



MASSIMILIANO BARONI

Il Whistleblower Tool dell'AI Act come infrastruttura di enforcement, tra garanzie e criticità

(Appunti per una lettura costituzionalmente orientata del nuovo canale di segnalazione dell'EU AI Office)

L'AI Act Whistleblower Tool, lanciato dalla Commissione europea tramite l'EU AI Office, riflette il più ampio passaggio dell'Unione verso un'infrastruttura di enforcement del diritto digitale costruita non soltanto su norme sostanziali, ma anche su procedure, canali di segnalazione e flussi informativi concepiti per sostenere un'azione amministrativa efficace. Il contributo propone un'analisi costituzionalmente orientata del Tool, inteso come infrastruttura procedurale capace di incidere sulle condizioni di accesso all'enforcement e sul controllo del potere tecnologico. Ne esamina il ruolo nell'emersione anticipata dei rischi e nel favorire la cooperazione tra differenti livelli regolatori, affrontando al contempo le tensioni tra la tutela delle libertà fondamentali e interessi contrapposti. L'analisi considera inoltre l'impatto del Tool sui sistemi interni di compliance e le principali criticità applicative, tra cui anonimato, gestione dei dati, obblighi di follow-up e overlap normativi. In conclusione, si sostiene che il Tool non debba essere considerato un mero portale di segnalazione, bensì una componente potenzialmente essenziale dell'evoluzione dell'infrastruttura europea di enforcement digitale.

AI Act – EU AI Office – Whistleblowing – Costituzionalismo digitale – Fonti

The AI Act Whistleblower Tool as an enforcement infrastructure: safeguards and critical issues

(Notes towards a constitutionally oriented reading of the EU AI Office's new reporting channel)

The AI Act Whistleblower Tool, launched by the European Commission through the EU AI Office, reflects the Union's broader shift toward an enforcement infrastructure for digital law built not only on substantive rules, but also on procedures, reporting channels and information flows designed to support effective administrative action. This article offers a constitutionally oriented analysis of the Tool as a procedural infrastructure shaping access to enforcement and the oversight of technological power. It examines its role in the early detection of risks and in fostering cooperation across regulatory levels, while also addressing the tensions arising between the protection of fundamental freedoms and competing interests. The analysis further considers the Tool's impact on internal compliance systems and key implementation challenges, including anonymity, data management, follow-up obligations, and regulatory overlap. It concludes that the Tool should be understood not merely as a reporting portal, but as a potentially essential component of the EU's evolving digital enforcement architecture.

AI Act – EU AI Office – Whistleblowing – Digital constitutionalism – Sources of law

L'Autore è Ricercatore a tempo determinato in Diritto costituzionale; Università degli Studi di Parma, Dipartimento di Giurisprudenza, Studi politici e internazionali. Attualmente Visiting Fellow – Public Law and Policy Program, UC Berkeley Law

SOMMARIO: 1. Premessa. Dal testo alla macchina dell'enforcement: il whistleblowing come infrastruttura regolatoria. – 2. Il Whistleblower Tool nell'architettura dell'AI Act: presupposto e destinatari; funzioni. – 2.1. Whistleblowing e governance dell'AI: perché un canale settoriale? – 3. Dalla Direttiva 2019/1937 all'innesto nell'AI Act: la tutela dei segnalanti nella transizione dalla governance al discorso sui diritti fondamentali. – 4. Riflessi applicativi per gli operatori nella compliance: integrazione con canali interni, governance dei flussi informativi, gestione del rischio. – 4.1. Il rischio di "over-reporting". – 5. Un canale, più livelli: alcune criticità. – 6. Dalla disciplina procedurale ad un profilo *para*-costituzionale?

1. Premessa. Dal testo alla macchina dell'enforcement: il whistleblowing come infrastruttura regolatoria

È ormai noto come l'attuale stagione del diritto europeo dell'innovazione¹ sia (sempre più) caratterizzata da una a dir poco elevata densità normativa, tale da aver suscitato tra gli addetti ai lavori non poche perplessità², di vario tipo e genere: in punto di opportunità, innanzitutto, oltre che per le difficoltà interpretative giocoforza derivanti da siffatto dedalo giuridico e, inevitabilmente, dall'efficacia dell'azione regolatoria unionale. Aspetti di cui sembra essersi ultimamente avveduto – pur con colpevole ritardo, d'altronde i segnali v'erano già tutti, da tempo³ – anche il legislatore europeo, come dimostrerebbe la recente virata oggi ben rappresentata dal pacchetto Omnibus.

Al di là del versante meramente quantitativo vi è tuttavia – a ben vedere – un secondo aspetto della c.d. *simplification* europea che merita maggiore attenzione, individuabile nella progressiva traslazione dalla mera enunciazione del precetto normativo alla volontà di costruirne le condizioni per una piena effettività.

Il lessico (pur spesso abusato) della "compliance by design" e della "governance", infatti, rappresenta non soltanto una specifica tecnica regolatoria, quanto anche una scelta di metodo: il diritto sostanziale non è più ritenuto sufficiente, cosicché l'Unione – consapevole della fragilità dell'enforcement tradizionale (e la prima decade di vita del GDPR è lì a dimostrarlo)⁴, specie in presenza di catene del valore transnazionali e opache⁵ – investe su infrastrutture innanzitutto procedurali che producono informazione, la qualificano, la rendono azionabile⁶. In questo quadro il Whistleblower

1. La pubblicazione si inserisce nell'ambito dello Spoke 1 "Human, social and legal aspects" dell'iniziativa "Security and Rights in CyberSpace – SERICS" a valere sulle risorse del NextGenerationEU – Piano Nazionale Ripresa e Resilienza (PNRR) - Missione 4 Componente 2 Investimento 1.3 – Avviso N. 341 del 15 marzo 2022 del Ministero dell'Università e della Ricerca; Codice progetto PE 0000014. Progetto "Human-centered approach and Regulatory Dimension in Developing an Interoperable and Secure Cyberspace" acronimo "HARD DISC" CODICE BANDO IIT SERICS SPOKE 1 N.1, CUP: B53C22003950001.

2. Su tutti RAFFIOTTA 2023.

3. Più approfonditamente POLLICINO 2025, sulla necessaria ridefinizione delle fonti del diritto nell'ecosistema digitale.

4. NOYB 2024, per cui "the GDPR has increased awareness of privacy but failed to establish a genuine European culture of compliance".

5. Sia consentito il rinvio a BARONI 2024-A.

6. Prendendo una strada solo apparentemente contraria a quella della semplificazione: introdurre, accanto alle norme primarie, ulteriori strumenti di applicazione del diritto europeo, non significa infatti necessariamente

Tool dell'AI Act⁷, lanciato dalla Commissione europea il 24 novembre 2025 e gestito dall'EU AI Office, merita di essere letto non come un mero accessorio tecnico-pratico o come un orpello documentale-organizzativo, ma come un diretto precipitato di questo nuovo corso: un dispositivo amministrativo-procedurale (di derivazione normativa) esemplificativo di una precisa idea di regolazione, attraverso cui inspessire il *fil rouge* che vuole legare l'innovazione⁸ alla fiducia di cui la medesima ha conclamato bisogno⁹. D'altronde, il modello europeo di regolazione del digitale c.d. "*rights-driven*"¹⁰ (nel quale il whistleblowing si inserisce a pieno titolo) assume quale preconditione necessaria che la fiducia dei cittadini nei sistemi di AI dipenda innanzitutto dalla capacità unionale di garantire il rispetto da parte di tali sistemi dei fondamentali principi etici e giuridici¹¹.

Cosa evidenzia, quindi, l'istituzione di questo canale di segnalazione settoriale dedicato all'AI? In primo luogo, che il rischio – soprattutto quello che incide su diritti fondamentali, salute e sicurezza – non può essere governato solo *ex post*, pretendendo di inseguire l'illecito dopo che esso ha già prodotto i suoi effetti. In secondo luogo, che l'informazione utile a intercettare le violazioni non è, il più delle volte,

manifesta e/o pubblicamente disponibile: piuttosto, essa si annida nei processi interni delle organizzazioni, nelle decisioni di progettazione, nelle scelte di deployment, e dunque in quell'area grigia in cui l'ordinamento è giocoforza capace di vedere solo se qualcuno vi fa luce. Il whistleblowing, da questo punto di vista, diviene una tecnologia giuridica di emersione e, di riflesso, anche un luogo – pur metaforico – di bilanciamento, tra un potenziale "denunciante" (spesso debole) e un potenziale incolpato (spesso forte), in cui si chiede all'autorità di trasformare una narrazione in un fatto giuridicamente rilevante non solo per il singolo, bensì anche a livello sistemico (non solo di private enforceability si tratta, insomma). Il che, inevitabilmente, passa anche attraverso garanzie procedurali e sostanziali.

Il tema non è in sé e per sé nuovo. Le note pronunce in materia della Corte europea dei diritti dell'uomo, da *Guja*¹² a *Heinisch*¹³ fino a *Gawlik*¹⁴, hanno mostrato come la segnalazione di illeciti ben possa intrecciarsi con l'art. 10 CEDU, specie alla luce dell'interesse pubblico alla notizia (e alla violazione di cui la segnalazione è vettore)¹⁵, imponendo di considerare il whistleblower non come un delatore, ma come un soggetto che – in certi contesti, in presenza di specifiche regole e principi¹⁶ – esercita una forma

appesantire il carico normativo dell'Ue, quanto piuttosto tentare di rendere Regolamenti e Direttive più facilmente azionabili e, in ultima istanza, più agili. Nello stesso disegno pare potersi inscrivere altresì il frequente ricorso – come dimostra esemplificativamente l'AI Act – a "nuove" forme regolatorie ibride pubblico-private: sul punto di recente DE MINICO 2025, che rievoca a tal fine il *droit doux* di J. Chevallier.

7. AI Act Whistleblower Tool: <https://ai-act-whistleblower.integrityline.app>.

8. Sempre più messa in crisi dall'azione regolatoria europea: sul punto si v. recentemente LONGO 2025, p. 294, che sottolinea il necessario rapporto bilaterale tra complessità tecnologica e governance flessibile.

9. "Lo strumento [così la Commissione in "La Commissione lancia uno strumento per gli informatori per la legge sull'IA"] offre un modo sicuro per segnalare potenziali violazioni della legge che potrebbero pregiudicare i diritti fondamentali, la salute o la fiducia dei cittadini", mirando "a promuovere l'innovazione e l'adozione dell'IA nell'UE, affrontando al tempo stesso i potenziali rischi per la salute, la sicurezza e i diritti fondamentali delle persone e salvaguardando la democrazia e lo Stato di diritto".

10. BRADFORD 2023, p. 105 ss.

11. DONATI 2025, p. 6, evidenzia la distanza (pratica e concettuale) tra il modello di matrice europea, quello cinese e quello made in USA.

12. *Guja v. Moldova*, n. 14277/04, 12 febbraio 2008.

13. *Heinisch v. Germany*, n. 28274/08, 21 luglio 2011.

14. *Gawlik v. Liechtenstein*, n. 23922/19, 16 febbraio 2021.

15. A mente del principio per cui "in uno Stato democratico governato dalla rule of law l'uso di metodi impropri da parte della pubblica autorità è precisamente il tipo di questione su cui il pubblico ha il diritto di essere informato". Lo ricorda PONTE 2021, p. 2712.

16. TEDESCO 2023, p. 748 ss., per cui trattasi di un vero e proprio "sotto-sistema, a sé stante, precipuamente dedicato al contesto lavorativo ed ispirato alla protezione dei dipendenti segnalanti", pienamente colto dalla Raccomandazione

di libertà di informazione a tutela dell'interesse generale. Permettendo così al whistleblowing di emanciparsi dalla propria condizione originaria di (mero) strumento di governance per divenire strumento *human rights oriented*, di rafforzamento della *rule of law* nelle molte dimensioni in cui questa si declina oltre che nel suo rapporto con democrazia e diritti fondamentali della persona¹⁷.

Il legislatore unionale, con la Direttiva UE 2019/1937, ha poi sistematizzato questa intuizione dentro un disegno di enforcement europeo, riconoscendo che le segnalazioni costituiscono uno degli "elementi a monte" (*upstream component*)¹⁸ dell'efficacia del proprio impianto normativo e, al tempo stesso, che di tale efficacia la protezione contro ritorsioni e la presenza di canali affidabili sono elementi indifferibili.

Il passaggio ulteriore – e, per certi versi, più interessante – consiste oggi nell'aggiuntivo) innesto settoriale: non più soltanto una tutela orizzontale, ma canali dedicati, capaci di permettere il dialogo con autorità ad elevata specializzazione (e non potrebbe essere diversamente in regimi giuridici – e per settori – complessi, nei quali la prova dell'illecito è sovente parte di documenti, dataset, log, modelli, e dunque di materiale che esige competenze e procedure di verifica adeguate). La protezione dei segnalanti, così, non è più solo sinonimo di mera tutela individuale, neppure solo di una garanzia "sociale", ma si intreccia anzi con i nuovi percorsi di regolazione, per divenire una delle leve applicative dell'enforcement digitale europeo.

L'obiettivo delle brevi riflessioni che seguiranno è duplice: da un lato, collocare il Whistleblower Tool nel rinnovando ecosistema normativo europeo, in particolare evidenziandone (e circoscrivendone)

la funzione sistemica nell'architettura dell'AI Act; dall'altro, interrogare lo strumento da una prospettiva specializzata, chiedendosi quali garanzie esso presupponga e quali rischi produca, considerando il bilanciamento tra tutela dei segnalanti, diritti degli operatori economici, protezione dei dati, segreto industriale e giusto procedimento. L'ipotesi di fondo è che qui si giochi una partita più ampia: la trasformazione dell'enforcement in sistema (anche) procedurale, e la necessità dunque che tale sistema resti compatibile con un orizzonte di legalità, evitando che il delicato equilibrio tra efficacia e garanzie penda verso un solo piatto della bilancia.

2. Il Whistleblower Tool nell'architettura dell'AI Act: presupposto e destinatari; funzioni

Occorre fin d'ora delimitare forma (e confini) dell'istituto, anche per coglierne con maggiore precisione la collocazione sistematica rispetto alla Direttiva 2019/1937¹⁹. Quest'ultima, come noto, delinea una pluralità di vie protette – segnalazione interna, segnalazione esterna e, alle condizioni più rigorose previste dal relativo art. 15, divulgazione pubblica – lasciando al segnalante la possibilità di rivolgersi direttamente all'autorità competente. Il previo ricorso al canale interno è sì incoraggiato, ma soltanto quando la violazione possa essere affrontata efficacemente all'interno dell'organizzazione e non vi sia ragione di temere ritorsioni (e la *ratio* del modello è evidente: preservare, da un lato, la capacità dell'ente di conoscere e correggere tempestivamente la deviazione; evitare, dall'altro, che il soggetto cui la violazione è imputabile conservi il controllo esclusivo sulla sua emersione).

del Comitato dei Ministri del Consiglio d'Europa del 30 aprile 2014 (Recommendation CM/Rec(2014)7 of the Committee of Ministers to member States on the protection of whistleblowers).

17. PARISI 2020, p. 12.

18. "A livello di Unione le segnalazioni e le divulgazioni pubbliche degli informatori costituiscono uno degli elementi a monte dell'applicazione del diritto e delle politiche dell'Unione. Essi forniscono ai sistemi di contrasto nazionali e dell'Unione informazioni che portano all'indagine, all'accertamento e al perseguimento dei casi di violazione delle norme dell'Unione, rafforzando in tal modo i principi di trasparenza e responsabilità" (Direttiva UE 2019/1937, Considerando n. 2).

19. Giova evidenziare come sia attualmente in corso presso la Commissione una valutazione della Direttiva 2019/1937: nel luglio 2026 è stata pubblicata la sintesi fattuale della consultazione pubblica (Staff working document on the evaluation of Directive (EU) 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report breaches of Union law (the 'Whistleblower Protection Directive'), doc. Ares(2026) 6975335); la valutazione complessiva è programmata per l'ultimo trimestre del 2026.

Il Whistleblower Tool dell'AI Act si innesta unicamente sul versante esterno di tale architettura, ma occorre sin d'ora circoscriverne natura e funzionamento, soprattutto alla luce della progressiva articolazione dei canali attraverso cui le informazioni relative a possibili violazioni dell'AI Act possono giungere all'amministrazione europea. Il Whistleblower Tool non rappresenta, infatti, l'unica via di accesso all'AI Office, dovendolo distinguere – *in primis* – dall'AI Act complaints tool²⁰: pur condividendo la funzione – *lato sensu* informativa – di portare all'attenzione dell'autorità fatti potenzialmente rilevanti ai fini dell'enforcement, si tratta di strumenti che rispondono a presupposti soggettivi, fondamenti normativi e logiche di tutela differenti; e proprio il raffronto tra i due canali consente di comprendere con maggiore precisione la specificità del whistleblowing nel disegno dell'AI Act.

Guardando al fondamento giuridico dei due strumenti, come si avrà modo di meglio specificare, l'art. 87 dell'AI Act non istituisce in senso proprio il Whistleblower Tool, né definisce autonomamente una procedura europea di segnalazione; operando piuttosto quale norma di raccordo, stabilendo l'applicabilità della Direttiva 2019/1937 alla segnalazione delle violazioni dell'AI Act e alla protezione delle persone che le segnalano. Ne deriva che l'identità giuridica del whistleblower continua a dipendere dal paradigma proprio della Direttiva e, in particolare, dall'acquisizione delle informazioni in un contesto lavorativo, nel significato ampio fatto proprio dal relativo art. 4 (non soltanto lavoratori subordinati, dunque, ma anche autonomi, azionisti, componenti degli organi societari, volontari, tirocinanti, e in generale gli stakeholder coinvolti nella catena del valore). Di più: coerentemente con tale disegno, infatti (e come recentemente²¹ chiarito dalla Commissione), l'attuale configurazione operativa del Whistleblower Tool ne delimita la platea alle persone *professionalmente* collegate ai provider di modelli di AI per finalità generali o ai provider di sistemi di AI rientranti

nella competenza di enforcement dell'AI Office. Il segnalante si trova dunque collocato in posizione qualificata rispetto all'organizzazione e, proprio per questo, è potenzialmente titolare di informazioni che l'autorità difficilmente potrebbe acquisire mediante la sola osservazione esterna del mercato. A tale asimmetria informativa corrisponde una specifica architettura di garanzia: la segnalazione può essere effettuata in forma anonima, in qualsiasi lingua ufficiale dell'Unione, mentre una secure inbox consente inoltre di mantenere il dialogo con l'AI Office, ricevere aggiornamenti e rispondere a eventuali richieste di chiarimento senza che venga meno l'anonimato: confidenzialità come precipitato funzionale della *ratio* antiritorsiva che attraversa l'intero modello della Direttiva 2019/1937.

Differente invece la struttura dell'AI Act complaints tool, che – qui sì – in diretta attuazione dell'art. 85 AI Act, ne istituisce una specifica modalità di esercizio, attribuendo – a qualunque persona fisica o giuridica che abbia motivo di ritenere sussistente una violazione del Regolamento relativa a sistemi di AI ricadenti nella competenza esclusiva dell'Ufficio – la possibilità di presentare un reclamo all'autorità di vigilanza del mercato competente: non richiedendo dunque una connessione professionale con il segnalato né con le modalità di apprensione dell'informazione, e – soprattutto – non prevedendo possibilità di segnalazioni anonime²².

In buona sintesi, mentre per il complaints tool assume rilievo in via principale l'accessibilità dell'amministrazione alla violazione, solo con il Whistleblower Tool il fulcro si sposta sulla protezione della fonte informativa: l'ordinamento attribuisce valore alla particolare posizione epistemica del soggetto inserito, in senso ampio, nel contesto professionale, e tenta di neutralizzare il costo individuale che potrebbe derivare dall'emersione dell'informazione (valorizzando il principio per cui conoscenze altrimenti destinate a rimanere interne alla struttura organizzativa possano essere

20. AI Act Complaints Tool, in digital-strategy.ec.europa.eu, aggiornato al 31 luglio 2026.

21. AI Act Whistleblower Tool, in digital-strategy.ec.europa.eu, aggiornato al 31 luglio 2026.

22. Il reclamo viene trattato confidenzialmente dall'AI Office e, qualora non ricada nella sua competenza, l'interessato ne viene informato. Previo consenso del reclamante e ove appropriato, la questione può inoltre essere trasmessa alla competente autorità nazionale di vigilanza del mercato o ad altra autorità responsabile dell'applicazione di obblighi unionali connessi alla tutela dei diritti fondamentali.

trasferite all'autorità senza che il loro autore sopporti un eccessivo rischio individuale).

Così letto, il Whistleblower Tool costituisce un canale esterno rispetto all'ente presso il quale l'informazione è stata acquisita e, al contempo, sovranazionale quanto alla sede cui la comunicazione è destinata; operando in certa qual misura come proiezione operativa della norma di raccordo (l'art. 87 AI Act), più che come assimilazione *tout court* tra questo e il sistema di protezione predisposto dalla Direttiva 2019/1937. Altrimenti detto, la Direttiva protegge la scelta tra canale interno ed esterno, mentre l'AI Act, attraverso il Tool, istituzionalizza e rafforza il secondo, collocandolo direttamente sul piano unionale, incidendo insieme sull'autonomia organizzativa dell'impresa, sulla posizione del segnalante (dal 2 agosto 2026)²³ e sulle modalità attraverso cui l'informazione privata entra nel circuito dell'enforcement pubblico²⁴.

Tanto più che l'AI Act Whistleblower Tool – lanciato come detto il 24 novembre 2025 – nasce in un momento, a ben vedere, peculiare per l'attuazione dell'AI Act: un tempo in qualche modo di transizione, caratterizzato da una applicabilità “per fasi” di molti degli obblighi previsti dal testo; davanti al quale l'Unione ha comunque ritenuto essenziale l'attivazione (anche) di strumenti di raccolta e trattamento delle informazioni utili. Il principale elemento d'interesse risiede non tanto nella descrizione funzionale dello strumento

– che, a conti fatti, non possiede caratteri di particolare novità – quanto nella decisione di ricondurre segnalazioni specializzate presso un neonato attore, l'AI Office appunto, costituito in seno al Regolamento e costruito proprio per concentrare competenze e capacità di intervento su un tema in rapido movimento, così creando un ecosistema di porte di ingresso specializzate, differenziate sulla base della posizione del segnalante, della natura dell'informazione e della competenza esercitata²⁵.

L'AI Office, istituito con decisione della Commissione del 24 gennaio 2024 e collocato all'interno della DG CNECT²⁶, pensato come centro di expertise e di coordinamento per l'implementazione dell'AI Act (e, come di recente modificato attraverso il Digital Omnibus, avente tra le altre cose competenza esclusiva sulla vigilanza e sul controllo applicativo dei sistemi di IA costruiti su modelli di IA di uso generale, a determinate condizioni)²⁷, ha invero il compito tra le altre cose di sostenere lo sviluppo e l'uso di un'AI che sia e diventi – lo si diceva *supra* – affidabile²⁸. La scelta di dirigere *ivi* le segnalazioni è dunque evidenza della volontà di dotare l'enforcement europeo di un punto di ingresso privilegiato (almeno per alcune categorie di violazioni) che sia capace di convogliare informazioni, avviare approfondimenti, attivare cooperazioni con le autorità competenti, e così – anche indirettamente – alimentare la produzione di *soft law* (linee guida, best practices²⁹), sempre

23. Il punto merita una precisazione: ben noto l'intervento nelle more del c.d. Digital Omnibus – e dei relativi effetti di modifica e rinvio dell'efficacia di alcune disposizioni dell'AI Act – è pur vero che il differimento di alcune disposizioni dell'AI Act non ha prorogato, in quanto tale, il regime transitorio della protezione del segnalante. Il Regolamento 2026/1744 non ha infatti toccato l'art. 87 AI Act (selettivamente rinviando come noto l'applicazione delle sezioni 1, 2 e 3 del Capo III, senza spostare *in quanto tale* la data generale del 2 agosto 2026). In dettaglio si v. *infra*, nota 67.

24. Ne deriva ovviamente un assetto potenzialmente concorrente, in cui la medesima vicenda può essere portata, in tempi ravvicinati o persino simultaneamente, all'attenzione dell'organizzazione, dell'autorità nazionale competente e dell'AI Office. Se non pare possibile – pena la violazione della Direttiva – stabilire una priorità astratta tra canale interno (comunque incoraggiato ad essere percorso in prima battuta) ed esterno, rimane evidente come tale assetto renda essenziale la necessità di garantire che la pluralità dei canali non si traduca in una frammentazione delle garanzie per il segnalante.

25. Al Whistleblower Tool e all'AI Act complaints tool si aggiunge, infatti, anche lo specifico canale previsto per i downstream providers di sistemi che incorporano modelli GPAI altrui, ai sensi dell'art. 89, par. 2 AI Act.

26. Directorate-General for Communication Networks, Content and Technology.

27. PAPPALARDO 2026 e *The enforcement framework of the AI Act*, in digital-strategy.ec.europa.eu.

28. *Ufficio europeo per l'IA*, in digital-strategy.ec.europa.eu.

29. Come già dimostrato tramite le Linee Guida Etiche per un'intelligenza artificiale affidabile elaborate dal Gruppo di esperti ad alto livello sull'AI; o tramite il Libro bianco sull'intelligenza artificiale del 2020.

con l'obiettivo di contribuire attivamente ad una AI che sia effettivamente "affidabile" (*trustworthy*) e, dunque, che ingeneri fiducia (e non timore) nei cittadini/consumatori³⁰.

Da un punto di vista sistematico, il Tool svolge almeno quattro funzioni.

La prima potrebbe dirsi epistemica: raccogliere segnali deboli, cioè informazioni che difficilmente emergerebbero attraverso gli ordinari canali di vigilanza, specie quando trattasi di violazioni relative a processi interni (gestione dei dati, addestramento, valutazioni di impatto, controlli di conformità).

Ben potrà avere poi una funzione di deterrenza: l'esistenza stessa del canale, purché credibile nella sua efficacia, contribuirà presumibilmente ad un abbassamento della soglia di rischio accettabile, auspicabilmente agendo positivamente in termini general-preventivi³¹.

Lato organizzativo, potrà avere una funzione di pungolo per il coordinamento tra autorità: inserita in un ecosistema multilivello, la segnalazione può essere la leva attraverso cui attivare la cooperazione tra AI Office, autorità nazionali competenti e, quando rilevante, organismi settoriali (privacy, sicurezza dei prodotti, cybersecurity)³².

La quarta, infine, è una funzione di legittimazione: offrire una via di accesso diretta (e non mediata) all'amministrazione europea contribuisce ad avvicinare agli addetti ai lavori il nuovo quadro regolatorio, che altrimenti rischierebbe di apparire come un insieme di prescrizioni tecniche a sé stanti, meramente teoriche e "lontane" dalla pratica quotidiana.

2.1. Whistleblowing e governance dell'AI: perché un canale settoriale?

Anche alla luce delle brevi considerazioni di cui sopra, è facile comprendere come tra le ragioni per cui l'AI Act, più di altri testi, sembra necessitare della specificità di un proprio whistleblower tool vi sia, probabilmente, la struttura stessa dell'illecito. Quando trattasi del funzionamento – o dell'analisi del rischio – di un sistema di AI, molte delle violazioni rilevanti non sono immediatamente osservabili dall'esterno: non è detto che producano un danno manifesto, possono consistere in omissioni documentali, in risk-assessment svolti in modo non corretto, in scelte di progettazione che solo chi lavora nel ciclo di vita del sistema può ricostruire (e forse neppure, con tutto ciò che ne consegue in punto di tutela rispetto ai possibili effetti avversi, anche a livello sistemico)³³. In un simile ordine di ipotesi, l'enforcement del framework regolatorio non può dipendere solo dall'effettuazione di un controllo a campione *ex post* (ad esempio tramite audit), né tantomeno solo dalla tutela giudiziale (sinonimo di una criticità già concretizzatasi in forma patologica); né infine tramite gli "ordinari" canali di segnalazione (non specializzati): occorre, piuttosto, un sensore distribuito che possa funzionare attivamente lungo la catena del valore dell'AI, crocevia di questioni sociali, etiche, di compliance o di pura responsabilità. Se è vero che la costruzione di un governo dell'AI ruota – e non potrebbe essere diversamente – attorno a tre elementi principali (il panorama tecnico, cioè limiti e portata dell'AI; la governance ideale, ovvero sia i potenziali percorsi per facilitare la cooperazione tra le parti; e la politica dell'AI, quindi le dinamiche

30. I risvolti positivi della regolazione tramite strumenti di *soft law* sono del resto noti: maggiore velocità rispetto alle fonti di *hard law*; maggiore capacità di adattamento; maggiore flessibilità e, di riflesso, una maggiore capacità di rispondere adeguatamente alle differenze tra i modelli sociali europei, nonché infine una non trascurabile maggiore attrattività agli occhi degli attori privati, il cui coinvolgimento nella costruzione e nell'elaborazione di standard e linee guida, codici di condotta e impegni etici rappresenta sovente la via migliore per bilanciare esigenze di ricerca e sviluppo da un lato e obblighi di tutela del consumatore dall'altro, sin dalla progettazione dei sistemi (dunque con un approccio by design e by default). Sia consentito sul punto il rinvio a BARONI 2024-B.

31. Si v. anche TARTAGLIA POLCINI 2022.

32. A puro titolo esemplificativo può vedersi il Protocollo d'intesa siglato tra AGCM e ANAC del 30 luglio 2024, che prevede diverse forme e canali di collaborazione e comunicazione bilaterale tra le Autorità per la gestione delle segnalazioni "Whistleblowing" (art. 2(6) del Protocollo).

33. Lo SAPIO 2021, p. 117, ricorda che "la consapevolezza del funzionamento dello strumento informatico si rivela necessaria anche per poter operare, a monte, la scelta del modello algoritmico, ovvero del tipo di sistema di I.A. (specie se acquistato presso fornitori terzi), trattandosi di una scelta non neutra ma destinata ad incidere sulla stessa legittimità delle singole decisioni".

politiche che influenzano le parti interessate)³⁴, il meccanismo di governance approntato dall'UE tramite il whistleblower tool dell'AI Act risponde a una visione organica: in cui la collaborazione interdisciplinare e il coinvolgimento degli stakeholder sono fondamentali per affrontare le sfaccettature multifattoriali e le sfide etiche associate allo sviluppo e all'implementazione della tecnologia³⁵ (e non è un caso che l'architettura disegnata dal Regolamento sia in sé collaborativa³⁶, richiedendo interazioni continue tra livelli e competenze, necessitando pertanto di canali capaci di trasformare l'informazione diffusa in decisione a rilevanza pubblica).

3. Dalla Direttiva 2019/1937 all'innesto nell'AI Act: la tutela dei segnalanti nella transizione dalla governance al discorso sui diritti fondamentali

Il Whistleblower Tool non nasce, ad ogni modo, dal nulla. Negli ultimi anni la Commissione ha

sperimentato strumenti simili in altri segmenti del diritto digitale: basti pensare al canale per segnalare violazioni della normativa sulla concorrenza³⁷ o, più recentemente, alle iniziative legate al DSA e al DMA³⁸. Non si tratta di convergenza solo organizzativa: piuttosto, deve leggersi come un indice della tendenza a costruire un ecosistema in cui il diritto digitale europeo si dota di porte di ingresso, meccanismi di raccolta e filtri di prioritizzazione delle segnalazioni, quasi a replicare – in chiave pubblica – le logiche di *incident management* proprie dell'ingegneria del software³⁹ (tanto è vero che, per l'appunto, la novità del tool in questione è che l'oggetto della segnalazione è spesso un sistema complesso, non un singolo contenuto o una condotta di mercato facilmente descrivibile; con ciò imponendo un salto di qualità nella capacità di valutare, verificare e contestualizzare la fondatezza della notizia e in ultima istanza portando ad un livello ulteriore l'impianto di diritto originario).

34. BIRKSTEDT–MINKKINEN–TANDON–MÄNTYMÄKI 2023, pp. 133-167.

35. LI 2023.

36. Scrive dell'AI Act come esempio di *collaborative governance* CANCELA-OUTEDA 2024. Più in dettaglio sul concetto di collaborative governance si v., *ex multis*, SCOTT–THOMAS 2017, pp. 191-214. In generale, forme di governance collaborativa nascono in risposta a numerosi fattori, tra cui la complessità e l'interconnessione di problemi che le istituzioni non sono in grado di risolvere di per sé sole, l'incapacità dello Stato di affrontare tali problemi, la necessità di valorizzare le conoscenze, i contributi e il coinvolgimento a livello locale e la necessità di affrontare problemi ad elevato grado di incertezza (ne sono un esempio le sfide poste dal cambiamento climatico, per cui v. BRYSON–CROSBY–MIDDLETON STONE 2006, pp. 44-55).

37. Lanciato, in continuità con il programma di c.d. *leniency*, per offrire anche alle persone che sono a conoscenza dell'esistenza o del funzionamento di un cartello o di altri tipi di violazioni delle norme antitrust l'opportunità di contribuire a porre fine a tali pratiche (*Whistleblowers and Informants*, in competition-policy.ec.europa.eu). Vi ha poco dopo fatto eco, a livello nazionale, l'AGCM, che nel febbraio 2023 ha introdotto un tool specifico (riorganizzando al contempo la propria struttura: prevedendo una direzione istruttoria dedicata alla lotta contro i cartelli, alla leniency e alla gestione delle segnalazioni whistleblowing: "Direzione Cartelli, Leniency e Whistleblowing", incardinata nel Dipartimento per la Concorrenza. Si v. la press release *Italian Competition Authority: whistleblowing platform introduced*, in agcm.it).

38. Rispettivamente come noto funzionali a permettere alle persone di fornire, senza timore di ritorsioni, informazioni che consentano di identificare e scoprire pratiche dannose delle piattaforme online di grandi dimensioni (VLOP) o dei motori di ricerca (VLOSE) designati ai sensi del DSA, o eventuali violazioni degli obblighi dei gatekeeper ai sensi del DMA (*Commission launches Whistleblower Tools for Digital Services Act and Digital Markets Act*, in digital-markets-act.ec.europa.eu). Sul punto più diffusamente l'analisi di HINCK–VAN DEN BOOM 2026.

39. Può farsi rimando all'art. 55 dell'AI Act. Un modello parzialmente affine, pur diverso per oggetto e finalità, è rappresentato dalla *coordinated vulnerability disclosure* prevista dal Cyber Resilience Act. Anche in quel caso l'ordinamento costruisce canali attraverso cui informazioni tecniche non immediatamente osservabili vengono trasferite a un soggetto capace di verificarle per poi intervenire attivamente; a differenza del whistleblowing, tuttavia, la CVD non presuppone necessariamente una violazione normativa né un contesto lavorativo, essendo innanzitutto orientata a diagnosi e correzione delle vulnerabilità.

A livello di portata generale, è con la Direttiva 2019/1937 che l'Unione ha inteso elevare il whistleblowing a componente strutturale dell'effettiva applicazione dei propri strumenti normativi. Non è un caso che i considerando del testo insistano su trasparenza e responsabilità: il segnalante alimenta sì i sistemi di enforcement e ne aumenta l'efficacia, ma perché questo avvenga occorre creare un quadro di flussi informativi aperti, fatto di canali interni ed esterni, di termini per il follow-up, di protezione dell'identità e di divieti di ritorsione. Agendo, in estrema sintesi, non solo e non tanto sulla funzionalità dell'istituto, quanto anche (e, a livello concettuale, prioritariamente) sulla tutela del segnalante come colui che, a proprio – potenziale – discapito, sceglie di agire in tutela di interessi fondamentali dell'Unione, primi tra tutti quello derivante dall'integrità del mercato interno e dal buon funzionamento della *res publica*: viene positivamente così la transizione del whistleblower da mero strumento procedurale a parte integrante (e attiva, o “prosociale”⁴⁰) di quel modello organizzativo che caratterizza il costituzionalismo contemporaneo e le organizzazioni internazionali, volto a disciplinare l'esercizio del potere pubblico e a renderlo strutturalmente controllabile. Ne discende una possibile concezione dell'istituto come componente essenziale dei sistemi democratici, in quanto capace di accrescerne l'*accountability* nei confronti dei governati e di riconoscere, in capo a questi ultimi, il pieno esercizio di diritti fondamentali – quali la libertà di coscienza e di pensiero, la libertà di espressione e il diritto all'informazione – per fini di controllo dell'azione pubblica.

Vi ha poi fatto seguito una rilettura delle coordinate interpretative del bilanciamento fra la condizione del whistleblower e la protezione dell'interesse pubblico. Così, si è assistito ad

una valorizzazione dell'azione del whistleblower: non più esercizio di una mera *facoltà* individuale, quanto piuttosto spazio di attivazione, da parte del singolo, delle proprie libertà fondamentali, da cui un restringimento delle possibilità di limitazione all'esercizio di tali diritti. Nonché, di conseguenza, del whistleblower stesso, per il quale l'azionabilità di un diritto fondamentale determina la recessione “o comunque [il divenire] meno qualificante, il collegamento con l'ambiente di lavoro [subordinato], reputato invece presupposto imprescindibile quando il whistleblowing è strumento esclusivamente di governance”⁴¹.

Assistiamo, in buona sintesi – e il whistleblower tool dell'AI Act ne è l'ultima dimostrazione – ad un vero e proprio cambio di paradigma, marcatamente distante da quanto accadeva in passato: se prima vi era (doveva esservi) perfetta corrispondenza tra la condizione di whistleblower e quella di membro subordinato dell'organizzazione oggetto della segnalazione⁴², ora vi si ricomprende una più vasta pluralità di relazioni professionali (addirittura si è arrivati a ipotizzare l'esistenza – a riprova di tale mutato ruolo funzionale del whistleblowing nella tutela, quasi realizzazione, del bene comune – di un vero e proprio “dovere” di segnalazione, pur a certe ben determinate condizioni)⁴³.

Ben si comprende allora il ruolo della Direttiva 2019/1937 quale strumento che ha inteso *ab initio* essere (e apparire) come “game changer”⁴⁴ nel ripensamento dell'istituto e – successivamente, come precipitato procedurale di un più profondo cambiamento teorico – nella costruzione di standard minimi comuni di protezione. Non sono certo mancate criticità (soprattutto relative, da un lato, all'ambizione di superare la frammentazione intra-europea; dall'altro, alle persistenti zone grigie e le opportunità mancate, che rischiano di

40. Si è parlato del whistleblowing come “pratica prosociale, mirata alla promozione della giustizia all'interno di un'organizzazione” (MICELI-NEAR 1992; come cit. in CEVA 2024, p. 76).

41. PARISI 2020, p. 14.

42. NEAR-MICELI 1985, pp. 1-16.

43. CEVA 2024, p. 77 ss.

44. ABAZI 2016, pp. 1061-1072, e ABAZI 2020, pp. 640-656, che non manca di sottolinearne alcune ombre, tra cui l'eccessivo spazio lasciato ai singoli sistemi nazionali (ad esempio, quando si consente alle autorità nazionali di decidere a loro discrezione se dare seguito alle segnalazioni anonime quando tali segnalazioni vengono affidate a canali esterni – art. 6(2) Dir. 2019/1937); o la mancanza di enforcement tools adeguati (a differenza di quanto avrebbe potuto prevedersi – ad esempio – tramite un Regolamento).

indebolire l'effettività dell'impianto di *discovery*); pur essendo vero, tuttavia, che alcune (forse le più rilevanti) di tali preoccupazioni si sono rivelate *ex post* infondate, e anzi persino portatrici di risvolti positivi: come avvenuto per la presenza di più basi giuridiche a fondamento dell'istituto del whistleblowing, inizialmente indicata come possibile fonte di contestazione della Direttiva e in realtà poi utilizzata come *passee-partout* in sede giudiziale, per agganciare la funzione del whistleblower a principi e diritti fondamentali dell'Unione.

Primo tra tutti il diritto alla libertà di espressione, che conosce così una istituzionalizzazione procedurale: a un ingranaggio della macchina di *compliance* si sostituisce, si diceva, il titolare di un diritto di rango primario a "impartire informazioni" su fatti di evidente interesse pubblico (quali l'opacità dei sistemi di AI ad alto rischio, l'elusione dei requisiti di *data governance* o l'assenza di adeguati *risk assessment* su modelli che impattano su diritti fondamentali e processi democratici), così divenendo parte attiva nella tutela della libertà di espressione⁴⁵. Se insomma il paradigma tradizionale della libertà di espressione è oggi messo alla prova dai nuovi media e dalle infrastrutture digitali, tanto da imporre una rilettura delle categorie classiche alla luce della *documerialità* e del potere delle piattaforme⁴⁶, il Whistleblower Tool dell'AI Act è allora suscettibile di esser letto come una specifica risposta di sistematizzazione digitale da

parte del legislatore del valore dell'informazione⁴⁷: tramite cui reimpostare, in chiave europea, il controllo diffuso del potere tecnologico, indicando la segnalazione come esercizio protetto di libertà – e non come eccezione tollerata – all'interno di un'infosfera pubblica capace di (ma contemporaneamente obbligata a) prescindere dai mediatori tradizionali⁴⁸.

Similmente accade per la tutela del mercato interno. Se è vero infatti che la creazione di un'AI affidabile non è un fine in sé, quanto piuttosto un mezzo per l'incentivazione dello sviluppo del contesto imprenditoriale europeo⁴⁹, è altrettanto innegabile che la presenza di canali di whistleblowing impone una ridefinizione dei limiti degli obblighi di segreto aziendale, d'ufficio, professionale o scientifico-industriale⁵⁰, e non v'è chi non veda come nel settore AI la dialettica tra emersione del rischio e tutela del segreto industriale sia particolarmente aspra: i dataset di addestramento, le architetture dei modelli, le tecniche di mitigazione, i processi di valutazione, costituiscono spesso il cuore competitivo dell'impresa. L'Unione ha già dovuto affrontare il nodo – in prima battuta – con la Direttiva 2016/943 sui segreti commerciali, il cui considerando n. 20 contiene una clausola di salvaguardia a favore delle rivelazioni effettuate per svelare una condotta illecita, un illecito professionale o un comportamento scorretto, a condizione che la divulgazione sia necessaria per proteggere l'interesse generale⁵¹. Bilanciamento di

45. VITARELLI 2023, pp. 813 ss.; in commento a *Halet v. Luxembourg*, n. 21884/18, 14 febbraio 2023. Pur al netto naturalmente del giudizio di proporzionalità (CORSARO 2023, p. 745 ss.) invero proprio di ogni operazione di bilanciamento tra diritti e interessi contrapposti (e di pari valore) (MORRONE 2008, pp. 185-204).

46. MAGNANI 2018.

47. PITRUZZELLA 2018.

48. D'altronde, tramite Internet ognuno può essere gratuitamente "stampatore, editore e direttore di sé stesso" (FROSINI 2020).

49. Da ultimo, il Rapporto Draghi ben evidenzia il necessario coordinamento intersettoriale e la condivisione dei dati per accelerare l'integrazione dell'IA nell'industria europea (in particolare intervenendo nei dieci settori strategici in cui i modelli di business dell'Ue trarranno il massimo vantaggio dalla rapida introduzione dell'IA: automotive, produzione avanzata e robotica, energia, telecomunicazioni, agricoltura, aerospaziale, difesa, previsioni ambientali, farmaceutica e sanità). *Draghi Report*, Parte A, Cap. 2, p. 34 ss.

50. VALENTE-GHERA 2021, pp. 245-258.

51. "Le misure, le procedure e gli strumenti di tutela previsti dalla presente direttiva non dovrebbero limitare la denuncia delle irregolarità. La tutela dei segreti commerciali, pertanto, non dovrebbe estendersi ai casi in cui la divulgazione di un segreto commerciale serve l'interesse pubblico, nella misura in cui siano rivelate condotte scorrette, irregolarità o attività illecite direttamente pertinenti. Ciò non dovrebbe essere inteso come un impedimento per le competenti autorità giudiziarie ad autorizzare una deroga all'applicazione delle misure,

per sé difficoltoso⁵², il cui sentiero viene dal Whistleblower Tool dell'AI Act imboccato (deve consentire di trasferire informazioni sensibili all'autorità, garantendo confidenzialità ma senza trasformarsi in un canale di *leakage* incontrollato), estendendone tuttavia a ben vedere il contesto istituzionale e materiale di applicazione, sulla scorta di quanto indicato dalla Direttiva 2019/1937: il segnalante dell'AI Act, infatti, non è – lo si diceva – necessariamente un lavoratore in senso stretto; potendo essere un consulente, un fornitore, un partner, un ricercatore, un soggetto che ha accesso a informazioni rilevanti in ragione della propria posizione nella filiera.

In questa prospettiva, l'elasticità soggettiva del Whistleblower Tool dell'AI Act, rivolgendosi insomma a chiunque sia *professionalmente connesso* alla filiera, è suscettibile di esser letta come la coerente trasposizione, nel nuovo spazio digitale regolato, della nozione ampia di “persona protetta” che la Direttiva 2019/1937 costruisce all'art. 4, includendo lavoratori subordinati, autonomi, azionisti, membri degli organi di gestione e controllo, volontari, tirocinanti, nonché chi operi “sotto la supervisione e direzione di appaltatori, subappaltatori e fornitori”. La formula del “contesto lavorativo” già presuppone, infatti, una rete di relazioni professionali non più riconducibile al solo rapporto di impiego tipico, ma estesa alle molteplici forme di collaborazione che popolano le catene globali del valore; ed è precisamente questa intuizione che il canale dell'AI Act riprende e porta a più ampie conseguenze nel dominio

dell'AI generativa e dei GPAI, dove la conoscenza del rischio è spesso distribuita tra attori diversi dal datore di lavoro in senso stretto (centri di ricerca, laboratori esterni, *safety teams* in outsourcing, partner industriali). La scelta di agganciare il Tool alla Direttiva (la cui copertura positiva risiede nell'art. 87 del Reg. 2024/1689)⁵³ e, quindi, di enfatizzare una platea estensibile di potenziali segnalanti, esplicita la vocazione del modello originario: il whistleblowing non nasce per proteggere una figura in senso stretto (il dipendente), bensì per garantire che chiunque, proprio in quanto inserito in un contesto professionale, possa trasferire informazioni sensibili all'autorità competente. Il segnalante diviene così un attore di *accountability democratica*, leva strutturale di responsabilizzazione nei circuiti di *open government*⁵⁴, valevole a sottolineare come la protezione non riguardi più una mera facoltà di collaborazione, bensì un vero e proprio diritto alla segnalazione, precipitato della libertà di espressione e di comunicazione delle informazioni *ex art. 10 CEDU* in funzione dell'interesse pubblico⁵⁵. Nell'utilizzare – lo si diceva *supra* – il diritto dell'Unione per costruire garanzie strutturali contro gli squilibri di potere nelle infrastrutture digitali⁵⁶, l'estensione dell'area applicativa del meccanismo di whistleblowing all'ecosistema dell'AI Act appare allora come un tassello essenziale: il legislatore europeo assegna alla segnalazione protetta una funzione dinamica, poiché – pur non mancando questa ampliata applicazione soggettiva di suscitare alcuni interrogativi⁵⁷ – la erige

delle procedure e degli strumenti di tutela, laddove il convenuto abbia tutti i motivi per credere, in buona fede, che la sua condotta abbia rispettato i criteri pertinenti di cui alla presente direttiva”.

52. FUMAGALLI-LORUSSO 2024.

53. Occorre precisare come non sia direttamente l'art. 87 AI Act ad ampliare la platea dei segnalanti, operando unicamente come norma di raccordo che evidenzia il fil rouge tra la segnalazione delle violazioni del Regolamento e la protezione dei relativi autori alla Direttiva 2019/1937. L'ampiezza soggettiva di che trattasi discende dunque dall'art. 4 della Direttiva e dalla relativa nozione di contesto lavorativo. Nell'ecosistema AI, ad ogni modo, il contesto lavorativo è potenzialmente esteso: può includere dipendenti, ricercatori, valutatori esterni, componenti di *safety teams*, appaltatori e fornitori. Tale estensione, pur non derivando direttamente dall'AI Act *in se*, caratterizza (e amplia) la portata applicativa del Tool, applicando il modello derivante dalla Direttiva a una filiera caratterizzata da una distribuzione orizzontale di conoscenza e responsabilità. Sull'art. 87 si v. nello specifico BRIGHI-BESSONE 2026.

54. DONINI 2024, p. 228 ss.

55. DONINI 2022, scrive del whistleblowing come di un atto di “responsabilità democratica”.

56. SPINIELLO 2025, p. 579 ss.

57. Si è detto che questa elasticità è coerente con la Direttiva 2019/1937, che adotta una nozione ampia di persone protette; ma l'innesto settoriale pone domande ulteriori: quali informazioni sono davvero utili? Come si valuta

a condizione di effettività del *digital regulatory framework* (AI Act, quindi, ma anche le altre regolamentazioni digitali – come DMA e DSA) e, dunque, a strumento attraverso cui guardare dentro i luoghi – spesso opachi – in cui si decide dell'uso dell'intelligenza artificiale e, con esso, della tenuta dei diritti fondamentali nello spazio digitale.

4. Riflessi applicativi per gli operatori nella compliance: integrazione con canali interni, governance dei flussi informativi, gestione del rischio

L'inaugurazione di un canale (esterno e centralizzato in seno all'AI Office) inedito impone una riconsiderazione dell'ecosistema dell'assetto organizzativo aziendale: un canale interno inefficace, percepito come non affidabile o ritorsivo, aumenterà infatti inevitabilmente la probabilità che la segnalazione venga effettuata tramite il canale esterno diretto all'AI Office, incentivando sì l'azione amministrativa in funzione correttiva, ma al tempo stesso sottraendo all'organizzazione la possibilità di correggere la deviazione prima di una sua escalation negativa. Inoltre, la Direttiva 2019/1937 aveva già introdotto specifici obblighi per molte organizzazioni (in particolare, soggetti con almeno 50 dipendenti) in materia di canali interni, follow-up e protezione dei segnalanti, mentre le legislazioni nazionali di attuazione – in Italia, ad esempio, la legge 179/2017 e poi⁵⁸ il d.lgs. 24/2023 – hanno ulteriormente dettagliato procedure e garanzie. L'arrivo del canale settoriale di Bruxelles non sostituisce tali obblighi, purtuttavia li risignifica. Innanzitutto, il nuovo canale diventa uno strumento di gestione del rischio regolatorio europeo, elevandosi, come visto, a perno funzionale di parte dell'architettura che guida la Strategia digitale europea in materia di AI. Soprattutto, però, può ben credersi che si candiderà a funzionare da benchmark: se l'AI Office si dimostrerà in grado di garantire anonimato, una determinata facilità di accesso e un certo livello di

confidenzialità, anche le singole organizzazioni si vedranno giocoforza costrette a offrire standard analoghi, se vorranno trattenere le segnalazioni *in-house*. Non solo: se è vero che anche gli obblighi già derivanti dalla Direttiva e dal d.lgs. n. 24/2023 imponevano l'adozione di più o meno predeterminate soluzioni procedurali adeguate caso per caso, non deve dimenticarsi che, in ambito AI, la compliance si risolve inevitabilmente in un'architettura informativa (documentazione tecnica, registri, log, report di valutazione, modelli di governance del dato), e l'istituto del whistleblowing interagisce con questa architettura in due modi: potendo svelare l'assenza di adeguata documentazione (o la sua natura meramente formale), ovvero potendo trasformare documenti interni in potenziali strumenti di indagine esterna. Ne discenderà per tutti un incentivo ulteriore alla cura del *record keeping*, come (non più mero adempimento burocratico, *ma*) vero presidio di difesa: se una segnalazione sarà sufficientemente dettagliata o comunque non manifestamente infondata, la capacità di farvi fronte dimostrando la conformità dei processi valutativi interni dipenderà dalla qualità e integrità delle evidenze dimostrabili.

4.1. Il rischio di “over-reporting”

Il tool in questione si candida quindi ad avere, nelle direzioni sopra viste, una funzione in qualche modo propulsiva, certamente da salutarsi con favore a livello ordinamentale, pur se – innegabilmente – presumibilmente scomoda al livello organizzativo-gestionale per gli attori economici privati. E, sotto tale profilo, non possono ignorarsene altresì i (possibili) risvolti pratici derivanti da un ipotetico abuso del tool, e dunque dai possibili effetti di “over-reporting”: la presenza di canali esterni facilmente accessibili può comprensibilmente incentivare segnalazioni eccessive, difensive, talvolta strumentali. In organizzazioni ad alta complessità tecnica, tanto più, il confine tra

la buona fede? Quali sono gli standard minimi di circostanziamento richiesti per attivare un follow-up? Residua poi il tema della specializzazione del canale esterno: la Direttiva del 2019 prevede canali esterni presso autorità competenti; qui, tuttavia, il canale esterno è addirittura europeo e settoriale, costruito attorno a un centro di expertise tecnica. Si tratta di un passaggio significativo: rafforza il ruolo dell'amministrazione europea come garante dei diritti nel digitale, ma aumenta anche il fabbisogno di garanzie procedurali, perché l'intervento di un'autorità sovranazionale su situazioni spesso radicate in contesti nazionali può incidere su posizioni giuridiche soggettive e generare conflitti di competenza. La scelta, insomma, non è neutra (*infra*, par. 4).

58. MAGRI 2019.

“errore” e “violazione” può non sembrare immediatamente intellegibile, ovvero essere addirittura sfumato, facendo sì che una errata percezione del rischio possa portare a segnalare anche condotte che richiederebbero, piuttosto, un intervento formativo o organizzativo.

Vero è che un simile rischio potrebbe, allo stato, apparire più teorico che empiricamente dimostrato. Deporrebbero per tale tesi, a titolo esemplificativo, gli evidentemente ridotti volumi registrabili dall’impiego a livello europeo di canali analoghi: il report di attività 2024 della Commissione registra circa 20 segnalazioni DMA, divenute poi 152 dal 2024 ad aprile 2026⁵⁹; mentre i canali antitrust anonimi ricevono, dal 2017, una media di circa 100 messaggi l’anno⁶⁰. Vero tuttavia, allo stesso tempo, che simili considerazioni non possono non tenere conto di un sistema regolatorio – quello europeo – a penetrazione tradizionalmente lenta: non per caso era la stessa Commissione a evidenziare, nel proprio report di attività 2024, un “significativo ritardo” da parte degli Stati membri nella trasposizione della Direttiva nel diritto nazionale, limitando per tale ragione l’efficacia del report ai soli Stati ove questa era invece già stata implementata alla data del 17 dicembre 2023⁶¹. Una situazione, d’altronde, coerente con un contesto caratterizzato da una scarsissima alfabetizzazione sul tema⁶².

Non sembra quindi azzardato ipotizzare un reale, futuro aumento delle segnalazioni, come da evidente auspicio del legislatore unionale e come, altresì, sembrerebbe confermare uno sguardo verso sistemi più maturi: nel programma della Securities and Exchange Commission statunitense sono state ricevute circa 24.980 segnalazioni nell’esercizio 2024 e circa 27.000 nel 2025⁶³. Certo il dato non è *sic et simpliciter* meccanicamente trasponibile al tool europeo – per la maggiore ampiezza soggettiva del canale statunitense⁶⁴ e per la presenza, in quel modello, di incentivi monetari – eppure dimostra come l’istituzione di un canale digitale esterno ben possa determinare un aumento progressivo dei flussi. La stessa SEC individuava, nel 2025, un significativo aumento delle segnalazioni – a conferma della puntualità delle misure richieste dall’Inspector General già nel 2022, finalizzate tra l’altro a incrementare l’efficienza del programma⁶⁵.

Anche tale ultimo profilo appare rilevante, considerando come il rischio di over-reporting non dovrebbe essere identificato solo con la mera eventualità di un numero elevato di segnalazioni, né – ancor meno – con la presunzione che una parte significativa di esse sia semplicemente mendace. Con tale rischio intendiamo, infatti, anche quello derivante da una relazione di possibile sproporzione tra il volume, la ripetitività o la ridotta

59. EU Commission Staff Working Document accompanying the document Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the Review of Regulation (EU) 2022/1925 of the European Parliament and of the Council on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act), in accordance with Article 53 thereof, doc SWD(2026) 123, 28 aprile 2026.

60. HINCK-VAN DEN BOOM 2026. Analogamente *Competition: EU Commission extends scope of anonymous whistleblower tool*, in ieu-monitoring.com, 9 gennaio 2023.

61. Alla data di pubblicazione del report erano per tale ragione in corso procedure di infrazione nei confronti di Austria, Belgio, Grecia, Lituania, Olanda e Slovacchia; mentre alla data di gennaio 2024 Estonia e Polonia ancora non avevano trasposto la Direttiva.

62. Nel 2018, solamente il 15% dei cittadini europei si dichiarava a conoscenza dell’esistenza di norme a protezione dei segnalanti (*Whistleblower protection fact sheet*, in ec.europa.eu).

63. SEC, Securities and Exchange Commission Office of the Whistleblower Annual Report to Congress for Fiscal Year 2024 e 2025 in sec.gov.

64. Che (pur comunque limitato alle segnalazioni provenienti da sole persone fisiche) non richiede necessariamente una connessione lavorativa con il segnalato.

65. United States Securities and Exchange Commission, *SEC’s Whistleblower Program: Additional Actions Are Needed To Better Prepare for Future Program Growth, Increase Efficiencies, and Enhance Program Management*, Report No. 575, 9 dicembre 2022 in sec.gov.

qualità informativa delle comunicazioni ricevute⁶⁶, da un lato, e la capacità dell'amministrazione di sottoporle a uno screening e a una valutazione tecnicamente competenti e sufficientemente accurati, dall'altro. Sembra fornirne conferma la stessa Direttiva 2019/1937, ove – contemplando espressamente il possibile afflusso elevato di segnalazioni, le violazioni chiaramente minori e le comunicazioni reiterate prive di elementi nuovi – consente agli Stati membri di autorizzare le autorità competenti a chiudere queste ultime e, in presenza di flussi particolarmente consistenti, a trattare prioritariamente le violazioni gravi o concernenti disposizioni essenziali: il considerando n. 70 giustifica tali previsioni proprio con la necessità di alleviare gli oneri gravanti sulle autorità incaricate del follow-up, a riprova di come la gestione selettiva dei flussi informativi sia parte integrante dell'architettura normativa del modello.

Allo stato dell'arte, in buona sintesi, è ragionevole affermare che in tale materia siano l'ambito soggettivo e materiale, nonché il grado di maturazione, la struttura degli incentivi e la riconoscibilità dell'illecito a determinare il successo di un canale di segnalazione; non consentendo, pertanto, di escludere che il Whistleblower Tool dell'AI Office possa conoscere una dinamica differente da canali analoghi una volta completata l'applicazione del Regolamento e divenuta pienamente operativa, come è dal 2 agosto 2026, la protezione antiritorsiva collegata alle violazioni dell'AI Act⁶⁷. D'altronde, lo strumento è stato costruito precisamente per ridurre i costi pratici e personali della segnalazione (consente l'anonimato, l'impiego di qualsiasi lingua ufficiale dell'Unione, il deposito di documenti e una

comunicazione successiva protetta con l'AI Office): la maggiore accessibilità del canale e il rafforzamento della fiducia del potenziale segnalante si pongono come condizioni dichiarate del suo (auspicabile) successo; che dovrebbe riflettersi, almeno tendenzialmente, in una crescita delle segnalazioni.

A ben vedere, quindi, lungi dal porsi come un mero effetto collaterale marginale del Whistleblower Tool dell'AI Act, il rischio di over-reporting rappresenta il rovescio della medaglia del *chilling effect* che la Corte EDU ha tentato di neutralizzare nella nota parabola che da *Guja* conduce, passando per *Heinisch*, alla più recente *Halet c. Lussemburgo*; mentre un flusso eccessivo di segnalazioni poco circostanziate, ripetitive o solo indirettamente rilevanti può saturare l'attività amministrativa di intervento, rallentando la capacità dell'AI Office di individuare e trattare i casi realmente critici, determinando – in ultima istanza – una diluizione dell'attenzione regolatoria sui rischi più gravi per diritti fondamentali, salute o sicurezza (e, se così fosse, l'over-reporting potrebbe ragionevolmente amplificare il rischio di pregiudizi reputazionali per i soggetti segnalati, soprattutto nelle fasi preliminari e informali dell'istruttoria, rischiando di alterare il delicato bilanciamento tra compliance e mercato, nonché favorire un uso strategico o opportunistico del canale all'interno di conflitti professionali o competitivi lungo la filiera dell'AI). Lo si può indirettamente leggere dalla sentenza Gawlik: in assenza di criteri chiari di valutazione e di filtri procedurali adeguati, che nello specifico del Whistleblower Tool dovranno necessariamente essere stabiliti in futuro alla luce del successo – o meno – dello strumento, quest'ultimo rischierebbe

66. La stessa SEC precisa che circa 12.000 delle segnalazioni 2025 provenivano da due sole persone, e che le stesse due persone ne avevano prodotte oltre 14.000 nel 2024.

67. Come *supra* si diceva, nota 23), il differimento di alcune disposizioni dell'AI Act non ha prorogato, in quanto tale, il regime transitorio della protezione del segnalante. Il Regolamento 2026/1744 non ha infatti toccato l'art. 87 AI Act (selettivamente rinviando come noto l'applicazione delle sezioni 1, 2 e 3 del Capo III, senza spostare *in quanto tale* la data generale del 2 agosto 2026): è da tale data quindi che la Direttiva 2019/1937 trova pertanto applicazione alle segnalazioni delle violazioni dell'AI Act. Il rinvio produce, nondimeno, un effetto indiretto sul perimetro di garanzia, non potendo certo essere qualificata come violazione del Regolamento la mera inosservanza di un obbligo che, per scelta sopravvenuta del legislatore europeo, non sia ancora applicabile. Alla piena operatività formale della protezione non corrisponde, dunque, l'immediata possibilità di segnalare l'intero apparato sostanziale dell'AI Act, con conseguente apparente (inedita) asimmetria: non più temporale tra esistenza del canale e tutela del segnalante (come era per un tool invero lanciato il 24 novembre 2025, ma per le cui segnalazioni si prevedeva tutela solo a far data dal 2 agosto 2026) bensì materiale, tra applicabilità della garanzia e progressiva espansione dell'insieme delle condotte suscettibili di integrare una violazione del Regolamento.

di trasformarsi da mezzo di *accountability* e tutela dell'interesse pubblico in fattore di opacità decisionale e di instabilità istituzionale, indebolendo proprio quella fiducia nel framework europeo dell'AI che il Tool è chiamato a rafforzare. Sul punto, può ricordarsi come proprio recentemente la Grande Camera della Corte EDU abbia ribadito i sei criteri di valutazione della segnalazione⁶⁸, insistendo su una loro applicazione complessiva e non meccanica, senza gerarchie prestabilite, all'esito di un bilanciamento effettivo degli interessi in gioco: non basta evocare l'interesse pubblico, occorrendo verificare la qualità e la ragionevole veridicità delle informazioni, la proporzionalità del pregiudizio inflitto al soggetto inciso e la congruità della reazione sanzionatoria⁶⁹. Che non possa allora rinvenirsi qui l'antidoto (quantomeno teorico) all'over-reporting: la forte protezione riconosciuta al whistleblower non si traduce in una immunità generalizzata del segnalante, ma è condizionata alla presenza di un *fumus* di fondatezza, di un interesse pubblico reale e di un sacrificio ragionevole delle posizioni antagoniste (segreto industriale, reputazione commerciale, interessi economici). In questo senso, il rischio di saturazione del canale viene affrontato non restringendo *ex ante* l'area del whistleblowing, quanto piuttosto rafforzando *ex post* l'architettura del bilanciamento: criteri chiari di valutazione, tracciabilità delle decisioni di archiviazione o di apertura delle indagini, proporzionalità tra il contenuto della segnalazione e gli effetti (anche reputazionali) sull'ente segnalato. In altre parole, un disegno che, nel solco di *Halet*, sappia tenere insieme due esigenze solo apparentemente inconciliabili: evitare che il timore di ritorsioni disincentivi i flussi informativi ed evitare, al tempo stesso, che la tutela del segnalante si trasformi in un incentivo istituzionale al leakage incontrollato di informazioni sensibili, sganciate da un effettivo interesse (costituzionalmente) apprezzabile.

5. Un canale, più livelli: alcune criticità

L'AI Act, come ormai noto, disegna un sistema di governance multilivello: agli Stati membri spetta

istituire autorità competenti, coordinarsi attraverso l'European Artificial Intelligence Board, e garantire ordinaria applicazione al Regolamento; alla Commissione e all'AI Office competono funzioni di coordinamento, produzione di atti attuativi, e – specie per i modelli GPAI – poteri di supervisione diretta. Il principale rischio implicito nel Regolamento è, come tipico di ogni sistema multilivello, la frammentazione derivante dalla moltiplicazione di sedi, procedure, standard e tempi di attuazione, con conseguente erosione della certezza del diritto e aumento del carico di compliance (specie sui piccoli attori)⁷⁰. L'AI Act Whistleblower Tool può diventare, paradossalmente, antidoto e al tempo stesso acceleratore di tale frammentazione.

Da un lato, offrendo un punto di raccolta centralizzato, consente all'AI Office di ridurre il rischio che informazioni rilevanti vadano perdendosi. Di converso, la presenza di un canale europeo specializzato, se incentiverà i segnalanti a rivolgersi direttamente a Bruxelles, comporterà un possibile spostamento verso il livello unionale del primo punto di emersione di vicende che presentano anche profili di competenza nazionale o settoriale (soprattutto allorché l'AI Office diventi sinonimo di maggiore sicurezza, *supra* par. 4), con il rischio – non così remoto – di overlap applicativi a livello unionale. Tanto più a mente della considerazione per cui l'AI non è certo un campo d'azione isolato: una segnalazione relativa a un sistema ad alto rischio potrà implicare contemporaneamente profili di protezione dei dati (GDPR), sicurezza informatica (NIS2), sicurezza dei prodotti (Nuovo quadro di sicurezza dei prodotti, Cyber Resilience Act), tutela dei consumatori. Il rischio dell'apertura di procedimenti paralleli in seno a differenti autorità diviene allora concreto, inscrivendosi nel novero di quelle sfide derivate che emergono dalla sovrapproduzione di norme regolamentari all'interno dell'Unione, e per le quali – ad oggi – non sembra a ben vedere esservi reale ed immediata soluzione (a mente non solo della loro proliferazione numerica, ma anche delle

68. Canale scelto, interesse pubblico, autenticità delle informazioni, buona fede, danno arrecato al datore di lavoro e ponderazione di quel danno rispetto all'interesse pubblico, severità della sanzione.

69. ANDREADAKIS-KAFTERANIS 2023.

70. "We claim to promote innovation, but we continue to add regulatory burdens to European companies, which are particularly costly for SMEs and self-defeating for those in the digital sectors", Draghi Report, Part A, 9.

complessità interpretativo-applicative generate da tale abbondanza normativa, per cui un eccesso quantitativo si traduce in ultima analisi in una debolezza qualitativa della legislazione)⁷¹. Pur se la Commissione – nelle sue pagine sulla governance dell'AI Act – insiste sulla necessità di cooperazione e coordinamento (lo stesso Digital Omnibus riconosce il rischio di overlap normativi, introducendo clausole di equivalenza regolatoria⁷² o presunzioni di conformità “incrociate”⁷³, che vanno così ad affiancarsi alle già esistenti disposizioni⁷⁴ finalizzate a favorire il coordinamento tra autorità), pare naturale nutrire più di qualche dubbio, sul piano applicativo, sulla capacità di tale cooperazione nel venir tradotta in protocolli e prassi capaci di garantire quella ricerca di soluzioni e standard uniformi – quasi una “esigenza di ‘forzare’ le singole amministrazioni anche ad ordinamento autonomo” – che una piena digitalizzazione richiederebbe⁷⁵. Ciò non significa, ovviamente, che l'AI Act non preveda strumenti *ad hoc*⁷⁶, quanto piuttosto che – di per sé considerate – tali soluzioni non eliminano

il problema che il tool rischia di rendere evidente: una singola segnalazione ben potrà generare flussi istruttori distinti, richieste documentali parzialmente sovrapponibili e decisioni adottate con scansioni temporali e secondo parametri differenti.

Occorre poi considerare che ogni canale di segnalazione sconta in sé un problema strutturale, derivante dalla scarsità delle risorse rispetto al – si diceva – potenziale numero di segnalazioni che possono pervenire. L'efficacia del canale dipende, allora, dalla capacità di filtrare e prioritizzare, ovvero di decidere – in buona sintesi – quali segnalazioni meritino un approfondimento e quali no. A *fortiori* in un contesto come quello dell'AI Act, tale scelta incide sulla tutela effettiva dei diritti: pur in presenza di strumenti normativi di cooperazione, la mancanza di un effettivo coordinamento può portare a decidere di non intervenire su una segnalazione potenzialmente relativa a una pratica vietata e così equivalere, nella sostanza, a una potenziale lesione di quegli interessi e libertà fondamentali che sono alla base della valutazione del rischio posta

71. RAFFIOTTA 2025 scrive a tal proposito di *multi-level enforcement concerns* (distinguendoli dai *multi-regulatory concerns*, di cui alle potenziali criticità derivanti dalla presenza di un numero eccessivo di atti legislativi e strumenti normativi all'interno del quadro normativo dell'Unione: *sic et simpliciter* dalla sovrabbondanza di norme e strumenti giuridici, intesa principalmente in termini quantitativi).

72. Art. 2, par. 13 AI Act, come modificato dal Regolamento 2026/1744, che consente di limitare l'applicazione di specifici obblighi dell'AI Act quando la normativa di armonizzazione settoriale inclusa nell'Allegato I assicuri un livello equivalente o superiore di protezione della salute, della sicurezza o dei diritti fondamentali.

73. Considerando 42.

74. Si v. l'obbligo degli Stati membri, previsto dall'art. 74, par. 10, di agevolare il coordinamento fra autorità AI e autorità competenti per altre discipline rilevanti; o il (anch'esso recentemente modificato) obbligo, *ex art.* 77, per le autorità di vigilanza del mercato e le autorità incaricate della tutela dei diritti fondamentali di stretta cooperazione e di assistenza reciproca, comprensivo dello scambio delle informazioni necessarie all'esercizio dei rispettivi mandati e finalizzato espressamente alla coerenza applicativa e alla semplificazione delle procedure. Su tale specifico articolo del Regolamento si v. GIANNELLI 2025.

75. GIANNELLI 2026, che ricorda l'apporto di Corte cost. n. 15/2010.

76. L'art. 70 impone agli Stati membri di designare almeno un'autorità di notifica e un'autorità di vigilanza del mercato, individuando altresì un'autorità di vigilanza quale punto di contatto unico, prevedendo che ogniquale volta le autorità nazionali competenti intendono fornire orientamenti e consulenza in relazione a un sistema di IA in settori disciplinati da altre disposizioni di diritto dell'Unione, siano consultate le autorità nazionali competenti a norma di tali disposizioni di diritto dell'Unione. A livello unionale, poi, gli artt. 65 e 66 affidano al Consiglio europeo per l'intelligenza artificiale il compito di favorire il coordinamento tra le autorità nazionali, condividere conoscenze e migliori pratiche, armonizzare le prassi amministrative e cooperare con le reti europee operanti, fra l'altro, nei settori della protezione dei dati, della cibersicurezza, della sicurezza dei prodotti, della concorrenza e della tutela dei consumatori, sottraendo di fatto la necessità di coordinamento alla sola convergenza spontanea delle amministrazioni nazionali. A ciò si aggiungono i già citati e più specifici criteri dell'art. 74 e dell'art. 77.

a fondamento della disciplina⁷⁷. Manca quindi, potrebbe altresì dirsi, una disciplina trasversale che trasformi la cooperazione tra autorità in coordinamento unitario del singolo caso segnalato. Situazione che – lo si può già ipotizzare senza timore di smentita – diverrà inevitabilmente più intricata con la stagione della robotica⁷⁸: se è vero – come è – che la convergenza fra safety, cybersecurity e tutela dei diritti fondamentali è destinata ad accentuarsi, val la pena allora chiedersi come sarà possibile assicurare che l'apprendimento post-market avvenga solo in ambiente sicuro (c.d. *safe environment*) e senza compromettere la “sicurezza” (cyber e fisica) del macchinario (tanto più se questa dipenderà dall'uso che ne farà l'utente o dall'ambiente in cui sarà inserito); e se – in presenza di criticità derivanti da situazioni solamente ipotetiche – potrà farsi uso di canali come l'AI Whistleblower Tool (occorrerà stabilire cosa sia “AI Act-relevant”, ad esempio tramite la pubblicazione, in forma almeno aggregata, di statistiche e criteri di priorità, in linea con una logica di accountability amministrativa).

6. Dalla disciplina procedurale ad un profilo para-costituzionale?

La collocazione del Whistleblower Tool nel perimetro dell'AI Act rende evidente un dato: l'enforcement dell'AI non è (più) (solo) un fatto tecnico-amministrativo; ma un campo nel quale diritti e principi giocano a pieno titolo la propria partita. Si è visto che il segnalante, se tutelato, non è soltanto un soggetto che fornisce informazioni utili all'autorità, quanto una cartina di tornasole dell'efficacia dell'*agere* pubblico⁷⁹ e del relativo impianto regolatorio.

Simile logica implica però un bilanciamento rigoroso. La tutela del segnalante si intreccia, come accennato, con la libertà di espressione e di informazione (art. 11 CDFUE; art. 10 CEDU), con la tutela del lavoro (in particolare contro le possibili ritorsioni), entrando allo stesso tempo in tensione anche con la libertà d'impresa e la tutela del segreto (art. 16 CDFUE), con il diritto al rispetto della vita privata e dei dati personali (artt. 7 e 8 CDFUE), oltre che con il diritto a un ricorso effettivo e a un giudice imparziale (art. 47 CDFUE) e al diritto a una buona amministrazione (art. 41 CDFUE).

In questa prospettiva, il whistleblowing si conferma come istituto intrinsecamente condizionato, la cui legittimazione non deriva dall'utilità dell'informazione in quanto tale, ma dalla sua collocazione entro un quadro procedurale idoneo a garantire proporzionalità, correttezza e imputabilità delle decisioni. La giurisprudenza della Corte EDU – culminata nella sentenza *Halet* – mostra chiaramente come la tutela del segnalante non si traduca in una immunità generalizzata, ma presupponga una valutazione complessiva dell'interesse pubblico perseguito, della qualità e attendibilità delle informazioni, del pregiudizio arrecato agli interessi contrapposti e della proporzionalità delle conseguenze. È precisamente questa logica di bilanciamento, tipica dello Stato di diritto, che deve essere trasposta – e resa operativa – anche nel contesto del Whistleblower Tool dell'AI Act.

La conseguenza è che la proceduralizzazione dell'istituto non rappresenta un mero corollario tecnico, bensì una vera e propria condizione affinché l'enforcement digitale si realizzi in conformità ai principi dello Stato di diritto e ai diritti fondamentali dell'Unione. Le garanzie che presidiano il

77. Da qui l'esigenza di criteri trasparenti e controllabili, almeno nelle linee generali, per evitare che la selezione diventi opaca e discrezionale.

78. In cui sicurezza funzionale (*safety*) e cybersicurezza (*security*) convergono: casi in cui una vulnerabilità può generare incidenti non solo cyber-, ma anche “fisicamente” pericolosi. L'AI Act si concentra non per caso – nella propria classificazione del rischio – sugli impatti sociali e sui diritti fondamentali delle persone, ma è a tutti evidente che in futuro “robot” intelligenti ben potrebbero avere impatti niente affatto trascurabili sull'utente finale; tanto più che il Regolamento Macchine – 2023/1230 (che troverà applicazione dal 20 gennaio 2027) impone che il risk assessment consideri rischi che possono sorgere lungo tutto il ciclo di vita del robot (inclusi quelli “prevedibili” come evoluzione intenzionale di comportamenti self-evolving); e similmente fa l'AI Act, che per i sistemi high-risk richiede un risk management system lungo tutto il ciclo di vita (e dunque gestione del rischio anche post-market).

79. PARISI 2019 accosta il whistleblowing all'accesso civico, evidenziandone il ruolo di controllo diffuso del cittadino sull'agire della pubblica amministrazione in funzione di garanzia della trasparenza dell'azione pubblica.

trattamento delle segnalazioni – dalla confidenzialità alla protezione dei dati, dalla tracciabilità delle decisioni all'esistenza di criteri di valutazione verificabili, fino alle conseguenze derivanti dall'abuso del canale – si avvicinano, per struttura e funzione, ai requisiti del giusto procedimento amministrativo e, in senso lato, del giusto processo⁸⁰. Un sistema di segnalazioni che non garantisca un corretto trattamento dei dati, che non offra un livello minimo di trasparenza sugli esiti, o che non distingua tra segnalazioni fondate e infondate, rischierebbe inevitabilmente di perdere credibilità sia presso i segnalanti (che non vedranno tutela) sia presso gli operatori (che lo percepiranno come arbitrario). Ne offrono diretta conferma anche i novelli artt. 75 a-d AI Act, ove illustrano come l'esercizio dei poteri dell'AI Office in sede ispettiva (la cui attivazione ben potrà nascere da segnalazioni whistleblowing) non potrà essere disgiunto dalla sua legalità procedurale.

Il giusto procedimento, dunque, non è solo una garanzia per i destinatari finali, ma un presupposto funzionale del ruolo: il passaggio della tecnica regolatoria europea da una logica liberale di mercato a una logica costituzionalmente orientata di tutela dei diritti e di contenimento dei poteri⁸¹ determina la traslazione degli strumenti di enforcement da semplici apparati tecnici a strumenti procedurali passibili di una lettura, in senso materiale, funzionale all'inveramento della strategia europea di governo del digitale; dato il loro intervento diretto nel determinare quali diritti sono effettivamente garantiti (e come, dunque, quanto e a quali condizioni siano esercitabili). Naturalmente sono molteplici le attività amministrative che incidono su beni costituzionalmente protetti senza acquisire per ciò solo una diversa collocazione nel sistema delle fonti: è però vero che – come si diceva – quando trattasi di intelligenza artificiale l'enforcement normativo dipende anche e soprattutto, per la natura opaca della tecnologia e per la settorializzazione delle competenze, dalla disponibilità di conoscenze che l'autorità non è in grado di acquisire attraverso la sola osservazione esterna del mercato (si pensi a violazioni che possono essere incorporate nei processi di addestramento, nella selezione dei dati, nella documentazione tecnica,

nei risultati delle valutazioni interne, nelle decisioni di deployment o nelle omissioni organizzative che accompagnano il ciclo di vita del sistema). Il tool, allora, assume una funzione di mediazione tra informazione, attivazione del potere pubblico ed effettività della tutela. Quando l'infrastruttura determina le condizioni concrete al ricorrere delle quali un potere diviene osservabile e una pretesa di tutela può essere portata all'attenzione dell'autorità, le scelte relative al suo disegno non sono più esclusivamente tecniche, ma sistemiche.

In ultima analisi, il Whistleblower Tool dell'AI Act è utile a raccontare qualcosa del modo in cui l'Unione sta cercando di governare il digitale: non più soltanto mediante testi legislativi, ma con infrastrutture di controllo e di informazione capaci (rendendo l'illecito sanzionabile in quanto osservabile e segnalabile) di redistribuire orizzontalmente le istanze di tutela e gli obblighi di accountability. Compiendo un passo utile, forse, anche per tentare di porre un limite al nascente diritto, avente genesi privatistica, in cui la responsabilità politica si dissolve e le scelte fondamentali vengono spostate verso attori economici forti o verso la tecnica stessa.

Tuttavia, proprio perché incide sulle forme di controllo del potere, non si tratta di una scelta – né di una infrastruttura – neutra: se l'enforcement si automatizza, rischiando di risolversi in una decisione opaca o scarsamente sindacabile, il pericolo è quello di un ulteriore indebolimento della legalità procedimentale e sostanziale, già messa sotto pressione dalla proliferazione di fonti atipiche, *soft law* e standard tecnici. In altri termini, l'effettività della regolazione non può essere perseguita a tutti i costi: la *rule of law* non può essere sostituita da una più efficiente *rule of enforcement* o, peggio, da una *rule of tech*. Un nuovo approccio partecipato alla costruzione del framework digitale europeo può certamente rafforzare la tutela dei diritti, rendendo più permeabili e controllabili gli spazi, tradizionalmente opachi, del potere tecnologico; ma non è tutto oro quel che luccica: potendosi anche produrre nuove asimmetrie informative e, di riflesso, contribuire con un ulteriore strappo⁸² alla crisi di effettività che attanaglia il diritto dell'Unione.

80. EDPS 2019, p. 7 ss.

81. DE GREGORIO 2021, pp. 41-70.

82. AZZARITI 2003.

Il compito del giurista, in questo snodo, sembra essere eminentemente progettuale: contribuire a un design istituzionale nel quale gli strumenti di enforcement restino saldamente ancorati ai principi fondamentali della cosa pubblica – separazione dei poteri, imputabilità delle decisioni, sindacabilità giurisdizionale, proporzionalità – affinché la ricerca dell'effettività non si traduca in una nuova forma di opacità decisionale. Solo in questa cornice il Whistleblower Tool dell'AI Act può svolgere una funzione che potrebbe dirsi *para*-costituzionale,

derivante cioè non da una sua costituzionalità intrinseca (ovviamente assente), quanto dalla sua componente funzionale: il modo in cui l'informazione viene resa conoscibile, selezionata e trasformata in azione pubblica condiziona l'effettività dei diritti e i limiti al potere; divenendo – da semplice leva di applicazione del diritto positivo – un pur naturalmente circoscritto presidio alla tenuta democratica e garantista del governo europeo dell'intelligenza artificiale; tramite garanzie procedurali adeguate e una cultura di responsabilità diffusa.

Riferimenti bibliografici

- V. ABAZI (2020), *The European Union Whistleblower Directive: A 'Game Changer' for Whistleblowing Protection?*, in "Industrial Law Journal", vol. 49, 2020, n. 4
- V. ABAZI (2016), *Trade Secrets and Whistleblower Protection in the European Union*, in "European Papers", vol. 1, 2016, n. 3
- S. ANDREADAKIS, D. KAFTERANIS (2023), *Halet v Luxembourg: The Final Act of the Luxleaks Saga*, in "Oxford Business Law Blog", 2023
- G. AZZARITI (2003), *La superiore legalità costituzionale è in crisi? (editoriale)*, in "Costituzionalismo.it", 2003, n. 3
- M. BARONI (2024-A), *Costituzione polemica. Orientarsi nello stato di crisi permanente per riscoprire la dignità e l'autorealizzazione individuale*, Editoriale Scientifica, 2024
- M. BARONI (2024-B), *L'AI Act e il rischio dell'iper-regolamentazione: usiamo di più la "soft-law"*, in "Agendadigitale.eu", 2024
- T. BIRKSTEDT, M. MINKKINEN, A. TANDON, M. MÄNTYMÄKI (2023), *AI governance: themes, knowledge gaps and future agendas*, in "Internet Research", vol. 33, 2023, n. 7
- A. BRADFORD (2023), *Digital Empires*, Oxford University Press, 2023
- R. BRIGHI, L.B. BESSONE (2026), *Articolo 87. Segnalazione delle violazioni e protezione delle persone segnalanti*, in P. Perri, G. Vaciago, G. Ziccardi (a cura di), "Commentario AI Act e Legge Nazionale sull'intelligenza artificiale", Lefebvre Giuffrè, 2026
- J.M. BRYSON, B.C. CROSBY, M. MIDDLETON STONE (2006), *The Design and Implementation of Cross-Sector Collaborations: Propositions from the Literature*, in "Public Administration Review", vol. 66, 2006, n. 1
- C. CANCELA-OUTEDA (2024), *The EU's AI act: A framework for collaborative governance*, in "Internet of Things", vol. 27, 2024
- E. CEVA (2024), *Il whistleblowing come dovere di etica pubblica*, in "Etica Pubblica", 2024, n. 2
- G.E. CORSARO (2023), *Il whistleblower come guardiano dei valori democratici: la sentenza Halet della Corte europea*, in "Diritti umani e diritto internazionale", 2023, n. 3
- G. DE GREGORIO (2021), *The rise of digital constitutionalism in the European Union*, in "International Journal of Constitutional Law", vol. 19, 2021, n. 1
- G. DE MINICO (2025), *Le fonti del diritto: un argine all'intelligenza artificiale?*, in "Rivista AIC", 2025, n. 3
- F. DONATI (2025), *La protezione dei diritti fondamentali nel regolamento sull'intelligenza artificiale*, in "Rivista AIC", 2025, n. 1

- V.M. DONINI (2024), *La simbiosi tra “accountability” e “whistleblowing” nel quadro del governo aperto: un’analisi delle dinamiche e delle interazioni*, in “federalismi.it”, 2024, n. 16
- V.M. DONINI (2022), *La tutela del whistleblower tra resistenze culturali e criticità legislative*, in “Penale Diritto e Procedura”, 2022
- EDPS (2019), *Guidelines on processing personal information within a whistleblowing procedure*, 2019
- T.E. FROSINI (2020), *Il costituzionalismo nella società tecnologica*, in “Consulta Online”, 2020
- M. FUMAGALLI, M. LORUSSO (2024), *Il d.lgs. del 10 marzo 2023*, n. 24, Key editore, 2024
- M. GIANNELLI (2026), *L’ordine invisibile. Cybersicurezza, poteri, diritti*, Giappichelli, 2026
- M. GIANNELLI (2025), *Art. 77. Poteri delle autorità che tutelano i diritti fondamentali*, in A. Mantelero, G. Resta, G.M. Riccio (a cura di), “Intelligenza artificiale commentario”, Wolters Kluwer, 2025
- S. HINCK, J. VAN DEN BOOM (2026), *Closing information gaps in DMA enforcement – Expectations, incentives, and the role of anonymity for whistleblowers*, in “Journal of Antitrust Enforcement”, vol. 14, 2026, n. 2
- N. LI (2023), *Ethical considerations in artificial intelligence: a comprehensive discussion from the perspective of computer vision*, in “SHS Web of Conferences”, vol. 179, 2023, n. 1
- E. LONGO (2025), *Regulatory Sandboxes: A wager on Europe’s AI strategy*, in “BioLaw Journal – Rivista di BioDiritto”, 2025, n. 3
- G. LO SAPIO (2021), *La black box: l’esplicabilità delle scelte algoritmiche quale garanzia di buona amministrazione*, in “federalismi.it”, 2021, n. 3
- C. MAGNANI (2018), *Libertà di espressione e fake news, il difficile rapporto tra verità e diritto. Una prospettiva teoretica*, in “Costituzionalismo.it”, 2018, n. 3
- M. MAGRI (2019), *Il whistleblowing nella prospettiva di una disciplina europea armonizzata: la legge n. 179 del 2017 sarà (a breve) da riscrivere?*, in “federalismi.it”, 2019, n. 18
- M.P. MICELI, J.P. NEAR (1992), *Blowing the Whistle: The Organizational and Legal Implications for Companies and Employees*, Lexington Books, 1992
- A. MORRONE (2008), *Bilanciamento (giustizia costituzionale)*, in “Enciclopedia del diritto. Annali”, volume II, tomo II, 2008
- J.P. NEAR, M.P. MICELI (1985), *Organizational Dissidence: The Case of Whistle-Blowing*, in “Journal of Business Ethics”, vol. 4, 1985, n. 1
- NOYB (2024), *GDPR: a culture of non-compliance? Numbers of evidence-based enforcement efforts*, European Center for Digital Rights, 2024
- M. PAPPALARDO (2026), *Il Digital Omnibus sull’Intelligenza Artificiale in GUUE: cosa cambia per le imprese*, in Altalex.com, 27 luglio 2026
- N. PARISI (2020), *La funzione del whistleblowing nel diritto internazionale ed europeo*, in “Lavoro, Diritti, Europa”, 2020, n. 2
- N. PARISI (2019), *La trasparenza amministrativa, strumento di prevenzione della corruzione*, in Avviso Pubblico (a cura di), “La trasparenza (im)possibile. Tutto quello che c’è da sapere sul diritto d’accesso in Italia”, Altreconomia, 2019
- G. PITRUZZELLA (2018), *La libertà di informazione nell’era di Internet*, in “MediaLaws”, 2018, n. 1
- F. PONTE (2021), *La Corte EDU sul whistleblowing: la libertà di espressione, i canali pubblici di segnalazione e una prospettiva non (ancora) italiana*, in “DPCE online”, 2021, n. 2

- O. POLLICINO (2025), *Regolazione e innovazione tecnologica nell'ordinamento della rete*, in "Rivista AIC", 2025, n. 2
- E.C. RAFFIOTTA (2025), *Shaping Europe's digital future through a change of perspective: the necessary rethinking of the relationship between regulation and innovation in the EU*, in "BioLaw Journal – Rivista di BioDiritto", 2025, n. 4
- E.C. RAFFIOTTA (2023), *Dalla self-regulation alla over-regulation in ambito digitale: come (e perché) di un necessario cambio di prospettiva*, in "Osservatorio sulle fonti", 2023, n. 2
- T.A. SCOTT, C.W. THOMAS (2017), *Unpacking the Collaborative Toolbox: Why and When Do Public Managers Choose Collaborative Governance Strategies?*, in "Policy Studies Journal", vol. 45, 2017, Special Issue n. 1
- C. SPINIELLO (2025), *È (ancora) tempo di costituzionalismo digitale*, in "Rivista Italiana di Informatica e Diritto", 2025, n. 1
- G. TARTAGLIA POLCINI (2022), *Il whistleblowing come strumento di prevenzione della corruzione: la protezione delle reporting person nel multilateralismo giuridico*, in "Archivio Penale", 2022, n. 3
- V.M. TEDESCO (2023), *Whistleblowing, alla luce del d.lgs. n. 24/2023, attuativo della dir. UE n. 2019/1937 sulla protezione dei segnalanti*, in "Massimario di giurisprudenza del lavoro", 2023, n. 4
- L. VALENTE, E. GHERA (2021), *Sulla riforma degli obblighi di segreto contenuta nella legge sul whistleblowing*, in "Rivista Italiana di Diritto del Lavoro", 2021, n. 3
- F. VITARELLI (2023), *Il whistleblowing come "baluardo" della libertà di espressione. La grande camera riconosce ampia tutela al whistleblower nell'ottica di prevenire chilling effect delle sanzioni sul diritto di informare la collettività su fatti di interesse pubblico*, in "RIDPP – Rivista Italiana di Diritto e Procedura Penale", vol. 66, 2023, n. 2