Disinformazione digitale

Modelli algoritmici e valori democratici

a cura di
Licia Califano
Federica Fabrizzi
Giovanni Sartor

Collana di Diritto

FrancoAngeli

SAGGI E RICERCHE

 ISBN 9788835191209

Copyright © 2026 by FrancoAngeli s.r.l., Milano, Italy. ISBN 9788835191209

Disinformazione digitale

Modelli algoritmici e valori democratici

a cura di
Licia Califano
Federica Fabrizzi
Giovanni Sartor

FrancoAngeli

Collana di Diritto

SAGGI E RICERCHE

 ISBN 9788835191209

Questo volume rappresenta il risultato conclusivo del Progetto PRIN 2022 PNRR "DAFNE" (*Democratic governance of Automated system for Fake News*), finanziato dall'Unione europea – Next Generation EU nell'ambito del PNRR, Missione 4 Componente 2, Investimento 1.1, Bando "PRIN 2022 PNRR" del MUR emanato con Decreto Direttoriale n. 1409 del 14 settembre 2022, Codice MUR P2022R7RS9, CUP H53D23010930001, e raccoglie alcune delle relazioni e degli interventi presentati al Convegno "Poteri pubblici e poteri privati nel governo dell'informazione digitale: la scelta europea dell'accountability" tenutosi presso l'Università degli Studi di Urbino Carlo Bo il 24-25 ottobre 2025.

Isbn e-book Open Access: 9788835191209

Copyright © 2026 by FrancoAngeli s.r.l., Milano, Italy. ISBN 9788835191209

Indice

Sezione II
Informazione digitale
e tutela dell'individuo

Sezione III
Comunicazione elettorale e politica

 ISBN 9788835191209

Riflessioni conclusive

AI e nuovi modelli di garanzia dei diritti fondamentali: criticità e prospettive applicative

*Giulia Vasino**

Sommario: 1. Introduzione. Dalla disinformazione digitale ai nuovi meccanismi di tutela – 2. La Fria: potenzialità e caratteristiche costitutive di un nuovo strumento di tutela – 2.1. Le difficoltà teorico-applicative e i modelli esistenti di riferimento – 2.2. Possibili parametri operativi coerenti con un approccio giuridico e qualitativo alla tutela dei diritti fondamentali – 2.3. Alcune prospettive di concreta implementazione della Fria – 3. Conclusioni.

1. Introduzione. Dalla disinformazione digitale ai nuovi meccanismi di tutela

L'obiettivo di definire e bloccare la circolazione di contenuti disinformativi – inquadrati, appunto, come nocivi per la libera dialettica democratica e idonei potenzialmente a determinare un rischio sistemico – ha assunto una solidità normativa senza precedenti[1]. Quest'ultima si estrinsecherebbe in un insieme di regole dotate di una forza performativa – un solido combinato disposto di atti – e una ampiezza – la pluralità di oggetti e destinatari – di inedita portata[2].

In merito al problema della disinformazione sono in particolare il Dsa e l'Aia a costruire una sinergia regolatoria all'apparenza formalmente impeccabile. Ambedue gli atti, com'è noto, si innervano su una logica di *accountability*

* Ricercatrice di Diritto costituzionale e pubblico, Università degli Studi di Urbino Carlo Bo.

1. Sull'evoluzione della sfera pubblica cfr. M. Manetti, *Internet e i nuovi pericoli per la libertà d'informazione*, in *Quaderni costituzionali*, n. 2/2023, p. 534.

2. Per una ricostruzione dell'articolato panorama normativo e le fonti più rilevanti in materia cfr. A. Iannuzzi, *Le fonti del diritto dell'Unione europea per la disciplina della società digitale*, in F. Pizzetti, S. Calzolaio, A. Iannuzzi, E. Longo, M. Orofino, (a cura di), *La regolazione europea della società digitale*, Giappichelli, 2024, spec. p. 13 e 23 ss.

 ISBN 9788835191209

graduata sulla base dell'entità del soggetto (Dsa)[3] o dei rischi potenziali posti da una determinata tipologia di AI (Aia)[4] con caratteristiche peculiari.

In forza di tale quadro, i destinatari delle prescrizioni devono assicurare di aver posto in essere i rigidi adempimenti individuati dall'atto normativo affinché possano essere considerati esenti da responsabilità per i contenuti ospitati ovvero per i danni potenziali prodotti da una determinata tecnologia. Tali obblighi si accompagnano poi a peculiari vincoli di trasparenza che in ambedue le discipline fungono da perno dell'impianto regolatorio nella sua totalità[5].

È tuttavia nel pilastro rappresentato dal *risk-based approach* che in entrambi gli atti emergono i profili più rilevanti con riferimento al peculiare problema della disinformazione. Quest'ultimo acquisisce per la prima volta all'interno del Dsa un suo autonomo ed esplicito inquadramento, qualificandosi come prodotto collaterale connesso all'utilizzo dei servizi delle *tech companies*[6]. Gli artt. 34-35 del Dsa, più nello specifico, individuando una soluzione per contrastare gli effetti negativi connaturati al mezzo, invitano a prestare particolare attenzione alle ipotesi in cui le modalità d'uso contribuiscano a «diffondere o amplificare contenuti fuorvianti o ingannevoli, compresa la disinformazione»[7].

L'AI Act tenta di perfezionare l'ambizioso scopo di mantenere un perfetto equilibrio fra la tutela dei diritti dei singoli e la sicurezza del prodotto[8], vista la sua primaria base giuridica, l'art. 114 Tfue[9]. Il regolamento, nonostante l'oggetto che lo caratterizza, mantiene infatti la continuità con le fonti preesistenti e quelle in fase di elaborazione, prediligendo lo stesso *humus* valoriale e approccio metodologico proiettato alla trasparenza, alla responsabilizzazione dell'operatore, e al principio cardine del controllo umano[10].

Benché le aspirazioni sottese all'Aia debbano forse essere ancora sondate e

3. Si fa riferimento, in particolare, al noto sistema a responsabilità condizionata (cfr. Capo III, Sez. V, Dsa).

4. Cfr. art. 6 e l'Allegato III dell'atto normativo.

5. Cfr., *ex multis*, l'art. 14 par. 1 del *Digital Services Act*. Emerge, inoltre, una diretta relazione fra trasparenza e dieta informazionale come evidenziano gli ulteriori obblighi di trasparenza posti per i sistemi di raccomandazione (art. 27) e i requisiti posti per le pubblicità online (art. 39).

6. Art. 34, par. 1. Tale correlazione si evince, inoltre, dal Considerando n. 2 del Dsa il quale si riferisce al contrasto a «contenuti illegali, disinformazione online e altri rischi per la società».

7. *Considerando* n. 84.

8. Per un ampliamento di questi profili cfr. M. Almada, N. Petit, *The EU AI Act: a medley of product safety and fundamental rights?*, in *SSRN Scholarly Paper*, Rochester, 2023.

9. Si veda il *Considerando* n. 1 dell'atto e l'art. 1 dell'Aia. In merito al punto specifico si rimanda a F. Fabrizzi, L. Durst, *Controllo e predittività. Le nuove frontiere del costituzionalismo nell'era dell'algoritmo*, Editoriale Scientifica, 2024, p. 7 ss.

10. Si vedano, in particolare, gli artt. 9 e 13 dell'atto normativo nonché gli artt. 43, 50 e 55.

analizzate in tutte le loro potenzialità[11], con riferimento alla disinformazione, elementi regolatori rilevanti possono trarsi dall'innalzamento degli obblighi di trasparenza previsti per i *deepfake*[12] e dalla specifica disciplina dedicata ai sistemi di intelligenza artificiale con finalità generali, con particolare riferimento al *genus* rappresentato dai sistemi di Gen-AI e quelle tecnologie che pongono rischi sistemici[13].

In merito all'elevazione dei requisiti di trasparenza posti con riferimento ai sistemi di AI generativa che producano un contenuto audio, immagine, video o testuale manipolato artificialmente, la prescrizione emergente dall'atto consiste nell'obbligo di marcatura del contenuto, con previsione analoga per quegli specifici prodotti elaborati «allo scopo di informare il pubblico su questioni di interesse pubblico»[14]. Tale previsione colpisce, oltre che per la sua ambivalenza e vaghezza definitoria, poiché trasmetterebbe una posizione apparentemente neutra nei confronti del prodotto manipolato per il quale non è prevista alcuna "etichetta di falsità" ma soltanto una certificazione di trasparenza che renda il fruitore della notizia consapevole della natura del contenuto. Si fugherebbe così l'ipotesi che il legislatore continentale prediliga di per sé orientamenti "paternalistici" in relazione a contenuti non autentici[15].

Tuttavia, al di là delle prescrizioni puntualmente dedicate alla disinformazione, è l'approccio regolatorio, nella sua generalità e innovatività, a invitare a una riflessione più radicale sul rapporto fra *fenomeno nocivo-rimedio-effettività della garanzia predisposta*. Di fronte a tale panorama normativo in fase di progressiva espansione e perfezionamento, si ha l'impressione, infatti, che il raffinamento della disciplina in chiave *human centric* – dichiaratamente modellata sul rafforzamento della posizione dell'individuo di fronte alla piattaforma o al fornitore/operatore di AI che si concreta anche in un articolato e convincente «*digital process of law*»[16] e in un'emersione di istanze assimilabili a un diritto

11. Sul punto si veda A. Simoncini, *La co-regolazione delle piattaforme digitali*, in *Rivista trimestrale di diritto pubblico*, n. 4, 2022, p. 1031 ss.; e, in senso critico, O. Pollicino, *Regolazione e innovazione tecnologica nell'"ordinamento della rete"*, *Relazione al XXXIX Convegno annuale dell'Associazione Italiana dei Costituzionalisti*, 15-16 ottobre 2024, pp. 52-53; F. Donati, *La protezione dei diritti fondamentali nel regolamento sull'intelligenza artificiale*, in *Rivista AIC,* n. 1, 2025, p. 1 ss.

12. Cfr. in part. art. 50, par. 4. Per una disamina approfondita cfr. M. Cazzaniga, *Una nuova tecnica (anche) per veicolare disinformazione: le risposte europee ai deepfakes*, in *MediaLaws*, n. 1, 2023, p. 170 ss.

13. Si v. art. 51 e i Considerando nn. 97 e 99.

14. Art. 50, par. 4.

15. Si v. G.E. Vigevani, *Potere politico e mezzi di comunicazione*, *Relazione al XXXIX Convegno annuale dell'Associazione Italiana dei Costituzionalisti*, 15-16 ottobre 2024, p. 33.

16. C. Caruso, *Il tempo delle istituzioni di libertà. Piattaforme digitali, disinformazione e discorso pubblico europeo*, in *Quaderni costituzionali*, n. 3, 2023, p. 559.

alla tutela giurisdizionale di fronte al privato – moltiplichi, piuttosto che attenuare, gli interrogativi teorici posti alla base.

Quest'ultimi, nella complessità del dibattito scientifico a essi dedicato, si sono da un lato ampiamente orientati verso la dimensione soggettiva. Parte della qualificata *querelle* giuridica sul tema ha infatti riservato uno spazio rilevante al "*chi*", scandagliando la natura problematica degli attori coinvolti[17]. Le voci critiche si sono concentrate maggiormente sui profili inerenti ai titolari, focalizzandosi, da un lato, sul famoso dilemma della cessione, nelle mani del soggetto privato, di un obbligo di implementazione della disciplina eurounitaria che vada ben oltre la mera applicazione di una previsione[18]; e, dall'altro, ponendo l'accento sulla pericolosa ricaduta che tale rinnovato rapporto privato-soggetto pubblico crei nei confronti di quest'ultimo[19].

Più chiaramente, se, da un canto, non è stato taciuto il rischio di uno scivolamento orizzontale che metta il privato nella pericolosa posizione di maneggiare i diritti fondamentali[20], si è insistito, dall'altro, sull'esigenza di non rimaner meno vigili sul soggetto statuale, ricordando le fondamenta delle garanzie di libertà e il paradigma classico autorità-individuo intorno al quale sono cuciti i diritti della persona.

In tal modo, in una prima ottica, è apparso ragionevolmente preoccupante lasciare alle piattaforme valutazioni sostanziali sui contenuti da eliminare, attuando spesso più che una delega di *enforcement* ma una vera e propria delega sostanziale[21]; contestualmente è stato altresì rilevato in modo altrettanto pertinente l'espansione della discrezionalità dell'operatore privato in presenza di tecnologie di AI.

Nel caso dell'AI Act, le perplessità del fronte soggettivo si fondono in modo

17. Fra i molti scritti dedicati dalla dottrina si rimanda a K. Klonick, *The new governors: the people, rules, and processes governing online speech*, in *Harvard Law Review*, vol. 131, n. 6, 2018, p. 1662 ss.; T. Gillespie, *Custodians of the Internet: platforms, content moderation, and the hidden decisions that shape social media*, Yale University Press, 2018, p. 5; J. Grimmelmann, *The Virtues of Moderation*, in *Yale Journal of Law and Technology*, n. 17, 2015, p. 61 ss.

18. M. Bassini *Fundamental rights and private enforcement in the digital age*, in *European Law Journal*, vol. 25, n. 2, 2019, p. 187.

19. Cfr. J. Balkin, *Old-school/new-school speech regulation*, in *Harvard Law Review*, 2014S. F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, in *University of Pennsylvania Law Review*, 2006, p. 13 ss.

20. Tale premessa critica oramai classica di tali studi è ricostruita da M. Betzu, *I baroni del digitale*, Editoriale Scientifica, 2022.

21. Sul punto cfr. M. Monti, *Privatizzazione della censura e Internet platforms: la libertà d'espressione e i nuovi censori dell'agorà digitale*, in *Rivista italiana di informatica e diritto*, n. 1, 2019, p. 35 ss.; O. Grandinetti, *Le piattaforme digitali come "poteri privati" e la censura online*, in *Rivista italiana di informatica e diritto*, n. 1, 2022, p. 181.

ancora più rappresentativo con la dimensione oggettiva della questione. Difatti, la natura inedita dei fenomeni e delle tecnologie con le quali il legislatore della sfera digitale è costretto a confrontarsi spinge sempre più a domandarsi se la normativa euro-unitaria riesca oggi a superare quell'implicito "stress test" anche rispetto alle modalità e alle condizioni individuate per effettuare concretamente, e in modo effettivo, la tutela dei diritti coinvolti. In particolare, il problema della disinformazione – quale fenomeno sistemico "tipico" dell'era digitale poiché esacerbato dai sistemi di AI – induce a riflettere, in una dimensione più ampia, sugli anticorpi garantistici disposti dall'atto normativo nella loro intrinseca struttura e finalità.

Fra gli stessi si distinguono quelle valutazioni di impatto le cui potenzialità appaiono ancora da sondare ed esplorare. In particolare, risalta, proprio per la sua aspirazione universalistica, il primo comma dell'art. 27 dell'Aia. La previsione stabilisce che «i *deployer* che sono organismi di diritto pubblico o sono enti privati che forniscono servizi pubblici e i *deployer* di sistemi di IA ad alto rischio di cui all'allegato III, punto 5, lettere *b* e *c*, effettuano una valutazione dell'impatto sui diritti fondamentali che l'uso di tale sistema può produrre».

Partendo da tale prospettiva più *macro*, si cercherà di porre in luce brevemente la relazione problematica fra la "spasmodica" centralità ricoperta dall'obiettivo di tutelare i diritti fondamentali di fronte all'avanzamento delle nuove tecnologie, testimoniata dalla moltiplicazione delle norme e degli strumenti messi sul campo, e la intrinseca problematicità di tali modelli, costretti a "far i conti" con l'elemento tecnologico posto alla base.

A tale fine si utilizzerà la Fria come strumento esemplificativo di tale difficile obiettivo di rimodellamento delle tutele. I tentativi del legislatore unionale di coniugare l'avanzamento tecnologico con soluzioni normative che riecheggiano istituti garantistici classici "scontano", infatti, le difficoltà scaturenti dall'ibridazione fra meccanismi di natura privatistica e la vocazione universalistica e di tutela propria del costituzionalismo liberal-democratico. Tali aporie e problematicità connaturate allo strumento potrebbero essere parzialmente attenuate solo attraverso un'oculata traduzione applicativa del test valutativo: una sfida che oggigiorno appare ancora in fase di costruzione.

2. La Fria: potenzialità e caratteristiche costitutive di un nuovo strumento di tutela

L'art. 27, imponendo un'analisi preventiva per i sistemi di AI ad alto rischio a garanzia dei diritti fondamentali – che si affianca, senza sostituirsi, a quella prevista dall'articolo 35 del Gdpr – appare esemplificativo delle aspirazioni

dell'intero atto normativo. Le stesse sembrano infatti tradotte anche a livello più strettamente operativo: imponendo al *deployer* una valutazione *ex ante*, il *Fundamental rights impact assessment,* ci si focalizzata sui danni potenziali che una determinata tecnologia potrebbe porre per i diritti del singolo all'interno di un regolamento che fa del rischio il proprio criterio ordinatore e la propria ossatura-guida.

La Fria, inoltre, si affianca a ulteriori test valutativi, contribuendo alla edificazione di un sistema di filtri particolarmente serrato anche di fronte a fenomeni suscettibili di esercitare un impatto di carattere generale[22]. Tale griglia valutativa comprende:

- il generale RMS imposto dall'articolo 9, un *Risk Management System* che interessa tutte le tecnologie ad alto rischio e che si concentra sulla previa individuazione ed eventuale mitigazione relativa a beni quali la sicurezza e la salute[23];
- il CA posto dall'art. 43, un *Conformity Assessment*, che esige una valutazione di *compliance* per le tecnologie ad alto rischio indicate nell'All. III «in order to ensure a high level of trustworthiness of high-risk AI systems».

Mentre il CA avrebbe una declinazione maggiormente "tecnica" e ricade sul provider, il Fria non a caso ricade sul *deployer* e si concentra sull'impatto operativo di un'AI che si "interfaccia con il mondo reale", con una particolare attenzione ai gruppi marginalizzati[24].

Infine, centrale, soprattutto in relazione a fenomeni quali la disinformazione, risulta lo SRA, imposto dall'art. 55, un *Systemic Risk Assessment* per i sistemi di AI con finalità generali che pongono rischi sistemici quali le Gen-AI[25]. Mentre il Fria si concentra sul *deployment-specific risk*, lo Sra si concentra sui rischi sistemi-

22. Sull'intreccio fra i vari test di valutazione anche esterni all'atto, quali il Dpia cfr. anche O. Pollicino, F. Paolucci, *Regulating AI Autonomy: A Constitutional Framework for the Digital Era*, SSRN Paper, 31 dicembre 2024, disponibile su: *papers.ssrn.com/sol3/papers.cfm?abstract_id=5098433*.

23. Si tratta, infatti di operatori pubblici o operatori privati che offrono un servizio pubblico (ad es. salute, educazione), operatori che per conto di autorità pubbliche sono chiamati a valutare l'ammissibilità delle persone fisiche nell'accesso a servizi essenziali (ad. es. sanità e credito).

24. In senso critico sulla distinzione di responsabilità fra *deployer* e *provider* cfr. F. Paolucci, *Shortcomings of the AI Act*, in *Verfassungsblog*, 14 marzo 2024.

25. Sulla natura dello Sra il quale è, in prima battuta, imposto dal Dsa e solo indirettamente incide anche sulla tecnologia utilizzata si rimanda all'analisi comparativa fra gli strumenti indicati di P. Chiara, F. Galli, *Normative Considerations on Impact Assessments in EU Digital Policy*, in *MediaLaws*, n. 1, 2024, p. 86; sull'ampia discrezionalità concessa al decisore politico circa

ci connessi a determinate tecnologie che hanno un impatto sociale significativo e, per questo, è sovente oggetto di critiche a livello teorico: si argomenta, infatti, come il *provider* vada a concentrarsi sulla tecnologia in sé in senso atecnico, ragionando su rischi di natura astratta che neppure potrebbero mai manifestarsi[26].

In realtà, da un punto di vista della strutturazione dello strumento e del suo livello di dettaglio regolatorio, il legislatore unionale pare esser senza dubbio pervenuto a un livello di sofisticatezza notevole soprattutto nel modo in cui si tende a coniugare l'equilibrio fra dato tecnico e tutela dei diritti. Il *Fundamental rights impact assessment* (Fria) è finalizzato infatti proprio a effettuare una valutazione preventiva del rischio in relazione ai diritti fondamentali attraverso, in estrema sintesi, tre momenti fondamentali:

1. individuare il rischio;
2. calcolarne l'entità;
3. prevenire e individuare misure di mitigazione (laddove possibili).

Gli elementi essenziali dell'*iter* valutativo appaiono tracciati con chiarezza ed efficacia a partire dal fronte definitorio, come si vede dalla qualificazione di "sistema ad alto rischio" fino alla enunciazione dei passaggi che costituirebbero il cuore operativo della Fria: i *deployer* effettuano una valutazione che comprende, in particolare, gli elementi seguenti:

1. una descrizione dei processi del *deployer* in cui il sistema di IA ad alto rischio sarà utilizzato in linea con la sua finalità prevista;
2. una descrizione del periodo di tempo entro il quale ciascun sistema di IA ad alto rischio è destinato a essere utilizzato e con che frequenza;
3. le categorie di persone fisiche e gruppi verosimilmente interessati dal suo uso nel contesto specifico;
4. i rischi specifici di danno che possono incidere sulle categorie di persone fisiche o sui gruppi di persone;
5. una descrizione dell'attuazione delle misure di sorveglianza umana, secondo le istruzioni per l'uso;
6. le misure da adottare qualora tali rischi si concretizzino, comprese le disposizioni relative alla governance interna e ai meccanismi di reclamo[27].

la definizione di rischio sistemico si esprime in senso critico F. Donati, *La protezione dei diritti fondamentali nel regolamento sull'intelligenza artificiale*, cit., p. 16.

26. Sul punto cfr. A. Mantelero, *The Fundamental Rights Impact Assessment (Fria) in the AI Act: roots, legal obligations and key elements for a model template*, in *Computer Science Law Review*, vol. 54, 2024.

27. Art. 27, par. 1, Aia.

Le dettagliate indicazioni contenute all'art. 27 par. 2 vengono inoltre arricchite, in prospettiva ermeneutica, dalle ulteriori specifiche contenute al *Considerando* n. 96, il quale estende l'aspetto definitorio soprattutto nell'inquadrare lo strumento sotto il profilo finalistico, soggettivo e temporale. Si ribadisce, infatti, come l'obiettivo della valutazione d'impatto sui diritti fondamentali è consentire al *deployer* di determinare i rischi specifici per i diritti delle persone o dei gruppi di persone che potrebbero essere interessati e di individuare le misure da adottare al concretizzarsi di tali rischi.

Viene chiarito, inoltre, come l'*impact assessment,* per essere realmente una cartina di tornasole efficace, dovrebbe essere svolto prima del primo impiego del sistema di IA ad alto rischio e dovrebbe essere aggiornata quando il *deployer* ritenga che uno qualsiasi dei fattori costitutivi sia cambiato. Nell'effettuare tale esame, l'operatore dovrebbe tenere conto delle informazioni pertinenti per un'adeguata valutazione dell'impatto, comprese, tra l'altro, le informazioni trasmesse dal fornitore del sistema di IA ad alto rischio nelle istruzioni per l'uso. Nel valorizzare la sinergia con il fornitore, il sopraccitato *Considerando* pone l'accento sull'importanza del coinvolgimento dei soggetti istituzionali chiave – dall'Autorità di vigilanza all'ufficio europeo per l'AI – a cui deve essere notificato il risultato, rafforzando in tal modo le prescrizioni previste dai parr. 3 e 5 dell'art. 27 Aia[28].

La valutazione di impatto deve infine rendere chiari i processi in cui verrà adoperata una determinata tecnologia, con quale finalità e con quale frequenza. Sulla base delle informazioni fornite dal *provider*, devono essere individuati in modo accurato i soggetti che potrebbero essere negativamente condizionati dall'uso di quella tecnologia e dimostrare consapevolezza circa le misure che intendono adottare in caso in cui quel rischio di materializzi effettivamente. Inoltre, il *deployer* deve assicurare, ancora una volta, un sistema interno di reclami e gestione degli stessi, con la garanzia della supervisione umana, non solo per consentire al singolo di esercitare un ruolo nelle procedure che attengono alla sua posizione soggettiva ma anche perché tali garanzie possono essere strumentali ad assicurare una mitigazione del rischio in concreto.

Questo processo dovrebbe, sulla carta, favorire un concreto *empowerment* degli individui sia per il livello di dettaglio sia per la sua natura preventiva nonché per il costante controllo dell'autorità pubblica, designata, ai sensi del pa-

28. Risulta poi centrale, all'interno del *Considerando* n. 96, il riferimento alla società civile: si evidenzia come, qualora i sistemi di IA siano utilizzati nel settore pubblico, risulterebbe di estrema efficacia il coinvolgimento di portatori di interessi pertinenti, compresi i rappresentanti di gruppi di persone che potrebbero essere interessati dal sistema di IA, gli esperti indipendenti e le organizzazioni della società civile.

ragrafo 3 dell'art. 27, come soggetto destinatario della scheda valutativa: una scelta che testimonia, ancora una volta, un accoglimento di quel percorso di valorizzazione del principio di trasparenza e responsabilità alla luce del faro rappresentato dai diritti.

Nonostante ciò, la necessaria esigenza di implementazione di questo strumento di tutela ha inevitabilmente posto l'attenzione su come realmente rendere concreto ed effettivo un modello che muove delicatamente a cavallo fra la carica assiologica dei diritti e il fronte tecnico e sfaccettato dei sistemi di AI. La sfida attiene sia alla dimensione teorica e di principio, da un lato, sia alla dimensione più strettamente operativa, dall'altro, mancando attualmente un'univoca "griglia" e strategia applicativa che possa garantire uno standard uniforme a livello europeo.

2.1. Le difficoltà teorico-applicative e i modelli esistenti di riferimento

Con riferimento alla prima prospettiva si osserva come il Fria – ancora una volta e maggiormente come il Dpia, data la massimizzazione della sua vocazione garantistica – sugella il ponte teorico-operativo fra la dimensione del rischio e quella dei diritti, fra un modello che presuppone la quantificabilità e una sfera, quale quella dei diritti fondamentali, caratterizzata da una netta dimensione qualitativa, la quale ha sempre faticato a coniugarsi con ambiti governati dal *quantum*[29].

A maggior ragione questa ambizione si rivela ancora più problematica se affidata a una valutazione *ex ante* che esige, quindi, che si prospetti e si soppesi l'entità delle possibili conseguenze negative prima che il danno si sia effettivamente prodotto. Con riferimento a quest'ultimo, la definizione contenuta nel *Considerando* n. 75 del Gdpr di «rischio per i diritti fondamentali» rimane comunque determinante ancorché vaga, qualificandolo lo stesso come «*a composite result of the probability and severity of certain data processing activities, which could lead to physical, material or non-material damage*». In tal modo, sembra confermarsi l'idea che il danno (*harm*) costituisca il vero ponte, a livello concettuale prima che giuridico, fra un parametro probabilistico e quantitativo, da un lato, e la dimensione ampia e socialmente situata dei diritti, dall'altro.

29. La cui delicata dissertazione teorica si lega chiaramente all'ampio dibattito classico sulla proporzionalità e sulla bilanciabilità dei diritti fondamentali; sul punto cfr. R. Alexy, *Constitutional Rights, Balancing, and Rationality*, in *Ratio Juris*, vol. 16, n. 2, 2003, pp. 133 ss.; A. Barak, *Proportionality. Constitutional Rights and their Limitations*, Cambridge University Press, 2012, pp. 460; J. Silva Sampaio, *Proportionality in Its Narrow Sense and Measuring the Intensity of Restrictions on Fundamental Rights*, in D. Duarte, J. Silva Sampaio (a cura di), *Proportionality in Law: An Analytical Perspective*, Springer Verlag, 2018, p. 71 ss.

Da un canto, non vanno taciuti gli elementi positivi di una valutazione *risk-based,* la quale manifesta un'indubbia praticità di metodo nel momento in cui consente all'interprete, in astratto, di fare affidamento anche su una lista di rischi possibili già quantificati a livello legislativo e giurisprudenziale: in questo modo si procede ad anticipare l'evento fissando una linea di demarcazione di non tollerabilità oltre la quale il danno prodotto sarebbe ritenuto non più accettabile[30].

Dall'altra parte, tuttavia, anche un'osservazione preliminare sembra far sorgere più opacità teoriche e perplessità. Si è insistito, infatti, non solo sulla delicatezza di quantificare *a priori* un danno ma anche sulla aggravata difficoltà che la ponderazione numerica porrebbe ulteriormente soprattutto quando ci si sposta dalla sfera dei diritti individuali "classici" verso la sfera dei principi, come nel caso della dignità individuale. La menzionata difficoltà di quantificazione verrebbe inoltre aggravata dalla percezione e graduazione che i diritti assumono a seconda dei contesti valoriali e normativi di riferimento, i quali sono soggetti a una sensibile variazione nel panorama unionale.

Alla luce di queste premesse critiche, dunque, ancor prima di approfondire l'operatività in concreto di un inedito sistema di garanzia valutandone le prospettive applicative, i primi commentatori hanno paventato la limitata idoneità di un meccanismo di tal natura di poter ridurre la complessità della stessa categoria di diritto fondamentale, appiattendo la stessa a una arida bidimensionalità[31]. Da quest'ultima discenderebbe, di conseguenza, soltanto la valutabilità di danni economicamente calcolabili e non il più multisfaccettato e poli-dimensionale impatto che un sistema di AI potrebbe avere sulla sfera della persona[32].

Alla sfiducia di un modello *risk-based* – fondata proprio sulla asserita inadeguatezza radicale fra dimensione dei diritti e valutazione del rischio affidata, per di più, a un operatore economico – si è contrapposta pertanto la più rassicurante prospettiva mitigatoria di un modello *rights-based*[33]. Quest'ultimo, a sua volta, tenta di eliminare le sfumature quantitative e adotta un approccio normativo che misura la violazione sulla base di un parametro violato: o vi è una violazione o non vi è. Tale linea direttiva, che pone la norma al centro dei

30. Si v. G. De Gregorio, P. Dunn, *The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age*, in *Common Market Law Review*, vol. 59, n. 2, 2022, pp. 473-500.

31. Una polemica teorica già ampiamente consolidata in dottrina già prima dell'entrata in vigore dell'atto, cfr. S. Engle Merry, *The Seductions of Quantification, Measuring Human Rights, Gender Violence, and Sex Trafficking*, The University of Chicago Press, 2016.

32. A. Mantelero, *Beyond Data. Human Rights, Ethical and Social Impact Assessment in AI*, Springer, 2022, pp. 54-55.

33. Sul punto A. Mantelero, *The Fundamental Rights Impact Assessment (Fria) in the AI Act: roots, legal obligations and key elements for a model template*, in *Computer Science Law Review*, vol. 54, 2024.

diritti fondamentali, ha il pregio di evitare prospettive riduzioniste ma, dall'altra, avrebbe a sua volta una intrinseca forma di rigidità che si scontra con la formulazione testuale del Fria, la quale menziona esplicitamente la probabilità e la severità del rischio[34].

Con riferimento alla seconda dimensione, la principale difficoltà dipende dall'assenza di un *framework* applicativo seppur minimale. Da un lato, l'Aia, in merito alla Fria, fa esplicito riferimento ai principi della Carta di Nizza (Cfreu), la cui applicazione per via giurisprudenziale esige il rispetto del principio di proporzionalità nell'interferenza/limitazione di ciascun diritto fondamentale (art. 52).

Purtuttavia, a parte il riferimento a una *human centric AI* e alcuni specifici richiami al principio di trasparenza e sorveglianza umana (artt. 13, par. 3, e art. 14, par. 2) il test offerto dall'Aia rimane piuttosto generico, non offrendo un adeguato *benchmark* di riferimento, contribuendo così a determinare incoerenza e episodicità[35].

Infine, anche il problema dell'inquadramento tecnico risulta determinante. La reticenza normativa sotto questo profilo è esacerbata dalla specificità delle tecnologie contemplate dall'Aia che spaziano dall'apprendimento automatico alla robotica fino al ragionamento simbolico. A questo si aggiunge l'interdipendenza fra modelli e il riaddestramento continuo che rischia di trasformare ciclicamente gli eventuali standard elaborati con fatica in riferimenti obsoleti. Questi fattori contribuiscono a rendere anche le esigenze di conformità un "obiettivo mobile"[36].

A ogni modo, il Fria rappresenta l'evoluzione e uno strumento complementare rispetto ad altri sistemi/test di valutazione di impatto sui diritti umani a cui inevitabilmente i primi commentatori hanno guardato nella ricerca di standard adeguati, dallo *Human Rights Democracy and Rule of Law Impact Assessment* (Huderia) allo *Human Rights Impact Assess*ments (Hria) elaborato in seno alle Nazioni Unite senza ovviamente tralasciare il ruolo chiave del già citato *Data Protection Impact Assessment* (Dpia) introdotto dal Gdpr[37].

34. G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, in *Computer Law & Security Review*, n. 56, 2025.

35. Cfr. L. Gatt *et al*., *Fria implementation model according to the AI Act*, in *EJPLT*, n. 2, 2024.

36. Sul punto cfr. P. Ceravolo *et al*., *HH4AI: a methodological framework for ai human rights impact assessment under the EU AI Act*, disponibile su: *arxiv.org/abs/2503.18994*.

37. Per una introduzione generale a tali strumenti si rimanda a N. Götzmann, *Introduction to the Handbook on Human Rights Impact Assessment: Principles, methods and approaches*, in N. Götzmann (ed.), *Handbook on Human Rights Impact* Assessment, Edward Elgar Publishing, 2019, p. 2 ss.

Con riferimento al supporto offerto dal diritto internazionale non va inoltre tralasciato il ruolo chiave svolto, almeno in una prima fase transitoria, dagli organismi di normazione. L'Organizzazione Internazionale per la Normazione (Iso) e la Commissione Elettrotecnica Internazionale (Iec), attraverso il comitato congiunto Iso/Iec Jtc 1/SC 42, hanno sviluppato un ampio insieme di standard dedicati all'intelligenza artificiale[38]. Gli stessi tentano di individuare un primo "pacchetto" di linee guida di riferimento per la valutazione dell'impatto dell'IA sui diritti fondamentali che tenga conto della gestione del rischio, della qualità dei dati, della robustezza dei modelli e della governance organizzativa[39].

Tra gli standard più rilevanti si colloca Iso/Iec 23894 che fornisce linee guida per la gestione del rischio nei sistemi di IA. Questo standard definisce un processo strutturato per l'identificazione, l'analisi, la valutazione e il trattamento dei rischi associati all'IA, ponendo l'accento sull'interazione tra fattori tecnici, organizzativi e contestuali. L'approccio proposto è coerente con la logica *risk-based* adottata dall'AI Act e può essere efficacemente integrato nei sistemi di gestione del rischio richiesti dall'art. 9 del Regolamento[40].

Accanto agli standard Iso/Iec, anche l'*Institute of Electrical and Electronics Engineers* (Ieee) ha sviluppato una serie di standard e linee guida che pongono una maggiore enfasi sulle dimensioni etiche e sociali dell'intelligenza artificiale. In particolare, la famiglia di standard Ieee 7000 affronta temi quali la trasparenza, la responsabilizzazione, la mitigazione dei *bias* e l'allineamento dei sistemi di IA ai diritti umani[41].

38. La cui rilevanza è richiamata e ricostruita in P. Ceravolo *et al*., *HH4AI: a methodological framework for ai human rights impact assessment under the EU AI Act*, 23 marzo 2025, p. 7 ss., disponibile su: *arxiv.org/abs/2503.18994*.

39.Per informazioni più dettagliate cfr. *www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0*.

40. Altri standard rilevanti includono l'Iso/Iec 23053, il quale fornisce un quadro di riferimento per l'IA basata su apprendimento automatico, e Iso/Iec 25059, che affronta la qualità dei sistemi di IA in termini di accuratezza, affidabilità e robustezza. Nel complesso, gli standard Iso/Iec offrono un insieme coerente di strumenti per affrontare le dimensioni tecniche della valutazione dell'IA. Tuttavia, essi tendono a concentrarsi prevalentemente sugli aspetti ingegneristici e organizzativi, lasciando in secondo piano una valutazione esplicita e sistematica dell'impatto sui diritti umani. Per un approfondimento si rimanda ugualmente al sito disponibile su: *www.iso.org/committee/6794475.html*.

41. Lo standard Ieee 7000 propone un modello per integrare considerazioni etiche nel processo di progettazione dei sistemi, promuovendo un approccio "*value-based*" allo sviluppo tecnologico. Altri standard della stessa famiglia, come Ieee 7001 (trasparenza), Ieee 7002 (data privacy) e Ieee 7003 (mitigazione dei bias), forniscono indicazioni operative per affrontare rischi specifici legati all'impatto dell'IA sui diritti fondamentali. Sul punto si rimanda al sito disponibile su: *standards.ieee.org/industry-connections/activities/ieee-global-initiative/*.

Tali linee guida risultano particolarmente rilevanti nel contesto degli *impact assessment* e offrono strumenti concreti per tradurre principi etici astratti in requisiti progettuali verificabili. Tuttavia, analogamente agli standard Iso/Iec, essi non sono concepiti specificamente per soddisfare gli obblighi giuridici dell'AI Act e non forniscono una metodologia integrata per la conduzione di una Fria completa[42].

In primo luogo, tali parametri sono spesso frammentati, ciascuno focalizzato su un aspetto specifico del sistema di IA, senza offrire una visione complessiva dell'impatto sui diritti fondamentali. In secondo luogo, gli standard tecnici tendono a privilegiare una prospettiva *ex ante* e progettuale, mentre l'AI Act richiede una valutazione continua lungo l'intero ciclo di vita del sistema, inclusa la fase di utilizzo concreto e di monitoraggio *post-deployment*. Inoltre, la maggior parte degli standard omette di affrontare in modo esplicito il bilanciamento tra diritti fondamentali potenzialmente confliggenti, creando così un supporto applicativo discontinuo e parziale[43].

2.2. Possibili parametri operativi coerenti con un approccio giuridico e qualitativo alla tutela dei diritti fondamentali

Alla luce di tali criticità, è apparso subito necessario un approccio metodologico integrato che sia in grado di combinare i contributi degli standard tecnici con una precomprensione giuridica adeguata.

Da un punto di vista teorico risultano di rilievo quelle riflessioni che si sono orientate verso la valorizzazione, in fase di implementazione del modello, del concetto di "interferenza" con i diritti fondamentali piuttosto che di "violazione".

L'interferenza, in primo luogo, è un concetto che sfugge alla logica dicotomica della violazione soggiacendo a una logica di gradualità. Inoltre, l'apporto definitorio offerto dalla giurisprudenza della Corte di Giustizia UE e dalla Corte Edu risulta determinante trovando all'interno della stessa un ricorrente riferimento al concetto di «*seriousness of interferences*»[44]. Entrambe le Corti

42. P. Ceravolo *et al*., *HH4AI: a methodological framework for ai human rights impact assessment under the EU AI Act*, cit., p. 9.

43. Ivi, p. 6.

44. Così la Corte Edu nel noto caso *Roman Zakharov v. Russia*, 2015, par. 232 in cui si ribadisce la necessità di individuare la «the seriousness of the interference with an applicant's right to respect for his or her private life» chiarendo come «the assessment depends on all the circumstances of the case, such as the nature, scope and duration of the possible measures, the grounds required for ordering them, the authorities competent to authorise, carry out and supervise them, and the kind of remedy provided by the national law. The Court has to determine

hanno infatti affermato che le interferenze costituiscono una scala graduabile, misurabile in termini di severità. Più chiaramente, è stato evidenziato come, ai fini di valutare la proporzionalità di un'interferenza, bisogna altresì misurarne la severità.

Se dunque ci si apre, a livello teorico, al necessario apporto di un modello *right-based* – il quale supererebbe i limiti dei meccanismi del rischio esclusivamente quantitativi – e, al contempo, la rigidità del concetto di violazione – tipico del modello *rights-based* – in favore di un concetto più mediato quale quello di interferenza, è opportuno individuare alcuni parametri.

Allo stato attuale, appaiono efficaci, in termini di approccio organico ed efficacia della valutazione, l'individuazione congiunta di alcuni parametri che, se applicati congiuntamente, risulterebbero oggettivi senza "inaridire" le implicazioni assiologiche sottese alla valutazione di un'interferenza con un diritto fondamentale. Fra le stesse si segnala un modello di azione integrato che tenga conto di:

a. un parametro oggettivo, ossia una valutazione della gravità dell'interferenza valutata normativamente;
b. un parametro soggettivo, finalizzato a soppesare il "valore sociale" di quella interferenza ovvero come la stessa venga percepita a livello individuale e collettivo in un determinato ambito;
c. un parametro basato sull'osservazione delle conseguenze eventualmente prodotte, volto a identificare l'effetto reale causabile da quella interferenza, contemplato in termini economici e di benessere[45].

Con riferimento al parametro a) esso indubbiamente risulta, agli occhi del giurista, il criterio più affidabile in quanto si fonda, in primo luogo, sulla valutazione di una violazione congiunta di fonti primarie (previsioni costituzionali, della Carta dei diritti fondamentali, Convenzione Edu) e secondarie attuative di un determinato diritto fondamentale. Nell'ambito di tale criterio possono risultare determinanti anche ulteriori sub-criteri "interni" quali:

1. le indicazioni, contenute all'interno dello stesso atto normativo/previsioni coinvolte, utili a classificare e inquadrare l'intensità di una violazione, le

whether the procedures for supervising the ordering and implementation of therestrictive measures are such as to keep the "interference" to what is "necessary in a democratic society» (§232). Con riferimento alla Corte di Giustizia, si richiama, a titolo meramente esemplificativo, i passaggi dedicati delle note sentenze *Digital Rights Ireland* (§61) e *Google Spain* (§81).

45. G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, in *Computer Law & Security Review*, n. 56, 2025, p. 5 ss.

quali offrono indicazioni circa la severità della interferenza, come ad esempio indicato all'art. 83 Gdpr[46];
2. l'interpretazione offerta dalla giurisprudenza delle Corti. Difatti, benché manchi una "gerarchia" fra diritti, non solo la Corte Edu si riferisce esplicitamente ad alcune disposizioni quali «*most fundamental provisions*», com'è noto, ma definisce la presenza di diritti non derogabili, nemmeno nelle situazioni emergenziali[47].

Rientrano all'interno di una valutazione oggettiva volta a stabilire la "severità", la valutazione della reversibilità del danno e la durata della violazione.

Con riferimento alla reversibilità, si tratta di un criterio più solido perché si intende soppesare l'entità della violazione al di là della percezione soggettiva del danno ma in termini di "obiettiva rimediabilità". In tale prospettiva, ad esempio, si considera irreversibile una mancata assunzione di un individuo per il suo orientamento sessuale in grado di violare il diritto antidiscriminatorio europeo (Directive 2000/78/EC) e l'art. 21 della Carta dei diritti fondamentali, mentre si potrebbe considerare di "lieve" severità, in termini di reversibilità, una mancata applicazione del principio di trasparenza nella fornitura di un servizio di intermediazione digitale (art. 15 Dsa) in grado di incidere negativamente sulla libertà di espressione[48]. La valutazione della durata della violazione

46. L'art. 83, par. 2, Gdpr afferma infatti che in ogni singolo caso si debba tener conto di alcuni elementi, fra i quali, ad esempio, «a) la natura, la gravità e la durata della violazione tenendo in considerazione la natura, l'oggetto o a finalità del trattamento in questione nonché il numero di interessati lesi dal danno e il livello del danno da essi subito; b) il carattere doloso o colposo della violazione; c) le misure adottate dal titolare del trattamento o dal responsabile del trattamento per attenuare il danno subito dagli interessati; d) il grado di responsabilità del titolare del trattamento o del responsabile del trattamento tenendo conto delle misure tecniche e organizzative da essi messe in atto».

47. Gli artt. 2 e 3 della Convenzione vengono qualificati come disposizioni fondamentali in quanto idonee a rappresentare i valori fondativi del Consiglio d'Europa e della società democratica. Alcuni diritti, inoltre (artt. 15, 3, 4) sono esplicitamente inderogabili anche in circostanze emergenziali. Si tratta, tuttavia, di un sub criterio da implementare, eventualmente, con un certo grado di attenzione, considerando che alcuni diritti, pur non essendo espressamente qualificati come "fondamentali", potrebbero essere declassati su un gradino inferiore (ad. es art. 14) e, soprattutto, che altri diritti contengono, analogamente al menzionato art. 83 del Gdpr, delle linee interne volte a stabilire le condizioni e i limiti apponibili al diritto stesso, come dimostra la stessa libertà di espressione. Il compito di individuare quali siano le circostanze che giustificano la violazione, seppur tassativamente indicate dall'articolo, comporterebbero che in fase di valutazione preventiva si prefiguri un bilanciamento fra diritti fondamentali (la classica questione fra libertà di espressione e tutela della dignità individuale). Oltretutto, il collegamento fra disposizioni e legislazione secondaria attuativa non è sempre così diretto, come per alcune categorie di diritti fondamentali.

48. Così G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, cit., p. 10.

emerge, invece, come requisito che consente di attribuire al tempo, inteso sia in termini di svolgimento prolungato sia di frequenza dell'accadimento, un peso in termini di entità della interferenza[49].

Infine, risulta determinante nell'integrazione del primo criterio la valutazione della sussistenza di una violazione congiunta di diritti fondamentali, alla luce della logica generale secondo la quale una potenziale violazione contestuale di più previsioni concernenti i diritti fondamentali dovrebbe accrescere l'entità dell'interferenza.

La mera sussistenza di un criterio oggettivo, tuttavia, è apparsa sotto alcuni profili parziale. La violazione congiunta di più diritti fondamentali può rivelarsi, ad esempio, un sub-criterio di valutazione che non resiste alle accuse di eccessivo formalismo, considerando che la medesima obbligazione può essere contenuta in più articoli: questi ultimi potrebbero stabilire la stessa prescrizione sotto il profilo sostanziale senza che questo legittimi, come affermato dalla stessa Cgue, a effettuare un "*double-counting*" nel momento dell'esame dell'entità di un'interferenza[50].

Dall'altra parte, l'avvicinamento verso una "classificazione" di valore dei diritti fondamentali, letto attraverso l'ottica offerta dalla variegata giurisprudenza chiamata a effettuare un centrale filtro ermeneutico anche in fase di costruzione di una valutazione preventiva, risulta un'operazione delicata. Non si tratta soltanto di porre l'attenzione sul noto e delicatissimo problema teorico della costruzione di una gerarchia fra diritti, evidenziato anche dalla nostra giurisprudenza interna[51] ed esasperato in una dimensione europea; quanto piuttosto attribuire ulteriormente al soggetto privato non solo il faticoso compito di valutare *a priori* il "peso" di un diritto ma anche quello di dover svolgere un bilanciamento a tutto tondo[52].

Per tale ragione, è stato opportunamente suggerito di integrare il criterio oggettivo con un parametro soggettivo che lasci spazio anche alla percezione individuale e collettiva di una determinata violazione. Tale operazione consisterebbe in uno sforzo di "traslazione normativa" del significato sociale attri-

49. Si fa riferimento, ad esempio, a una privazione prolungata di una libertà o la ricorrente violazione di un parametro in termini di contesto, soggetti coinvolti, modalità della violazione.

50. Come ribadito dalla stessa Corte di Giustizia, cfr. CJEU, C-741/21, 11 aprile 2024, in cui esclude che la violazione congiunta di più previsioni del Gdpr determini una plurima violazione del diritto, riconoscendo l'identità del medesimo bene tutelato.

51. Fra le tante pronunce riconducibili al tema si richiama la paradigmatica sentenza della Corte costituzionale n. 85/2013, punto 9 del *Considerato in diritto*.

52. Si pensi all'articolo 10 che tutela la libertà di espressione, il compito di individuare quali siano le circostanze che giustificano la violazione, seppur tassativamente indicate dall'articolo, comporterebbero che in fase di valutazione preventiva si prefiguri un bilanciamento fra diritti fondamentali (la classica questione fra libertà di espressione e tutela della dignità individuale).

buito a una violazione, attribuendo così un ruolo chiave, in relazione a una potenziale lesione di un diritto fondamentale, alla «*perception prevailing in society at a given time*»[53].

Dietro tale sforzo integrativo, che tenta di introdurre una «*normative assessment of social significance*» emergerebbe il tentativo di flessibilizzare il parametro oggettivo, dando alla Fria quella ulteriore elasticità che le consentirebbe maggior completezza in termini di precomprensione del danno potenziale[54]. Si tratta, d'altra parte, della costruzione di un criterio estremamente delicata la quale, come osservato, può facilmente sovrapporsi con la valutazione delle conseguenze negative sul benessere del singolo. La granularità di tale parametro emerge maggiormente quando il parametro soggettivo rileva anche in chiave individuale, al fine di cogliere la percezione prettamente personale del danno eventuale.

Il parametro basato sulla valutazione degli effetti (c) supera la dimensione normativo-oggettiva e quella soggettiva e propone di classificare la «*severity of the expected consequences*» ossia il potenziale pregiudizio che un determinato sistema potrebbe causare nell'esercizio di un diritto fondamentale, utilizzando come valore di riferimento il benessere individuale, valutato alla luce della salute fisica e mentale e delle possibilità economiche eventualmente pregiudicate[55].

53. Si richiama, in tal modo, l'articolato intervento dell'Avvocato Generale Sánchez-Bordona, v. *Opinion of Advocate General Campos Sánchez-Bordona delivered on 6 October 2022, UI v Österreichische Post AG*, in cui si afferma «I am in no doubt that there is a fine line between mere upset (which is not eligible for compensation) and genuine non-material damage (which is eligible for compensation) and I am also aware of how complicated it is to delimit, in the abstract, the two categories and apply them to a particular dispute. That difficult task falls to the courts of the Member States, which will probably be unable to avoid in their rulings the perception prevailing in society at a given time regarding the permissible degree of tolerance where the subjective effects of infringement of a provision in this area do not exceed a de minimis level».

54. Così G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, cit., p. 12. Al fine del corretto inquadramento, la dottrina a supporto individua la consultazione di esperti indipendenti del settore come una via chiave soprattutto perché offrirebbero anche il supporto empirico ovvero basato su una valutazione effettuata sul gruppo maggiormente colpito rispetto a un determinato diritto. A tal scopo si richiama il principio 18, par. 1, lett. *b*, dei *Guiding principles on Business and Human Rights* tracciati dalle Nazioni Unite, in cui si afferma che «in order to gauge human rights risks, business enterprises should identify and assess any actual or potential adverse human rights impacts» sia opportuno effettuare «meaningful consultation with potentially affected groups and other relevant stakeholders, as appropriate to the size of the business enterprise and the nature and context of the operation».

55. G. Malgieri, C. Santos, *Assessing the (Severity of) Impacts on Fundamental Rights*, cit., p. 17.

2.3. Alcune prospettive di concreta implementazione della Fria

Proprio in forza di tali preliminari presupposti teorico applicativi tracciati e delle criticità segnalate, un'ulteriore riflessione dottrinaria ha condotto alcuni rilevanti tentativi di implementazione concreta di tale modello: è emersa progressivamente un'ottica semi-quantitativa, la quale sembra tener conto proprio della peculiarità dell'oggetto da valutare, a cavallo fra dato tecnico e sfumature di valore.

Fra gli stessi si segnala, da un lato, l'interessante proposta volta a costruire un modello di Fria in tre momenti operativi[56]. Il primo step consisterebbe in un inquadramento della tecnologia adoperata: un'analisi finalizzata a comprendere se la stessa possa qualificarsi come sistema ad alto rischio, ai sensi delle previsioni del regolamento. In caso di esito positivo, si passerebbe a due ulteriori momenti di valutazione[57].

Il primo, più strettamente tecnico, servirebbe a produrre il cosiddetto Qri, ovvero una valutazione del rischio sotteso al sistema di AI testato, individuato sulla base di un *range* da 1 a 10[58]. Il Questionario serve a comprendere le caratteristiche intrinseche della tecnologia utilizzata, la sua applicazione e le sue caratteristiche operative[59]. Per far ciò, vengono dunque selezionati *key risk factors* come *fairness, transparency, human oversight* e assegnato un punteggio che va da uno a 10 sulla base di una classifica basata su una scala predefinita (Qri). Se in questa fase emerge un valore oltre una determinata soglia la valutazione si blocca perché per il sistema quel parametro è già troppo alto e quindi concepito come inaccettabile. Questa prima valutazione, in sintesi, pone delle basi numeriche concernenti il potenziale rischio intrinseco della tecnologia e della sua applicazione sulle quali poi si edificherà la matrice che guarda ai diritti fondamentali[60].

Quest'ultima consiste nel secondo momento operativo chiave. La valutazione si struttura come precipuo test per comprendere l'impatto della tecnolo-

56. Il modello sperimentale elaborato da A. Cosentini *et al.*, *Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights*, in *MediaLaws*, 4 marzo 2025, spec. par. 3 ss.

57. Un modello di analisi ripreso e sviluppato analogamente in S. Bertaina *et al.*, *Fundamental rights and artificial intelligence impact assessment: A new quantitative methodology in the upcoming era of AI Act*, in *Computer Law & Security Review*, vol. 56, 2025, p. 5 ss.

58. Cfr. A. Cosentini *et al.*, *Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights*, cit., par. 3 ss.

59. Proprio perché può essere somministrato prima o dopo il *deployment*, lo stesso quindi si sdoppia in un questionario elaborato in fase di sviluppo della tecnologia e poi in fase operativa e applicativa.

60. Cfr. A. Cosentini *et al.*, *Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights*, cit., par. 3 ss.

gia sui diritti, così come definiti dalla Carta dei diritti fondamentali dell'Unione. Esige, dunque, un metodo semiquantitativo che si fonda sull'attribuzione di *Impact Scores* (IS) rispetto al singolo diritto fondamentale esaminato. Per essere efficace la matrice deve:

1. andare oltre i dati di *training* del sistema;
2. avere come panorama di individui presi a riferimento un numero adeguato e vario di persone potenzialmente capaci di subire l'impatto, come previsto dall'articolo 10 dell'Aia.

Nello svolgimento di tale analisi, vengono presi in considerazione tutti i diritti fondamentali previsti dalla Cfreu ma viene svolta una valutazione separata per ciascun diritto. Per ogni diritto fondamentale della Carta viene elaborato un IS score che va da 0 a 100 (con 1 indicante il minimo rischio e 100 indicante il massimo rischio).

Infine, il test tecnico e la valutazione sui diritti fondamentali devono operare congiuntamente per offrire un risultato sinottico: il questionario offre informazioni qualitative e una comprensione situazionale della macchina, la matrice offre un'analisi semiquantitativa relativa a tutti i diritti fondamentali[61].

La Fria può dirsi così conclusa nel momento in cui si perviene a un valore finale che sia il complesso dei valori raggiunti che combino il Qri con il IS in un unico indice[62].

Di analogo rilievo e di simile approccio appare la metodologia sperimentale HH4AI (*Human Rights & Human-centered AI Impact Assessment*), concepita come un quadro strutturato in linea con i requisiti dell'AI Act ma di diretta ispirazione ai principali *framework* internazionali in materia di diritti umani. Essa nasce dall'esigenza di superare la frammentazione degli approcci esistenti, integrando dimensioni giuridiche, etiche e tecniche in un unico processo coerente[63].

La metodologia si fonda su un approccio di valutazione progressiva "a strati" (*gate-based*), che consente di filtrare progressivamente i rischi rilevanti e di concentrare l'analisi approfondita solo sugli aspetti che presentano un poten-

61. Se il punteggio è inferiore a 50 il sistema ha passato il test; se il punteggio è uguale o superiore a 75 è necessario richiedere una seria analisi delle sezioni interessate; se il punteggio è medio (fra 50 e 75) bisogna mettere in atto necessari interventi e riprogrammare le fasi 2 e 3.

62. Nel calcolare il risultato finale si attribuisce un peso maggiore (del 70%) al valore IS perché è reputata maggiormente idonea nel cogliere la granularità e pervasività del danno potenziale e ha una base quantitativa più oggettiva. Cfr. A. Cosentini *et al*., *Assessing the Impact of Artificial Intelligence Systems on Fundamental Rights*, cit. par. 3.3.

63. P. Ceravolo *et al*., *HH4AI: a methodological framework for ai human rights impact assessment under the EU AI Act*, cit. p. 9 ss.

ziale impatto significativo sui diritti fondamentali. Tale struttura risponde a un'esigenza di proporzionalità e di efficienza, evitando valutazioni eccessivamente onerose per sistemi che non presentano criticità rilevanti[64].

Il processo HH4AI si articola in quattro fasi principali. La prima consiste nella raccolta di informazioni generali sul sistema di IA oggetto di valutazione. Questo step ha l'obiettivo di delineare il contesto di utilizzo, le sue finalità, le principali funzionalità e gli attori coinvolti. Questa analisi consente di determinare se la tecnologia rientri tra quelli classificabili come ad alto rischio ai sensi dell'AI Act e di individuare, in via preliminare, i diritti fondamentali potenzialmente coinvolti[65].

La seconda fase introduce una *checklist* sui diritti fondamentali, concepita come uno strumento di *screening* preliminare concernente i beni tutelati. Tale disamina si basa sui principali diritti riconosciuti dalle Carte internazionali ed europee, tra cui la Carta dei Diritti Fondamentali dell'Unione Europea, la Convenzione Europea dei Diritti dell'Uomo e la Dichiarazione Universale dei Diritti Umani.

L'obiettivo è identificare rapidamente se, e in che misura, il sistema di IA possa incidere su specifici sfere e principi oggetto di tutela. La analisi consente di individuare quali diritti richiedano un'analisi approfondita nella fase successiva e quali, invece, possano essere esclusi dalla valutazione dettagliata, riducendo così la complessità del processo.

La terza fase rappresenta il cuore della metodologia HH4AI ed è dedicata alla valutazione approfondita dell'impatto del sistema di IA sui diritti fondamentali individuati come rilevanti. In questa fase, l'analisi si concentra su specifiche aree principali di impatto, quali, ad esempio, *accountability* e trasparenza, per le quali vengono stimate la probabilità e la gravità del potenziale impatto negativo sui diritti fondamentali, tenendo conto del contesto di utilizzo e delle misure di mitigazione già previste[66].

L'approccio adottato appare coerente con la logica *risk-based* dell'AI Act e consente di distinguere tra rischi accettabili, rischi che richiedono misure correttive e rischi non accettabili che potrebbero rendere il sistema incompatibile con i requisiti normativi.

Infine, l'ultima fase della metodologia HH4AI produce un *output* strutturato che sintetizza i risultati della valutazione e individua le misure di mitigazione necessarie per ridurre i rischi individuati. Quest'ultime possono includere misure quali raccomandazioni tecniche fino a interventi di natura più pret-

64. Ivi, p. 10.
65. Ivi, p. 9.
66. *Ibidem.*

tamente giuridica che supportino il *deployer* nell'adempimento degli obblighi stabiliti dal regolamento.

3. Conclusioni

Il percorso di riflessione sopradescritto promuove una prospettiva di ricomposizione teorica e di implementazione concreta fra le peculiari sfide del contesto virtuale e la natura immutabile dei diritti fondamentali, utilizzando la Fria come modello di un più generale orientamento normativo.

Tale andamento racchiude con sé alla radice, a tutta evidenza, l'ormai quasi decennale dibattito sul costituzionalismo digitale. Quest'ultimo, al di là delle sfaccettate definizioni emerse in dottrina, persegue l'ambiziosa finalità di traslare gli obiettivi sottesi alla legalità costituzionale e ai principi della *rule of law* nella dimensione dell'innovazione tecnologica e digitale[67]. Il modello di *impact assessment* sopra analizzato non sarebbe altro, infatti, che un tentativo ulteriore, dopo le prescrizioni del Dsa[68], di implementazione, sempre più articolata e sofisticata, di istituti, principi e paradigmi garantistici tipici del costituzionalismo "classico" in presenza di attori, contesti e fenomeni differenti[69].

Dietro tale direzione legislativa emergerebbe, tuttavia, quella premessa imprescindibile secondo la quale i fenomeni che il decisore politico contemporaneo si trova a dover affrontare non costituirebbero problematiche nuove quanto piuttosto criticità e interrogativi sempre esistenti che la globalizzazione e l'affermazione della dimensione virtuale e tecnologica avrebbero semplicemente esasperato, sfruttandone la forza iper-generativa[70].

67. L. Gill, D. Redeker, U. Gasser, *Towards Digital Constitutionalism? Mapping Attempts to Craft an Internet Bill of Rights, in Berkman Center Research Publication*, n. 1, 2015, p. 1 ss.; G. De Gregorio, *The rise of digital constitutionalism in the European Union*, in *International Journal of Constitutional Law*, vol. 19, n. 1, 2021, p. 41 ss. In termini definitori ugualmente centrale è stata la riflessione di E. Celeste, *Digital constitutionalism: a new systematic theorisation*, in *International Review of Law, Computers & Technology*, vol. 33, n.1, 2019, p. 81; per un apporto a tale dibattito più risalente cfr. V. Karavas, *Governance of Virtual Worlds and the Quest for a Digital Constitution*, in C.B. Graber, M. Burri-Nenova (eds.), *Governance of Digital Game Environments and Cultural Diversity: Transdisciplinary Enquiries*, Elgar, 2010.

68. G. De Gregorio, *The Digital Services Act: A Paradigmatic Example of European Digital Constitutionalism*, in *Diritti Comparati*, 17 maggio 2021.

69. Sull'evoluzione di tale percorso, ancora, cfr. G. De Gregorio, *Digital constitutionalism across the Atlantic*, in *Global constitutionalism*, n. 11, 2022, p. 297 ss. e il già citato lavoro monografico dello stesso autore cfr. Id., *Digital constitutionalism in Europe*, cit., p. 41 ss.

70. Sul punto cfr. A. Jr. Golia, *The Critique of Digital Constitutionalism*, in *Max Planck Institute for Comparative Public Law & International Law – MPIL*, Research Paper, 2022, p. 12.

Questo presupposto logico ha portato con sé una conseguenza: se la società non è costretta a confrontarsi con un problema intrinsecamente e sostanzialmente differente la risposta regolatoria può essere strutturalmente identica, ossia quella ereditata dalle tradizioni giuridiche degli Stati membri e accolta anche nella Carta dei diritti fondamentali dell'Unione la quale attinge, appunto, a strumenti garantistici del costituzionalismo europeo[71].

Purtuttavia, l'adattamento dello strumentario di tutele proprie della dimensione *offline* continua a esser accompagnato, a oggi, da irrisolte criticità. Non si tratta soltanto di evidenziare come l'azione intrapresa dal legislatore unionale di circoscrivere l'impatto delle nuove tecnologie e i nuovi poteri privati mediante un quadro strutturato e organico di garanzie poste a tutela della persona rimanga esso stesso limitato e talvolta schiacciato sull'orizzonte procedurale, senza un adeguato supporto sostanziale[72]; quanto, più precisamente, fare i conti con il fatto che alcuni istituti tipici dello Stato di diritto costituzionale che presiedono la limitazione di un diritto fondamentale appaiono solo parzialmente rimodellabili e implementabili, come dimostra il *framework* preso in esame.

Sebbene l'AI Act compia, in una visione sinottica, passi ulteriori nell'espandere e dettagliare le modalità di massimizzazione della tutela dei diritti[73], come già accennato, ugualmente presenta in alcune sue previsioni-chiave una debolezza insuperabile che intacca elementi costitutivi della *rule of law*, nonostante i tentativi di ibridazione e gli sforzi di traduzione applicativa emersi alla luce dell'ampia riflessione dottrinaria ricostruita.

Tali criticità operative concernenti il corretto inquadramento delle tecnologie e l'adozione delle idonee misure preventive imposte dal Regolamento verrebbero soprattutto esasperate dalla difficoltà di non riuscire a disporre di modelli di implementazione soddisfacenti di tali ambiziosi *impact assessments* che

71. O. Pollicino, *Libertà di espressione, piattaforme digitali e cortocircuiti di natura costituzionale*, in *Privacy&*, n. 1, 2021, p. 7.

72. Per lo sviluppo di questa tesi si veda l'intervento di L. Califano, *La strategia normativa dell'Unione europea per un nuovo ordine digitale*, nel fascicolo speciale a cura di L. Califano, F. Fabrizzi, G. Sartor, *Information disorder e sistema democratico*, in *federalismi.it*, n. 15, 2025.

73. Sul punto, *ex multis,* cfr. ancora la ricostruzione multi prospettica contenuta nel volume a cura di F. Fabrizzi, L. Durst, *Controllo e predittività*, cit., spec. pp. 7-69. In chiave prospettica si veda altresì M. Orofino, *Obiettivi, ambito di applicazione e principi fondamentali dell'AI Act*, in F. Pizzetti, S. Calzolaio, A. Iannuzzi, E. Longo, M. Orofino (a cura di), *La regolazione europea della società digitale*, Giappichelli, 2024, p. 35 ss.; si veda anche A. Adinolfi, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno alla loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione*?, vol. I, il Mulino, 2022, p. 156.

risultino uniformi ed efficaci rispetto alla materia trattata. La Fria e le sue prime vicende e ipotesi adattive testimoniano la essenziale e primigenia difficoltà di coniugare modelli semiquantitativi di giudizio con l'ambito ontologicamente qualitativo dei diritti fondamentali, caratterizzati per loro natura dal loro essere situati e proni a sfumature di valore[74].

Di fronte a tale quadro che pare sempre più connotato da slanci regolatori e tentativi pioneristici di costruzione di soluzioni applicative univoche, è difficile, a ogni modo, lasciarsi dietro la convinzione che tale paradigma possa reputarsi una via imperfetta ma più facilmente percorribile, data l'edificazione della stessa su radici garantistiche familiari e consolidate, le quali rassicurerebbero in merito alla costruzione di un adeguato ancorché imperfetto sistema di salvaguardia dei diritti fondamentali nel cyberspazio[75].

Dall'altra parte, tuttavia, le opacità teoriche e la difficile traduzione applicativa che numerosi di tali istituti mostrano rendono sempre meno distanti quelle elaborazioni alternative che interrogano le asserite "camaleontiche" potenzialità del costituzionalismo: si pone in dubbio, in sintesi, la sua imperitura capacità di modellarsi, lasciando intatta la sua forza di reazione a nuove forme di potere.

Le numerose incertezze che avvolgono tale re-impianto valoriale in uno scenario con attori e fenomeni nuovi contribuirebbero a rafforzare, ad esempio, le contro argomentazioni che suggeriscono di vagliare le potenzialità del *societal constitutionalism*, alla base del quale giacerebbe proprio la convinzione dell'inadeguatezza dei paradigmi tradizionali nel poter reagire alle svolte prodotte dalla globalizzazione nell'epoca contemporanea. In quest'ultima la frammentazione dei tradizionali rapporti di potere imporrebbe di andare oltre i classici modelli fondati sulla *rule of law*, sulla separazione dei poteri e sul ruolo garantistico dell'autorità giudiziaria[76] in quanto sistemi reputati inadatti a far fronte a una sfera pubblica "polverizzata" in cui si ergono nuove forme di dominio nei confronti della persona. Tali fenomeni si identificherebbero ben oltre i tradizionali poteri politici e sociali ma si configurerebbero, nell'ottica di tale teorica, quali nuovi "centri di potere": un mutamento di attori che

74. Sul profilo di metodo cfr. E. Longo, *La disciplina del "rischio digitale"*, in F. Pizzetti, S. Calzolaio, A. Iannuzzi, E. Longo, M. Orofino (a cura di), *La regolazione europea della società digitale*, Giappichelli, 2024, p. 75 ss.

75. M. Betzu, *Poteri pubblici e poteri privati nel mondo digitale*, in *Rivista "Gruppo di Pisa"*, n. 2, 2021, pp. 180-181; G. Palombella, *È possibile una legalità globale?,* il Mulino, 2012, p. 161.

76. Sui "nuovi" principi guida dell'era digitale, si veda la interessante proposta di E. Longo, A. Pin, *Oltre il costituzionalismo? Nuovi principi e regole costituzionali per l'era digitale*, in *Diritto pubblico comparato ed europeo*, n. 1, 2023, p. 12 ss.

esigerebbe di tagliare quel legame indissolubile fra costituzionalismo e soggetto pubblico statuale[77].

Più chiaramente, dunque, in un contesto in cui l'ordinamento giuridico è ormai ben distante dalla sua formulazione originaria, e in uno scenario in cui si intersecano inedite forme di coercizione, molto più sfumate, diffuse e policentriche, nonché sistemi normativi sovrapposti, non è facile non volgere lo sguardo verso teorie antitetiche che fanno appello a quelle intrinseche capacità di produzione normativa di un sistema sociale.

Si immaginerebbe, in sintesi, la costruzione alternativa di una costituzionalizzazione dei processi che non passa necessariamente attraverso lo Stato e la sua architettura tradizionale ma si nutre dell'interazione fra nuovi attori che, attivamente o passivamente, divengono protagonisti di nuove forme di dominio.

In tale scenario, quindi, appare sempre più lecito contemplare, anche in un'ottica di integrazione fra modelli di pensiero, la convinzione che istituti, regole e garanzie non possano configurarsi soltanto in modo "responsivo" di fronte a fenomeni inediti, affrettandosi a adattare all'occorrenza l'armamentario garantistico a disposizione del costituzionalismo liberale; ma, in una prospettiva nuova, che siano e debbano risultare normativamente e attivamente mobili e reattivi in presenza di un esistente mutato[78].

77. Per un ulteriore approfondimento teorico cfr. A. Jr Golia, G. Teubner, *Societal Constitutionalism in the Digital World: An Introduction*, *Max Planck Institute for Comparative Public Law & International Law – MPIL Research Paper*, n. 11, 2023, p. 1 ss. Per una collocazione della teorica nel più ampio dibattito sul global constitutionalism, cfr. A. Wiener *et al*., *Global constitutionalism: Human rights, democracy and the rule of law*, in *Global Constitutionalism*, vol. 1, n. 1, 2012, p. 3 ss.

78. A. Jr. Golia, *The Critique of Digital Constitutionalism*, cit., p. 8.