sistema con il quale si interagisce; il diritto a una spiegazione dei risultati dell’IA, quindi una *AI explainability*; il diritto a un controllo umano sul funzionamento del sistema di IA, definito come *human oversight*, il quale si connette al principio di non-esclusività dell’IA nella decisione⁹. Tali potenziali nuovi principi giuridici si andrebbero a collocare nel seno del c.d. *European Constitutional Law of Technology*. Nel successivo paragrafo ci si concentrerà, in particolare, sull’*human oversight*, dunque l’intervento e controllo dell’Uomo sull’operato dell’IA.

3. *Human first* e l’IA

L’Uomo fin dalle origini si è prodigato di migliorare le proprie condizioni di vita con l’uso della tecnica. Inizialmente, quest’ultima rimaneva strutturalmente uno strumento asservito agli scopi del proprio creatore. Tuttavia, lo sviluppo della tecnica impone un sempre maggiore grado di autonomia delle invenzioni per incrementare il grado di soddisfazione della *societas*. Ciononostante, doveroso tentare un contenimento di tale “slittamento esistenziale” attraverso lo strumento del diritto, il quale rimane una delle opere più longeve, potenti e umane. È stato già evidenziato che la tecnica è presente nella vita dell’Uomo fin dai suoi albori. Purtuttavia, la *téchne* ha una gemella, la quale nacque in un periodo contiguo e, pressoché, in contemporanea¹⁰. Ci si riferisce allo *ius*, il quale è già rinvenibile nelle prime regole di condotta di un dato gruppo di uomini rispetto ai propri membri (su un piano interno) e agli altri gruppi (su di un piano esterno). In altri termini, la tecnica come massima espansione delle ambizioni e potenze, alle volte incontrollate, dell’umanità; al contrario,

9. Così V.G. Sapuppo, *Etica e IA*, in M. Iaselli (a cura di), *AI Act. Principi, regole e applicazioni pratiche del Reg. UE 1689/2024*, Maggioli, 2024.

10. Si consulti N. Irti, E. Severino, *Dialogo su diritto e tecnica*, Laterza, 2001.

lo *ius* come limite e guida verso un determinato fine¹¹. Negli ultimi decenni tale rapporto non è mutato, tuttavia la tecnica ha subito una netta accelerazione, e il diritto non è riuscito sempre ad adattarsi. Invero, sono addirittura state avanzate tesi che vorrebbero la sostituzione di uno *ius* umano con un c.d. *ius* sintetico, in quanto oggettivo e imparziale. In tale ultimo scenario, la stessa esistenza del diritto rischierebbe di essere elisa. Dunque, si può cogliere la necessità di ripudiare la tesi che vorrebbe un diritto assolutamente neutrale¹² allo scopo di favorire la tecnica, poiché una delle sue essenze è proprio quella di controbilanciarla e di consentirne la manifestazione alla condizione di salvaguardare i diritti fondamentali degli individui. In particolare, non occorre un diritto neutrale, bensì un buon diritto¹³.

Detto ciò, quello che potrebbe essere ammesso è l'enucleazione di un insieme di norme coordinate su di un certo territorio per evitare disallineamenti – coerente a tale prospettiva l'approccio dell'UE con l'*AI Act*. Questa tensione tra il dominio umano sulle proprie opere e la ricerca continua di miglioramenti della propria esistenza impone un temperamento che solamente il diritto può svolgere.

In tale elaborato si accoglierà la tesi che impone la preferenza di un Uomo-*dominus* rifiutando una deriva totale che imponga di mutare la visione di una umanità al centro del mondo. Quest'ultima vorrebbe che si passasse da una prospettiva tendenzialmente umano-centrica a una tecno-centrica. In tal guisa, la visione umano-centrica è stata cristallizzata anche dall'*AI Act* che ha voluto imporre una netta preferenza. L'idea di Uomo al comando dell'IA è il fulcro dei diritti tratteggiati nel precedente paragrafo, poiché essi presuppongono che l'umanità sia sempre il principio e la finalità ultima dei sistemi artificiali.

Il cittadino, dunque, dovrà mantenere un certo grado di controllo sui sistemi di IA e sul loro operato se vorrà garantire il rispetto dei diritti costituzionali. Questo criterio di gestione dell'IA ha differenti gradi, ossia la pervasività del controllo rispetto alle operazioni compiute dall'IA aumenterà in virtù del concreto scenario.

Tali categorie sono conosciute con le espressioni: *human in command*, *hu-*

11. Questa considerazione impone, altresì, che il diritto intervenga a monte senza dover inseguire lo sviluppo costante e imprevedibile dell'IA. In altri termini, la sua portata e la forza deve inserirsi in un approccio *by design* all'IA che riesca a giocare d'anticipo evitando, laddove possibile, di giungere al tradizionale approdo violazione-sanzione.

12. Si veda R. Abbott, *The Reasonable Robot. Artificial Intelligence and the Law*, Cambridge University Press, 2020.

13. Così M. Angelone, *Alla ricerca del paradigma ottimale. La governance dell'IA tra legislazione e regolazione*, in *Tecnologie e Diritto*, n. 2/2024.

man in the loop, *human on the loop* e *human out of the loop*¹⁴. L'insieme di tali espressioni contribuiscono alla formazione del principio dell'*human oversight*, ossia il controllo dell'Uomo sull'IA.

Incominciando dalla categoria più debole, ossia l'*Houtl*, si può comprendere come tale categoria, nell'ipotesi in cui l'Uomo non svolga alcuna attività nella decisione dell'IA, non potrà trovare applicazione nel caso in cui siano in gioco diritti delle persone. In un simile scenario, vi è una assenza totale dell'intervento umano, ovvero sia il *dominus* assoluto è l'IA. Di conseguenza, un simile livello di pervasività dell'IA appare, in via astratta, incompatibile con i trattamenti che riguardino prerogative degli individui.

Procedendo nell'analisi emerge l'*Hotl*. In tale caso, l'IA ha l'iniziativa e la gestione del compito da svolgere, mentre l'Uomo assume il compito di sorvegliare i suoi procedimenti potendo intervenire in una fase *ex post* rispetto all'*output* del sistema. Questo approccio potrà essere perseguito laddove il sistema di IA non comporti pericoli rilevanti per i diritti fondamentali delle persone. Invece, un approccio all'*human oversight* che consenta un più penetrante controllo dell'essere umano è quello dell'*Hitl*. Quest'ultimo implica una gestione-controllo dell'essere umano dell'attività da svolgere in cui l'IA assiste, non decide. Questo modello pare compatibile con la gestione della maggior parte dei dati che riguardino diritti fondamentali e che comportino un rischio medio-alto. Inoltre, il quadro normativo attualmente vigente sostiene la necessità di applicare *Hitl* in tali situazioni in base al *risk-based approach*¹⁵. In ultimo luogo, si trova l'*Hic*. In tale ipotesi, l'IA avrà un ristretto ambito applicativo essendo l'intero procedimento controllato dall'Uomo. In tale caso è proprio l'Uomo a essere il *dominus*.

Tale rigore troverà applicazione laddove le attività da svolgere possono coinvolgere e influenzare diritti fondamentali con un rischio alto-inaccettabile. Dunque, il diritto allo *human oversight* verrà declinato, nella sua ampiezza e profondità, in base allo stato tecnologico e all'interesse oggetto di bilanciamento. Da un lato, più trasparenza e spiegabilità avrà un sistema di IA e mag-

14. Una parte della dottrina ha enucleato una ulteriore categoria, ossia lo *human post the loop* (*Hptl*). Quest'ultima si porrebbe tra lo *Houtl* e lo *Hotl* in termini di grado di controllo umano. Così M. Durante, *Potere computazionale. L'impatto delle Ict su diritto, società, sapere*, Meltemi, 2019.

15. Tale impostazione è particolarmente sostenuta dal legislatore europeo. Invero, già il Gdpr era fondato su ciò. Ovviamente la tutela dei diritti fondamentali e l'approccio basato sul rischio necessitano anch'essi di una continua opera di contemperamento, poiché quest'ultimo è tipicamente uno strumento di stampo privatistico-aziendale orientato più all'utile del mercato che alla tutela delle libertà. In altri termini, nessun diritto può essere considerato assoluto e tiranno (Corte Costituzionale, n. 85/2013), tuttavia ciò non equivale alla possibilità di una sua erosione oltre una minima soglia.

giori saranno le possibilità per seguire un approccio *HotI*; viceversa, laddove il problema della *black box* non venga sufficientemente contenuto, risulterà opportuno incrementare il controllo umano. Dall'altro lato, assumendo una differente prospettiva, all'aumento di criticità del diritto oggetto di attenzione da parte del sistema di IA, vi sarà la necessità di utilizzare sistemi trasparenti e spiegabili.

Questa tensione tra stato tecnologico e diritti dovrà essere oggetto di un contemperamento continuo da parte del legislatore, dei giudici e degli operatori del diritto. In ogni caso, e prescindere dalla etichetta, tali espressioni possono essere tradotte con il diritto-dovere a una sorveglianza umana sull'IA. Invero, si impone la regola secondo la quale in caso di applicazione di sistemi di IA, in settori giudicati come critici, occorrerà la supervisione dell'Uomo – c.d. *collaborative intelligence*¹⁶. Questo criterio ha avuto già nel *Gdpr* una sua iniziale, benché parziale, enunciazione all'articolo 22. A una prima lettura, tale previsione risulta essere perentoria pur potendosi applicare solamente nei casi in cui l'IA utilizzi dati personali. Purtuttavia, tale enunciazione subisce importanti deroghe ai successivi commi rendendo, *de facto*, il divieto di trattamenti automatizzati non la regola, bensì l'eccezione. Ovvero, ai sensi del comma 2, il divieto non si applicherà laddove vi sia un contratto, una legge o il consenso del soggetto destinatario di tale trattamento. In particolare, la deroga relativa all'accordo dell'interessato acuisce il rischio del *blind consent*, ossia la tendenza di prestare il proprio consenso ai fornitori dei servizi *internet* senza effettuare una sostanziale lettura delle condizioni del trattamento¹⁷. Quindi, la previsione del *Gdpr* è particolarmente lacunosa.

Alla luce di quanto osservato, appare fondata la visione secondo la quale non si potesse ritenere esistente nel *Gdpr* un diritto sufficientemente solido al controllo umano nei trattamenti automatizzati da parte dell'IA.

Il regolamento europeo sull'IA (n. 1689/2024) tenta di offrire una soluzione a tali aspetti problematici. In particolare, l'articolo 14 dell'*AI Act* – rubricato *Human oversight*. In tale sede, al comma 1, si prevede che al fine di garantire la supervisione dell'Uomo sull'IA rientrante nella categoria ad alto rischio, quest'ultima dovrà essere ideata, progettata e realizzata in un modo compatibile con le capacità di comprensione e di gestione dell'essere umano (*human friendly by design*).

16. Si veda S. Quintarelli *et al.*, *AI: profili etici. Una prospettiva etica sull'Intelligenza Artificiale: principi diritti e raccomandazioni*, in *Rivista di Biodiritto*, n. 3/2019.

17. Il consenso, quanto meno in certi casi, è divenuto una finzione, ovvero un consenso consapevolmente disinformato. Così M. Barberis, *Separazione dei poteri e giustizia digitale*, Mimesis, 2023.

La seconda parte di tale articolo fissa l'obiettivo di tale *human oversight*, ovvero sia quello di prevenire i rischi per i diritti fondamentali oggetto del trattamento da parte dell'IA. Precisa la norma che tale valutazione, effettuata *ex ante*, terrà conto dell'uso "naturale" di tale sistema, o al più di quello che si poteva ragionevolmente prevedere in base alle caratteristiche del caso concreto – *reasonably foreseeable misuse*. Tale precisazione è foriera di conseguenze, poiché ritorna il tema della trasparenza e dell'esplicabilità, ossia per riuscire ad anticipare un uso ragionevole del sistema, occorre conoscere e comprendere il funzionamento dello stesso. In assenza di un sufficiente livello di tali caratteristiche, tale norma non si potrà rispettare. Già queste prime considerazioni sull'articolo 14 rivelano l'intima connessione tra la *transparency*, la *explainability* e lo *human oversight*: essi sono tutti delle *conditio sine qua non*. Questa triade deve essere considerata unitaria, ovvero un unico "meta-diritto"¹⁸ a una *governance* algoritmica umano-centrica nel contesto dell'IA.

Proseguendo, il comma 3 affronta il tema dell'intensità del controllo umano sul funzionamento del sistema di IA cercando di individuare un concreto contemperamento tra efficienza e diritti. La profondità della sorveglianza umana dipenderà dai diritti in discussione, dai rischi nei loro confronti, dal livello di autonomia del sistema di IA e in ultimo dal contesto di riferimento e di funzionamento del sistema. Quindi, l'effettiva applicazione di tale regola dovrà tenere in considerazione quattro caratteristiche. Da ultimo, opportuno osservare il comma 4. Quest'ultimo elenca le effettive misure che dovrebbero essere offerte all'utilizzatore finale per concretizzare il diritto al controllo umano sull'IA. Queste sono principalmente quelle legate alla possibilità di discostarsi dall'*output* prodotto dal sistema nella consapevolezza che lo stesso può essere errato o comunque impreciso. Purtroppo, non solo l'*AI Act* affronta tali principi, bensì anche il *Digital Service Act* (n. 2065/2022). In particolare, l'articolo 17 impone di offrire una motivazione umana (c.d. *human oversight* rinvenibile anche negli articoli 20 e 21) laddove un determinato contenuto venga moderato nel seno di una piattaforma privata digitale. Questa disposizione si lega strutturalmente, rappresentandone un'altra manifestazione, ai principi di *AI Disclosure* ed *Explainability*. L'esigenza sorge al fine di evitare l'inquinamento del dibattito pubblico e la diffusione di disinformazione (articolo 34 del *Dsa*).

Tale consapevolezza è fondamentale e sarà uno dei principali punti problematici nell'uso dei sistemi di IA ad alto rischio, poiché il rischio di appiattimento della decisione umana su quanto suggerito dal sistema è concreto e attuale.

18. La sua natura implica che l'assenza di un solo componente annulla gli altri, e ciò lo eleva da un mero pacchetto di diritti a un "meta-diritto". Per esempio, il diritto a un ricorso (*oversight*) è inutile se non si sa perché una decisione è stata presa (*explainability*).

In dottrina si è parlato di un *effet moutonnier*¹⁹, ossia l'incapacità dell'Uomo di analizzare in maniera critica la decisione dell'IA in forza del mito dell'infallibilità dell'IA, il quale cattura l'Uomo non consentendogli più una libera scelta. In altre parole, l'opzione individuata dal sistema diverrebbe la *default-option*, la quale godrebbe di un plusvalore di significato, poiché l'individuo considererebbe l'IA superiore. Invero, salvo palesi errori, per discostarsene servirebbe un intenso sforzo²⁰. L'*output* dell'IA assume un simile valore poiché riesce ad alleggerire il peso della motivazione²¹.

Questa tematica sarà una fondamentale perché solamente un vaglio umano della decisione renderà quest'ultima accettabile dalla prospettiva dei consociati. Osservando l'impianto dell'articolo 14 dell'*AI Act* si comprende l'ampiezza potenziale della sua applicazione, poiché i sistemi ad alto rischio occupano una percentuale importante dei sistemi di IA. Il rischio di una interpretazione eccessivamente rigida potrebbe comportare alcuni problemi e il richiamo del comma 3 alle "circostanze" di utilizzo dell'IA potrebbe non essere sufficiente per un equo temperamento tra le esigenze di tutela dei diritti e le necessità di sviluppo tecnologico dell'UE.

I nuovi diritti fondamentali che dovranno essere riconosciuti all'interno del sistema costituzionale per consentire un governo umano-centrico dell'IA sono rinvenibili sia nella legislazione europea sia in quella nazionale. Tali diritti fanno parte di una costellazione di principi intimamente legati tra di loro, ovvero servirà una continua interazione tra gli stessi per consentire all'Europa di ergersi a guida capace di concentrare i valori occidentali volti a un equilibrio tra l'ambizione della *téchne* e la tutela dell'Uomo. Tale temperamento è frutto dell'intervento del diritto nella sua dimensione astratta, ma si evidenzia la necessità che le norme si trasformino da *law in the books* in *law in action*²². In caso contrario, non si potrà che ammettere l'abdicazione assoluta dello *ius* dal suo ruolo di co-protagonista – affianco alla tecnica – e la sua riduzione a mera comparsa nella storia dell'umanità. Un simile scenario dovrà essere assolutamente evitato per riuscire a mantenere l'essere umano *in command*.

19. Così A. Garapon, J. Lassègue, *Justice digitale. Révolution graphique et rupture anthropologique*, Puf, 2018.

20. Tuttavia, viste le competenze richieste, questo scenario appare essere di difficile attuazione.

21. Si veda A. Simoncini, *Amministrazione digitale algoritmica. Il quadro costituzionale, in Il diritto dell'amministrazione pubblica digitale*, Giappichelli, 2020.

22. Così R. Pound, *Law in Books and Law in Action*, in *American Law Review*, vol. XXXIV, 1910. Risulta necessario sia il riconoscimento di tali diritti fondamentali sia concrete normative mirate, ad esempio, all'alfabetizzazione digitale o a interventi di *design* per contrastare *bias* cognitivi.

4. Dai poteri pubblici ai *private governors*: lotta alla disinformazione come banco di prova per i nuovi diritti

L'analisi fin qui svolta ha consentito di definire i contorni giuridici dei nuovi diritti e di vederne un potente riconoscimento nell'*AI Act* e nel *Dsa*. Tuttavia, una delle sfide più pressanti per i sistemi democratici proviene dall'ecosistema informativo dominato da attori privati. La diffusione di *fake news* non è un fenomeno astratto, ma il risultato di precise architetture tecnologiche gestite da quelli che sono stati efficacemente definiti *private governors*. Infatti, le piattaforme *online* che moderano, curano e amplificano i contenuti esercitando, *de facto*, un potere normativo sulla sfera pubblica. Questa condizione va affrontata con urgenza, poiché trascurare tali aspetti rischia di indebolire l'impianto del sistema giuridico.

In questo contesto, l'*AI disclosure*, l'*explainability* e l'*human oversight* diventano strumenti essenziali per una cittadinanza attiva e democratica. In primo luogo, la trasparenza impone che l'utente venga informato dell'utilizzo di determinati sistemi di IA su tali piattaforme per il trattamento dei suoi dati e preferenze. Già tale circostanza è essenziale per segnalare, fin dal principio all'utente, il contesto e le potenzialità critiche. Successivamente, il diritto all'*explainability* si trasforma nel diritto dell'utente-elettore di comprendere perché un contenuto informativo è stato rimosso o la sua visibilità ridotta (*shadow banning*), oppure, e al contrario, perché una notizia falsa viene promossa algoritmicamente raggiungendo milioni di persone. Tale risultato sarà raggiunto laddove l'utente venisse informato, in un linguaggio comprensibile, delle motivazioni che hanno portato il sistema di IA a effettuare certe scelte. Infine, il diritto allo *human oversight* diventa il presidio contro la censura automatizzata garantendo un ricorso umano effettivo contro le decisioni che, nel tentativo di arginare la disinformazione, rischiano di sopprimere il dissenso legittimo o l'informazione veritiera, benché sgradita²³. Lo stesso *effet moutonnier* si manifesta quando milioni di cittadini delegano acriticamente il proprio giudizio agli algoritmi dei *social media* diventando così manipolabili. Questo risultato minerebbe una delle precondizioni fondamentali in una democrazia, per cui il combinato disposto del *Gdpr*, dell'*AI Act* e del *Dsa* si ergono a guida a un utilizzo responsabile dell'IA. Tuttavia, occorre un proficuo impegno per concretizzare tali garanzie, poiché altrimenti il rischio dell'*effet moutonnier* non vorrebbe arginato rimanendo le stesse una mera finzione giuridica.

Riconoscere questi principi come diritti fondamentali significa, pertanto, affrontare il problema della disinformazione alla radice. Ciò come una que-

23. Tale diritto è un presupposto indispensabile per un consenso libero dei cittadini.

stione di equilibrio di potere dotando il cittadino degli anticorpi cognitivi e giuridici necessari per difendere la propria autonomia di giudizio e partecipare consapevolmente al dibattito democratico.

5. Conclusioni

Nel presente elaborato si è cercato di evidenziare l'origine e il significato dei diritti fondamentali e, conseguentemente, di comprenderne il valore e la necessaria presenza. In questa *societas* digitalizzata, il diritto alla conoscibilità, spiegabilità e al controllo umano innanzi all'IA non sono una questione marginale, bensì centrale al fine di garantire le libertà degli individui sia nella loro sfera privata sia in quella pubblica. I sistemi di IA potrebbero essere sia lo scudo sia la spada in relazione alle *fake news*: essi possono aiutare nella lotta alle stesse, ovvero amplificarle. I principi analizzati sono il fondamento di tale discussione, ossia solo se correttamente applicati si potrà giungere a dei risultati positivi. Tale ricostruzione basata sui diritti integra e rafforza l'approccio basato sul rischio del *Dsa* e dell'*AI Act*, offrendo una tutela di rango costituzionale. Tuttavia, l'approccio basato sul rischio di derivazione privatistica è uno consequenzialista, ossia valuta i diritti in termini di "impatti" da mitigare. Invece, l'approccio basato sui diritti fondamentali è uno deontologico: li considera come limiti invalicabili a prescindere dal calcolo dei rischi. La sfida è, come detto, un temperamento tra tali due prospettive con la consapevolezza dell'inesistenza di valori assoluti.

L'ordinamento deve vegliare sul loro rispetto con lo scopo di evitare abusi. A tal fine, tali diritti sono stati positivizzati in appositi testi legislativi, ma gli stessi, ancora più a monte, si ritengono presenti nel quadro delle libertà già sancite nel testo Costituzionale italiano. Questo alla luce di una lettura innovativa dell'articolo 2, il quale viene posto come fondamento per l'elaborazione di nuovi diritti fondamentali legati alla *governance* algoritmica. Di conseguenza, con una simile prospettiva nemmeno laddove determinati testi legislativi fossero modificati-abrogati si potrebbero negare gli stessi, in quanto oramai immanenti e definitivamente affermati.