LUISS

Corso di laurea in Giurisprudenza

Cattedra di Artificial Intelligence And International Law

L'intelligenza artificiale tra normativa e impieghi militari: profili giuridici e applicativi

Prof. Ruoppo Roberto

RELATORE

Prof. Brozzetti Filiberto

CORRELATORE

Marco Parisi (matr. 162973)

CANDIDATO

Anno Accademico 2024/2025

INDICE

INTRODUZIONE

CAPITOLO I

NORMATIVA SOVRANAZIONALE IN MATERIA DI IA

CAPITOLO II

LA TECNOLOGIA DI RICONOSCIMENTO FACCIALE NEL CONTESTO GIURIDICO EUROPEO E INTERNAZIONALE

CAPITOLO III

I SISTEMI D'ARMA AUTONOMI E IL CASO ISRAELO-PALESTINESE

CONCLUSIONI

INTRODUZIONE

L'intelligenza artificiale

L'intelligenza artificiale rappresenta oggi una delle tecnologie trasformative più significative della nostra epoca, perché ha ridefinito i confini tra possibilità tecnologiche e responsabilità umane, in modi che erano impensabili solo pochi decenni fa. Questa rivoluzione digitale non si limita a modificare i processi produttivi o le modalità di comunicazione, ma penetra nelle strutture fondamentali della società contemporanea, influenzandone decisioni che toccano la vita quotidiana di miliardi di persone. Dalle raccomandazioni algoritmiche che orientano le nostre scelte di consumo ai sistemi predittivi che supportano le decisioni giudiziarie, dall'automazione industriale ai veicoli autonomi, l'intelligenza artificiale si configura come una tecnologia pervasiva che richiede un ripensamento profondo delle categorie giuridiche, etiche e delle politiche tradizionali.[1]

La natura stessa dell'intelligenza artificiale solleva interrogativi fondamentali, che vanno ben oltre la dimensione puramente tecnica. Cosa significa delegare a un algoritmo decisioni che fino a ieri erano prerogativa esclusiva dell'intelligenza umana? Come possiamo garantire che sistemi progettati per ottimizzare specifici parametri non compromettano valori più ampi come l'equità, la trasparenza e la dignità umana? In che modo le società democratiche possono governare tecnologie che evolvono a una velocità superiore a quella dei tradizionali processi normativi? Questi interrogativi assumono particolare urgenza, quando l'intelligenza artificiale viene impiegata in settori sensibili come la sicurezza pubblica, la giustizia penale e, soprattutto, il dominio militare, dove le conseguenze degli errori algoritmici possono tradursi in violazioni dei diritti fondamentali o, nei casi più estremi, in perdite di vite umane.

Il panorama contemporaneo dell'intelligenza artificiale è caratterizzato da una tensione costante tra innovazione e regolamentazione, tra le promesse di efficienza e produttività offerte da queste tecnologie e i rischi sistemici che esse comportano per

[1] Atabekov, A., "*Artificial Intelligence in Contemporary Societies: Legal Status and Definition, Implementation in Public Sector across Various Countries*", *Social Sciences*, 12(3), Art. 178, 2023. Ricognizione comparata su definizioni e *status* giuridico dell'IA e sulle strategie di implementazione nel settore pubblico in ordinamenti di diverse tradizioni; utile per l'inquadramento generale sulla pervasività e sulla *governance* pubblica richiamato nell'introduzione. Disponibile su: https://www.mdpi.com/2076-0760/12/3/178

la stabilità sociale e la protezione dei diritti individuali. Le capacità predittive degli algoritmi di machine learning, alimentate da quantità di dati senza precedenti, hanno dimostrato di poter superare le prestazioni umane in numerosi campi, dalla diagnosi medica al riconoscimento di pattern complessi. Tuttavia, questa potenza di calcolo è spesso accompagnata da una scarsa trasparenza, che rende difficile capire come un sistema di intelligenza artificiale arrivi alle sue conclusioni.[2]

La questione della trasparenza algoritmica, si intreccia strettamente con quella della responsabilità. Quando un sistema autonomo commette un errore o produce un risultato discriminatorio, a chi deve essere attribuita la responsabilità? Al programmatore che ha scritto il codice, all'organizzazione che ha raccolto e preparato i dati di addestramento, all'autorità che ha autorizzato l'impiego del sistema, o all'operatore che ne ha supervisionato il funzionamento? Questa distribuzione della responsabilità lungo catene decisionali complesse, rappresenta una delle sfide più significative per i sistemi giuridici contemporanei, tradizionalmente costruiti attorno al concetto di responsabilità individuale e di causalità diretta.

L'Unione Europea ha tentato di rispondere a queste sfide, attraverso l'adozione di vari strumenti; infatti nel primo capitolo di questa ricerca si esaminerà il quadro normativo sovranazionale in materia di intelligenza artificiale, con particolare attenzione all'Artificial Intelligence Act dell'Unione Europea e alla Convenzione quadro del Consiglio d'Europa. L'analisi si concentrerà sulla classificazione risk-based dei sistemi di IA, sui meccanismi di governance e controllo previsti dalla normativa europea, e sul ruolo della *soft law* nella definizione di standard e principi condivisi a livello internazionale, nonché ai principi dell'OCSE per un'IA affidabile, e una molteplicità di iniziative di *soft law*, che testimoniano la complessità della governance di queste tecnologie.[3]

---

[2] Sakubu, D., "*Challenges of Artificial Intelligence today and future implications for society and the world*", *World Journal of Advanced Research and Reviews*, 26(1), 2025, pp. 3045–3054. Rassegna *peer-reviewed* delle sfide sistemiche dell'IA, con attenzione a prestazioni dei sistemi di *machine learning* e problemi di trasparenza/responsabilità; pertinente al passaggio sulle capacità predittive e sull'opacità dei processi decisionali. DOI: 10.30574/wjarr.2025.26.1.1380. PDF disponibile su: https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-1380.pdf

[3] Commissione europea, "*European approach to artificial intelligence*". Sintesi ufficiale dell'approccio dell'UE all'intelligenza artificiale, fondato sui principi di eccellenza e fiducia, con rinvii all'*Artificial Intelligence Act* e agli strumenti di *soft law*; rilevante per il passaggio sull'architettura di *governance* europea. Disponibile su: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Non è un caso che parallelamente al processo di regolamentazione europea, altri attori globali hanno sviluppato approcci differenti alla governance dell'intelligenza artificiale. Gli Stati Uniti hanno privilegiato un modello basato prevalentemente su strumenti di *soft law* e iniziative settoriali, riflettendo una cultura giuridica che valorizza la flessibilità normativa e l'innovazione tecnologica. La Cina ha adottato un approccio pragmatico, che combina elementi di regolamentazione vincolante con meccanismi di controllo sociale, in linea con il proprio modello di governance autoritaria. Questa diversità di approcci normativi, riflette non solo differenze culturali e politiche, ma anche visioni contrastanti sul ruolo che l'intelligenza artificiale dovrebbe svolgere nella società e sui meccanismi più appropriati per governarne lo sviluppo e l'impiego.[4]

La presente ricerca si concentra inoltre su due ambiti applicativi dell'intelligenza artificiale, che presentano particolare rilevanza per la tutela dei diritti fondamentali e per la stabilità dell'ordine giuridico internazionale.

Il primo ambito di indagine riguarda i sistemi di riconoscimento facciale e, più in generale, le tecnologie biometriche impiegate per finalità di sorveglianza e controllo sociale. Questi sistemi, caratterizzati da una capacità di identificazione e tracciamento senza precedenti, pongono sfide inedite per la protezione della privacy, della libertà di movimento e di espressione, e per il principio di non discriminazione. La natura probabilistica di queste tecnologie, unita alla loro tendenza a riprodurre e amplificare i bias presenti nei dati di addestramento, può tradursi in forme di discriminazione algoritmica che colpiscono in modo sproporzionato le minoranze e i gruppi vulnerabili.

Il secondo ambito di indagine concerne l'impiego dell'intelligenza artificiale nel dominio militare e, in particolare, lo sviluppo di sistemi d'arma autonomi capaci di selezionare e ingaggiare obiettivi senza intervento umano diretto. Questa evoluzione tecnologica rappresenta una trasformazione paradigmatica nella conduzione della guerra, sollevando interrogativi fondamentali sulla compatibilità di

[4] Alic, Dalia, *The Role of Data Protection and Cybersecurity Regulations in Artificial Intelligence Global Governance: A Comparative Analysis of the European Union, the United States, and China Regulatory Framework*, Tesi di laurea magistrale, Central European University, Vienna, giugno 2021. Studio comparato su UE/USA/Cina che mostra differenze di approccio e di tecnica regolatoria; fonda il paragrafo sulla diversità degli approcci globali. PDF disponibile su: https://www.etd.ceu.edu/2021/alic_dalia.pdf

tali sistemi con i principi del diritto internazionale umanitario, in particolare con i principi di distinzione, proporzionalità e precauzione. La possibilità che macchine possano prendere decisioni di vita o di morte in modo autonomo, sfida le fondamenta etiche e giuridiche su cui si basa la regolamentazione internazionale dei conflitti armati.

L'analisi di questi due domini applicativi, non può prescindere dall'esame di casi concreti che illustrano le implicazioni pratiche dell'impiego di sistemi di intelligenza artificiale in contesti ad alto rischio. Il caso Clearview AI ha dimostrato come la raccolta massiva di dati biometrici da fonti pubbliche, possa creare archivi di sorveglianza di dimensioni e capacità senza precedenti, sollevando questioni cruciali sulla legittimità di tali pratiche e sui meccanismi di controllo necessari per prevenirne l'abuso. La sentenza Glukhin c. Russia della Corte Europea dei Diritti dell'Uomo, ha stabilito importanti precedenti giurisprudenziali sui limiti dell'impiego del riconoscimento facciale da parte delle autorità pubbliche, definendo standard rigorosi per la giustificazione di tali misure e per le garanzie procedurali che devono accompagnarle.

Sull'aspetto del dominio militare, si vedrà come il conflitto israelo-palestinese ha assunto una dimensione tecnologica senza precedenti, caratterizzandosi come quello che alcuni analisti hanno definito "la prima guerra di intelligenza artificiale su larga scala". L'impiego di sistemi come Lavender, Gospel e Where's Daddy per l'identificazione e il *targeting* automatizzato di obiettivi militari, rappresenta un'evoluzione qualitativa nella conduzione delle operazioni belliche, con implicazioni profonde per la protezione dei civili e per l'applicazione del diritto internazionale umanitario. Questi sistemi, progettati per operare in ambienti urbani complessi, caratterizzati da un'elevata densità demografica, sollevano interrogativi cruciali sulla capacità dell'intelligenza artificiale di effettuare le distinzioni morali e giuridiche richieste dal diritto dei conflitti armati. Nel terzo capitolo infatti, verrà affrontata la questione dei sistemi d'arma autonomi, esaminando le definizioni concettuali, i modelli di interazione umana, e le implicazioni per il diritto internazionale umanitario. L'analisi si concentra sui concetti di controllo umano significativo e di responsabilità nella catena decisionale, nonché sulle sfide poste dall'automazione delle decisioni letali. Il caso del conflitto israelo-palestinese viene

utilizzato come laboratorio per comprendere le implicazioni pratiche dell'impiego di sistemi di intelligenza artificiale in contesti bellici reali, con particolare attenzione ai sistemi di *targeting* automatizzato e di sorveglianza biometrica.

L'approccio metodologico adottato in questa ricerca privilegia l'analisi interdisciplinare, riconoscendo che la comprensione delle implicazioni giuridiche dell'intelligenza artificiale richiede una conoscenza approfondita, non solo delle norme e dei principi giuridici, ma anche delle caratteristiche tecniche di questi sistemi e delle dinamiche organizzative che ne governano lo sviluppo e l'impiego. La validità e l'efficacia delle regole giuridiche, dipendono infatti dalla loro capacità di tradursi in procedure operative concrete e verificabili, capaci di orientare il comportamento degli attori coinvolti e di fornire strumenti per l'attribuzione della responsabilità in caso di malfunzionamenti o abusi.

La dimensione comparativa assume particolare rilevanza in questo contesto, considerando che l'intelligenza artificiale è una tecnologia intrinsecamente globale che trascende i confini nazionali e richiede forme di coordinamento internazionale, per essere efficacemente governata. L'analisi delle diverse strategie normative adottate dall'Unione Europea, dagli Stati Uniti, dalla Cina e da altri attori internazionali, offre spunti preziosi per comprendere i vantaggi e i limiti dei diversi approcci alla regolamentazione dell'IA, nonché le possibilità di convergenza verso standard comuni.

L'obiettivo di questa ricerca non è quello di fornire risposte definitive a questioni che sono ancora in rapida evoluzione, ma piuttosto di contribuire al dibattito scientifico ed etico, offrendo strumenti concettuali e analitici per comprendere la complessità delle sfide poste dall'intelligenza artificiale alla governance contemporanea. Nel complesso, il presente elaborato afferma il principio secondo cui l'efficacia della regolamentazione dell'IA, dipende dalla capacità di coniugare principi giuridici solidi con una comprensione approfondita delle caratteristiche tecniche di questi sistemi e delle dinamiche organizzative che ne governano l'impiego. Solo attraverso questo approccio integrato sarà possibile sviluppare forme di governance che sappiano cogliere le opportunità offerte dall'intelligenza artificiale, senza compromettere i valori fondamentali su cui si basano le società democratiche.

CAPITOLO I

NORMATIVA SOVRANAZIONALE IN MATERIA DI IA

Sommario: 1. L'Artificial Intelligence Act; 1.1. Sistemi di intelligenza artificiale a basso rischio; 1.2. Sistemi di intelligenza artificiale a rischio limitato; 1.3. Sistemi di intelligenza artificiale ad alto rischio; 1.4. Sistemi di intelligenza artificiale a rischio sistemico; 1.5. Sistemi di intelligenza artificiale proibiti; 2. La Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto; 2.1. I rapporti tra la Convenzione quadro e l'AI Act; 2.2. Sviluppo negoziale e criticità sistemiche della Convenzione quadro sull'IA; 3. La *soft law* nell'intelligenza artificiale a livello internazionale; 3.1. L'approccio dell'OCSE: principi per un'IA affidabile; 3.2. L'Europa: oltre l'AI Act, un "ecosistema di *soft law*"; 3.3. Gli Stati Uniti: un "mosaico di iniziative di *soft law*"; 3.4. La Cina: un approccio pragmatico tra *soft law* e regolamentazione emergente.

1. L'Artificial Intelligence Act

L'AI Act, ossia "Artificial Intelligence Act", è il primo regolamento al mondo sull'intelligenza artificiale e mira ad assicurare che i sistemi di IA immessi sul mercato europeo siano sicuri e rispettino i diritti fondamentali e i valori dell'Unione Europea. Il Parlamento Europeo lo ha approvato il 13 marzo 2024, il Consiglio dell'UE ha dato la sua approvazione definitiva il 21 maggio 2024 ed è entrato in vigore il 1° agosto 2024. L'AI Act si basa su un approccio fondato sul rischio e classifica i sistemi di intelligenza artificiale in quattro categorie: sistemi a basso rischio, a rischio limitato, ad alto rischio e sistemi di intelligenza artificiale proibiti.

Il Regolamento sull'Intelligenza Artificiale rappresenta un punto di svolta nel panorama giuridico internazionale, configurandosi come il primo intervento normativo organico e settoriale adottato da una grande organizzazione sovranazionale in materia di intelligenza artificiale. Il regolamento si caratterizza per una struttura articolata e per una significativa innovazione metodologica, in quanto

delinea un quadro regolatorio uniforme per tutti gli Stati membri, con efficacia diretta e immediata sulle imprese e sulle amministrazioni che sviluppano, immettono sul mercato o utilizzano sistemi di IA nell'Unione Europea.

Una delle principali novità dell'AI Act consiste nella sua ambizione di combinare la promozione dell'innovazione con la necessità di garantire un elevato livello di tutela dei diritti fondamentali, della sicurezza e dell'ordine pubblico. Il regolamento mira a coniugare la competitività dell'industria europea dell'IA con la protezione effettiva degli individui, riconoscendo la natura trasversale e trasformativa di queste tecnologie. In questo senso, l'AI Act si pone non solo come riferimento per il mercato interno, ma anche come modello potenzialmente replicabile in altri contesti regionali e internazionali, rafforzando il ruolo dell'Unione Europea come "*normative power*" nel settore digitale.[5]

Dal punto di vista sistematico, l'AI Act si distingue per l'introduzione di una disciplina fortemente orizzontale, destinata a integrarsi con la normativa settoriale già vigente (ad esempio in materia di sicurezza dei prodotti, protezione dei dati personali, responsabilità civile), e per l'attenzione riservata ai meccanismi di enforcement e ai poteri delle autorità di vigilanza. L'istituzione di un AI Office europeo e di una rete di autorità nazionali competenti rafforza infatti la capacità di monitorare, indirizzare e coordinare l'applicazione della disciplina, a garanzia di un'effettiva uniformità interpretativa e operativa tra i diversi Stati membri.[6] L'importanza dell'AI Act risiede anche nel suo approccio dinamico e "*future-proof*", ossia nella capacità di adattarsi alle evoluzioni tecnologiche e di prevedere meccanismi di aggiornamento e revisione, che rendono il regolamento uno strumento aperto e resiliente rispetto alle sfide poste dalla rapida trasformazione dell'IA.[7]

---

[5] L. Floridi et al., "*AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*", AI4People, 2018, pp. 23-28. Gli autori sottolineano come l'Unione Europea, attraverso l'adozione di un quadro etico e normativo avanzato per l'intelligenza artificiale, si ponga come modello di riferimento ("*reference model*") e potenziale *leading example* per la comunità internazionale, promuovendo uno sviluppo tecnologico responsabile, orientato ai diritti fondamentali, all'inclusività e al benessere sociale. In particolare, AI4People auspica che l'approccio europeo venga assunto come standard condiviso anche in altri contesti giuridici e culturali. Disponibile su: https://ai4people.org/PDF/AI4People_Ethical_Framework_For_A_Good_AI_Society.pdf

[6] Regolamento (UE) 2024/1689 sull'intelligenza artificiale (AI Act), artt. 64-71.

[7] M. Veale, F. Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, in *Computer Law Review International*, vol. 22, 2021, pp. 97–106. Gli autori offrono un'analisi dettagliata dell'*AI Act*, sottolineando la struttura orizzontale del regolamento, la sua capacità di

Nei prossimi paragrafi approfondiremo quindi i vari livelli di rischio disciplinati dall'AI Act.

1.1. Sistemi di intelligenza artificiale a basso rischio

I sistemi di intelligenza artificiale a basso rischio rappresentano una categoria di applicazioni che, pur utilizzando l'intelligenza artificiale, non comportano per i diritti fondamentali o per la sicurezza degli utenti alcun rischio significativo. Tali sistemi, che includono tipologie di IA come filtri spam, videogiochi o chatbot di uso ordinario, non incidono in maniera rilevante sulle decisioni degli utenti e, pertanto, sono considerati a rischio minimo dal regolamento.

La disciplina dei sistemi a basso rischio è limitata all'art. 95 dell'AI Act, che incoraggia l'elaborazione di codici di condotta volontari. Questi strumenti, pur privi di natura vincolante, sono concepiti per estendere, su base non obbligatoria, l'applicazione ai sistemi a basso rischio dei requisiti che il Regolamento impone ai sistemi ad alto rischio nel Capo III, Sezione 2. In tal modo, il legislatore europeo favorisce la diffusione di standard più elevati in materia di trasparenza, qualità dei dati, documentazione e supervisione umana, così da garantire un utilizzo responsabile dell'intelligenza artificiale anche al di fuori del perimetro degli obblighi giuridici stringenti.

Nel comma 1 si stabilisce che l'Ufficio per l'intelligenza artificiale e gli Stati membri sono incaricati di incoraggiare e facilitare l'elaborazione di tali codici di condotta. Questi devono promuovere l'applicazione volontaria di alcuni o tutti i requisiti stabiliti nel Capo III, Sezione 2 del regolamento, che si riferisce ai requisiti generali per i sistemi di IA ad alto rischio. È importante notare che questa applicazione deve tenere conto delle specifiche soluzioni tecniche e delle migliori pratiche nel settore, assicurando che i requisiti siano praticabili e adattabili alla tipologia di sistema di IA in questione. Il comma 1 enfatizza, quindi, la necessità di

---

integrarsi con la normativa settoriale vigente (in particolare in materia di sicurezza dei prodotti, protezione dei dati personali e responsabilità civile), e la centralità dei meccanismi di *enforcement* e delle autorità di vigilanza, tra cui l'istituzione di un *AI Office* europeo e di una rete di autorità nazionali competenti. L'articolo evidenzia inoltre la volontà del legislatore europeo di assicurare una disciplina flessibile e aggiornata, in grado di adattarsi al rapido progresso tecnologico attraverso strumenti normativi di revisione periodica, pdf disponibile su: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3896852

un approccio pragmatico che rispetti le capacità tecniche e le peculiarità dei diversi sistemi di IA.

Il comma 2 estende la possibilità di applicazione volontaria a requisiti specifici che possono riguardare tutti i sistemi di IA, non limitandosi a quelli a basso rischio. Tali requisiti devono essere misurabili tramite obiettivi chiari e indicatori di prestazione. Questo comma fornisce un'ulteriore struttura di orientamento, che include diverse aree specifiche per migliorare l'affidabilità e l'impatto sociale dei sistemi di IA. Tra gli esempi proposti vi sono:

a) gli elementi applicabili previsti negli orientamenti etici dell'Unione per un'IA affidabile;

b) valutare e ridurre al minimo l'impatto dei sistemi di intelligenza artificiale sulla sostenibilità ambientale, anche per quanto riguarda la programmazione efficiente dal punto di vista energetico e le tecniche per la progettazione, la formazione e l'uso efficienti dell'intelligenza artificiale;

c) promuovere la conoscenza dell'IA, in particolare delle persone che si occupano dello sviluppo, del funzionamento e dell'uso dell'IA;

d) facilitare una progettazione inclusiva e diversificata dei sistemi di intelligenza artificiale, anche attraverso l'istituzione di team di sviluppo inclusivi e diversificati e la promozione della partecipazione delle parti interessate a tale processo;

(e) valutare e prevenire l'impatto negativo dei sistemi di intelligenza artificiale sulle persone vulnerabili o sui gruppi di persone vulnerabili, anche per quanto riguarda l'accessibilità per le persone con disabilità, nonché sulla parità di genere".[8]

Il comma 3 definisce i soggetti legittimati a redigere i codici di condotta, che possono essere i singoli fornitori o utilizzatori di sistemi di IA, ma anche organizzazioni rappresentative di tali attori, comprese organizzazioni della società civile e istituzioni accademiche. I codici di condotta possono riguardare uno o più sistemi di IA, a seconda della somiglianza degli scopi di tali sistemi, assicurando che le linee guida siano adeguate al contesto e agli obiettivi specifici.

Infine, il comma 4 sottolinea l'importanza di prendere in considerazione le esigenze specifiche delle PMI (Piccole e Medie Imprese), incluse le start-up, nell'elaborazione dei codici di condotta. Questa disposizione riflette la necessità di

[8] Regolamento (UE) 2024/1689 sull'intelligenza artificiale (AI Act), art 95 c. 2

garantire che le piccole imprese, che potrebbero avere risorse limitate, non siano eccessivamente penalizzate o sopraffatte da requisiti troppo onerosi. In questo modo, l'articolo cerca di promuovere una partecipazione inclusiva, che permetta anche alle realtà più piccole di adottare pratiche di IA responsabili, senza compromettere la loro competitività.

In sintesi, l'art. 95 dell'AI Act mira a incentivare l'adozione volontaria di codici di condotta per i sistemi di IA a basso rischio, promuovendo un approccio flessibile, pratico e inclusivo. Pur non imponendo obblighi severi, la disposizione incoraggia a estendere, su base non vincolante, l'applicazione ai sistemi a basso rischio dei requisiti sanciti dal Capo III, Sezione 2, originariamente previsti per i sistemi ad alto rischio. In questo modo, si favorisce la diffusione di standard più elevati di trasparenza, affidabilità e sostenibilità, assicurando che anche tali sistemi siano sviluppati in modo responsabile e rispettoso dei diritti fondamentali.

L'Articolo 95 dell'AI Act riprende e adatta la logica dei codici di condotta già prevista dall'art. 40 GDPR, sebbene vi siano differenze significative quanto all'approvazione e al monitoraggio di tali codici. Secondo la dottrina,[9] i codici dell'AI Act, a differenza di quelli del GDPR, non necessitano di approvazione da parte di un'autorità europea, né di un sistema di monitoraggio esterno. Tuttavia, l'esperienza maturata nell'ambito della privacy fornisce un utile riferimento per comprendere i possibili effetti di questi strumenti di *soft law* nel contesto dell'intelligenza artificiale.

1.2. Sistemi di intelligenza artificiale a rischio limitato

I sistemi di intelligenza artificiale a rischio limitato rappresentano un gruppo di tecnologie disciplinate dal Regolamento sull'Intelligenza Artificiale (AI Act), che si trovano tra i sistemi a basso rischio e quelli ad alto rischio. Questi sistemi sono contraddistinti da una serie di specifiche normative destinate a mitigarne i potenziali rischi, i quali non sono di tale entità da giustificare una regolazione tanto stringente

[9] Ceyhun Necati Pehlivan, Nikolaus Forgó, Peggy Valcke (a cura di), *The EU Artificial Intelligence (AI) Act: Commentary*, Oxford University Press, Oxford, 2024, p. 1317, par. 3.5.1 Il volume offre un'analisi sistematica delle differenze tra i codici di condotta previsti dall'AI Act (art. 95) e quelli disciplinati dal GDPR (art. 40), evidenziando che i primi non richiedono approvazione né monitoraggio esterno da parte di un'autorità europea, a differenza del modello adottato in ambito privacy. PDF disponibile su:

quanto quella applicata ai sistemi ad alto rischio, ma sono comunque significativi e necessitano di un quadro di protezione adeguato per salvaguardare i diritti degli individui e la sicurezza collettiva.

Nel contesto del AI Act, per i sistemi a medio rischio ci rifacciamo all'articolo 50, il quale stabilisce l'obbligo per i fornitori di IA di garantire la trasparenza. Questo principio di trasparenza è fondamentale per ridurre i rischi di inganno e manipolazione che potrebbero derivare dall'uso di sistemi automatizzati in contesti in cui l'interazione tra l'uomo e la macchina può sembrare indistinguibile da quella con un altro essere umano, come nel caso di chatbot avanzati o sistemi di IA generativa. La trasparenza non è solo una questione di rendere noti all'utente i funzionamenti di base del sistema di IA, ma implica anche l'obbligo di informarlo su come vengono elaborati i dati, quali decisioni siano influenzate dall'IA e come i risultati generati possano impattare la sua vita.

L'Art. 50 chiarisce che l'utente deve essere esplicitamente informato che sta interagendo con un sistema di IA, impedendo che si crei confusione o una falsa percezione di interazione con un essere umano; il principio si applica a una vasta gamma di sistemi, dai chatbot utilizzati nei servizi clienti fino ai sistemi di raccomandazione di contenuti online. In particolare, la necessità di trasparenza è particolarmente rilevante in contesti in cui l'IA è utilizzata per influenzare decisioni, come nel caso dei *deep fake* o dei sistemi di raccomandazione, dove la non conoscenza della natura algoritmica del sistema potrebbe comportare per l'utente rischi di manipolazione o distorsione della percezione.

Interessante notare come nel febbraio 2019, il Parlamento ha adottato una risoluzione sull'intelligenza artificiale che trattava una serie di questioni, tra cui i requisiti di trasparenza. La risoluzione adottata dal Parlamento poneva l'accento sull'importanza della trasparenza nelle applicazioni dell'intelligenza artificiale, stabilendo che ogni forma di interazione tra un utente e un sistema di IA dovesse essere esplicitamente riconoscibile. In particolare, veniva chiesto che fosse introdotto un meccanismo obbligatorio che consentisse agli utenti di distinguere con chiarezza quando stavano comunicando o fruendo di contenuti generati artificialmente. Nell'ambito delle produzioni digitali, la risoluzione sollecitava la Commissione ad imporre ai creatori di *deep fake*, ossia video e immagini sintetiche realizzate

mediante IA, l'obbligo di segnalare in modo inequivocabile la natura artificiale di tali materiali, garantendo così una maggiore tutela rispetto a possibili manipolazioni o fraintendimenti da parte del pubblico,[10] preoccupazione evidentemente ripresa e normata dall'AI Act all'articolo 50.

1.3. Sistemi di intelligenza artificiale ad alto rischio

I sistemi di intelligenza artificiale ad alto rischio sono disciplinati principalmente dal Capitolo II dell'AI Act, in particolare dagli Articoli 6 e 7, che definiscono con chiarezza quali siano le applicazioni che rientrano in questa categoria. Tali sistemi sono considerati ad alto rischio in quanto operano in settori che, sebbene cruciali per il funzionamento della società, presentano un rischio significativo per la sicurezza, la salute e i diritti fondamentali delle persone. La necessità di un quadro normativo rigoroso per questi sistemi nasce proprio dalla gravità delle conseguenze che il loro malfunzionamento o abuso potrebbe provocare.

L'Articolo 6 definisce con precisione i criteri per identificare i sistemi di IA ad alto rischio. Si tratta di applicazioni utilizzate in ambiti sensibili, come sanità, giustizia, trasporti, educazione e servizi finanziari, dove le decisioni automatizzate potrebbero influire direttamente sulla vita delle persone. Un esempio lampante riguarda i sistemi di IA che supportano diagnosi mediche o la gestione dei dispositivi sanitari, dove errori potrebbero mettere in pericolo la salute dei pazienti. Analogamente, in ambito giuridico, l'utilizzo di IA per l'analisi di prove o per determinare sentenze legali può incidere gravemente sui diritti degli individui, come nel caso della valutazione del rischio di recidiva per le decisioni giudiziarie.

Oltre alla specificità dei settori, l'Articolo 6 introduce anche il concetto di rischio sistemico, un aspetto che si riferisce agli effetti che un sistema di IA può avere sul mercato in generale. Tali effetti, sebbene non immediatamente percepibili, potrebbero alterare significativamente il funzionamento della società, come nel caso

[10] Ivi, p. 784, par. 2.2. Gli autori evidenziano come la risoluzione del Parlamento europeo del febbraio 2019 abbia sottolineato la necessità di introdurre specifici obblighi di trasparenza nell'impiego di sistemi di intelligenza artificiale nelle interazioni con gli utenti. In particolare, la risoluzione ha invitato la Commissione europea a prevedere un meccanismo di etichettatura che renda sempre riconoscibili i contenuti generati tramite IA, con particolare attenzione ai deep fake e ai materiali sintetici, allo scopo di rafforzare la tutela dei cittadini contro i rischi di manipolazione e disinformazione digitale.

dei sistemi di IA che operano nel settore finanziario e che, se mal progettati, potrebbero destabilizzare interi mercati.

Inoltre, l'Articolo 7 attribuisce alla Commissione Europea la possibilità di aggiornare o modificare la lista dei settori e delle applicazioni considerate ad alto rischio. Questa flessibilità è fondamentale per rispondere all'evoluzione rapida della tecnologia e alle nuove sfide sociali ed economiche. L'aggiornamento costante consente di integrare nella regolamentazione anche nuove applicazioni di IA che emergono nel tempo, ma che non erano previste all'inizio del processo normativo. In questo modo, l'AI Act rimane uno strumento vivo, capace di adattarsi alle trasformazioni tecnologiche senza risultare obsoleto.

L'approccio normativo delineato dall'AI Act per i sistemi di IA ad alto rischio, prevede una serie di obblighi stringenti per i fornitori di questi sistemi. Prima che tali sistemi possano essere immessi sul mercato, devono sottoporsi a una valutazione completa del rischio. Questa valutazione è cruciale per determinare l'impatto che l'IA avrà sugli utenti finali, non solo in termini di funzionalità, ma anche rispetto ai potenziali danni che potrebbero derivare da errori di funzionamento, malintesi o manipolazioni da parte di terzi. Gli sviluppatori devono quindi garantire che il sistema sia progettato in modo da ridurre al minimo il rischio di danni, adottando misure preventive adeguate.

Una volta immesso nel mercato, il sistema di IA deve essere sottoposto a un monitoraggio costante, al fine di rilevare tempestivamente eventuali criticità o anomalie che possano emergere nel tempo. A tal fine, i fornitori sono obbligati a implementare meccanismi di controllo per verificare che il sistema continui a rispettare i principi di trasparenza, equità e sicurezza. Questi obblighi si estendono per tutta la durata del ciclo di vita del sistema, includendo anche eventuali aggiornamenti o modifiche che potrebbero alterare il suo funzionamento iniziale.

In sintesi, la regolamentazione dei sistemi di IA ad alto rischio, come delineato dagli Articoli 6 e 7, risponde a una necessità di proteggere l'individuo e la collettività dagli effetti negativi che l'intelligenza artificiale potrebbe avere in contesti particolarmente sensibili. L'AI Act, attribuendo un ruolo fondamentale alla valutazione preventiva dei rischi e alla supervisione continua dei sistemi, si propone come uno strumento di tutela contro possibili abusi, errori o malfunzionamenti che

potrebbero compromettere la sicurezza, la salute e i diritti degli individui. Grazie all'introduzione di un sistema di aggiornamento dinamico dei settori ad alto rischio, il regolamento si configura come un quadro normativo in grado di adattarsi rapidamente all'evoluzione delle tecnologie, mantenendo alta la protezione degli utenti.

Un aspetto centrale della disciplina europea relativa ai sistemi di intelligenza artificiale ad alto rischio consiste nell'imposizione, a carico dei fornitori, di una serie articolata di obblighi in materia di gestione dei dati, prevenzione delle distorsioni algoritmiche e salvaguardia dell'equità nei processi decisionali automatizzati. Il regolamento richiede che l'intero ciclo di vita del sistema, dalla progettazione, passando per l'addestramento, fino all'implementazione operativa, sia improntato a criteri rigorosi di qualità, rappresentatività e pertinenza dei dati utilizzati. Viene attribuita specifica rilevanza all'individuazione, valutazione e correzione di potenziali *bias*, ossia di quegli elementi che, se non adeguatamente gestiti, potrebbero determinare esiti discriminatori o pregiudizievoli a danno di persone appartenenti a categorie protette o, più in generale, lesivi dei diritti fondamentali. In questa prospettiva, la normativa prescrive controlli approfonditi sia sulle modalità di raccolta e annotazione delle informazioni, sia sulle assunzioni sottese alla loro selezione e aggregazione, imponendo la predisposizione di processi di audit e di meccanismi documentati per la verifica costante dell'assenza di errori, lacune o distorsioni significative.[11]

Sul versante della sicurezza e della robustezza, il legislatore europeo pretende l'adozione di soluzioni tecniche e organizzative idonee a garantire la resilienza dei sistemi rispetto a guasti, alterazioni, tentativi di manipolazione interna o esterna e, in particolare, rispetto ai rischi connessi ai cosiddetti circuiti di feedback, ossia a quelle situazioni in cui gli output generati dall'IA possano influenzare ciclicamente i dati di input, amplificando progressivamente effetti distorsivi. Tali obblighi si riflettono non solo nella fase di progettazione, ma lungo tutto l'arco della vita del sistema, imponendo aggiornamenti continui, strategie di monitoraggio e prassi di risposta

[11] Il regolamento europeo sull'intelligenza artificiale richiede che le distorsioni nei dati dei sistemi ad alto rischio siano individuate e prevenute, in particolare se possono influire su salute, sicurezza o diritti fondamentali, prevedendo specifici obblighi in tema di rappresentatività e trasparenza dei dati utilizzati (art. 10, par. 2-5 AI Act).

rapida alle vulnerabilità emerse, così da garantire che le prestazioni dichiarate siano effettivamente mantenute e che la protezione dei diritti e della sicurezza degli interessati non venga compromessa.[12]

Il Regolamento sull'Intelligenza Artificiale dell'Unione Europea stabilisce anche una serie di fasi applicative per la gestione dei sistemi di IA ad alto rischio, con l'intento di garantire la conformità, la trasparenza e la sicurezza nel loro impiego. Queste fasi sono particolarmente rilevanti per le tecnologie che possono avere un impatto significativo sulla sicurezza e sui diritti fondamentali degli individui. Le disposizioni normative relative alle fasi applicative dell'IA ad alto rischio sono distribuite tra diversi articoli dell'AI Act, che regolano la valutazione della conformità, la marcatura CE, la registrazione, la vigilanza post-mercato, e altre misure di monitoraggio.

La prima fase applicativa fondamentale per l'IA ad alto rischio è la valutazione della conformità, disciplinata dall'articolo 43 del regolamento. Questa fase è fondamentale perché impone ai fornitori di sistemi di IA l'obbligo di accertarsi che i propri prodotti siano conformi ai requisiti tecnici e di sicurezza previsti dal Regolamento. L'articolo 43, infatti, impone che il fornitore di un sistema di IA ad alto rischio esegua una valutazione completa che confermi che il sistema soddisfa tutti i criteri stabiliti, in relazione agli aspetti di gestione del rischio e trasparenza. La valutazione di conformità deve essere documentata e può essere soggetta a ispezioni o revisioni da parte delle autorità competenti. In caso di esito positivo, il fornitore è autorizzato a proseguire con le fasi successive.

Una volta effettuata la valutazione di conformità, la fase successiva prevede l'apposizione della marcatura CE. Secondo l'articolo 48, la marcatura CE è obbligatoria per i sistemi di IA ad alto rischio che abbiano superato la valutazione di conformità. Questa marcatura attesta che il sistema è conforme alle disposizioni del Regolamento sull'Intelligenza Artificiale e può essere immesso sul mercato dell'Unione Europea. La marcatura CE è un requisito fondamentale per garantire la

[12] Gli obblighi di accuratezza, robustezza e cibersicurezza per i sistemi di IA ad alto rischio sono dettagliati all'art. 15 AI Act, che impone ai fornitori di garantire che i sistemi "siano progettati e sviluppati in modo tale da conseguire un adeguato livello di accuratezza, robustezza e cibersicurezza", nonché di adottare misure tecniche e organizzative capaci di assicurare resistenza a errori, guasti, manipolazioni e "circuiti di feedback" che possano produrre distorsioni sistemiche o vulnerabilità informatiche durante tutto il ciclo di vita del sistema.

legalità della commercializzazione dei sistemi di IA, e rappresenta un simbolo di fiducia per i consumatori e gli utenti, in quanto indica che i prodotti sono stati sottoposti a controlli rigorosi in merito alla loro sicurezza e compatibilità con la legislazione europea.

Subito dopo l'immissione sul mercato, un'altra fase rilevante è la registrazione del sistema nell'UE AI Database, come indicato negli articoli 49 e 71. I fornitori dei sistemi di IA ad alto rischio devono assicurarsi che i loro modelli siano registrati in una banca dati centralizzata, gestita dalle autorità competenti, che consente una tracciabilità continua di tutte le applicazioni di IA potenzialmente pericolose. La registrazione è essenziale per consentire alle autorità di vigilare sull'utilizzo e sullo sviluppo di queste tecnologie, raccogliendo informazioni cruciali sulla loro implementazione e sui rischi connessi. Inoltre, la registrazione serve a monitorare i sistemi per possibili modifiche che potrebbero alterare la loro sicurezza o il loro impatto sui diritti fondamentali.

La fase finale di gestione post-immissione riguarda la vigilanza post-mercato, regolata dall'articolo 72. Questo processo implica un monitoraggio continuo dei sistemi di IA ad alto rischio, una volta immessi sul mercato. I fornitori sono obbligati a raccogliere dati sull'uso dei sistemi, verificarne le performance e intervenire subito in caso di malfunzionamenti o problematiche di sicurezza. La vigilanza post-mercato, pertanto, consente di rilevare tempestivamente eventuali rischi legati all'utilizzo dei sistemi, e garantisce che i sistemi continuino a rispettare gli standard di sicurezza e di protezione previsti dal Regolamento. In aggiunta, la Commissione Europea e le autorità nazionali sono incaricate di monitorare i mercati, effettuando ispezioni e verifiche per accertare che i fornitori rispettino gli obblighi di sicurezza.

Per concludere, il percorso applicativo dei sistemi di IA ad alto rischio delineato nell'AI Act, prevede fasi interconnesse che mirano a garantire la sicurezza e la trasparenza di queste tecnologie. Dalla valutazione di conformità iniziale alla vigilanza post-mercato, ogni fase contribuisce a proteggere i diritti dei cittadini e a favorire l'adozione responsabile delle tecnologie avanzate nell'Unione Europea.

1.4. Sistemi di intelligenza artificiale a rischio sistemico

Gli articoli 51 e 52 del Regolamento sull'Intelligenza Artificiale dell'Unione

Europea introducono una disciplina inedita per i cosiddetti modelli di IA di uso generale, ossia quelle architetture algoritmiche sviluppate senza una destinazione d'uso predeterminata, ma destinate ad essere integrate e riutilizzate in una pluralità di applicazioni e settori. Sebbene tali sistemi non rientrino ex se nella categoria delle IA ad alto rischio, il legislatore europeo ha previsto che, in presenza di determinate condizioni, possano essere classificati come modelli di "IA a rischio sistemico".[13]

L'articolo 51, infatti, attribuisce alla Commissione Europea il potere di designare come "modello di IA con rischio sistemico" qualsiasi sistema di intelligenza artificiale di uso generale il cui impiego, in ragione di caratteristiche tecniche, ampiezza di diffusione o modalità di utilizzo, comporti "impatti gravi su larga scala" sulla salute, la sicurezza, i diritti fondamentali, l'ambiente, la democrazia o lo Stato di diritto.[14] Si tratta di una valutazione che deve tenere conto non solo delle potenzialità della tecnologia, ma anche del concreto contesto applicativo, basti pensare al caso di un *foundation model* impiegato in ambito sanitario o finanziario, dove un errore algoritmico o un malfunzionamento potrebbero produrre conseguenze dannose a catena, interessando migliaia di utenti e determinando effetti dirompenti a livello sociale o economico.[15] Il regolamento precisa che, ai fini della classificazione, la Commissione dovrà valutare fattori come la portata degli effetti negativi, la possibilità di incidenti sistemici, la vulnerabilità a manipolazioni o abusi e l'interazione con infrastrutture critiche o dati sensibili.

Una volta effettuata la designazione, scatta in capo al fornitore del modello una serie di obblighi rafforzati: l'articolo 52 impone, infatti, che sia trasmessa tempestivamente una notifica alla Commissione Europea entro due settimane dalla consapevolezza del rischio, fornendo tutte le informazioni necessarie a giustificare la classificazione e a dimostrare le misure di mitigazione adottate.[16] Tra gli obblighi più stringenti vi sono quelli relativi alla valutazione di conformità, all'apposizione della

[13] Gli articoli 51 e 52 del Regolamento (UE) 2024/1689 disciplinano espressamente i modelli di intelligenza artificiale di uso generale e prevedono che possano essere classificati come "modelli a rischio sistemico" in presenza di impatti gravi su larga scala.

[14] L'art. 51, par. 1 e 2, attribuisce alla Commissione Europea il potere di valutazione e designazione sulla base di rischi per la salute, la sicurezza, i diritti fondamentali, l'ambiente, la democrazia o lo Stato di diritto. Il testo precisa i criteri e le modalità di classificazione.

[15] Sui criteri di valutazione della portata degli effetti negativi, del rischio sistemico e della possibilità di incidenti a catena, vedasi par. 108 del preambolo e l'art. 51, par. 2

[16] L'articolo 52 prevede obblighi di notifica, registrazione, marcatura CE e sorveglianza post-mercato, da attuarsi entro due settimane dalla consapevolezza della sussistenza dei rischi.

marcatura CE, alla registrazione del modello negli appositi database europei e all'adozione di un sistema di sorveglianza e vigilanza post-mercato specificamente calibrato sul livello di rischio individuato. Il regolamento, inoltre, riconosce al fornitore la facoltà di presentare motivazioni tecniche o documentali per contestare la classificazione, ovvero di chiedere la revisione della stessa qualora siano venuti meno i presupposti di rischio originari, ad esempio per l'introduzione di nuove misure di sicurezza o per una diversa destinazione d'uso del modello.[17]

Questa disciplina innovativa mira a prevenire che tecnologie altamente versatili e pervasive, come i grandi modelli generativi (es. *GPT, Llama, Gemini*), possano produrre effetti negativi di portata sistemica in assenza di adeguati meccanismi di controllo, trasparenza e responsabilizzazione. Il regolamento prevede dunque una vigilanza rafforzata sui *foundation models*, riconoscendo che il loro impatto potenziale può superare i limiti delle singole applicazioni e incidere in modo profondo sull'ecosistema digitale europeo, sulla protezione dei diritti fondamentali e sulla stessa stabilità democratica delle istituzioni.[18]

1.5. Sistemi di intelligenza artificiale proibiti

Il Capitolo III dell'AI Act, specificamente all'articolo 5, definisce con chiarezza i sistemi di intelligenza artificiale che sono espressamente vietati nell'Unione Europea, a causa dei rischi estremamente elevati che comportano per la sicurezza, la dignità e i diritti fondamentali degli individui. La scelta di vietare determinate tecnologie deriva dalla consapevolezza che tali sistemi potrebbero causare danni irreparabili alla società, alla privacy e alla libertà personale.

L'Articolo 5 stabilisce una lista di applicazioni di IA che sono categoricamente proibite, sottolineando che l'uso di queste tecnologie è considerato inaccettabile per le implicazioni etiche, legali e morali che comportano. Un esempio centrale di tecnologia vietata è l'uso di sistemi di IA per il controllo sociale massivo, come quelli impiegati in contesti di sorveglianza biometrica in tempo reale.

[17] La possibilità di contestazione della classificazione o di richiesta di revisione è disciplinata dall'art. 52, par. 2 e 3

[18] Il par. 109 del preambolo ribadisce l'obiettivo di garantire la sicurezza, la trasparenza e la responsabilità dei modelli a rischio sistemico, con particolare riferimento alla tutela della democrazia e dei diritti fondamentali.

L'adozione di sistemi di riconoscimento facciale in ambienti pubblici senza il consenso esplicito delle persone, ad esempio, è considerata una violazione dei diritti alla privacy e alla protezione dei dati personali. Questi sistemi, infatti, presentano un rischio inaccettabile di abusi, soprattutto in contesti autoritari, dove potrebbero essere utilizzati per reprimere la libertà di espressione o per la sorveglianza indiscriminata di intere popolazioni.

Un altro settore vietato dall'Articolo 5 riguarda i sistemi di IA che manipolano comportamenti e opinioni degli utenti, come nel caso di applicazioni di IA utilizzate per la creazione di deepfake o altre forme di manipolazione video e audio che possono ingannare deliberatamente l'opinione pubblica. Tali sistemi presentano il rischio concreto di creare disinformazione su larga scala, influenzando le decisioni politiche e sociali degli individui. L'uso di IA per la manipolazione emotiva, ad esempio in contesti pubblicitari o politici, è un'altra applicazione che rientra in questa categoria di rischio proibito, poiché viola il principio di autonomia individuale e può minare la fiducia nelle istituzioni democratiche.

Come abbia avuto modo di analizzare L'Artificial Intelligence Act si fonda su un approccio dichiaratamente "*risk-based*", principio secondo cui più elevato è il rischio, più stringenti sono le regole applicabili. In tale cornice, il legislatore europeo ha espressamente vietato pratiche ritenute lesive della dignità e dei diritti fondamentali tra le quali spiccano i sistemi di social scoring applicati dalle autorità pubbliche ai cittadini. Tale scelta risponde al principio di proporzionalità sancito dall'articolo 5, paragrafo 4, TUE, volto a garantire un equilibrio tra innovazione e tutela dei valori fondamentali.[19]

Il confronto con l'esperienza cinese è istruttivo. Il *Social Credit System* è stato costruito su un'infrastruttura di sorveglianza estremamente capillare, che comprende l'installazione di oltre duecento milioni di telecamere a circuito chiuso e meccanismi di monitoraggio in tempo reale degli acquisti e dei comportamenti, con l'obiettivo di attribuire ai cittadini punteggi di affidabilità. La letteratura segnala

[19] M. Ebers, "*Truly Risk-based Regulation of Artificial Intelligence: How to Implement the EU's AI Act*", in *European Journal of Risk Regulation*, vol. 16, 2025, p. 684–685. L'autore illustra l'architettura dell'AI Act, fondato su un approccio basato sul rischio e sul principio di proporzionalità dell'art. 5, par. 4, TUE. PDF disponibile su https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/truly-riskbased-regulation-of-artificial-intelligence-how-to-implement-the-eus-ai-act/E526C1D0D7368F9691082220609D60F4

inoltre come, nel discorso pubblico occidentale, tale assetto sia stato descritto come una forma di sorveglianza quasi totale, scarsamente rispettosa della protezione dei dati personali e dello Stato di diritto.[20]

Il funzionamento del sistema poggia su architetture dati interconnesse. Accanto a registri e blacklist gestiti da autorità pubbliche, tra cui quelle della Corte Suprema, si collocano moduli creditizi e commerciali. Nel 2015 la Banca popolare di Cina ha autorizzato grandi operatori a sperimentare schemi di valutazione, fra cui Sesame Credit del gruppo Alibaba e i servizi collegati a WeChat. L'alimentazione del punteggio attinge così alle transazioni di centinaia di milioni di utenti e si traduce in vantaggi o restrizioni tangibili, fra cui limiti alla mobilità, all'accesso a strutture ricettive o a specifiche attività economiche.[21]

Un elemento decisivo è il ruolo dell'intelligenza artificiale, che consente l'identificazione biometrica in spazi pubblici, il tracciamento dei comportamenti offline e online e procedure di graduazione dei cittadini. Le fonti descrivono il sistema come sempre attivo. Flussi continui di tracce comportamentali alimentano modelli che generano premi o sanzioni, articolati in gruppi che incidono su settori, qualifiche, accessi e consumi, con conseguenze dirette sulla vita quotidiana.[22]

La stessa letteratura evidenzia una tensione strutturale fra opacità e trasparenza. Un sistema totalmente opaco impedirebbe ai cittadini di trarre apprendimento dai feedback, mentre una trasparenza totale favorirebbe il gioco opportunistico delle regole. Ne deriva una scelta di semi trasparenza, che però solleva problemi ulteriori di prevedibilità, riuso dei dati e rimedi effettivi, mostrando

[20] Q. V. Nguyen, S. Lafrance, C. T. Vu, *China's Social Credit System. A Challenge to Human Rights*, in «The Law, State and Telecommunications Review», vol. 15, n. 2, 2023, pp. 102 e 103. Si descrivono l'estensione della rete CCTV, il monitoraggio in tempo reale di acquisti e azioni e le critiche occidentali in termini di sorveglianza ampia, scarsa tutela della privacy e carenza di Stato di diritto. PDF disponibile su: https://pdfs.semanticscholar.org/b4ed/781a2f58e2a115197b7a0cef60d25c41c814.pdf

[21] Ibid., p. 103. Si ricostruiscono le componenti operative del sistema, dalle blacklist giudiziarie e amministrative alle piattaforme private con licenze della banca centrale dal duemilaquindici, inclusi Sesame Credit e integrazioni con WeChat, nonché le restrizioni pratiche su viaggi, alloggi e attività economiche.

[22] Ivi, pp. 104 e 105. Si evidenzia il ruolo dell'intelligenza artificiale, in particolare riconoscimento facciale, analisi dei comportamenti e procedure di citizen grading, e la natura sempre attiva del sistema con premi e sanzioni che incidono su settori, qualifiche, accessi e consumi.

i rischi di un meccanismo che, pur presentato come strumento di fiducia sociale, tende a trasformare la cittadinanza in un punteggio permanente.[23]

Diverso è invece il tema del credit scoring, affrontato dalla Corte di giustizia dell'Unione europea nella sentenza *Schufa* (C-634/21). In questo caso la Corte ha stabilito che la mera attribuzione di uno "score" da parte di un sistema algoritmico è sufficiente a far scattare l'applicazione dell'art. 22 GDPR, anche qualora la decisione finale venga formalmente confermata da un essere umano. Tale approdo ha superato l'orientamento tradizionale espresso in passato dalla giurisprudenza tedesca, che riteneva sufficiente l'intervento umano a posteriori per escludere la qualificazione della decisione come pienamente automatizzata.[24] La pronuncia *Schufa* rafforza dunque la tutela degli interessati nei confronti delle decisioni basate sul credit scoring, mostrando come anche pratiche non vietate dall'AI Act possano sollevare rilevanti problemi di compatibilità con i diritti fondamentali.

L'Articolo 5 proibisce inoltre sistemi di IA destinati al social scoring da parte delle autorità pubbliche e pratiche di profilazione automatizzata prive di garanzie, applicazioni che possono concretamente sfociare in una valutazione predittiva del comportamento criminale o nel rischio di recidiva, con conseguenti discriminazioni e violazioni dei principi di uguaglianza e presunzione di innocenza.

Un caso paradigmatico del rischio connesso ai sistemi di social scoring e profilazione automatizzata è stato affrontato dalla Corte di Giustizia dell'Unione Europea nella recente sentenza SCHUFA, dove la Corte ha stabilito che la mera attribuzione di uno "*score*" da parte di un sistema automatizzato è sufficiente per far scattare l'applicazione dell'art. 22 GDPR**,** anche se la decisione finale viene poi confermata da un essere umano. Questo supera l'orientamento tradizionale, espresso

[23] Ibid., pp. 104 e 105. Si segnala la tensione fra opacità e trasparenza, la scelta di una semi trasparenza per evitare il gioco opportunistico e i correlati rischi di prevedibilità, riuso dei dati e debolezza dei rimedi.

[24] Ceyhun Necati Pehlivan, Nikolaus Forgó, Peggy Valcke (a cura di), *The EU Artificial Intelligence (AI) Act: Commentary*, Oxford University Press, Oxford, 2024, p. 135 par. 2.4.2. Qui si esamina il passaggio giurisprudenziale tra la posizione tradizionale della Corte Suprema tedesca, secondo cui l'esistenza di una decisione finale da parte di un essere umano escludeva la configurazione di una decisione "automatizzata" ai sensi della normativa privacy, e la successiva interpretazione fornita dalla Corte di Giustizia dell'Unione Europea nel caso SCHUFA. La CJEU afferma che la semplice attribuzione di uno "score" da parte di un sistema automatizzato è sufficiente per far scattare le tutele dell'art. 22 GDPR, anche qualora la decisione venga successivamente confermata o rivista da un soggetto umano, ampliando così il perimetro di protezione contro la profilazione automatizzata.

in passato anche dalla Corte Suprema tedesca,[25] secondo cui l'intervento umano poteva escludere la qualificazione della decisione come pienamente automatizzata.

Nel contesto delle tecnologie di intelligenza artificiale proibite dall'AI Act, il tema della tutela effettiva dei diritti fondamentali e della responsabilità degli attori coinvolti assume una rilevanza cruciale, anche alla luce della giurisprudenza europea in materia di protezione dei dati personali e di responsabilità delle istituzioni dell'Unione.

Riguardo l'IA e le varie responsabilità che si vanno a configurare nell'immissione di questi sistemi nel mercato europeo, è utile ricordare che la responsabilità extracontrattuale dell'Unione europea si fonda su un impianto giurisprudenziale che ne definisce con precisione l'ambito e i requisiti applicativi. In particolare, la Corte di giustizia ha stabilito che il diritto al risarcimento per i danni cagionati dalle istituzioni dell'Unione può essere riconosciuto solo qualora ricorrano tre presupposti: innanzitutto, occorre che la norma violata sia preordinata a conferire diritti ai singoli; inoltre, la violazione deve risultare manifestamente grave rispetto ai limiti posti al potere discrezionale delle istituzioni; infine, è necessario che esista un nesso causale diretto tra l'inadempimento e il danno lamentato dal ricorrente.

Questo orientamento, consolidato dalla nota sentenza Bergaderm[26], pone l'accento sulla necessità di una tutela effettiva dei diritti individuali, richiedendo che la responsabilità dell'Unione sia ancorata non solo alla gravità della violazione, ma anche all'esistenza di un legame immediato tra condotta e pregiudizio subìto. Secondo la Corte, la disciplina della responsabilità extracontrattuale dell'Unione non può, in assenza di ragioni particolari, differire da quella prevista per la responsabilità

[25] Ivi, p. 135 par. 2.4.2. Qui si esamina il passaggio giurisprudenziale tra la posizione tradizionale della Corte Suprema tedesca, secondo cui l'esistenza di una decisione finale da parte di un essere umano escludeva la configurazione di una decisione "automatizzata" ai sensi della normativa privacy, e la successiva interpretazione fornita dalla Corte di Giustizia dell'Unione Europea nel caso SCHUFA. La CJEU afferma che la semplice attribuzione di uno "score" da parte di un sistema automatizzato è sufficiente per far scattare le tutele dell'art. 22 GDPR, anche qualora la decisione venga successivamente confermata o rivista da un soggetto umano, ampliando così il perimetro di protezione contro la profilazione automatizzata.

[26] Corte di Giustizia dell'Unione europea, 4 luglio 2000, C-352/98 P, *Laboratoires Pharmaceutiques Bergaderm SA e Goupil c.* Commissione, spec. parr. 41-44. Il testo della sentenza è consultabile all'indirizzo:
https://curia.europa.eu/juris/document/document.jsf?docid=45097&doclang=IT.

degli Stati membri rispetto ai danni causati ai singoli dalla violazione del diritto europeo.[27]

Questa equiparazione garantisce una protezione uniforme, indipendentemente dalla natura dell'autorità responsabile, e rappresenta oggi un punto di riferimento anche per le nuove sfide poste dall'intelligenza artificiale, specie rispetto ai rischi connessi a sistemi vietati o ad alto rischio.

La disciplina delle IA proibite si collega anche al quadro delineato dal *General Data Protection Regulation* (GDPR), in materia di trattamento lecito, proporzionalità e limiti ai trattamenti dei dati automatizzati, soprattutto laddove tali sistemi implichino decisioni automatizzate o trattamenti su larga scala di dati sensibili.[28] La giurisprudenza della Corte di Giustizia ha più volte sottolineato che qualsiasi restrizione ai diritti fondamentali derivante dal trattamento automatizzato deve essere prevista da una norma chiara, essere necessaria e proporzionata allo scopo perseguito, e prevedere adeguati rimedi giurisdizionali[29] . In particolare, la Corte di giustizia dell'Unione europea ha più volte ribadito che la tutela offerta dalla normativa europea in materia di dati personali deve essere effettiva e garantita anche in presenza di trattamenti automatizzati e di raccolte massive di dati, soprattutto quando non sono previste misure di controllo o limitazioni sufficienti[30].

---

[27] Ivi, par. 43, che sottolinea come la tutela riconosciuta ai singoli dal diritto dell'Unione non può variare in base all'autorità (nazionale o comunitaria) cui è imputabile il danno.

[28] Sulle condizioni di liceità e i limiti ai trattamenti automatizzati di dati personali e di categorie particolari di dati, v. Regolamento (UE) 2016/679, artt. 9 e 22: il primo pone un divieto generale al trattamento dei dati sensibili salvo specifiche eccezioni, mentre il secondo tutela il diritto dell'interessato a non essere sottoposto a decisioni basate unicamente su trattamenti automatizzati, inclusa la profilazione. Regolamento (UE) 2016/679, artt. 9 e 22, https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A32016R0679.

[29] Corte di Giustizia dell'Unione Europea, sentenza 16 luglio 2020, causa C-311/18, *Schrems II*, ECLI:EU:C:2020:559, punto 175. «Occorre aggiungere, a quest'ultimo riguardo, che il requisito secondo cui qualsiasi limitazione nell'esercizio dei diritti fondamentali deve essere prevista dalla legge implica che la base giuridica che consente l'ingerenza in tali diritti deve definire essa stessa la portata della limitazione dell'esercizio del diritto considerato» (par. 175). Testo disponibile in italiano su https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:62018CJ0311.

[30] Ivi, parr. 175-183

Nel caso delle IA proibite, come la sorveglianza biometrica di massa, tali requisiti risultano spesso insoddisfatti, in quanto l'assenza di base legale chiara, la mancata proporzionalità e la difficoltà di garantire rimedi effettivi ai soggetti interessati configurano una violazione dei principi cardine sanciti sia dall'AI Act sia dalla giurisprudenza europea.

In conclusione, la normativa dell'AI Act relativa ai sistemi di IA proibiti, rappresenta un passo decisivo verso una regolamentazione equilibrata e responsabile dell'intelligenza artificiale. Con l'esplicito divieto di applicazioni che violano i diritti fondamentali delle persone, come il riconoscimento facciale indiscriminato o la manipolazione delle opinioni pubbliche, il legislatore europeo intende garantire che l'intelligenza artificiale rimanga uno strumento di progresso e non di abuso. Il rigoroso controllo sulle applicazioni vietate si configura quindi, come un'importante protezione nei confronti di potenziali rischi di discriminazione, sorveglianza non etica e disinformazione, assicurando che la tecnologia non venga utilizzata per scopi contrari agli interessi collettivi e alla libertà individuale.

2. La Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto

La Convenzione quadro sull'Intelligenza Artificiale (AI) del Consiglio d'Europa, rappresenta un'importante iniziativa per la regolamentazione internazionale dell'intelligenza artificiale, in particolare per quanto riguarda la protezione dei diritti umani e il rispetto dei principi etici. La Convenzione è un tentativo di armonizzare gli approcci legali tra i vari Stati membri, per garantire che l'uso e lo sviluppo delle tecnologie di IA siano condotti nel rispetto dei diritti fondamentali, della dignità umana e della giustizia sociale.

La Convenzione è stata firmata il 5 settembre 2024, in occasione della Conferenza informale dei Ministri della Giustizia tenutasi a Vilnius, in Lituania, distinguendosi come il primo trattato internazionale giuridicamente vincolante in

materia di IA, volto a garantire che l'utilizzo dei sistemi di intelligenza artificiale sia pienamente conforme ai diritti umani, ai principi democratici e allo Stato di diritto[31].

La Convenzione quadro sull'IA, ha il merito di definire linee guida e principi chiari per affrontare le problematiche globali e intercontinentali sollevate dall'introduzione dell'intelligenza artificiale nelle società moderne. Il documento si colloca all'interno di un quadro giuridico internazionale, volto a promuovere la cooperazione internazionale e a stabilire principi comuni tra gli Stati membri, mirando a creare una normativa globale che sia coerente, equa e centrata sui diritti umani. La Convenzione si concentra su una serie di principi che devono essere rispettati dagli Stati membri in relazione all'uso dell'IA, al fine di garantire che tale tecnologia non venga utilizzata per scopi dannosi. In particolare, si pone l'attenzione sulla protezione dei diritti fondamentali degli individui e sulla necessità di mantenere la trasparenza e la responsabilità nelle applicazioni di IA. Questi principi sono ampiamente discussi negli articoli del documento.

Sotto il profilo sostanziale, la Convenzione si compone di un preambolo e VIII capi, per un totale di 36 articoli. Il Preambolo chiarisce sin da subito l'obiettivo fondamentale: garantire che le attività connesse al ciclo di vita dei sistemi di IA rispettino pienamente i principi della democrazia, dei diritti umani e dello Stato di diritto, riconoscendo al contempo il potenziale di tali sistemi nel promuovere la prosperità umana, il benessere sociale, lo sviluppo sostenibile e l'emancipazione femminile, nonché la necessità di stabilire un quadro giuridico applicabile a livello globale che definisca regole comuni per disciplinare le attività basate sull'IA, così da preservare i valori condivisi e sfruttarne i vantaggi ai fini di un'innovazione responsabile[32].

Nel Preambolo si evidenzia inoltre il carattere "quadro" della Convenzione, che potrà essere integrata da strumenti aggiuntivi volti ad affrontare questioni specifiche inerenti al ciclo di vita dei sistemi di IA; tali protocolli aggiuntivi dovranno incoraggiare, tra l'altro, la valutazione dei rischi emergenti legati a queste

[31] Consiglio d'Europa, *Council of Europe opens first ever global treaty on AI for signature*. Comunicato stampa, Strasburgo, Maggio 2024. Disponibile su: https://www.coe.int/en/web/portal/-/council-of-europe-opens-first-ever-global-treaty-on-ai-for-signature

[32] Consiglio d'Europa, *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law* (CETS n. 225), Vilnius, 5 settembre 2024 (testo adottato, non in vigore). Disponibile su: https://rm.coe.int/1680afae3c

tecnologie, inclusi quelli per la salute umana, per l'ambiente e per aspetti socioeconomici quali il lavoro e l'occupazione[33].

La Convenzione quadro adottata dal Consiglio d'Europa affronta in modo organico le complesse questioni poste dall'impiego dei sistemi di intelligenza artificiale, introducendo principi e obblighi specifici per garantire una tutela avanzata dei diritti fondamentali e delle istituzioni democratiche. In particolare, il testo definisce chiaramente l'intelligenza artificiale come un sistema basato su macchine capace di produrre, a partire dagli input ricevuti, output in grado di influenzare ambienti fisici o virtuali, con diversi livelli di autonomia e adattabilità[34].

Gli Stati aderenti alla Convenzione sono chiamati a sviluppare un rigoroso sistema di gestione dei rischi legati all'utilizzo dell'intelligenza artificiale, basato su una valutazione preventiva e continuativa, orientata da un principio di precauzione proporzionato alla gravità e probabilità dei possibili effetti negativi. Questo approccio consente una modulazione efficace delle misure a seconda dei rischi specifici riscontrati nei diversi contesti applicativi[35]. L'ambito di applicazione principale riguarda le attività condotte direttamente da autorità pubbliche o da soggetti privati che agiscono per loro conto, tuttavia è prevista la possibilità per gli Stati di estendere la disciplina convenzionale anche a privati indipendenti, mediante apposita dichiarazione e nel rispetto degli obiettivi generali stabiliti[36].

Pur promuovendo una regolamentazione ampia, la Convenzione introduce alcune precise eccezioni. In particolare, esclude espressamente dal proprio campo di applicazione le attività relative alla sicurezza nazionale, incluse quelle di natura militare, purché esse siano condotte nel rispetto degli obblighi internazionali relativi ai diritti umani e alle istituzioni democratiche[37]. Sono altresì escluse le attività di ricerca e sviluppo di sistemi IA non ancora operativi, tranne nei casi in cui tali attività comportino potenziali rischi significativi per i diritti fondamentali, la

[33] Ibid
[34] Ibid., art. 2, p. 3, che fornisce la definizione ufficiale di sistemi di IA.
[35] Ivi, art. 1, par. 2, p. 2, relativo al principio precauzionale e alla modulazione delle misure sulla base della gravità e della probabilità degli impatti.
[36] Ivi, art. 3, par. 1 lett. a-b, p. 3, riguardo all'ambito di applicazione principale e alla sua possibile estensione ai privati indipendenti.
[37] Ibid., art. 3, par. 2, p. 3, che prevede l'esclusione esplicita per le attività di sicurezza nazionale, inclusa la difesa, con vincolo al rispetto dei diritti umani.

democrazia o lo Stato di diritto[38]. Questa impostazione consente di mantenere un bilanciamento tra la tutela dei diritti individuali e la salvaguardia della sovranità nazionale.

Gli Stati aderenti devono inoltre assicurare che tutte le attività che implicano l'uso di sistemi di IA siano conformi agli obblighi internazionali di protezione dei diritti umani, garantendo un'applicazione uniforme e coerente degli standard di tutela[39]. Specifiche misure devono essere adottate per proteggere l'integrità delle istituzioni democratiche e garantire una partecipazione equa e libera ai dibattiti pubblici, nonché una formazione autonoma delle opinioni da parte dei cittadini[40].

La trasparenza e la supervisione dei sistemi automatizzati rivestono un ruolo fondamentale. La Convenzione stabilisce l'obbligo per gli Stati di assicurare che le decisioni automatizzate basate su sistemi di IA siano comprensibili e sottoposte a un efficace monitoraggio. Ciò implica che ai cittadini debba essere garantito l'accesso alle informazioni rilevanti per contestare eventuali decisioni automatizzate lesive dei loro diritti[41]. Nel contesto elettorale, tale principio è applicato con particolare rigore, esigendo che l'impiego di sistemi di IA avvenga in modo totalmente trasparente, senza limitare la libertà di espressione e la partecipazione politica, né alterare la volontà degli elettori[42].

In aggiunta, la Convenzione prevede strumenti procedurali concreti a tutela degli individui, tra cui la possibilità per i cittadini di contestare decisioni automatizzate attraverso l'accesso a informazioni chiare sul funzionamento degli algoritmi utilizzati [43] . Viene altresì sottolineata l'importanza di consultazioni pubbliche inclusive e di dibattiti informati sul tema dell'IA, coinvolgendo attivamente società civile e ONG[44].

---

[38] Ibid., art. 3, par. 3, p. 3, concernente l'esclusione delle attività di ricerca e sviluppo sperimentali, salvo che possano avere impatti negativi significativi sui diritti fondamentali.
[39] Ivi, art. 4, p. 4, sull'obbligo generale degli Stati di conformità agli standard internazionali di protezione dei diritti umani nell'uso dell'IA.
[40] Ibid., art. 5, par. 1-2, p. 4, inerente alla tutela delle istituzioni democratiche e alla garanzia di una partecipazione libera e paritaria al dibattito pubblico.
[41] Ibid., art. 8, p. 4, relativo all'obbligo di trasparenza e supervisione delle decisioni basate sull'IA.
[42] Ibid., art. 5, par. 2 e art. 8, p. 4, in riferimento alla trasparenza elettorale e alla tutela della libertà politica.
[43] Ivi, art. 14, par. 2 lett. a-c, p. 5, che prevede misure procedurali di contestazione delle decisioni automatizzate.
[44] Ivi, art. 19, p. 7, che sottolinea la necessità di coinvolgere la società civile in consultazioni pubbliche sul tema.

A supporto della corretta applicazione delle norme convenzionali, viene istituita una Conferenza delle Parti, con il compito di monitorare l'attuazione della Convenzione, facilitare la cooperazione internazionale e promuovere lo scambio di informazioni tra gli Stati[45]. Infine, ogni Stato parte è tenuto a istituire o designare autorità indipendenti, dotate delle risorse necessarie, per garantire la vigilanza e l'effettivo rispetto degli obblighi previsti dalla Convenzione stessa[46].

2.1. I rapporti tra la Convenzione quadro e l'AI Act

Dopo aver analizzato la struttura, i principi e il campo di applicazione della Convenzione quadro, risulta essenziale metterne in luce le differenze rispetto al modello regolatorio adottato dall'Unione europea con l'AI Act. Sebbene entrambi i testi condividano l'obiettivo di tutelare i diritti fondamentali nell'ecosistema digitale e di promuovere un uso responsabile dell'intelligenza artificiale, essi divergono profondamente nella tecnica normativa e nei criteri di protezione adottati.

La Convenzione quadro si caratterizza per un approccio orizzontale e generalista, nel quale i principi di non discriminazione, trasparenza, responsabilità e diritto a un ricorso effettivo trovano applicazione in modo uniforme a tutti i sistemi di IA, indipendentemente dal livello di rischio o dall'ambito di utilizzo. Questo impianto garantisce un nucleo di tutele inderogabili e una coerenza sostanziale nell'applicazione dei diritti su scala sovranazionale, senza prevedere distinzioni tra tipologie di applicazioni algoritmiche. L'intento è quello di evitare "zone grigie" nella tutela e assicurare che nessuna innovazione tecnologica possa essere sottratta al rispetto dei diritti fondamentali, della democrazia e dello Stato di diritto[47].

Di contro, l'AI Act si fonda su una logica strettamente graduata, ispirata al cosiddetto "*risk-based approach*" che abbiamo affrontato nei paragrafi precedenti. In tale prospettiva, le misure più stringenti sono previste solo laddove il rischio sia giudicato significativo, mentre alle applicazioni a basso impatto vengono imposti requisiti minimi di trasparenza o addirittura semplici codici di condotta volontari. Ne

[45] Ivi, art. 23, p. 7-8, relativo alla Conferenza delle Parti e ai suoi compiti.
[46] Ivi, art. 26, par. 1-2, p. 9, riguardo alla costituzione di autorità indipendenti per la vigilanza sull'applicazione della Convenzione.

[47] Consiglio d'Europa, Convenzione quadro sull'intelligenza artificiale, artt. 7-11 (sul nucleo dei diritti fondamentali tutelati dalla Convenzione;)

risulta un modello tecnico che privilegia l'innovazione e la proporzionalità, ma che lascia margini di differenziazione nella tutela effettiva dei diritti[48].

Un ulteriore elemento distintivo riguarda la portata e l'applicabilità dei due strumenti: mentre la Convenzione quadro è aperta all'adesione di Stati non membri dell'Unione europea, configurandosi come quadro giuridico potenzialmente globale, l'AI Act resta uno strumento tipicamente europeo, destinato agli operatori che intendano immettere sistemi di IA sul mercato unico. Questa diversità consente alla Convenzione di operare come standard minimo di riferimento, capace di uniformare le garanzie fondamentali in una prospettiva multilivello, e all'AI Act di definire standard più elevati e dettagliati per il contesto europeo.[49]

Infine, occorre segnalare che la coesistenza dei due strumenti non è priva di criticità interpretative: la letteratura giuridica più recente, evidenzia come la complementarità tra Convenzione quadro e AI Act richieda un costante lavoro di coordinamento normativo, per evitare sovrapposizioni o vuoti di tutela in fase di recepimento nazionale. In prospettiva, l'efficacia di entrambi gli strumenti dipenderà dalla loro capacità di dialogare e integrarsi, valorizzando i rispettivi punti di forza e assicurando una protezione multilivello dei diritti nell'era dell'IA.[50]

2.2. Sviluppo negoziale e criticità sistemiche della Convenzione quadro sull'IA

La Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale segna un passo pionieristico nel tentativo di delineare una cornice giuridica

[48] Regolamento sull'intelligenza artificiale (AI Act), 2024, artt. 5, 7-11 (sulla logica del *risk-based approach* e i requisiti progressivi di tutela nell'AI Act)

[49] Sul carattere orizzontale e multilivello della tutela dei diritti fondamentali (non discriminazione, privacy, dati personali) nella Convenzione quadro, si vedano i §§ 75–80 dell'*Explanatory Report* (Consiglio d'Europa, 2024), dove si chiarisce che tali garanzie si applicano a tutti i sistemi di IA, senza distinzioni di rischio o settore, e si ribadisce la vocazione internazionale della Convenzione. Testo consultabile su https://rm.coe.int/1680afae67

[50] Sul coordinamento e le criticità interpretative tra la Convenzione quadro del Consiglio d'Europa sull'IA e l'AI Act dell'UE, si vedano: C.-C. Kirin Chang, "*The First Global AI Treaty: Analyzing the Framework Convention on Artificial Intelligence and the EU AI Act*", *University of Illinois Law Review Online*, 2024, pp. 86–99 (spec. pp. 93–98, sulla necessità di evitare sovrapposizioni e lacune di tutela, promuovendo un dialogo normativo multilivello e protocolli condivisi tra gli strumenti); M. Presno & A. Meuwese, "Regulating AI from Europe: *a joint analysis of the AI Act and the Framework Convention on AI*", *Theory and Practice of Legislation*, vol. 13, n. 1, 2025, pp. 14–20 (sulla complementarità e il coordinamento pratico tra Convenzione e AI Act, e sulle prospettive di armonizzazione giuridica nel sistema europeo e globale). Chang, pdf completo su SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5069335; Presno & Meuwese, pdf disponibile su: https://www.tandfonline.com/doi/pdf/10.1080/20508840.2025.2492524.

vincolante per la regolamentazione dei profili più delicati dell'IA, con particolare riguardo alla tutela dei diritti umani, alla trasparenza e alla responsabilità pubblica. Il processo di negoziazione, avviato dall'Ad Hoc Committee on Artificial Intelligence (CAHAI)[51], si è distinto per il coinvolgimento di una pluralità di attori, organi intergovernativi, autorità di vigilanza, società civile e organizzazioni internazionali come UE, OCSE e UNESCO, riflettendo la necessità di un approccio multilivello e inclusivo nella costruzione del trattato quadro. Le fasi preparatorie, culminate nel *Feasibility Study* e nei *Possible Elements*, hanno portato in luce la presenza di importanti lacune nella disciplina internazionale, suggerendo la scelta di una convenzione generalista, rafforzata dalla previsione di futuri protocolli settoriali capaci di affrontare questioni tecniche emergenti.[52]

Uno degli aspetti di maggiore discussione riguarda l'equilibrio tra tutela uniforme dei diritti fondamentali e attenzione ai rischi specifici delle applicazioni di IA. Pur mantenendo un'impostazione generalista e orizzontale, che impone agli Stati l'adozione di obblighi inderogabili indipendentemente dal rischio connesso al sistema di IA, la Convenzione richiama la necessità di misure adeguate e proporzionate laddove emergano rischi particolari. Tuttavia, questa apertura, pur pensata per garantire flessibilità e aggiornamento, rischia di tradursi in implementazioni eterogenee tra i diversi Paesi, con possibili divergenze nella protezione effettiva dei diritti e nella possibilità di accesso a rimedi giurisdizionali uniformi. [53] Un ulteriore elemento critico deriva dalla mancanza di pieno allineamento tra le definizioni fondamentali adottate dalla Convenzione e quelle

[51] *Nuove regole per l'Intelligenza Artificiale: una sfida europea 16 Dicembre 2021, Politenico di Torino Magazine "L'Ad Hoc Committee on Artificial Intelligence (CAHAI) è un comitato istituito dal Consiglio dei Ministri del Consiglio d'Europa (CoE) per esaminare la fattibilità di un quadro giuridico per lo sviluppo, la progettazione e l'applicazione dell'Intelligenza Artificiale, basato sulle norme del Consiglio d'Europa in materia di diritti umani, democrazia e Stato di diritto. Il Comitato è stato istituito nel 2019 con un mandato di due anni…*" , citazione disponibile su: https://archivio-poliflash.polito.it/in_ateneo/nuove_regole_per_l_intelligenza_artificiale_una_sfida_europea

[52] Morawska, E.H., "*Council of Europe standards and activities related to AI: towards a Framework Convention on AI and human rights?*", in *Artificial intelligence and international human rights law*, Edward Elgar, 2024, pp. 36-37 e 39-40, dove si ricostruisce il processo negoziale dalla fase CAHAI fino al *Consolidated Working Draft*, sottolineando il ruolo dei protocolli settoriali come strumenti di adattamento flessibile. Disponibile su: https://www.elgaronline.com/edcollchap-oa/book/9781035337934/book-part-9781035337934-9.xml

[53] Ivi, pp. 40-41. L'autrice evidenzia come la scelta di un approccio generalista e la previsione di obblighi "proporzionati" possa produrre notevoli eterogeneità nell'attuazione nazionale, anche a causa della diversa capacità degli Stati membri di valutare e gestire i rischi specifici legati ai sistemi di IA.

presenti nell'AI Act europeo e nei principi OCSE, rischio che può favorire conflitti interpretativi e ostacolare l'affermazione di un quadro giuridico globale realmente coerente.[54]

Particolarmente controversa è anche la previsione di esenzioni per le attività di difesa e sicurezza nazionale. Questa scelta, pur riconoscendo la sovranità degli Stati membri, solleva interrogativi circa la possibile creazione di "zone d'ombra" prive di controllo democratico, specie nei settori della sorveglianza e della cyber-intelligence. In questi casi, la Convenzione si limita a un generico rinvio ai principi internazionali sui diritti umani, senza precisare criteri rigorosi di necessità e proporzionalità nell'adozione di misure restrittive.[55]

Infine, la struttura "quadro" del trattato, concepita per essere integrata da futuri protocolli tematici, rappresenta un elemento di flessibilità ma introduce anche il rischio di frammentazione e rallentamento attuativo. Se da un lato questa impostazione consente di adeguare rapidamente la disciplina alle innovazioni tecnologiche, dall'altro lato potrebbe causare ritardi e disomogeneità nei tempi e nei modi di recepimento da parte degli Stati, incidendo sull'effettività della tutela garantita.[56]

In conclusione, la Convenzione quadro si configura come un fondamentale primo passo verso la regolamentazione internazionale dell'IA, ma la sua concreta efficacia dipenderà dalla capacità degli Stati parte di colmare le lacune definitorie, assicurare meccanismi di supervisione realmente indipendenti e preservare la

[54] Ivi, pp. 38-39. Qui si documentano le criticità derivanti dalla mancanza di coerenza tra la definizione di "sistema di intelligenza artificiale" nel testo convenzionale rispetto a quella adottata dall'AI Act europeo e dai principi OCSE, con potenziale rischio di interpretazioni difformi a livello internazionale.

[55] Morawska, pp. 41-42, e cit. anche in Hildebrandt, M., "*National Security Exceptions in AI Regulation*", *International Journal of Law and Information Technology*, 32(2), 2024, pp. 116-138: la dottrina giuridica sottolinea i rischi connessi alle deroghe per sicurezza nazionale, evidenziando la necessità di maggiore chiarezza sui criteri di necessità e proporzionalità e sul ruolo del controllo democratico, specie nei settori ad alto rischio tecnologico.

[56] Morawska, pp. 42-43. L'autrice sottolinea il rischio di frammentazione normativa e di rallentamento attuativo insiti nella scelta di protocolli aggiuntivi, il cui iter di adozione può risentire delle complessità negoziali tra Stati membri.

coerenza sistemica, evitando che la nuova disciplina rimanga una mera "cornice regolatoria" priva di strumenti di tutela realmente operativi.[57]

3. La *soft law* nell'intelligenza artificiale a livello internazionale

L'avanzamento esponenziale dell'intelligenza artificiale ha sollevato questioni complesse relative alla sua governance, spingendo i governi e le organizzazioni internazionali a esplorare diverse forme di regolamentazione. Accanto alla cosiddetta "*hard law*", costituita da leggi e regolamenti vincolanti, un ruolo sempre più preminente è assunto dalla "*soft law*". Quest'ultima comprende un insieme di principi, linee guida, raccomandazioni e codici di condotta che, pur non avendo forza giuridica vincolante, influenzano significativamente lo sviluppo e l'applicazione dell'IA, promuovendo pratiche responsabili e etiche. La *soft law* offre flessibilità e agilità, consentendo di adattarsi rapidamente all'evoluzione tecnologica e di facilitare la cooperazione internazionale. Questo documento esplorerà il panorama della *soft law* sull'IA in diverse giurisdizioni e contesti internazionali, analizzando gli approcci adottati dall'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE), dagli Stati Uniti, dall'Europa (escludendo l'AI Act e la Convenzione quadro del Consiglio d'Europa) e dalla Cina, evidenziando le loro peculiarità e il loro impatto sulla governance globale dell'IA.

3.1. L'approccio dell'OCSE: principi per un'IA affidabile

L'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE) è stata una delle prime organizzazioni internazionali a formulare principi guida per l'intelligenza artificiale, riconoscendo la necessità di un approccio etico e responsabile allo sviluppo e all'uso di questa tecnologia. I Principi dell'OCSE sull'IA, adottati nel maggio 2019, rappresentano un esempio significativo di *soft law* a livello globale, fornendo un quadro di riferimento non vincolante ma influente per i governi e gli *stakeholder*.[58] Questi principi sono stati approvati da tutti i paesi membri

[57] Ivi, p. 43, dove si rimarca la necessità di meccanismi di supervisione indipendenti e di una costante cooperazione multilivello per evitare che la Convenzione rimanga priva di reale efficacia sul piano operativo.

[58] L'approccio dell'*OCSE* all'intelligenza artificiale si riflette nella *Recommendation of the Council on Artificial Intelligence*, adottata il 22 maggio 2019, uno dei più significativi esempi di *soft law* a livello

dell'OCSE e da altri paesi partner, dimostrando un ampio consenso internazionale sulla direzione da intraprendere per una governance responsabile dell'IA.

I Principi dell'OCSE si basano su cinque valori fondamentali e cinque raccomandazioni per i responsabili politici. I valori includono la crescita inclusiva, lo sviluppo sostenibile e il benessere; i sistemi di IA dovrebbero essere progettati in modo da beneficiare le persone e il pianeta, promuovendo la prosperità e la sostenibilità. La seconda dimensione è la centralità dell'uomo e l'equità; i sistemi di IA dovrebbero rispettare lo Stato di diritto, i diritti umani, i valori democratici e la diversità, garantendo un controllo umano significativo. La terza è la trasparenza e la spiegabilità; i sistemi di IA dovrebbero essere trasparenti e spiegabili per consentire alle persone di comprendere il loro funzionamento e di contestare le decisioni. La quarta è la robustezza, la sicurezza e la protezione; i sistemi di IA dovrebbero essere robusti, sicuri e protetti per evitare rischi imprevisti e usi impropri. Infine, la responsabilità; gli attori dell'IA dovrebbero essere responsabili del corretto funzionamento dei loro sistemi e del rispetto dei principi.[59]

Le raccomandazioni per i responsabili politici, invece, si concentrano sulla promozione di un ambiente favorevole all'innovazione responsabile, sulla promozione di un ecosistema di IA affidabile, sulla garanzia di politiche di IA inclusive, sulla facilitazione della cooperazione internazionale e sulla promozione della ricerca e dello sviluppo. L'OCSE ha sottolineato l'importanza della *soft law* in questo contesto, in quanto permette una maggiore flessibilità e adattabilità rispetto alla *hard law*, elementi cruciali in un settore in rapida evoluzione come quello dell'IA. La *soft law*, infatti, può essere implementata e modificata con relativa facilità, garantendo che le politiche rimangano pertinenti e efficaci.[60]

---

globale. Questo strumento, pur non essendo giuridicamente vincolante, fornisce un quadro di riferimento influente per governi e stakeholder, evidenziando la sua natura di raccomandazione del Consiglio. [OECD, *Recommendation of the Council on Artificial Intelligence*, 2019, Preambolo e Sezione I, pdf disponibile su: https://legalinstruments.oecd.org/public/doc/643/643.en.pdf]

[59] OLIVATO, Giulia, «OCSE – Raccomandazione sull'intelligenza artificiale: principi per la gestione responsabile di una AI affidabile e raccomandazioni agli Stati aderenti», *Biodiritto – AI Legal Atlas (AI Docs)*, 3 maggio 2024 (ult. mod. 20 maggio 2024). Disponibile su: https://www.biodiritto.org/AI-Legal-Atlas/AI-Docs/OCSE-Raccomandazione-sull-intelligenza-artificiale-principi-per-la-gestione-responsabile-di-una-AI-affidabile-e-raccomandazioni-agli-Stati-aderenti

[60] Gutierrez e Marchant (2021), nel documento *Soft law 2.0: Incorporating incentives and implementation mechanisms into the governance of artificial intelligence*, analizzano come la *soft law*, grazie alla sua flessibilità, possa colmare le lacune normative dell'*intelligenza artificiale*, a patto che siano previsti incentivi e meccanismi efficaci di implementazione. Il testo sottolinea l'importanza di

L'OCSE ha anche sviluppato l'Osservatorio sulle Politiche di IA (OECD.AI), una piattaforma che raccoglie e analizza le politiche e le iniziative sull'IA a livello globale, inclusi gli approcci di *soft law*. Questo osservatorio funge da strumento per la condivisione delle conoscenze e per il monitoraggio dell'implementazione dei Principi dell'OCSE, contribuendo a promuovere una maggiore coerenza e coordinamento tra le diverse giurisdizioni. L'impegno dell'OCSE nella *soft law* dell'IA riflette la convinzione che un approccio collaborativo e basato su principi condivisi sia essenziale per guidare lo sviluppo dell'IA verso un futuro che sia etico, responsabile e a beneficio di tutti.

3.2. L'Europa: oltre l'AI Act, un "ecosistema di *soft law*"

Sebbene l'Unione Europea sia nota per il suo approccio normativo all'intelligenza artificiale, incarnato dall'AI Act, il panorama europeo della governance dell'IA è arricchito da un vivace ecosistema di *soft law*. Queste iniziative, che precedono e affiancano la legislazione vincolante, svolgono un ruolo cruciale nel plasmare un approccio etico e responsabile all'IA, promuovendo la fiducia e l'innovazione. La *soft law* europea in materia di IA si manifesta attraverso linee guida, strategie nazionali, codici di condotta e iniziative di standardizzazione, che contribuiscono a creare un quadro normativo completo e multilivello.[61]

Un esempio fondamentale di *soft law* europea sono le Linee Guida Etiche per un'IA Affidabile, pubblicate dal Gruppo di Esperti ad Alto Livello sull'Intelligenza Artificiale (AI HLEG) della Commissione Europea nel 2019. Queste linee guida stabiliscono sette requisiti chiave per un'IA affidabile: intervento e sorveglianza umana, robustezza tecnica e sicurezza, privacy e governance dei dati, trasparenza; diversità, non discriminazione ed equità, benessere sociale e ambientale e responsabilità. Sebbene non siano giuridicamente vincolanti, queste linee guida

---

raccomandazioni politiche orientate all'innovazione responsabile, all'inclusività e alla cooperazione internazionale. [C.I. Gutierrez, G.E. Marchant, *Soft law 2.0: Incorporating incentives and implementation mechanisms into the governance of artificial intelligence*, OECD.AI, 2021, disponibile su: https://oecd.ai/en/wonk/soft-law-2-0]

[61] Sull'importanza della soft law nel modello europeo di governance dell'intelligenza artificiale e sulla complementarità tra strumenti giuridicamente vincolanti e strumenti flessibili (come linee guida, codici di condotta, standard tecnici e strategie nazionali), v. Commissione Europea, *White Paper on Artificial Intelligence – A European approach to excellence and trust*, 2020, pp. 18-22, pdf disponibile su: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065&from=EN

hanno avuto un'influenza significativa sul dibattito politico e normativo in Europa e a livello internazionale, e hanno ispirato molte delle disposizioni dell'AI Act.[62]

Inoltre, la Commissione Europea ha promosso lo sviluppo di codici di condotta volontari per l'IA, come strumento per incoraggiare l'adozione di pratiche responsabili da parte del settore privato. Questi codici, spesso sviluppati in collaborazione con le aziende e altri *stakeholder*, mirano a tradurre i principi etici in azioni concrete, fornendo orientamenti pratici su come sviluppare e utilizzare l'IA in modo etico e affidabile. Un esempio è il Codice di Condotta sull'IA e la Disinformazione, che mira a contrastare la diffusione di contenuti falsi e manipolatori generati dall'IA.[63]

A livello nazionale, molti Stati membri dell'UE hanno sviluppato le proprie strategie nazionali per l'IA, che spesso includono elementi di *soft law*. Queste strategie definiscono le priorità nazionali per lo sviluppo e l'uso dell'IA, e possono includere linee guida etiche, programmi di finanziamento per la ricerca e lo sviluppo, e iniziative per promuovere l'adozione dell'IA nel settore pubblico e privato. Ad esempio, la Germania ha pubblicato la sua Strategia Nazionale per l'Intelligenza Artificiale, che pone un forte accento sulla ricerca, lo sviluppo e l'applicazione dell'IA in modo etico e responsabile.[64]

Infine, le organizzazioni di standardizzazione europee, come il CEN e il CENELEC, stanno lavorando allo sviluppo di standard tecnici per l'IA. Questi standard, sebbene volontari, possono svolgere un ruolo importante nel garantire l'interoperabilità, la sicurezza e l'affidabilità dei sistemi di IA, e possono essere utilizzati come riferimento per dimostrare la conformità ai requisiti legali.

---

[62] Per l'individuazione dei sette requisiti essenziali di un'IA affidabile e il ruolo fondante svolto dalle Linee Guida Etiche nel dibattito normativo e nell'ispirare le prime previsioni dell'AI Act, si veda High-Level Expert Group on Artificial Intelligence (AI HLEG), *Ethics Guidelines for Trustworthy AI*, European Commission, 2019, pdf ufficiale disponibile su: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

[63] In merito allo sviluppo di codici di condotta volontari, con specifico riferimento agli strumenti adottati a livello europeo per contrastare la disinformazione e l'uso distorto dell'IA nel settore dell'informazione, cfr. European Commission, *2022 Strengthened Code of Practice on Disinformation*, documento e allegati disponibili su: https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

[64] Per un esempio di strategia nazionale contenente elementi di *soft law* e orientamenti etici per l'adozione responsabile dell'intelligenza artificiale, si veda Federal Government of Germany, *Artificial Intelligence Strategy of the German Federal Government*, update 2020, pdf disponibile su: https://www.ki-strategie-deutschland.de/files/downloads/Nationale_KI-Strategie_engl.pdf

L'ecosistema europeo di *soft law* sull'IA, quindi, è un complemento essenziale alla *hard law*, contribuendo a creare un ambiente normativo flessibile, adattabile e orientato al futuro, in grado di promuovere un'IA che sia non solo innovativa, ma anche etica e affidabile.

3.3. Gli Stati Uniti: un "mosaico di iniziative di *soft law*"

Negli Stati Uniti, l'approccio alla regolamentazione dell'intelligenza artificiale si distingue per la sua preferenza verso la *soft law*, in contrasto con l'orientamento più normativo dell'Unione Europea. Questa scelta riflette la cultura giuridica e politica statunitense, che privilegia l'innovazione e la flessibilità, affidandosi spesso a linee guida, standard volontari e codici di condotta piuttosto che a leggi vincolanti. L'assenza di una legislazione federale onnicomprensiva sull'IA ha portato a un mosaico di iniziative a livello statale e federale, spesso sotto forma di *soft law*, che mirano a guidare lo sviluppo e l'uso responsabile dell'IA.[65]

Una delle iniziative più significative a livello federale è *l'Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, emesso dal Presidente Biden nell'ottobre 2023. Sebbene sia un ordine esecutivo e non una legge, esso stabilisce una serie di direttive e principi che influenzano l'azione delle agenzie federali e incoraggiano il settore privato a seguire determinate pratiche. L'ordine promuove la sicurezza e la protezione dell'IA, la protezione della privacy, l'equità e i diritti civili, la promozione dell'innovazione e della concorrenza, e il rafforzamento della leadership americana nell'IA. Questo documento, pur non

[65] Marchant e Gutierrez (2023), nel loro studio *Soft Law 2.0: An Agile and Effective Governance Approach for Artificial Intelligence*, pubblicato sulla *Minnesota Journal of Law, Science & Technology*, evidenziano che negli Stati Uniti l'assenza di una legislazione federale organica in materia di *intelligenza artificiale* ha determinato un quadro caratterizzato da una pluralità di iniziative sia a livello statale sia federale. Queste iniziative, prevalentemente ascrivibili al novero della *soft law*, hanno la funzione di orientare lo sviluppo e l'impiego responsabile dell'*IA*. Gli autori sottolineano come la *soft law* costituisca attualmente il principale meccanismo di governance dell'*IA* negli Stati Uniti, sia per scelta sia per necessità, in assenza di una disciplina vincolante a livello federale, e che questa situazione è destinata a perdurare almeno nel prossimo futuro [MARCHANT, G.E., GUTIERREZ, C.I., *Soft Law 2.0: An Agile and Effective Governance Approach for Artificial Intelligence*, *Minnesota Journal of Law, Science & Technology*, vol. 24, n. 2, 2023, pp. 376-384, pdf disponibile su: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4473812].

essendo direttamente vincolante per le aziende private, funge da potente segnale e da catalizzatore per lo sviluppo di standard e best practice volontarie.[66]

Altre iniziative di *soft law* negli Stati Uniti includono i National Institute of Standards and Technology (NIST) AI Risk Management Framework, pubblicato nel gennaio 2023. Questo framework fornisce un approccio flessibile e volontario per la gestione dei rischi associati all'IA, aiutando le organizzazioni a incorporare la fiducia nei loro prodotti e servizi basati sull'IA. Il framework è stato sviluppato attraverso un processo collaborativo che ha coinvolto esperti del settore, accademici e rappresentanti della società civile, e mira a essere uno strumento pratico per le aziende che sviluppano e utilizzano l'IA.[67]

Anche a livello statale, diversi stati hanno iniziato a esplorare forme di *soft law* per l'IA. Ad esempio, alcuni stati hanno emesso linee guida per l'uso dell'IA nel settore pubblico o hanno istituito *task force* per studiare le implicazioni dell'IA e formulare raccomandazioni. Queste iniziative, pur essendo limitate al contesto statale, contribuiscono a creare un ambiente normativo più definito e a promuovere un approccio responsabile all'IA a livello locale.

La preferenza per la *soft law* negli Stati Uniti è spesso giustificata dalla rapidità con cui la tecnologia dell'IA si evolve. La *soft law*, essendo più agile, può essere aggiornata e adattata più facilmente rispetto alla *hard law*, che richiede processi legislativi più lunghi e complessi. Tuttavia, questa flessibilità comporta

---

[66] L'*Executive Order* 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, emanato dal Presidente degli Stati Uniti il 30 ottobre 2023, rappresenta uno degli interventi federali più rilevanti in materia di *IA*. Sebbene non costituisca una legge vincolante, l'ordine stabilisce una serie di direttive e principi, delineati nella sezione "Sec. 2. Policy and Principles" (p. 75192 del *Federal Register*), che guidano l'azione delle agenzie federali e incentivano il settore privato ad adottare standard responsabili. Tra i principi promossi figurano la sicurezza e la protezione dell'*IA*, la tutela della privacy, l'equità e i diritti civili, la promozione dell'innovazione e della concorrenza, nonché il rafforzamento della leadership americana nell'*IA*. [THE WHITE HOUSE, *Executive Order* 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, *Federal Register*, vol. 88, n. 211, 1° novembre 2023, p. 75192, pdf disponibile su: https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence].

[67] Il *National Institute of Standards and Technology* (*NIST*) ha pubblicato nel gennaio 2023 l'*Artificial Intelligence Risk Management Framework* (*AI RMF 1.0*), concepito come uno strumento flessibile e volontario per la gestione dei rischi associati all'*intelligenza artificiale*. Il framework, sviluppato attraverso un processo collaborativo e trasparente che ha coinvolto esperti del settore, accademici e rappresentanti della società civile, mira a fornire alle organizzazioni linee guida pratiche per favorire la fiducia nei prodotti e servizi basati sull'*IA*. [*NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY*, *Artificial Intelligence Risk Management Framework* (*AI RMF 1.0*), 2023, p. 2, pdf disponibile su: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf]

anche delle sfide, come la mancanza di uniformità e la potenziale assenza di meccanismi di applicazione coercitiva. Nonostante ciò, l'approccio statunitense alla *soft law* sull'IA continua a evolversi, cercando di bilanciare l'innovazione con la necessità di affrontare i rischi emergenti.

3.4. La Cina: un approccio pragmatico tra *soft law* e regolamentazione emergente

La Cina ha adottato un approccio distintivo alla governance dell'intelligenza artificiale, caratterizzato da una combinazione di *soft law* e una crescente tendenza verso la regolamentazione vincolante. Il paese mira a diventare un leader globale nell'IA entro il 2030, e la sua strategia di governance riflette la volontà di bilanciare l'innovazione tecnologica con la necessità di mantenere il controllo sociale e la stabilità. Sebbene la Cina abbia iniziato a introdurre leggi più stringenti sull'IA, la *soft law* ha giocato e continua a giocare un ruolo fondamentale nel guidare lo sviluppo e l'applicazione dell'IA.[68]

Uno degli esempi più rilevanti di *soft law* in Cina è il New Generation Artificial Intelligence Development Plan (AIDP), pubblicato nel 2017. Questo piano strategico a lungo termine delinea gli obiettivi nazionali per lo sviluppo dell'IA, promuovendo la ricerca e lo sviluppo, l'integrazione dell'IA in vari settori e la creazione di un ecosistema favorevole all'innovazione. Sebbene non sia una legge vincolante, l'AIDP ha fornito un quadro di riferimento per le politiche governative e gli investimenti nel settore dell'IA, incoraggiando le aziende e le istituzioni accademiche a contribuire alla visione nazionale.[69]

---

[68] Shen e Liu (2023), nel loro studio *China's Normative Systems for Responsible AI: From Soft Law to Hard Law*, esaminano come la governance cinese dell'*intelligenza artificiale* sia caratterizzata da una fase iniziale di forte utilizzo della *soft law* (standard tecnici, codici di condotta, linee guida etiche, opinioni governative e iniziative di autoregolamentazione) e da una successiva tendenza verso l'adozione di strumenti giuridicamente vincolanti. Gli autori sottolineano che, pur essendo in atto un rafforzamento della regolazione *hard law*, la *soft law* continua a svolgere un ruolo centrale nell'orientare lo sviluppo e l'applicazione dell'*IA* in Cina, soprattutto in attesa di una legislazione unitaria e sistematica, prevista entro il 2030. [SHEN, W., & LIU, Y., *China's Normative Systems for Responsible AI: From Soft Law to Hard Law*, 2023, in particolare pp. 150-153, pdf disponibile su: https://www.researchgate.net/publication/370274672_China's_Normative_Systems_for_Responsible_AI_From_Soft_Law_to_Hard_Law]

[69] Il *New Generation Artificial Intelligence Development Plan* (*AIDP*), pubblicato dal Consiglio di Stato della Repubblica Popolare Cinese nel 2017, rappresenta un esempio centrale di *soft law* in Cina. Il piano, non vincolante, definisce obiettivi strategici e linee guida per lo sviluppo dell'*IA* a livello

La Cina ha anche sviluppato una serie di linee guida etiche e standard tecnici per l'IA. Ad esempio, il *National Governance Committee for the New Generation Artificial Intelligence* ha pubblicato i *New Generation Artificial Intelligence Ethics Norms* nel 2021, che stabiliscono principi etici per lo sviluppo, l'uso e la gestione dell'IA, tra cui l'equità, la trasparenza, la sicurezza e la responsabilità. Questi principi, pur essendo non vincolanti, servono da riferimento per le aziende e i ricercatori, promuovendo un approccio etico all'IA.[70]

Inoltre, il governo cinese ha incoraggiato lo sviluppo di codici di condotta e standard industriali da parte di associazioni di settore e aziende leader. Questi strumenti di *soft law* mirano a promuovere l'autoregolamentazione e a garantire che le pratiche di sviluppo dell'IA siano allineate con gli obiettivi nazionali e i principi etici. L'approccio cinese si basa spesso su un modello di governance adattiva, in cui le linee guida e i principi vengono testati e affinati prima di essere eventualmente trasformati in regolamentazioni più rigide. Questo permette al governo di sperimentare e imparare dalle pratiche del settore prima di imporre obblighi vincolanti.

Nonostante la forte enfasi sulla *soft law*, la Cina ha recentemente introdotto regolamentazioni più specifiche e vincolanti in settori chiave dell'IA, come gli algoritmi di raccomandazione e l'IA generativa. Questo indica una transizione graduale da un approccio prevalentemente basato sulla *soft law* a un mix di *soft* e *hard law*, riflettendo la crescente maturità del settore e la necessità di affrontare

---

nazionale, con particolare attenzione all'innovazione, alla sicurezza e alla crescita dell'ecosistema tecnologico. [*STATE COUNCIL OF THE PEOPLE'S REPUBLIC OF CHINA*, *New Generation Artificial Intelligence Development Plan*, 2017, testo ufficiale (cinese) su: http://www.gov.cn/zhengce/content/2017-07/20/content_5211990.htm; traduzione inglese: https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf]

[70] Il *National Governance Committee for the New Generation Artificial Intelligence* ha pubblicato nel 2021 le *New Generation Artificial Intelligence Ethics Norms*, principali linee guida etiche cinesi in materia di *IA*. Il documento, adottato come *soft law* e privo di effetti vincolanti diretti, stabilisce una serie di principi fondamentali (art. 3), tra cui il progresso del benessere umano, l'equità e la giustizia, la tutela della privacy e della sicurezza, la controllabilità, la responsabilità e la crescita della cultura etica. Ulteriori articoli disciplinano la sicurezza e la trasparenza (art. 12), la prevenzione dei bias e delle discriminazioni algoritmiche (art. 13), nonché la possibilità di adattamento delle norme da parte di enti e organizzazioni (art. 24). Il testo ufficiale (in cinese) è disponibile su: http://www.gov.cn/zhengce/content/2021-09/26/content_5639459.htm; per la traduzione inglese integrale si veda: *Center for Security and Emerging Technology* (*CSET*), Georgetown University, *Ethical Norms for New Generation Artificial Intelligence Released*, ottobre 2021, disponibile su: https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/

rischi specifici. Tuttavia, la *soft law* continua a essere uno strumento essenziale per guidare l'innovazione e promuovere un'IA responsabile in Cina, fungendo da ponte tra i principi generali e le applicazioni pratiche.[71]

CAPITOLO II

LA TECNOLOGIA DI RICONOSCIMENTO FACCIALE NEL CONTESTO GIURIDICO EUROPEO E INTERNAZIONALE

Sommario: 1. Introduzione; 1.1. La tecnologia di riconoscimento facciale: natura, funzionamento e implicazioni giuridiche; 1.2. Criticità normative e lacune regolatorie nella disciplina delle FRT; 2. Il caso Clearview AI: il paradigma della sorveglianza e la necessità di bilanciamento con i diritti fondamentali; 3. La sentenza Glukhin c. Russia: le prime applicazioni giurisprudenziali della Corte EDU

1. Introduzione

La tecnologia di riconoscimento facciale (*facial recognition technology*, *FRT*), una delle più potenti e discusse applicazioni dell'Intelligenza Artificiale, si sta affermando come uno strumento di sorveglianza e identificazione dalle implicazioni epocali, sollevando questioni giuridiche e sociali di fondamentale importanza per le democrazie contemporanee. Questa tecnologia, che permette l'identificazione e la verifica automatizzata dell'identità di un individuo attraverso l'analisi computerizzata delle sue caratteristiche biometriche facciali, ha conosciuto una rapida e pervasiva diffusione in molteplici settori. Tuttavia, è nel suo impiego da parte delle forze dell'ordine che emergono le sfide più profonde e complesse per la tutela dei diritti fondamentali e per l'equilibrio dei poteri nello Stato di diritto. La crescente adozione

[71] Zou e Zhang (2024), nel loro studio *Navigating China's regulatory approach to generative artificial intelligence and large language models*, descrivono come la Cina affianchi alla *soft law* l'adozione di regolamenti vincolanti per l'*IA generativa* e altri settori strategici, mantenendo la *soft law* come ponte tra principi generali e applicazioni pratiche (pp. 2-6). [ZOU, M., ZHANG, L., *Navigating China's regulatory approach to generative artificial intelligence and large language models*, *Cambridge Forum on AI: Law & Governance*, 2024, disponibile su: https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/navigating-chinas-regulatory-approach-to-generative-artificial-intelligence-and-large-language-models/969B2055997BF42DE693B7A1A1B4E8BA]

di questi strumenti da parte delle autorità pubbliche, nel 2021 già 11 dei 27 Stati membri dell'Unione Europea ne avevano avviato l'utilizzo in ambito penale,[72] rende non più differibile un'analisi rigorosa e una riflessione critica sul quadro normativo esistente e futuro.

Il presente capitolo si propone di condurre un'analisi approfondita delle complesse interazioni tra la tecnologia di riconoscimento facciale e l'ordinamento giuridico, con un'attenzione privilegiata al diritto dell'Unione Europea, che si pone oggi come uno degli attori più avanzati nel tentativo di regolamentare questo fenomeno. L'indagine muoverà da un esame delle fondamenta tecniche della FRT, per comprenderne il suo funzionamento e la sua natura intrinsecamente probabilistica. Si tratta, infatti, di una tecnologia che non fornisce certezze assolute, ma esprime i suoi risultati in termini di probabilità di somiglianza, con un margine di errore che, seppur ridotto negli algoritmi più moderni, può avere conseguenze devastanti in termini di falsi positivi (erronee identificazioni) e falsi negativi (mancate identificazioni). Successivamente, verranno esplorate le sue applicazioni pratiche e le sfide che queste pongono ai principi democratici, analizzando come l'uso della FRT possa incidere sulla libertà di espressione, di riunione e, più in generale, sulla vita privata dei cittadini.

Un punto focale dell'analisi sarà dedicato al caso emblematico di Clearview AI, una società statunitense che ha esemplificato in modo paradigmatico i rischi connessi a un uso non regolamentato di questa tecnologia. Nel gennaio 2020, un'inchiesta del *New York Times* ha svelato come Clearview AI avesse segretamente costruito un database di dimensioni inaudite, raccogliendo miliardi di immagini di volti da Internet, per poi offrire i suoi servizi a centinaia di agenzie di polizia.[73]

---

[72] Simmler, M., e Canova, G., "*Facial recognition technology in law enforcement: Regulating data analysis of another kind*", Computer Law & Security Review, 56, 2025, p. 1. Gli autori, citando uno studio di F. Ragazzi et al. del 2021, evidenziano la rapida adozione della *FRT* in Europa, sottolineando che già nel 2021 "almeno 11 dei 27 Stati membri dell'UE" avevano iniziato a sfruttare le potenzialità di questa tecnologia nel contesto delle indagini penali.
Disponibile su: https://www.sciencedirect.com/science/article/pii/S0267364924001572

[73] Dul, C., "*Facial Recognition Technology vs Privacy: The Case of Clearview AI*", *Queen Mary Law Journal*, 3, 2022, p. 1. L'autrice introduce il caso partendo dall'articolo del *New York Times* del 18 gennaio 2020, che per primo ha svelato al grande pubblico l'esistenza e le pratiche di Clearview AI. Riporta inoltre che il database è cresciuto fino a 10 miliardi di foto nell'ottobre 2021. Disponibile su: https://www.qmul.ac.uk/law/research/journals/the-queen-mary-law-journal/media/law/docs/research/2022QMLJ1.pdf

Questo modello di business, fondato su una sorveglianza occulta e su una massiccia aggregazione di dati personali, rappresenta una violazione senza precedenti del diritto fondamentale alla privacy e ha sollevato un'ondata di critiche e azioni legali a livello globale. Il caso Clearview dimostra in modo lampante come le logiche di mercato e le motivazioni economiche possano entrare in rotta di collisione con la concezione tradizionale e la finalità stessa del diritto alla privacy, evidenziando l'inadeguatezza di un'applicazione puramente "verticale" dei diritti fondamentali, pensata per regolare i rapporti tra Stato e cittadino, in un cyberspazio dominato da attori privati dotati di un potere quasi-statale.[74]

Il trattamento di dati biometrici, categoria nella quale rientrano a pieno titolo le immagini facciali, è una delle questioni più delicate e centrali del diritto europeo della protezione dei dati. La legislazione dell'UE, consapevole della loro natura sensibile, ha predisposto un regime di protezione rafforzata.[75] I due pilastri di questo quadro normativo sono il Regolamento Generale sulla Protezione dei Dati (GDPR), applicabile ai trattamenti per finalità commerciali e generali, e la Direttiva 2016/680 (nota come Direttiva *Law Enforcement* o LED), che disciplina specificamente il trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati. L'articolo 10 della Direttiva LED è di cruciale importanza, poiché stabilisce che il trattamento di dati biometrici volto all'identificazione univoca di una persona fisica, è consentito solo se "strettamente necessario" e accompagnato da garanzie adeguate per i diritti e le libertà dell'interessato.[76]

---

[74] Ivi, p. 2. L'autrice sostiene che il modello di business di Clearview, che collega i ritratti nel suo database ad altre informazioni online, sfida in modo senza precedenti il diritto fondamentale alla privacy dei soggetti interessati. Il caso dimostra l'inefficacia di un'applicazione puramente "verticale" dei diritti fondamentali di fronte al potere esercitato da attori privati nel cyberspazio.

[75] Agenzia dell'Unione Europea per i Diritti Fondamentali (FRA), "*Facial recognition technology: fundamental rights considerations in the context of law enforcement*", FRA Focus, 2019, p. 5. Il diritto dell'UE riconosce le immagini facciali come "dati sensibili" e una forma di "dati biometrici", che richiedono una protezione rafforzata ai sensi della legislazione sulla protezione dei dati, in particolare il GDPR e la Direttiva Law Enforcement. Disponibile su: https://staging.fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf

[76] Simmler, M., e Canova, G., *op. cit.*, p. 5. L'articolo 10 della Direttiva 2016/680/UE (Direttiva *Law Enforcement*) è il fulcro della regolamentazione del trattamento di dati biometrici per finalità di contrasto. Esso impone che tale trattamento sia consentito solo se "strettamente necessario" e accompagnato da "garanzie adeguate" per la protezione dei diritti e delle libertà degli interessati.

È fondamentale comprendere che l'impiego della FRT non si esaurisce in un singolo atto, ma si articola in un processo a più fasi di trattamento dei dati, ciascuna delle quali comporta una potenziale e grave interferenza con i diritti fondamentali. Dalla raccolta iniziale delle immagini, che può avvenire all'insaputa dell'interessato, si passa all'analisi biometrica, che di per sé genera nuovi dati sensibili (il *template* facciale), per arrivare infine allo sfruttamento di tali informazioni a fini investigativi. Ogni singolo passaggio di questa "catena di esposizione" deve essere legittimato da una base giuridica chiara, precisa e prevedibile.[77] La Corte Europea dei Diritti dell'Uomo, nella sua storica sentenza sul caso *Glukhin c. Russia*, ha infatti stabilito che l'uso della FRT da parte delle autorità per identificare un manifestante pacifico, in assenza di un quadro normativo adeguato, costituisce una violazione dei diritti fondamentali, segnando un punto di non ritorno nella giurisprudenza continentale.[78]

In questo scenario di crescente complessità, si inserisce il recente *Artificial Intelligence Act* (AI Act) dell'UE, del quale abbiamo parlato del capitolo precedente, il primo tentativo organico a livello mondiale di regolamentare l'Intelligenza Artificiale. L'AI Act classifica i sistemi di FRT utilizzati in spazi accessibili al pubblico per finalità di contrasto come ad "alto rischio" o, in alcuni casi, ne vieta l'uso in tempo reale come previsto dall'art. 5, salvo eccezioni rigorosamente definite. Tuttavia, come questo capitolo si propone di dimostrare, sebbene l'AI Act introduca importanti novità e requisiti di trasparenza e robustezza, esso non fornisce di per sé le basi giuridiche necessarie a giustificare l'uso di tali tecnologie da parte delle forze dell'ordine, che dovranno comunque trovare legittimazione nella Direttiva LED e nelle legislazioni nazionali.[79] Persistono, dunque, significative lacune normative che richiedono un intervento legislativo mirato e una riflessione giurisprudenziale approfondita.

---

[77] Ivi, p. 4. Gli autori descrivono l'uso della FRT come una "catena di esposizione" (*exposure chain*). Questa metafora evidenzia efficacemente come l'individuo sia sottoposto a una serie di passaggi di trattamento dei dati, dalla raccolta all'analisi, fino all'utilizzo finale, ognuno dei quali costituisce un'interferenza distinta e potenzialmente grave con i suoi diritti fondamentali.

[78] Ivi, p. 1. Si fa riferimento alla sentenza *Glukhin c. Russia*, in cui la Corte Europea dei Diritti dell'Uomo ha ritenuto che l'uso della FRT da parte delle autorità per analizzare filmati e immagini dei social media per identificare un manifestante violasse i diritti fondamentali.

[79] Ivi, p. 9. Gli autori sostengono che, in assenza di un quadro giuridico completo e specifico, l'uso della FRT da parte delle forze dell'ordine non è consentito. L'AI Act, pur introducendo regole importanti, non colma questa lacuna, non fornendo di per sé una base giuridica sufficiente per l'impiego di tali tecnologie.

L'analisi che seguirà, pertanto, non si limiterà a una mera ricognizione della tecnologia, ma si addentrerà nelle sue più profonde implicazioni giuridiche, etiche e sociali. Attraverso l'esame della giurisprudenza, della dottrina e dei più recenti sviluppi normativi, questo capitolo intende offrire un contributo critico al dibattito, delineando i contorni di una delle sfide più urgenti per il futuro dello Stato di diritto e della democrazia liberale nell'era digitale.[80] La necessità di un quadro giuridico che bilanci efficacemente le esigenze di sicurezza con la tutela dei diritti fondamentali è stata ribadita con forza anche dagli organi consultivi e di controllo europei, che hanno richiesto un approccio basato sulla massima cautela e, in alcuni casi, un divieto generalizzato di tali pratiche negli spazi pubblici.[81]

1.1. La tecnologia di riconoscimento facciale: natura, funzionamento e implicazioni giuridiche

Le tecnologie di riconoscimento facciale (FRT) rappresentano una delle applicazioni più potenti e pervasive dell'Intelligenza Artificiale (IA), capaci di incidere profondamente sulle strutture sociali e giuridiche contemporanee. La loro funzione primaria consiste nell'identificazione o nell'autenticazione automatizzata di individui attraverso l'analisi delle loro caratteristiche facciali uniche.[82] Questo processo, apparentemente tecnico, si articola in una sequenza complessa di operazioni: inizia con il rilevamento di un volto (*face detection*), prosegue con l'allineamento e la normalizzazione dell'immagine, per poi culminare nell'estrazione

[80] Selwyn, N., Andrejevic, M., O'Neill, C., Gu, X. and Smith, G., "*Facial Recognition Technology: Key Issues and Emerging Concerns*", in R. Matulionyte, M. Zalnieriute (eds.), The Cambridge Handbook of Facial Recognition in the Modern State, 2024, p. 13. Gli autori discutono come l'ascesa della FRT abbia portato a un commento contrario, crescente e vigoroso, sui possibili danni sociali derivanti dal suo sviluppo e implementazione, con critiche che la definiscono una tecnologia profondamente discriminatoria e distorta. Pdf disponibile su: https://researchmgt.monash.edu/ws/portalfiles/portal/592017706/583251749_oa.pdf

[81] EDPB-GEPD, Parere congiunto 5/2021 sulla proposta di regolamento sull'intelligenza artificiale, 18 giugno 2021, p. 3. Nel loro parere congiunto, il Comitato Europeo per la Protezione dei Dati e il Garante Europeo della Protezione dei Dati chiedono un "divieto generale di qualsiasi utilizzo dell'IA a fini di riconoscimento automatico delle caratteristiche umane in spazi accessibili al pubblico", evidenziando i rischi elevati di intrusione nella vita privata e le ripercussioni sull'anonimato nei luoghi pubblici. PDF disponibile su: https://www.edpb.europa.eu/system/files/2021-10/edpb-edps_joint_opinion_ai_regulation_it.pdf

[82] Simmler, M., e Canova, G., "*Facial recognition technology in law enforcement: Regulating data analysis of another kind*", Computer Law & Security Review, 56, 2025, p. 2. Gli autori descrivono la FRT come una tecnologia biometrica classica, basata sull'identificazione delle caratteristiche biologiche o comportamentali di un individuo, al pari delle impronte digitali, dell'iride o della voce

di un *template* biometrico, ovvero una rappresentazione matematica del volto. È questo *template*, e non l'immagine in sé, a essere confrontato con un database per scopi di identificazione (confronto 1:n) o di verifica (confronto 1:1).[83] La natura intrinsecamente probabilistica di questo confronto è un elemento cruciale: il sistema non fornisce mai una certezza assoluta, ma un punteggio di somiglianza, un dato che espone l'intero processo a un inevitabile margine di errore, con il rischio di falsi positivi e falsi negativi.[84]

L'impiego di queste tecnologie da parte delle forze dell'ordine, in particolare, ha sollevato un dibattito acceso e polarizzato. Da un lato, vengono presentate come strumenti indispensabili per la lotta al crimine, il terrorismo e la protezione della sicurezza pubblica; dall'altro, sono percepite come una minaccia senza precedenti per i diritti fondamentali, in particolare per il diritto alla privacy, alla protezione dei dati e alla libertà di espressione e di riunione. Questa tensione è stata riconosciuta ai massimi livelli istituzionali. L'Agenzia dell'Unione Europea per i Diritti Fondamentali (FRA) ha evidenziato come l'uso della FRT, specialmente in modalità "*live*" e in spazi pubblici, possa avere un impatto sproporzionato sulla dignità delle persone e generare un "effetto raggelante" (*chilling effect*) sulla partecipazione alla vita democratica.[85] La consapevolezza di poter essere costantemente monitorati e identificati può, infatti, dissuadere i cittadini dall'esercitare il proprio dissenso o dal partecipare a manifestazioni pacifiche, minando così le fondamenta stesse di una società aperta e pluralista.

La dimensione giuridica del problema è resa ancora più complessa dalla natura dei dati trattati. Le immagini facciali, quando processate attraverso tecniche specifiche per consentire l'identificazione univoca di una persona, costituiscono "dati

[83] Ibid. Il processo tecnico viene illustrato attraverso una sequenza di fasi: rilevamento del volto, allineamento, generazione del template e confronto (*matching*). L'articolo include anche una rappresentazione grafica di questo processo.
[84] Agenzia dell'Unione Europea per i Diritti Fondamentali (FRA), "*Facial recognition technology: fundamental rights considerations in the context of law enforcement*", *FRA Focus*, 2019, p. 9. Il documento evidenzia la natura probabilistica degli algoritmi, che non restituiscono mai un risultato definitivo ma solo probabilità. Si sottolinea che un tasso di errore apparentemente basso (es. 0.01%) può tradursi in un numero molto elevato di persone erroneamente identificate in scenari di massa.
[85] Ivi, p. 4. La FRA mette in guardia sull'"effetto raggelante" (*chilling effect*) che l'uso della *FRT* può avere sulla libertà di riunione, poiché i cittadini potrebbero temere di essere identificati e quindi essere scoraggiati dal partecipare a manifestazioni.

biometrici".[86] Ai sensi della legislazione europea, in particolare del Regolamento Generale sulla Protezione dei Dati (GDPR) e della Direttiva *Law Enforcement* (*LED*), i dati biometrici rientrano nelle "categorie speciali di dati personali", il cui trattamento è, in linea di principio, vietato. L'articolo 10 della Direttiva LED, applicabile alle attività di polizia e giudiziarie, ammette il trattamento di tali dati solo se "strettamente necessario" e soggetto a garanzie specifiche e adeguate.[87] Questo standard di "stretta necessità" impone un onere argomentativo particolarmente rigoroso per le autorità che intendono avvalersi di tali strumenti, le quali devono dimostrare non solo l'utilità della misura, ma anche l'impossibilità di raggiungere il medesimo obiettivo con mezzi meno intrusivi.

Una ulteriore e fondamentale criticità risiede nella qualità e nella composizione dei database utilizzati per l'addestramento e il confronto. È stato ampiamente documentato come gli algoritmi di FRT presentino tassi di errore significativamente più elevati quando applicati a determinati gruppi demografici, in particolare donne e persone con la pelle scura.[88] Questo fenomeno, noto come *algorithmic bias*, non è un mero difetto tecnico, ma una conseguenza diretta della scarsa rappresentatività di questi gruppi nei dataset di addestramento. L'utilizzo di sistemi affetti da tali distorsioni in contesti di applicazione della legge può portare a discriminazioni sistemiche, con il rischio di false accuse e arresti ingiusti che colpiscono in modo sproporzionato le comunità già marginalizzate.[89]

La giurisprudenza europea ha iniziato a tracciare i confini di ammissibilità di queste tecnologie. Nella celebre sentenza *Glukhin c. Russia*, la Corte Europea dei Diritti dell'Uomo ha condannato l'uso della FRT per identificare e arrestare un

[86] Simmler, M., e Canova, G., op. cit., p. 5. Si chiarisce che le immagini facciali, quando utilizzate per l'identificazione univoca, diventano dati biometrici ai sensi della Direttiva (UE) 2016/680 (Direttiva Law Enforcement), e rientrano quindi nelle categorie speciali di dati personali.
[87] Ibid**.** L'articolo 10 della Direttiva LED è citato come la norma chiave che subordina il trattamento di dati biometrici a un criterio di "stretta necessità" e alla presenza di garanzie adeguate.
[88] Selwyn, N., et al., "*Facial Recognition Technology: Key Issues and Emerging Concerns*", in R. Matulionyte, M. Zalnieriute (eds.), The Cambridge Handbook of Facial Recognition in the Modern State, 2024, p. 13. Gli autori evidenziano come negli Stati Uniti si siano verificati casi regolari di discriminazione razziale guidata dalla FRT, con la polizia che ha utilizzato il riconoscimento facciale per avviare arresti ingiustificati e altri errori giudiziari a danno di gruppi sociali minoritari.
[89] Agenzia dell'Unione Europea per i Diritti Fondamentali (FRA), op. cit., p. 10. Il report della FRA discute esplicitamente il problema del bias algoritmico, riportando che i software di riconoscimento facciale sono stati spesso addestrati principalmente su immagini di uomini bianchi, risultando molto meno accurati su donne e persone appartenenti ad altri gruppi etnici.

manifestante pacifico, definendo tale pratica "incompatibile con gli ideali e i valori di una società democratica" in assenza di un quadro normativo chiaro, prevedibile e dotato di solide garanzie contro gli abusi.[90] La Corte ha sottolineato che l'intrusività di questi strumenti richiede il massimo livello di giustificazione e di controllo. Questo approccio è stato recepito, almeno in parte, dall'AI Act dell'Unione Europea, che vieta l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico per finalità di contrasto, pur prevedendo eccezioni significative per la ricerca di vittime di reati gravi, la prevenzione di minacce terroristiche imminenti e l'identificazione di sospetti per un catalogo di reati gravi.[91]

Tuttavia, anche l'AI Act presenta delle lacune. Non fornisce una base giuridica autonoma per il trattamento dei dati, rimandando alle normative nazionali e alla Direttiva LED, e lascia ampi margini di discrezionalità agli Stati membri nel definire le procedure di autorizzazione. Inoltre, il suo approccio basato sul rischio, che distingue tra uso "in tempo reale" (vietato con eccezioni) e uso "*ex post*" (considerato ad alto rischio e quindi soggetto a requisiti specifici), è stato criticato per non cogliere appieno la gravità dell'interferenza, che non dipende solo dalla temporalità dell'analisi, ma dalla raccolta e conservazione stessa dei dati biometrici su larga scala.[92]

In definitiva, l'inquadramento giuridico delle FRT si presenta come un sistema multilivello complesso e ancora in evoluzione, che intreccia la protezione dei dati, i diritti umani, la regolamentazione dell'IA e il diritto penale. La sfida per il legislatore e per l'interprete è quella di costruire un quadro normativo che, senza

---

[90] Simmler, M., e Canova, G., op. cit., p. 1. La sentenza Glukhin c. Russia (App. n. 11519/20, 4 luglio 2023) della CEDU viene citata come il primo giudizio fondamentale sull'uso della FRT da parte delle forze dell'ordine, in cui la Corte ha ritenuto che tale uso per identificare un manifestante violasse i diritti fondamentali.

[91] Paolucci, F., "*Enhancing Oversight and Addressing Gaps: Assessing the Impact of the AI Act on Biometric Identification Systems*", in N. Menéndez González, G. Mobilio (eds.), *Next Democratic Frontiers for Facial Recognition Technology (FRT)*, 2025, p. 79. L'autrice spiega come l'AI Act classifichi l'uso della FRT in tempo reale come pratica vietata (con eccezioni) e l'uso ex post come pratica ad alto rischio, delineando un approccio normativo basato sulla gravità dell'interferenza. Pdf disponibile su: https://link.springer.com/chapter/10.1007/978-3-031-89794-8_5

[92] Jasserand, C., "*Facial Recognition in Public Spaces and the Principle of Necessity*", in N. Menéndez González, G. Mobilio (eds.), *Next Democratic Frontiers for Facial Recognition Technology (FRT)*, 2025, p. 68. L'autrice critica l'AI Act per non aver giustificato adeguatamente perché l'uso ex post della FRT ponga meno rischi rispetto all'uso in tempo reale, suggerendo che la distinzione sia più politica che basata su una solida analisi dei rischi per i diritti fondamentali. Disponibile su: https://link.springer.com/content/pdf/10.1007/978-3-031-89794-8_4

rinunciare aprioristicamente ai potenziali benefici di queste tecnologie nella lotta alla criminalità, sappia implementare garanzie procedurali e sostanziali così robuste da impedire che lo strumento di sicurezza si trasformi in un'architettura di sorveglianza di massa, incompatibile con i principi fondamentali dello Stato di diritto democratico.[93]

1.2. Criticità normative e lacune regolatorie nella disciplina delle FRT

L'attuale assetto normativo che governa le tecnologie di riconoscimento facciale, si configura come un sistema complesso e multilivello, caratterizzato da significative lacune e da una frammentazione che fatica a tenere il passo con la rapida e pervasiva evoluzione tecnologica. Sebbene il diritto europeo abbia compiuto passi importanti, in particolare con l'adozione dell'AI Act, persistono aree di incertezza e criticità strutturali che rischiano di compromettere l'effettiva tutela dei diritti fondamentali. L'analisi di queste lacune non è un mero esercizio teorico, ma una necessità impellente per comprendere come il potere algoritmico della FRT possa insinuarsi nelle maglie di un quadro giuridico non ancora pienamente attrezzato a governarlo.

Una delle principali criticità risiede nella difficoltà di applicare i principi cardine della protezione dei dati, come la necessità e la proporzionalità, a sistemi intrinsecamente progettati per la sorveglianza di massa. L'uso della FRT da parte delle forze dell'ordine, specialmente negli spazi pubblici, implica per sua natura il trattamento indiscriminato dei dati biometrici di un numero enorme di individui, la stragrande maggioranza dei quali non ha alcun legame con attività criminali.[94]

---

[93] Levantino, F. P., "*From Identity to Emotional Dominance? "Early Warnings" on Emotion Recognition Uses by Police Forces*", in N. Menéndez González, G. Mobilio (eds.), *Next Democratic Frontiers for Facial Recognition Technology (FRT)*, 2025, p. 143. L'autore, discutendo il caso Glukhin, sottolinea come la Corte EDU abbia collegato le violazioni dei diritti a implicazioni più ampie per le società democratiche, in particolare attraverso gli effetti raggelanti generati da tali tecnologie di sorveglianza. Disponibile su: https://link.springer.com/content/pdf/10.1007/978-3-031-89794-8_8

[94] Gabrielli, G., "*The Use of Facial Recognition Technologies in the Context of Peaceful Protest: The Risk of Mass Surveillance Practices and the Implications for the Protection of Human Rights*", European Journal of Risk Regulation, 16, 2025, p. 518. L'autrice descrive l'uso della FRT in tempo reale come una pratica che implica la scansione di grandi folle per confrontare i volti con una watchlist, evidenziando come tale sorveglianza indiscriminata sia intrinsecamente problematica. PDF disponibile su: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/A4B2FABA8F32DDBC0217C86837CDBAC6/S1867299X25000261a.pdf/the-use-

Questa pratica di sorveglianza generalizzata, o a strascico (*dragnet surveillance*), entra in diretta collisione con la giurisprudenza consolidata della Corte di Giustizia dell'Unione Europea, la quale ha più volte affermato che misure di conservazione dei dati generalizzata e indifferenziata non sono compatibili con i principi di necessità e proporzionalità.[95] La stessa Corte Europea dei Diritti dell'Uomo, nel caso *Glukhin* già citato, ha censurato la legislazione russa proprio per la sua eccessiva ampiezza, che non prevedeva limiti chiari sulla natura delle situazioni, sulle finalità o sulle categorie di persone che potevano essere soggette a monitoraggio tramite FRT.

Un'altra lacuna fondamentale riguarda la definizione e la gestione dei database biometrici. Il caso Clearview AI ha dimostrato in modo emblematico come la distinzione tra dati "pubblicamente disponibili" e dati utilizzabili per finalità di sorveglianza, sia estremamente problematica. Clearview ha costruito il suo immenso archivio attraverso lo *scraping* di immagini da Internet, sostenendo di utilizzare solo informazioni pubbliche. Tuttavia, come sottolineato da numerose autorità garanti, la pubblicazione di una foto su un social network non costituisce un consenso implicito al suo inserimento in un database biometrico per finalità di polizia.[96] L'AI Act ha tentato di porre un argine a questa pratica, vietando la creazione di database facciali attraverso lo "*scraping* non mirato" da Internet o da filmati di videosorveglianza.[97] Ciononostante, la norma non risolve la questione dei database già esistenti o di quelli creati con altre modalità, né chiarisce a sufficienza i criteri di qualità e di non discriminazione che tali archivi dovrebbero rispettare per essere considerati legittimi.

---

of-facial-recognition-technologies-in-the-context-of-peaceful-protest-the-risk-of-mass-surveillance-practices-and-the-implications-for-the-protection-of-human-rights.pdf

[95] Rezende, I. N., "*Facial Recognition in Police Hands: Assessing the "Clearview case' from a European perspective*", New Journal of European Criminal Law, 11(3), 2020, nel paragrafo "Strict necessity test". L'autrice, analizzando il test di stretta necessità alla luce della giurisprudenza della Corte di Giustizia dell'UE, applica il principio secondo cui le misure di sorveglianza devono essere limitate e basate su criteri oggettivi, in contrasto con la sorveglianza generalizzata e indiscriminata che caratterizzerebbe l'uso non regolamentato di un database come quello di Clearview. Disponibile su: https://journals.sagepub.com/doi/epub/10.1177/2032284420948161

[96] Dul, C., "*Facial Recognition Technology vs Privacy: The Case of Clearview AI*", Queen Mary Law Journal, 3, 2022, p. 13. L'autrice spiega come la Commissione Nazionale francese per l'Informatica e le Libertà (CNIL) abbia ordinato a Clearview AI di cessare il riutilizzo delle fotografie disponibili su Internet, contestando la base giuridica della raccolta dati e violando l'articolo 6 del GDPR.

[97] Simmler, M., e Canova, G., op. cit., p. 9. L'articolo 5(1)(e) dell'AI Act viene citato come la norma che vieta "la messa in circolazione, la messa in servizio o l'uso di sistemi di IA che creano o espandono database di riconoscimento facciale attraverso lo scraping non mirato di immagini facciali da Internet o da filmati di *CCTV (closed circuit television)*

Questo ci conduce a una terza, e forse più insidiosa, criticità: il *bias* algoritmico e il rischio di discriminazione. È scientificamente provato che molti sistemi di FRT presentano tassi di errore significativamente più alti su determinati gruppi demografici, come donne e minoranze etniche, a causa di dataset di addestramento non rappresentativi. [98] Questa "imprecisione tecnica", come la definisce l'AI Act, non è un semplice difetto correggibile, ma una falla strutturale che può portare a discriminazioni sistemiche. Quando uno strumento del genere viene utilizzato dalle forze dell'ordine, il rischio è che gli errori del sistema si traducano in arresti ingiusti, false accuse e un rafforzamento dei pregiudizi esistenti, colpendo in modo sproporzionato le comunità già vulnerabili. Sebbene l'AI Act imponga requisiti di qualità per i dati di addestramento, la verifica della loro effettiva implementazione e l'efficacia delle misure contro il *bias* rimangono una sfida aperta, la cui gestione è cruciale per la legittimità democratica di questi strumenti.[99]

Una quarta lacuna emerge sul piano procedurale e delle garanzie. L'AI Act, pur stabilendo la necessità di un'autorizzazione *ex ante* per l'uso della FRT in tempo reale da parte delle forze dell'ordine, lascia agli Stati membri la scelta tra un'autorità giudiziaria o un'autorità amministrativa indipendente. Questa alternativa non è neutrale, l'intervento di un giudice offre, per sua natura, un livello di garanzia, indipendenza e terzietà superiore a quello di un organo amministrativo, come richiesto dalla giurisprudenza della CEDU per le misure di sorveglianza più intrusive.[100] L'assenza di un obbligo di autorizzazione esclusivamente giudiziaria, rappresenta una potenziale vulnerabilità nel sistema di tutela, che potrebbe portare a un abbassamento degli standard di protezione a livello nazionale.

---

[98] Agenzia dell'Unione Europea per i Diritti Fondamentali (FRA), "*Facial recognition technology: fundamental rights considerations in the context of law enforcement*", FRA Focus, 2019, p. 27. Il report della FRA evidenzia che i sistemi di FRT hanno mostrato tassi di errore più elevati su donne e persone di colore, producendo risultati distorti che possono portare a discriminazione.

[99] Selwyn, N., et al., "*Facial Recognition Technology: Key Issues and Emerging Concerns*", in R. Matulionyte, M. Zalnieriute (eds.), The Cambridge Handbook of Facial Recognition in the Modern State, 2024, p. 13. Gli autori sostengono che il bias algoritmico non è un problema puramente tecnico, ma un problema sociotecnico, radicato in dataset, modelli e contesti di utilizzo distorti, che finisce per esacerbare i danni già subiti dai gruppi socialmente marginalizzati.

[100] Paolucci, F., "*Enhancing Oversight and Addressing Gaps: Assessing the Impact of the AI Act on Biometric Identification Systems*", in N. Menéndez González, G. Mobilio (eds.), Next Democratic Frontiers for Facial Recognition Technology (FRT), 2025, p. 86. L'autrice sottolinea come l'AI Act lasci agli Stati membri la scelta tra un'autorità giudiziaria e una amministrativa per l'autorizzazione all'uso della FRT, evidenziando come questa scelta non sia neutrale e abbia implicazioni significative per le garanzie procedurali e l'indipendenza del controllo.

Infine, il quadro normativo attuale mostra una difficoltà strutturale nel governare la natura "duale" di queste tecnologie, che possono essere utilizzate sia per finalità civili che militari o di sicurezza. L'uso della FRT in contesti di conflitto armato per il *targeting*, che approfondiremo nel capitolo successivo, solleva questioni specifiche legate al diritto internazionale umanitario, come il rispetto dei principi di distinzione e proporzionalità, che non sono direttamente affrontate dalle normative sulla protezione dei dati o sull'IA.[101] Questa frammentazione tra regimi giuridici diversi (diritto della privacy, diritto penale, diritto internazionale umanitario) crea zone d'ombra in cui l'uso della tecnologia rischia di sfuggire a un controllo efficace, con potenziali effetti di *spillover* delle pratiche e delle logiche militari nei contesti di sicurezza interna. Anche per questo, organi consultivi europei hanno richiesto un divieto generalizzato di queste pratiche negli spazi pubblici, ritenendole incompatibili con i diritti fondamentali.[102]

In sintesi, nonostante i progressi normativi, la disciplina delle FRT presenta ancora lacune significative che richiedono un intervento mirato. È necessaria una maggiore chiarezza sui limiti della sorveglianza di massa, una regolamentazione più stringente sulla creazione e l'uso dei database biometrici, meccanismi più efficaci per contrastare il *bias* algoritmico e garanzie procedurali più solide, come l'obbligo di un controllo giurisdizionale preventivo. Solo colmando queste lacune, sarà possibile garantire che l'impiego delle tecnologie di riconoscimento facciale avvenga nel rispetto pieno e sostanziale dei diritti fondamentali e dei principi dello Stato di diritto.

---

[101] Rosenzweig, I., e Pacholska, M., "*The use of facial recognition for targeting under international law*", International Review of the Red Cross, 2025, p. 250. Gli autori analizzano l'uso della FRT nel contesto dei conflitti armati, evidenziando come una determinazione errata dell'obiettivo basata su un'identificazione facciale positiva possa essere considerata indiscriminata e arbitraria, violando i principi del diritto internazionale umanitario. PDF disponibile su: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/E8F47BBC8987E7E505C9152A3AADBCDE/S1816383124000705a.pdf/the-use-of-facial-recognition-for-targeting-under-international-law.pdf

[102] EDPB-GEPD, Parere congiunto 5/2021 sulla proposta di regolamento sull'intelligenza artificiale, 18 giugno 2021, p. 13. Il parere congiunto sollecita un "divieto generale di qualsiasi uso dell'IA a fini di riconoscimento automatico" in spazi pubblici, comprese caratteristiche come il volto, l'andatura, le impronte digitali e la voce, ritenendo tali pratiche incompatibili con i diritti fondamentali.

2. Il caso Clearview AI: il paradigma della sorveglianza e la necessità di bilanciamento con i diritti fondamentali

Il caso Clearview AI rappresenta un punto di svolta nel dibattito globale sulla tecnologia di riconoscimento facciale, fungendo da catalizzatore per una rinnovata e urgente riflessione sui limiti della sorveglianza e sulla protezione dei dati personali. Emerso con prepotenza sulla scena pubblica a seguito di un'inchiesta del *New York Times* del gennaio 2020, il caso ha svelato l'esistenza di una *start-up* tecnologica che, operando in una zona grigia normativa e con scarsa trasparenza, aveva sviluppato uno strumento di una potenza e pervasività senza precedenti.[103] A differenza dei sistemi di riconoscimento facciale tradizionali, storicamente limitati al confronto di immagini con database governativi preesistenti (come le foto segnaletiche), l'innovazione di Clearview AI è consistita nel combinare un sofisticato algoritmo di IA, con un database privato di dimensioni colossali, costruito attraverso una pratica aggressiva e controversa di *data scraping*.

L'azienda ha "raschiato" (*scraped*) miliardi di immagini di volti umani da Internet, attingendo a una vasta gamma di fonti pubbliche, tra cui social network come Facebook, YouTube e LinkedIn, siti di notizie, blog e piattaforme educative. Questo processo ha permesso a Clearview di assemblare un archivio biometrico che, al momento della sua scoperta, contava oltre tre miliardi di immagini, un numero poi cresciuto esponenzialmente fino a superare i dieci miliardi nell'ottobre 2021.[104] L'applicazione, venduta a centinaia di agenzie di polizia negli Stati Uniti e in altri paesi, permetteva ai suoi utenti di caricare una fotografia di un individuo e, in pochi secondi, ottenere una serie di immagini corrispondenti tratte dal suo vasto database, complete di link alle fonti originali. Questa funzionalità, di fatto, annullava

[103] Devany, B. E., "*Clearview AI's First Amendment: A Dangerous Reality?*", Texas Law Review, 101(2), 2022, p. 474. L'autrice introduce il caso facendo riferimento alla causa intentata dall'ACLU contro Clearview AI per la violazione del Biometric Information Privacy Act (BIPA) dell'Illinois, sottolineando come l'azienda abbia raccolto miliardi di foto personali, ma pubblicamente disponibili, da siti come Facebook e LinkedIn. PDF disponibile su: https://texaslawreview.org/wp-content/uploads/2023/01/Devany.Printer2-8.pdf

[104] Kohn, M., "*Clearview AI, TikTok, and the Collection of Facial Images in International Law*", Chicago Journal of International Law, 23(1), 2022, p. 199. L'articolo, citando una fonte del New York Times, descrive Clearview AI come un servizio di riconoscimento facciale con sede negli Stati Uniti, le cui pratiche commerciali sono state rese note da un'inchiesta giornalistica. Si menziona che Clearview "può raccogliere automaticamente immagini di volti di persone da tutto Internet". PDF disponibile su: https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1832&context=cjil

l'anonimato negli spazi pubblici e digitali, trasformando ogni volto in un potenziale punto di accesso a un dossier digitale sulla vita di una persona.

Il modello di business di Clearview AI si fonda su una premessa giuridicamente ed eticamente problematica: la presunta legittimità di raccogliere e utilizzare dati personali perché "manifestamente resi pubblici" dall'interessato. Tuttavia, questa interpretazione estensiva e unilaterale del concetto di "dato pubblico" si scontra frontalmente con i principi fondamentali della protezione dei dati, in particolare con il principio di finalità e con il concetto di consenso informato. Come evidenziato da numerose autorità garanti della protezione dei dati in Europa, il fatto che un'immagine sia accessibile online non implica un consenso implicito al suo *scraping*, alla sua indicizzazione in un database biometrico e al suo successivo utilizzo per finalità di sorveglianza da parte delle forze dell'ordine.[105] Le politiche sulla privacy della maggior parte delle piattaforme di social media, inoltre, vietano esplicitamente questo tipo di raccolta automatizzata di dati, come dimostrano le lettere di diffida (*cease-and-desist*) inviate a Clearview da giganti come Twitter e Google.

Il caso Clearview ha messo a nudo la profonda asimmetria di potere che caratterizza l'era digitale. Da un lato, i cittadini che condividono fotografie online per finalità sociali e personali; dall'altro, un'entità privata che capitalizza su questi dati, trasformandoli in uno strumento di sorveglianza venduto a enti governativi. Questa dinamica sfida la tradizionale distinzione tra sfera pubblica e privata, creando una forma di sorveglianza ibrida in cui attori privati svolgono una funzione quasi-statale, eludendo le garanzie legali che normalmente regolano l'accesso ai dati da parte dello Stato.[106] L'azienda, infatti, non agisce come un mero fornitore di tecnologia, ma come un intermediario di dati che ha creato il proprio archivio al di fuori di ogni controllo democratico e normativo.

---

[105] Ivi, p. 209. L'autrice riporta la denuncia presentata dal cittadino tedesco Matthias Marx presso il Garante della Privacy di Amburgo, che ha portato all'ordine di cancellazione dei dati biometrici (i valori hash) di Marx da parte di Clearview.

[106] Rezende, I. N., "*Facial Recognition in Police Hands: Assessing the "Clearview case' from a European perspective*", New Journal of European Criminal Law, 11(3), 2020, p. 378. L'autrice inquadra il caso Clearview nella tendenza globale del riutilizzo di dati raccolti dal settore privato per scopi di contrasto, ma ne sottolinea la specificità: i dati non vengono trasferiti su base occasionale, ma raccolti da un'azienda privata con il preciso intento di renderli disponibili alle agenzie governative.

Le implicazioni per i diritti fondamentali sono profonde e sistemiche. In primo luogo, vi è una palese violazione del diritto alla privacy e alla protezione dei dati personali, come sancito dal GDPR in Europa. La raccolta massiva e indiscriminata di immagini facciali, la creazione di *template* biometrici e la loro conservazione a tempo indeterminato, avvengono senza una base giuridica valida, in violazione degli articoli 5, 6 e 9 del GDPR.[107]; l'articolo 9 del GDPR classifica i dati biometrici utilizzati "allo scopo di identificare in modo univoco una persona fisica" come una categoria speciale di dati, il cui trattamento è di norma vietato. Una delle poche eccezioni applicabili in teoria a questo scenario è quella prevista dall'art. 9, par. 2, lett. e), che consente il trattamento di dati "resi manifestamente pubblici dall'interessato". In secondo luogo, l'esistenza stessa di un simile database genera un potente "*chilling effect*" sulla libertà di espressione e di riunione. La consapevolezza che ogni partecipazione a una manifestazione, ogni attività pubblica, possa essere registrata e archiviata, e che la propria identità possa essere svelata in qualsiasi momento, può scoraggiare la partecipazione democratica e il dissenso politico, minando le fondamenta di una società libera e aperta.[108]

Le azioni legali e le sanzioni intraprese da diverse autorità garanti europee, tra cui quelle di Francia, Italia, Grecia e Regno Unito, hanno dichiarato illegittime le pratiche di Clearview AI nel territorio dell'Unione, imponendo multe significative e ordinando la cancellazione dei dati dei cittadini europei.[109] Tuttavia, l'efficacia di queste decisioni si scontra con la natura globale e spesso elusiva delle operazioni delle aziende tecnologiche, sollevando complesse questioni di giurisdizione e di esecuzione transfrontaliera. Il caso Clearview, quindi, non è solo la storia di una singola azienda controversa, ma è il paradigma delle sfide che le democrazie liberali devono affrontare. Esso dimostra la necessità di un quadro normativo che non si

[107] Ivi, p. 380. Viene analizzata la conformità delle pratiche di Clearview con l'articolo 9 del GDPR. Si sostiene che la base giuridica dei "dati manifestamente resi pubblici" è difficilmente applicabile, poiché le policy dei social network spesso vietano lo scraping e gli utenti possono avere profili privati.
[108] Devany, B. E., op. cit., p. 475. L'autrice riporta le preoccupazioni degli attivisti per i diritti civili, i quali temono che "diffondere la paura e scoraggiare l'attivismo possa essere proprio lo scopo" dei nuovi sistemi di riconoscimento facciale come quello di Mosca, tracciando un parallelo con i rischi posti da Clearview.
[109] Kohn, M., op. cit., p. 211. L'articolo menziona le decisioni di diverse autorità di regolamentazione europee. In particolare, si fa riferimento alla decisione della CNIL (Francia) che ha ordinato a Clearview di cessare la raccolta di dati di individui in territorio francese e di cancellare i dati esistenti.

limiti a reagire *ex post* alle violazioni, ma che sia in grado di anticipare e prevenire i rischi, stabilendo limiti chiari e invalicabili all'uso di tecnologie così potenti.

Un secondo profilo di analisi giuridica attiene alla qualificazione del ruolo di Clearview AI ai sensi del diritto europeo. Quando l'azienda fornisce i suoi servizi alle forze dell'ordine di uno Stato membro, si pone il problema di definire se essa agisca come "titolare del trattamento" o "responsabile del trattamento". Sebbene Clearview si presenti come un semplice fornitore di servizi, la sua attività non si limita a fornire un software, ma include la gestione di un database proprietario, creato e alimentato autonomamente. Questo la colloca in una posizione ambigua. Se agisse come mero responsabile del trattamento per conto di un'autorità di polizia, dovrebbe operare esclusivamente su istruzione di quest'ultima. Tuttavia, è Clearview che determina le finalità e i mezzi essenziali della raccolta dati su larga scala, configurandosi piuttosto come un titolare autonomo o, al più, un contitolare del trattamento.[110] Questa distinzione è cruciale, poiché un titolare del trattamento è soggetto a tutti gli obblighi del GDPR, compresi quelli di trasparenza, liceità e correttezza, che, come visto, appaiono sistematicamente violati.

La difesa di Clearview AI, basata su un'interpretazione aggressiva del Primo Emendamento della Costituzione statunitense, ha ulteriormente complicato il quadro. L'azienda ha sostenuto di avere il diritto, protetto dalla libertà di parola, di raccogliere informazioni pubblicamente disponibili e di venderle.[111] Questa tesi, sebbene criticata da molti giuristi come "pericolosa" e "semplicistica",[112] si inserisce in una tendenza della giurisprudenza statunitense a interpretare in modo sempre più estensivo la protezione accordata al discorso commerciale. Tale scontro tra il diritto alla privacy, così come inteso in Europa, e la libertà di espressione, nella sua

[110] Rezende, I. N., op. cit., p. 381. nel paragrafo "Subsequent use of GDPR data in private–public partnerships". L'autrice analizza la complessa qualificazione giuridica di Clearview ai sensi del diritto europeo, distinguendo tra il ruolo di responsabile (processor) e quello di contitolare (joint controller). Si evidenzia che, qualora Clearview agisse come contitolare, determinando insieme alle forze dell'ordine le finalità e i mezzi del trattamento, acquisirebbe lo status di "autorità competente" ai sensi della Direttiva, con tutte le responsabilità che ne derivano.

[111] Devany, B. E., op. cit., p. 475. Si riporta la tesi difensiva di Clearview AI, secondo cui l'azienda avrebbe un diritto, protetto dal Primo Emendamento, di raccogliere dati e vendere il suo servizio di riconoscimento facciale.

[112] Ibid. L'articolo elenca le critiche mosse da diversi accademici a questa tesi, definita "priva di fondamento", "semplicistica", "in contrasto con la dottrina consolidata del Primo Emendamento" e persino "pericolosa".

accezione statunitense, evidenzia la profonda divergenza filosofica e giuridica tra i due continenti, e rappresenta un ostacolo significativo a qualsiasi forma di regolamentazione globale.[113]

Infine, il caso ha messo in luce la vulnerabilità delle infrastrutture digitali. Nonostante le affermazioni dell'azienda sulla sicurezza dei propri sistemi, Clearview AI ha subito una violazione dei dati che ha portato al furto della sua intera lista di clienti, esponendo le agenzie di polizia che utilizzavano il servizio.[114] Questo episodio ha dimostrato che, al di là delle questioni di legittimità della raccolta, la concentrazione di enormi quantità di dati biometrici sensibili in un unico database, crea un "bersaglio di alto valore" per attacchi informatici, con rischi incalcolabili per la sicurezza nazionale e individuale. La vicenda, quindi, non solo ha svelato le pratiche di un'azienda, ma ha anche acceso un faro sulle fragilità sistemiche dell'ecosistema digitale e sulla necessità di un approccio alla sicurezza informatica che sia integrato fin dalla progettazione (*security by design*) in qualsiasi sistema che tratti dati su larga scala.[115]

3. La sentenza *Glukhin c. Russia*: le prime applicazioni giurisprudenziali della Corte EDU

La sentenza della Corte Europea dei Diritti dell'Uomo (CEDU) nel caso Glukhin c. Russia, emessa il 4 luglio 2023, rappresenta una pietra miliare nella giurisprudenza continentale in materia di tecnologie di riconoscimento facciale. Per la prima volta, un tribunale internazionale di tale levatura si è pronunciato in modo specifico e dettagliato sulla compatibilità dell'uso della FRT da parte delle forze dell'ordine, con i diritti sanciti dalla Convenzione Europea dei Diritti dell'Uomo, in particolare nel contesto delle manifestazioni pacifiche.

---

[113] Ivi, p. 493. Viene analizzata la "tensione storica tra gli interessi della privacy e quelli della libertà di parola", citando il famoso articolo di Warren e Brandeis "*The Right to Privacy*" come punto di partenza di questo dibattito.

[114] Forbes, *Clearview AI, The Company Whose Database Has Amassed 3 Billion Photos, Hacked* By Kate O'Flaherty , l'articolo menziona la violazione dei dati subita da Clearview AI, che ha portato al furto della sua lista clienti, dimostrando la vulnerabilità di tali sistemi. Disponibile su: https://www.forbes.com/sites/kateoflahertyuk/2020/02/26/clearview-ai-the-company-whose-database-has-amassed-3-billion-photos-hacked/

[115] Devany, B. E., op. cit., p. 502. Nel proporre un modello legislativo per regolamentare la FRT, l'autrice suggerisce di "stabilire test di sicurezza prima di ingaggiare un fornitore di FRT" per minimizzare il rischio di violazioni di dati.

Il caso trae origine da una protesta solitaria e pacifica, inscenata dal cittadino russo Nikolaj Glukhin. Il 23 agosto 2019, egli si recò nella metropolitana di Mosca portando con sé una sagoma di cartone a grandezza naturale di un attivista politico, Konstantin Kotov, che era stato arrestato pochi giorni prima. La protesta del signor Glukhin, di per sé non violenta e non dirompente, fu documentata da fotografie e video che vennero successivamente pubblicati online su un canale Telegram. Fu proprio attraverso l'analisi di questo materiale digitale, che le autorità russe, avvalendosi di sistemi di FRT in modalità ex post, riuscirono a identificare il ricorrente. Successivamente, le stesse tecnologie furono impiegate in modalità "in tempo reale" (live) per localizzarlo e arrestarlo giorni dopo, sempre all'interno della rete metropolitana. A seguito dell'arresto, il signor Glukhin fu condannato per aver violato le procedure nazionali sulla conduzione di eventi pubblici, non avendo notificato preventivamente la sua manifestazione.[116]

Partendo da questa base fattuale, la Corte di Strasburgo ha condannato la Russia per la violazione degli articoli 10 (libertà di espressione) e 8 (diritto al rispetto della vita privata e familiare) della Convenzione. La decisione si fonda su un'analisi rigorosa che intreccia i principi di legalità, necessità e proporzionalità, stabilendo standard elevati per l'impiego di tecnologie così invasive.

Sotto il profilo della libertà di espressione (Art. 10), la Corte ha ritenuto che l'arresto e la successiva condanna del ricorrente per una manifestazione pacifica e non autorizzata costituissero un'ingerenza sproporzionata. Tuttavia, l'elemento di maggiore innovazione risiede nell'analisi condotta ai sensi dell'articolo 8. La Corte ha qualificato l'uso della FRT come "particolarmente intrusivo" nella vita privata degli individui, un'ingerenza che richiede un "elevato livello di giustificazione" per essere considerata "necessaria in una società democratica".[117]

---

[116] Zalnieriute, M., "*International Decisions: Glukhin v. Russia*", American Journal of International Law, 117(4), 2023, p. 695. L'autrice riassume i fatti del caso, descrivendo la protesta solitaria di Glukhin con la sagoma di cartone, la sua identificazione tramite FRT dopo la pubblicazione di video su Telegram e il suo successivo arresto, fornendo il contesto fattuale essenziale per la decisione della Corte. PDF disponibile su: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/AF9C0AAFA6E44CE9F1881B0542177D3C/S0002930023000520a.pdf/glukhin-v-russia-app-no-1151920-judgment.pdf

[117] Corte Europea dei Diritti dell'Uomo, "*Use of facial-recognition technology breached rights of Moscow underground protestor*", Comunicato Stampa ECHR 207 (2023), 4 luglio 2023, p. 3. Il comunicato stampa ufficiale della Corte riassume i punti chiave della sentenza, affermando che la Corte ha ritenuto le misure contro il Sig. Glukhin "particolarmente intrusive" e che la sua protesta

Il cuore della sentenza risiede nella valutazione della "qualità della legge" russa. La Corte ha riscontrato che la legislazione nazionale era formulata in termini eccessivamente ampi e vaghi, non fornendo garanzie adeguate contro il rischio di abusi e arbitrarietà. Secondo i giudici di Strasburgo, una normativa che regola l'uso della FRT deve essere chiara e precisa, definendo in modo dettagliato la natura delle situazioni che possono giustificare l'uso della tecnologia, le finalità perseguite, le categorie di persone che possono essere oggetto di sorveglianza e le procedure di autorizzazione, supervisione e conservazione dei dati.[118] In assenza di tali "garanzie minime", la legislazione russa non soddisfaceva il requisito di prevedibilità e legalità richiesto dalla Convenzione, rendendo l'interferenza con la vita privata del signor Glukhin illegittima. La Corte ha esplicitamente affermato che è "essenziale disporre di regole dettagliate che disciplinino la portata e l'applicazione di tali misure, nonché di solide garanzie contro il rischio di abusi".[119]

Un altro aspetto cruciale della sentenza è il riconoscimento del potente "effetto raggelante" (chilling effect) che l'uso della FRT può avere sull'esercizio dei diritti di libertà di espressione e di riunione. La consapevolezza di poter essere identificati e monitorati in modo sistematico durante una protesta può, infatti, dissuadere i cittadini dal partecipare alla vita pubblica e dal manifestare il proprio dissenso. La Corte ha concluso che l'uso di una tecnologia così intrusiva nel contesto dell'esercizio di un diritto fondamentale è "incompatibile con gli ideali e i valori di

---

pacifica non rappresentava un pericolo per l'ordine pubblico. PDF disponibile su: https://hudoc.echr.coe.int/app/conversion/pdf/?library=ECHR&id=003-7694109-10618091&filename=Judgment%20Glukhin%20v.%20Russia%20-%20use%20of%20facial-recognition%20technology%20against%20Moscow%20underground%20protestor.pdf

[118] Gabrielli, G., "*The Use of Facial Recognition Technologies in the Context of Peaceful Protest: The Risk of Mass Surveillance Practices and the Implications for the Protection of Human Rights*", European Journal of Risk Regulation, 16, 2025, p. 534. L'autrice, analizzando la sentenza Glukhin, sottolinea come la Corte abbia riscontrato che la legislazione russa "mancava di qualsiasi limitazione relativa alla natura delle situazioni che giustificano l'uso della FRT, agli obiettivi previsti, o alle categorie di persone prese di mira". PDF disponibile su: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/A4B2FABA8F32DDBC0217C86837CDBAC6/S1867299X25000261a.pdf/the-use-of-facial-recognition-technologies-in-the-context-of-peaceful-protest-the-risk-of-mass-surveillance-practices-and-the-implications-for-the-protection-of-human-rights.pdf

[119] Zalnieriute, M., op. cit., p. 697. L'autrice riporta che la Camera ha concluso che la legislazione russa sulla FRT non soddisfaceva il requisito della "qualità della legge" ai sensi dell'articolo 8, poiché non stabiliva adeguate garanzie contro il rischio di abusi e arbitrarietà.

una società democratica governata dallo Stato di diritto".[120] Questa affermazione, di portata storica, non sancisce un divieto assoluto, ma stabilisce una forte presunzione di incompatibilità, specialmente quando la tecnologia viene impiegata per monitorare attività politiche pacifiche.

La sentenza Glukhin ha implicazioni che vanno ben oltre il caso specifico e si proiettano direttamente sul dibattito in corso nell'Unione Europea riguardo all'AI Act. Sebbene la Corte non abbia imposto un divieto totale, ha fissato standard di protezione così elevati che molte delle eccezioni previste dall'AI Act per l'uso della FRT da parte delle forze dell'ordine potrebbero essere difficilmente compatibili con la Convenzione, se non interpretate in modo estremamente restrittivo e accompagnate da garanzie procedurali e sostanziali molto robuste. La decisione, quindi, funge da monito per i legislatori nazionali che saranno chiamati a implementare l'AI Act: qualsiasi normativa interna che autorizzi l'uso della FRT dovrà essere redatta con la massima precisione e prevedere un controllo, preferibilmente giudiziario, rigoroso e indipendente.[121]

CAPITOLO III

I SISTEMI D'ARMA AUTONOMI E IL CASO ISRAELO-PALESTINESE

Sommario: 1. I sistemi d'arma autonomi (AWS): sfide giuridiche e tecnologiche nel diritto internazionale umanitario (DIU); 1.1. Caratteristiche tecnologiche e livelli di autonomia; 1.2. AWS e diritto internazionale umanitario: principi fondamentali; 1.3. Profili di responsabilità e *accountability*; 1.4. Controllo umano significativo e supervisione; 1.5. Prospettive future e necessità di regolamentazione; 1.6. AI Act e

[120] Corte Europea dei Diritti dell'Uomo, op. cit., p. 3. Il comunicato stampa della Corte conclude che "l'uso della tecnologia di riconoscimento facciale nel caso del Sig. Glukhin era stato incompatibile con gli ideali e i valori di una società democratica governata dallo Stato di diritto".

[121] Palmiotto, F., e Menéndez González, N., "Facial recognition technology, democracy and human rights", Computer Law & Security Review, 56, 2025, p. 4. Gli autori commentano la sentenza Glukhin evidenziando che, sebbene la Corte non abbia vietato l'uso della FRT, ha stabilito standard così elevati che sarà difficile per gli Stati giustificarne l'uso, specialmente per il monitoraggio di attività politiche, e che ciò avrà un impatto significativo sul dibattito relativo all'AI Act. PDF disponibile su: https://www.sciencedirect.com/science/article/pii/S0267364923000675?via%3Dihub

sistemi d'arma autonomi: incidenza normativa a monte tra divieti biometrici, trasparenza e modelli di uso generale; 2. Sistemi di riconoscimento facciale e armi autonome nel contesto bellico: la trasformazione digitale del conflitto armato in Palestina; 2.1. *Red Wolf*: sistema di identificazione automatizzata nei checkpoint; 2.2. *Blue Wolf*: sorveglianza biometrica diffusa; 2.3. *Wolf Pack*: integrazione sistemica della sorveglianza; 3. I sistemi d'arma autonomi: *Lavender*, algoritmo della morte automatizzata; 3.1. *Gospel*: predizione comportamentale per il *targeting*; 3.2. *Where's Daddy*: *targeting* familiare automatizzato; 4. Implicazioni giuridiche e prospettive future: la sfida al DIU; 4.1. Il problema del controllo umano significativo; 4.2. Responsabilità e *accountability*; 4.3. Prospettive future e regolamentazione

1. I sistemi d'arma autonomi (*AWS*): sfide giuridiche e tecnologiche nel diritto internazionale umanitario (DIU)

Il panorama contemporaneo della tecnologia militare è caratterizzato da un'evoluzione senza precedenti, che ha portato all'emergere dei sistemi d'arma autonomi (AWS). Negli ultimi anni si è assistito a un'escalation significativa nel dibattito internazionale riguardo lo sviluppo e l'implementazione di sistemi d'arma dotati di intelligenza artificiale, capaci di operare con gradi variabili di autonomia sul campo di battaglia.[122] Questi sistemi, una volta programmati e attivati, si affidano all'intelligenza artificiale per adattarsi alle diverse circostanze operative e per identificare, selezionare e ingaggiare obiettivi militari senza necessitare di intervento umano diretto.[123]

Gli AWS in generale comprendono: droni, riconoscimento facciale, intelligenza artificiale e big data. In base al grado di autonomia delle armi si distinguono a livello accademico in tre categorie di armamenti:

[122] Seixas-Nunes, A., *The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective*, Cambridge University Press, 2022, p. 9. L'autrice evidenzia come negli ultimi anni si sia assistito a "una marcata escalation nel dibattito internazionale riguardo lo sviluppo e l'attivazione di sistemi d'arma dotati di intelligenza artificiale, noti come sistemi d'arma autonomi (AWS)". PDF disponibile su: https://www.cambridge.org/core/books/legality-and-accountability-of-autonomous-weapon-systems/FE880FD3F459B29A495D79D0C8347D79

[123] Ibidem, p. 9. L'autrice specifica che "questi sistemi, una volta programmati, dipenderanno dall'IA per adattarsi a diverse circostanze sul campo di battaglia e per identificare, selezionare e ingaggiare obiettivi militari senza la necessità di intervento umano".

-***Human in the loop weapons***: questi sono dei robot capaci di selezionare dei bersagli erogando una determinata forza solo con il comando di un essere umano. Nel sistema "human-in-the-loop", un operatore umano è direttamente coinvolto nel processo decisionale, in particolare per azioni critiche come la selezione e l'ingaggio dei bersagli. In questo modello, l'essere umano ha un ruolo attivo, garantendo che le decisioni siano basate su ragionamento umano e che *l'accountability* resti chiara, riducendo significativamente il rischio di azioni errate o non intenzionali da parte del sistema autonomo.

Queste tipologie di macchine presentano dei vantaggi etici, in quanto la presenza dell'operatore umano consente l'applicazione di ragionamento morale, essenziale per la valutazione del contesto e per rispettare i principi di distinzione e proporzionalità richiesti dal diritto internazionale umanitario. Le decisioni relative a situazioni complesse, come distinguere tra combattenti e civili, richiedono la capacità di giudizio umano che va oltre le capacità di un algoritmo.

Tuttavia, questo modello presenta alcune limitazioni pratiche. Ad esempio, la velocità delle operazioni militari moderne può superare i tempi di reazione umana. Inoltre, in operazioni prolungate, può compromettersi la capacità di prendere decisioni tempestive e accurate.

-***Human on the loop weapons***: Robot capaci di selezionare bersagli erogando una forza sotto la supervisione di un umano che può comunque annullare le azioni della macchina. L'umano quindi mantiene un ruolo di supervisore rispetto alle operazioni del sistema autonomo, ma non è obbligato a intervenire in ogni singola decisione. L'operatore monitora l'attività del sistema e può intervenire qualora venga rilevato un comportamento anomalo o problematico.

Il vantaggio si considera essere la combinazione tra la velocità operativa degli algoritmi autonomi e un controllo umano che può correggere la direzione qualora necessario, permettendo una risposta rapida in situazioni critiche. È più efficiente rispetto al HITL in scenari ad alta velocità, dove le decisioni devono essere prese in tempo reale.

Lo svantaggio principale è che l'operatore umano potrebbe non essere in grado di intervenire tempestivamente in scenari molto complessi o rapidi, a causa delle limitazioni di tempo e delle capacità cognitive. Inoltre, la supervisione passiva

può portare a una ridotta vigilanza dell'operatore, soprattutto in operazioni lunghe o monotone che possono aumentare il rischio di errori.

-***Human out of the loop weapons***: Questa categoria riguarda le macchine in grado di selezionare degli obiettivi erogando una forza senza l'intervento umano né tantomeno altre interazioni. In questo modello il sistema autonomo opera in modo completamente indipendente, prendendo decisioni e agendo senza alcun intervento umano. Una volta attivato, il robot esegue le sue operazioni in modo autonomo, basandosi su algoritmi, intelligenza artificiale o apprendimento automatico senza richiedere alcun input da parte dell'essere umano.

Questo sistema ha il vantaggio di operare con velocità e precisione che un umano non potrebbe eguagliare, soprattutto in ambienti dinamici o in scenari in cui è necessaria una rapida elaborazione di grandi quantità di dati. Il sistema non è soggetto a distrazioni o affaticamento, caratteristiche umane che potrebbero compromettere l'efficacia delle operazioni in scenari di lungo periodo.

Il principale svantaggio è l'assenza di controllo umano, il che pone seri problemi in termini di responsabilità legale ed etica. In caso di errori, danni o violazioni del diritto internazionale umanitario, sarebbe estremamente difficile determinare chi è responsabile. Senza il giudizio umano, i sistemi autonomi potrebbero violare principi cruciali del diritto internazionale, come la distinzione tra combattenti e civili, la proporzionalità dell'attacco e la necessità dell'uso della forza.

Queste categorizzazioni non sono state ufficializzate da trattati o convenzioni ma sono ampiamente accettate in ambito accademico per distinguere il livello di autonomia delle macchine. Questi sistemi bellici automatici pongono sfide significative a causa della loro natura ove il controllo umano molto spesso sfugge durante il funzionamento dell'arma.

La portata rivoluzionaria di questa tecnologia emerge chiaramente quando si considera che, per alcuni esperti del settore, gli AWS potrebbero rappresentare un cambiamento paradigmatico fondamentale nel rapporto tra armi e controllo umano, trasformando gli strumenti bellici da semplici mezzi a veri e propri agenti autonomi

operanti sul teatro delle operazioni.[124] Questa trasformazione solleva interrogativi profondi non solo dal punto di vista tecnologico, ma soprattutto sotto il profilo giuridico e etico, coinvolgendo una molteplicità di attori internazionali nel processo di definizione normativa.

Il dibattito accademico e istituzionale che circonda gli AWS coinvolge una varietà considerevole di stakeholder, includendo Stati nazionali, le loro industrie militari e della difesa, aziende private del settore tecnologico, organizzazioni della società civile, enti umanitari e per la tutela dei diritti umani, oltre a una vasta comunità di studiosi e ricercatori.[125] Questa molteplicità di voci riflette la complessità intrinseca della materia, che tocca simultaneamente aspetti tecnologici dell'autonomia, implicazioni per la conformità alle norme del diritto internazionale umanitario, e questioni fondamentali relative all'attribuzione della responsabilità per le azioni compiute da sistemi dotati di elevati livelli di autonomia.

1.1. Caratteristiche tecnologiche e livelli di autonomia

Dal punto di vista tecnico, l'autonomia nei sistemi d'arma rappresenta un concetto complesso che richiede una comprensione approfondita delle sue manifestazioni pratiche. Il Comitato Internazionale della Croce Rossa ha fornito una definizione operativa dell'autonomia come "la capacità del sistema di agire senza intervento umano diretto, sebbene costituisca un continuum caratterizzato da vari livelli e numerose aree grigie".[126] Questa definizione evidenzia la natura graduale dell'autonomia, che non si presenta come una caratteristica binaria ma piuttosto come uno spettro di capacità che può variare significativamente tra diversi sistemi e contesti operativi.

---

[124] Ibidem, p. 9. L'autrice osserva che "per alcuni, gli AWS potrebbero rappresentare un cambiamento di paradigma in termini di armi e controllo umano e diventare la “terza rivoluzione degli affari militari”, da strumenti a nuovi agenti sul campo di battaglia".

[125] Spadaro, A., "*Superior Responsibility for Autonomous Weapon Systems Crimes: Conceptual and Practical Challenges*", Journal of International Criminal Justice, Vol. 21, 2023, p. 1120. L'autore sottolinea che "negli ultimi anni, una varietà di stakeholder, inclusi stati, le loro industrie militari e della difesa, aziende private, società civile, organizzazioni umanitarie e per i diritti umani, e studiosi, sono stati coinvolti in dibattiti riguardanti lo sviluppo e la regolamentazione dei sistemi d'arma autonomi". PDF disponibile su: https://www.cambridge.org/core/books/legality-and-accountability-of-autonomous-weapon-systems/FE880FD3F459B29A495D79D0C8347D79

[126] Ibidem, nota 1. L'autore cita la definizione del Comitato Internazionale della Croce Rossa che definisce l'autonomia come "la capacità del sistema di agire senza intervento umano diretto, sebbene sia un continuum con vari livelli e molte aree grigie".

La comprensione tecnica dell'autonomia nei sistemi militari si basa sul riconoscimento che essa rappresenta fondamentalmente una forma avanzata di controllo automatizzato piuttosto che l'assenza totale di controllo umano. I sistemi di controllo basati su software utilizzati per ottenere funzionamento autonomo negli AWS sono essenzialmente computer specializzati che eseguono programmi progettati per controllare il sistema d'arma in sostituzione di un operatore umano.[127] Questi sistemi operano secondo il paradigma operativo "sense-think-act" (percepire-pensare-agire), che viene frequentemente utilizzato come definizione operativa di un robot e che descrive il processo ciclico attraverso cui il sistema raccoglie informazioni dall'ambiente, le elabora e agisce di conseguenza.[128]

Un aspetto cruciale da comprendere riguarda il fatto che anche i programmi software più complessi e sofisticati rimangono sostanzialmente insiemi di istruzioni predefinite elaborate da programmatori umani.[129] Questo principio è spesso oscurato nelle discussioni sui sistemi d'arma sofisticati, dove la complessità del comportamento può far apparire che il sistema stia effettuando scelte autonome. In realtà, anche quando un sistema sembra "scegliere" tra diverse opzioni, tale scelta è stata predeterminata dalla persona o dall'organizzazione responsabile della programmazione, con l'espressione di tale volontà che rimane in attesa nella memoria del sistema fino al verificarsi del trigger prestabilito, cioè l'evento o la combinazione di condizioni definita ex ante che, una volta rilevata dal sistema, attiva l'azione programmata.

La questione della complessità e dell'affidabilità dei sistemi software rappresenta una sfida significativa nel contesto degli AWS. I sistemi software complessi presentano caratteristiche di imprevedibilità che rendono estremamente difficile, se non impossibile, testare esaustivamente ogni possibile scenario operativo

---

[127] McFarland, T., *Autonomous Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law*, Cambridge University Press, 2020, p. 33. L'autore spiega che "i sistemi di controllo basati su software utilizzati per ottenere un funzionamento autonomo negli AWS e in altre macchine sono essenzialmente computer specializzati che eseguono programmi che controllano la macchina al posto di un operatore umano". PDF disponibile su: https://www.cambridge.org/core/books/autonomous-weapon-systems-and-the-law-of-armed-conflict/09BFF6BB5B88E34935678B5A0606A8A7

[128] Ibidem, L'autore descrive come questo processo sia "incapsulato nel ben noto paradigma "*sense-think-act*" che è spesso usato come definizione operativa di un robot".

[129] Ivi, p. 34. L'autore chiarisce che "il fatto che anche programmi molto complessi sono solo insiemi di istruzioni predefinite è spesso oscurato nelle discussioni sui sistemi d'arma sofisticati".

con risorse di sviluppo ragionevoli e in tempi accettabili.[130] Questa limitazione intrinseca solleva interrogativi importanti sulla capacità di garantire un comportamento completamente prevedibile e conforme alle norme giuridiche in tutte le circostanze operative.

1.2. AWS e diritto internazionale umanitario: principi fondamentali

L'integrazione degli AWS nel quadro normativo del diritto internazionale umanitario presenta sfide interpretative significative che richiedono un'analisi approfondita dei principi fondamentali che governano la condotta delle ostilità. La valutazione giuridica di questi sistemi deve necessariamente partire dal riconoscimento che il diritto internazionale umanitario non presenta barriere legali insormontabili al loro sviluppo e utilizzo, purché vengano rispettate le restrizioni esistenti sulla natura e l'uso delle armi che si applicano agli AWS come a qualsiasi altro tipo di armamento.[131]

Il principio di distinzione, pietra angolare del diritto internazionale umanitario,[132] assume particolare rilevanza nel contesto dei sistemi d'arma autonomi. Questo principio richiede che le parti in conflitto distinguano costantemente tra civili e combattenti, dirigendo le operazioni militari esclusivamente contro obiettivi militari legittimi. L'implementazione di questo principio attraverso sistemi autonomi solleva questioni complesse relative alla capacità di tali sistemi di effettuare valutazioni contestuali accurate in ambienti operativi dinamici e spesso ambigui.

---

[130] Ivi, p. 35. L'autore chiarisce che "il fatto che anche programmi molto complessi sono solo insiemi di istruzioni predefinite è spesso oscurato nelle discussioni sui sistemi d'arma sofisticati" e che "nessun computer è in grado di scegliere da solo se eseguire o meno un programma memorizzato nella sua memoria, o di esercitare discrezione sull'esecuzione di una particolare istruzione all'interno di un programma".

[131] Ivi, p. 175. L'autore conclude che "nel complesso, il DIU non presenta barriere insormontabili al continuo sviluppo e uso degli AWS. Le restrizioni esistenti sulla natura e l'uso delle armi continuano ad applicarsi agli AWS come si applicano ad altri tipi di armi".

[132] Fondamento normativo del principio di distinzione: v. Protocollo addizionale I alle Convenzioni di Ginevra del 12 agosto 1949, relativo alla protezione delle vittime dei conflitti armati internazionali (adottato 8 giugno 1977, in vigore 7 dicembre 1978), artt. 48 ("regola di base"), 51(2) (divieto di rendere la popolazione civile oggetto di attacco) e 52(2) (definizione di "obiettivi militari"); testo ufficiale in ICRC, IHL Treaties Database: https://ihl-databases.icrc.org/en/ihl-treaties/api-1977. Per i conflitti armati non internazionali, v. Protocollo addizionale II (adottato 8 giugno 1977, in vigore 7 dicembre 1978), art. 13; testo in: https://ihl-databases.icrc.org/en/ihl-treaties/apii-1977. Per la natura consuetudinaria del principio, v. ICRC, Customary IHL Database, Regola 1 ("Distinzione tra civili e combattenti") e Regola 7 ("Distinzione tra beni civili e obiettivi militari"): https://ihl-databases.icrc.org/en/customary-ihl/v1/rule1 ; https://ihl-databases.icrc.org/en/customary-ihl/v1/rule7

Analogamente, il principio di proporzionalità presenta sfide specifiche quando applicato agli AWS. Questo principio proibisce attacchi che possano causare perdite di vite umane civili, ferimenti ai civili o danni ai beni di carattere civile che sarebbero eccessivi rispetto al vantaggio militare concreto e diretto previsto. La valutazione della proporzionalità richiede tradizionalmente un giudizio umano complesso che considera molteplici fattori contestuali, sollevando interrogativi sulla capacità dei sistemi autonomi di replicare tale processo decisionale con la necessaria sensibilità etica e giuridica.

1.3. Profili di responsabilità e *accountability*

Una delle sfide più complesse poste dagli AWS riguarda l'attribuzione della responsabilità per le azioni compiute da questi sistemi, particolarmente quando tali azioni risultano in violazioni del diritto internazionale umanitario. La dottrina della responsabilità superiore è stata proposta come possibile soluzione a questo dilemma, diventando una sorta di panacea per affrontare le "ansietà" legate al cosiddetto "gap di responsabilità".[133]

La responsabilità per le azioni di un AWS ricade necessariamente sugli esseri umani presenti nella catena decisionale, includendo comandanti militari, operatori diretti, programmatori del sistema e produttori dell'hardware e software.[134] Tuttavia, l'identificazione precisa del soggetto responsabile dipende dal contesto specifico dell'incidente e dal livello di autonomia del sistema al momento dell'azione controversa. Questa distribuzione della responsabilità lungo la catena di comando e sviluppo presenta sfide significative per l'applicazione pratica dei meccanismi di *accountability* esistenti.

La responsabilità statale rappresenta un altro aspetto cruciale di questa problematica. Gli stati che sviluppano, acquisiscono e impiegano i sistemi d'arma autonomi mantengono obblighi specifici sotto il diritto internazionale, inclusi quelli

[133] Spadaro, A., op. cit., p. 1120. L'autore evidenzia come "la dottrina della responsabilità superiore è stata proposta come soluzione a questo problema, diventando una panacea proposta per curare le ansietà legate a questo percepito "gap di responsabilità".

[134] Seixas-Nunes, A., op. cit., p. 23. L'autrice analizza la distinzione fondamentale tra algoritmi deterministici e algoritmi di machine learning, evidenziando che questi ultimi "sono in grado di apprendere, di essere adattivi, ai nuovi dati raccolti dall'ambiente, e sono progettati per replicare i processi decisionali umani".

derivanti dall' I Protocollo Addizionale alle Convenzioni di Ginevra del 1949, che richiede la revisione legale delle nuove armi.[135] Tuttavia, l'applicazione di questi obblighi agli AWS presenta complessità particolari legate alla natura dinamica e adattiva di questi sistemi.

1.4. Controllo umano significativo e supervisione

Il concetto di "controllo umano significativo" è emerso come elemento centrale nel dibattito sulla regolamentazione degli AWS. Questo principio richiede che, nonostante il grado di autonomia del sistema, gli esseri umani mantengano un livello appropriato di controllo e supervisione sulle decisioni critiche, particolarmente quelle relative alla selezione e all'ingaggio degli obiettivi.[136]
L'implementazione pratica del controllo umano significativo presenta sfide tecniche e operative considerevoli. I sistemi autonomi sono progettati per operare in ambienti complessi e dinamici dove la comunicazione diretta con operatori umani può essere limitata o impossibile. Tuttavia, l'introduzione di sistemi d'arma altamente autonomi solleva questioni cruciali per il sistema di giustizia penale internazionale, poiché una capacità di comportamento autonomo nei sistemi d'arma potrebbe compromettere la capacità della Corte Penale Internazionale di perseguire efficacemente i reati definiti nello Statuto di Roma. Il problema principale risiede nel fatto che gli elementi mentali richiesti per stabilire la responsabilità penale potrebbero essere più difficili da soddisfare quando il sistema d'arma utilizzato è capace di operare in modo altamente autonomo.[137]

La progettazione di AWS deve quindi incorporare meccanismi che garantiscano la possibilità di intervento umano quando necessario, mantenendo al

---

[135] *La via italiana all'intelligenza artificiale per scopi militari*, 16 Marzo 2021 IRPA ISTITUTO DI RICERCHE SULLA PUBBLICA AMMINISTRAZIONE. Disponibile su: https://www.irpa.eu/la-via-italiana-allintelligenza-artificiale-per-scopi-militari/

[136] Seixas-Nunes, A., op. cit., p. 24. L'autrice evidenzia che "gli algoritmi di machine learning definiranno i propri processi, sviluppandosi passo dopo passo quando confrontati con nuovi input, rendendo difficile, se non impossibile, prevedere con certezza i risultati" e solleva la questione cruciale: "potrebbero gli algoritmi di machine learning operare senza controllo/intervento umano?".

[137] McFarland, T., op. cit., p. 140. L'autore analizza come "una capacità di comportamento autonomo nei sistemi d'arma potrebbe minacciare la capacità della ICC di perseguire con successo i reati rilevanti definiti nello Statuto di Roma" e evidenzia che "il problema principale identificato è che gli elementi mentali richiesti potrebbero essere più difficili da soddisfare quando il sistema d'arma utilizzato è capace di operare in modo altamente autonomo".

contempo la capacità operativa del sistema in condizioni di autonomia. Un sistema quindi "*human-in-the-loop*" sembrerebbe il più efficace per bilanciare risultati e finalità operative con il problema dell'*accountability* e nell'evitare l'errore algoritmico.

1.5. Prospettive future e necessità di regolamentazione

L'evoluzione tecnologica degli AWS procede a ritmo accelerato, rendendo urgente lo sviluppo di framework normativi adeguati che possano governare il loro sviluppo e impiego. Tuttavia, gran parte del dibattito è diventato polarizzato e paralizzato da comprensioni contrastanti del significato di "autonomo" applicato ai sistemi d'arma.[138] Gli stati con interesse nel dibattito dovrebbero urgentemente cercare di stabilire una comprensione dell'autonomia delle armi che supporti un'analisi approfondita dei suoi effetti legali e di altro tipo, poiché non possono sperare di raggiungere un accordo su una risposta significativa senza prima concordare su una comprensione del cambiamento che deve essere affrontato.

Le discussioni in corso presso la United Nations Convention on Certain Conventional Weapons (CCW) riflettono la complessità di questo bilanciamento. Mentre alcuni stati sostengono la necessità di un divieto preventivo sugli AWS completamente autonomi, altri preferiscono approcci normativi più graduali che permettano lo sviluppo tecnologico entro parametri giuridicamente definiti.[139]

La sfida principale consiste nel trovare un equilibrio delicato tra l'innovazione tecnologica e i benefici potenziali degli AWS da un lato, e la protezione dei valori fondamentali del diritto internazionale umanitario dall'altro, per prevenire una corsa agli armamenti destabilizzante.[140]

---

[138] Ivi, p. 176. L'autore sottolinea che "la comunità internazionale sta discutendo attivamente la necessità di stabilire norme comuni e principi condivisi per gli AWS, con l'obiettivo di garantire che un controllo umano significativo sia mantenuto sulle funzioni critiche".

[139] Spadaro, A., op. cit., p. 1135. L'autore osserva che "le discussioni presso la CCW riflettono posizioni divergenti tra stati che sostengono approcci proibitivi e quelli che preferiscono regolamentazioni più graduali".

[140] McFarland, T., op. cit., p. 178. L'autore evidenzia "la necessità di trovare un equilibrio delicato tra l'innovazione tecnologica e i benefici potenziali degli AWS da un lato, e la protezione dei valori fondamentali del diritto internazionale umanitario dall'altro"

Questo equilibrio richiede un approccio multidisciplinare che coinvolga esperti tecnici, giuristi e decisori politici in un dialogo costruttivo orientato alla definizione di standard internazionali condivisi.

La necessità di sviluppare tecniche di testing del software e standard tecnici appropriati per garantire l'efficacia delle revisioni delle armi come mezzo per prevenire violazioni del DIU attraverso gli AWS rappresenta una priorità immediata.[141] Solo attraverso un approccio coordinato e multidisciplinare sarà possibile garantire che lo sviluppo degli AWS proceda in conformità con i principi fondamentali del diritto internazionale e con il rispetto della dignità umana.

Nonostante non ci sia stata una codificazione internazionale, ci sono state però occasioni dove la stessa ONU si è espressa adottando risoluzioni come la "Risoluzione L56" del 1° novembre 2023 ove la Prima Commissione, dedicata al Disarmo, dell'Assemblea Generale delle Nazioni Unite ha adottato la prima Risoluzione mai discussa sulle armi autonome sottolineando la "necessità urgente per la comunità internazionale di affrontare le sfide e le preoccupazioni sollevate dai sistemi d'arma autonomi."[142] La Risoluzione A/RES/78/241 (il documento ufficiale che formalmente adotta la Risoluzione L56 nell'ambito della 78ª sessione) sugli AWS afferma che le sfide e le preoccupazioni connesse a questi sistemi devono essere affrontate con urgenza considerando insieme profili umanitari giuridici di sicurezza tecnologici ed etici.[143] Per dare seguito concreto la risoluzione incarica il Segretario Generale di raccogliere in modo ordinato le opinioni degli Stati e di altri soggetti includendo una specifica riflessione sul ruolo degli esseri umani nelle decisioni sull'uso della forza e di presentare alla settantanovesima sessione una

[141] Ivi, op. cit., p. 175. L'autore nelle raccomandazioni afferma che "sarà necessario sviluppare tecniche di testing del software e standard tecnici appropriati per garantire l'efficacia delle revisioni delle armi come mezzo per prevenire violazioni del DIU attraverso gli AWS".

[142] "*Storica risoluzione Onu contro i sistemi d'arma autonomi. Un risultato storico ha portato all'approvazione da parte di 164 Stati di una risoluzione volta ad arginare l'utilizzo di sistemi d'arma autonomi*." Rete Italiana Pace e Disarmo, 09 Novembre, Micromega, 2023. Disponibile su: https://www.micromega.net/storica-risoluzione-onu-contro-i-sistemi-darma-autonomi

[143] Assemblea Generale delle Nazioni Unite *Lethal autonomous weapons systems* A/RES/78/241 22 dicembre 2023 paragrafo operativo 1 testo che richiama l'urgente necessità di affrontare le sfide e le preoccupazioni poste dagli AWS nelle prospettive umanitarie giuridiche di sicurezza tecnologiche ed etiche. Disponibile su https://docs.un.org/en/A/RES/78/241

relazione sostanziale corredata da un allegato con i contributi pervenuti.[144] La partecipazione è ampliata poiché vengono invitati organizzazioni internazionali e regionali il Comitato internazionale della Croce Rossa la società civile l'accademia e l'industria a trasmettere valutazioni che aiutino a delineare opzioni normative operative e di policy.[145] A garanzia della continuità istituzionale l'Assemblea iscrive infine la voce Lethal autonomous weapons systems all'ordine del giorno della settantanovesima sessione così che il dibattito sugli AWS prosegua su basi documentate e comparabili.[146]

Importanti anche le considerazioni delle organizzazioni umanitarie, Amnesty International e Human Rights Watch, due delle principali organizzazioni che si occupano di diritti umani, le quali hanno sollevato preoccupazioni significative sull'uso dei sistemi d'arma autonomi (AWS). Amnesty, attraverso le sue campagne come "Stop Killer Robots", ha evidenziato come questi sistemi possano prendere decisioni di vita o di morte senza il controllo umano diretto. L'organizzazione sottolinea che l'automazione nella guerra potrebbe portare a violazioni devastanti del diritto internazionale umanitario e dei diritti umani, in quanto i macchinari non sarebbero in grado di applicare gli stessi principi etici e legali che una persona umana può applicare in scenari complessi.[147]

Human Rights Watch, d'altra parte, ha criticato il fatto che alcune potenze militari, come Russia e Stati Uniti, stiano ostacolando gli sforzi per creare un trattato internazionale che regoli o vieti l'uso di armi autonome. Secondo l'organizzazione, questi sistemi potrebbero portare a decisioni arbitrarie e inaccettabili sulla vita e la morte, senza che vi sia una responsabilità chiara. Anche l'adozione di misure volontarie, come i "codici di condotta", non sono sufficienti per fermare la

[144] Ivi paragrafo operativo 2 mandato al Segretario Generale di sollecitare le opinioni degli Stati e degli altri soggetti anche sul ruolo degli esseri umani nell'uso della forza e di presentare alla settantanovesima sessione una relazione sostanziale con allegato contenente i contributi ricevuti.
[145] Ivi paragrafo operativo 3 invito rivolto a organizzazioni internazionali e regionali al Comitato internazionale della Croce Rossa alla società civile all'accademia e all'industria a trasmettere contributi al Segretario Generale ai fini della relazione.
[146] Ivi paragrafo operativo 4 iscrizione della voce *Lethal autonomous weapons systems* all'ordine del giorno della settantanovesima sessione per assicurare seguito istituzionale.
[147] Global: "*A critical opportunity to ban killer robots – while we still can"*, November 2, 2021. Disponibile su: https://www.amnesty.org/en/latest/news/2021/11/global-a-critical-opportunity-to-ban-killer-robots-while-we-still-can/

proliferazione di tali armi, che potrebbero espandere enormemente l'uso della forza senza supervisione umana.[148]

Le perplessità nell'utilizzo di questi sistemi in guerra rimane cruciale anche a fronte delle guerre in corso. Nel prossimi paragrafi infatti approfondiremo l'utilizzo di sistemi IA da parte di Israele in Palestina per comprendere meglio la portata di questo fenomeno degli AWS e non solo.

1.6. AI Act e sistemi d'arma autonomi: incidenza normativa a monte tra divieti biometrici, trasparenza e modelli di uso generale

Gli AWS non rientrano direttamente nel perimetro applicativo dell'AI Act quando l'impiego è esclusivamente militare o di difesa o di sicurezza nazionale poiché il regolamento delimita in modo espresso la propria sfera e preserva le competenze degli Stati su tali ambiti. Questa esclusione non rende però irrilevante il regolamento rispetto agli AWS. Una parte significativa delle capacità che rendono possibile l'automazione delle funzioni d'arma nasce nel settore civile e spesso si sviluppa in contesti di *dual use*. Componenti software moduli di percezione e architetture di *machine learning* maturano in mercati non militari e solo successivamente entrano nei programmi di ricerca nella fornitura e nella manutenzione dei sistemi destinati alla difesa. In questa prospettiva l'AI Act incide a monte e ai margini del perimetro militare innalzando standard minimi di liceità trasparenza e *governance* per i mattoni tecnologici che possono essere riutilizzati in applicazioni legate alla sicurezza e al conflitto armato. L'effetto non consiste nel dettare regole d'ingaggio bensì nel fissare condizioni di qualità e tracciabilità delle componenti su cui gli AWS fanno affidamento.[149]

---

[148] *Killer Robots: Military Powers Stymie Ban*, December 19, 2021. Disponibile su: https://www.hrw.org/news/2021/12/19/killer-robots-military-powers-stymie-ban#:~:text=%28Geneva%2C%20December%2020%2C%202021%29%20%E2%80%93%20Major%20military%20powers,new%20international%20treaty%2C%20Human%20Rights%20Watch%20said%20today.

[149] Parlamento europeo e Consiglio dell'Unione europea, Regolamento (UE) 2024/1689 del 13 giugno 2024, Artificial Intelligence Act, art. 2, parr. 3 e 4 che escludono dal campo di applicazione i sistemi di IA posti sul mercato messi in servizio o utilizzati esclusivamente per finalità militari di difesa o di sicurezza nazionale e precisano la non incidenza sulle competenze nazionali in materia di sicurezza nonché art. 2, par. 12 sull'*open-source* quando operano gli obblighi di cui agli artt. 5 o 50 o sussistono condizioni di alto rischio. Disponibile su: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Un primo profilo di incidenza riguarda il catalogo delle pratiche vietate. Il regolamento proibisce la creazione o l'espansione di banche dati per il riconoscimento facciale tramite *web scraping* non mirato di immagini reperite su internet o attraverso reti di videosorveglianza. La misura colpisce alla radice la costruzione di *dataset* biometrici di massa e riequilibra il rapporto tra innovazione e diritti fondamentali. È vietato l'impiego di sistemi di *emotion recognition* nei luoghi di lavoro e nelle istituzioni scolastiche salvo specifiche finalità mediche o di sicurezza e non è consentita la *biometric categorisation* che deduce categorie sensibili come origini etniche opinioni politiche credo religioso o filosofico, appartenenza sindacale, vita o orientamento sessuale. Per l'identificazione biometrica a distanza in tempo reale negli spazi pubblici a fini di *law enforcement* la regola è il divieto con limitate eccezioni tipizzate soggette ad autorizzazioni garanzie e vincoli temporali e territoriali. Anche se concepite per l'ambito civile e di polizia queste scelte riducono l'area di liceità dei dati e dei moduli su cui si costruiscono pipeline addestrative con ricadute sull'intero ecosistema da cui gli AWS possono attingere.[150]

Un secondo profilo riguarda la trasparenza nei passaggi informativi più delicati. L'AI Act richiede che chi impiega un sistema di intelligenza artificiale segnali l'interazione quando non è evidente e impone di informare gli interessati in caso di utilizzo di *emotion recognition* o *biometric categorisation*. Stabilisce inoltre l'obbligo di etichettare in modo riconoscibile i contenuti sintetici e i *deepfake* quando sono destinati al pubblico. Questo nucleo regolatorio presidia la circolazione di contenuti manipolati contrasta pratiche di *deception* tecnologica e rende più chiara la provenienza dei materiali informativi. Ne deriva una riduzione dello spazio per la disinformazione e per quelle forme di pressione cognitiva che possono incidere sulla selezione dei bersagli e sulle soglie di impiego della forza soprattutto in contesti ibridi e in fasi pre-conflittuali.[151]

[150] Ivi, art. 5, par. 1, lett. e divieto di creare o ampliare banche dati facciali mediante *web scraping* non mirato da internet o CCTV f divieto di *emotion recognition* in lavoro e scuola salvo fini medici o di sicurezza g divieto di *biometric categorisation* che deduce categorie sensibili h e parr. 2–6 uso di *real-time remote biometric identification* in spazi pubblici per *law enforcement* limitato a eccezioni tipizzate con autorizzazioni garanzie e limiti.

[151] Ivi, art. 50 obblighi di trasparenza con segnalazione dell'interazione con l'IA quando non evidente informativa su *emotion recognition* e *biometric categorisation* etichettatura riconoscibile dei contenuti sintetici e dei *deepfake* destinati all'informazione del pubblico e art. 3, punto 60 definizione di *deep fake*.

Il terzo profilo concerne i *general-purpose AI models* vale a dire i modelli di uso generale incorporabili in una pluralità di sistemi. Ai fornitori sono imposti obblighi di documentazione tecnica e di comunicazione verso gli operatori *downstream* che intendano integrare il modello nonché la pubblicazione di una sintesi adeguata dei dati usati per l'addestramento e una politica sul diritto d'autore. Quando un modello presenta *systemic risk* il quadro si fa più stringente con valutazioni del modello *adversarial testing* anche mediante prove indipendenti quando opportuno misure di mitigazione e monitoraggio segnalazione degli incidenti e garanzie rafforzate di *cybersecurity*. Il regolatore europeo promuove inoltre *codes of practice* per dare struttura e proporzionalità agli adempimenti. Poiché gli AWS attingono spesso a componenti di uso generale sviluppate nel mercato civile questo regime produce un effetto di "igiene regolatoria" lungo la catena di fornitura incrementando tracciabilità, qualità della *governance,* tecnica e verificabilità dei sistemi riutilizzati in ambito militare. La tempistica applicativa è scandita su più date che incidono sul momento in cui gli effetti a monte cominciano a farsi sentire. I capitoli generali inclusi i divieti si applicano dal due febbraio 2025. Le disposizioni sui *general-purpose AI models* e l'architettura di *governance* si applicano dal due agosto duemilaventicinque. L'applicazione complessiva decorre dal due agosto 2026 mentre la regola classificatoria di cui all'articolo sei paragrafo uno si applica dal due agosto 2027. Questa scansione consente agli operatori di adeguare processi e *workflow* prima che gli obblighi ricadano a pieno regime anche sui componenti suscettibili di interagire con lo sviluppo degli AWS.[152]

Nel loro insieme queste scelte non disciplinano l'impiego degli AWS in combattimento che resta affidato al diritto internazionale umanitario e alla normativa nazionale di settore. Agiscono però sul terreno in cui gli AWS prendono forma con effetti sul modo in cui i dati sono raccolti organizzati e verificati e sul modo in cui i moduli intelligenti vengono documentati e controllati. L'innalzamento degli standard a monte riduce la disponibilità di pratiche illecite o opache consolida la tracciabilità

[152] Ivi, Capo V sui *general-purpose AI models* artt. 51–52 classificazione e procedura per i modelli con *systemic risk* art. 53 documentazione tecnica informazioni ai soggetti *downstream* sintesi dei dati di addestramento politica sul diritto d'autore art. 55 *adversarial testing* mitigazione e monitoraggio dei rischi *incident reporting cybersecurity* e art. 113 sulla decorrenza applicativa con capitoli I e II dal 2 febbraio 2025 Capo V e *governance* dal 2 agosto 2025 applicazione complessiva dal 2 agosto 2026 e art. 6, par. 1 dal 2 agosto 2027.

della catena informativa e impone una cultura della prova e del controllo che tende a riflettersi anche oltre il perimetro civile. Per gli ordinamenti che intendono governare seriamente i rischi degli AWS questo approccio non sostituisce il confronto sulle regole d'impiego ma lo integra offrendo un metodo verificabile per incidere su ciò che alimenta la capacità autonoma di osservare decidere e agire.

2. Sistemi di riconoscimento facciale e armi autonome nel contesto bellico: la trasformazione digitale del conflitto armato in Palestina

Il conflitto israelo-palestinese ha assunto, negli ultimi anni, una dimensione tecnologica senza precedenti, caratterizzata dall'integrazione massiva di sistemi di intelligenza artificiale nelle operazioni militari. Questa trasformazione rappresenta un paradigma completamente nuovo nella conduzione della guerra, dove algoritmi di machine learning e sistemi autonomi hanno progressivamente sostituito o affiancato il processo decisionale umano in operazioni critiche come l'identificazione dei bersagli, la sorveglianza di massa e l'esecuzione di attacchi mirati. La peculiarità del teatro operativo palestinese, caratterizzato da un'elevata densità demografica e dalla compresenza di combattenti e civili in spazi urbani ristretti, ha accelerato lo sviluppo e l'implementazione di tecnologie di intelligenza artificiale sempre più sofisticate, trasformando questo conflitto in quello che alcuni analisti hanno definito "la prima guerra di intelligenza artificiale su larga scala".[153]

L'evoluzione tecnologica del conflitto israelo-palestinese non può essere compresa senza considerare il contesto geopolitico più ampio in cui si inserisce. La Striscia di Gaza, con i suoi 2,3 milioni di abitanti concentrati in appena 365 chilometri quadrati, rappresenta uno dei territori più densamente popolati al mondo, creando sfide operative uniche per le forze militari. In questo contesto, l'esercito israeliano ha sviluppato e implementato una serie di sistemi di intelligenza artificiale progettati per operare in ambienti urbani complessi, dove la distinzione tra combattenti e civili risulta particolarmente difficile. Questi sistemi, che spaziano dal

[153] Grundy-Warr, C. e Sidaway, J.D., "*Gaza: The first full-scale AI war?*", Political Geography, vol. 118, 2025, p. 1 del PDF. Gli autori analizzano come l'integrazione dell'intelligenza artificiale generativa nelle operazioni militari stia trasformando radicalmente la natura del conflitto israelo-palestinese, definendolo "the first full-scale AI war" e introducendo nuove dimensioni di automazione nel processo decisionale militare. Articolo disponibile su: https://www.sciencedirect.com/journal/political-geography/vol/118/suppl/C

riconoscimento facciale automatizzato alla selezione autonoma dei bersagli, rappresentano una rivoluzione nel modo in cui vengono condotte le operazioni militari moderne, sollevando al contempo questioni fondamentali riguardo al rispetto del diritto internazionale umanitario e alla protezione della popolazione civile.[154]

La documentazione disponibile rivela come l'implementazione di sistemi di intelligenza artificiale nel conflitto israelo-palestinese abbia seguito un percorso di sviluppo accelerato, particolarmente intensificatosi durante l'operazione "Spade di Ferro" iniziata il 7 ottobre 2023. Durante questa operazione, l'utilizzo di sistemi automatizzati per l'identificazione e l'eliminazione di presunti militanti di Hamas ha raggiunto livelli di sofisticazione e scala precedentemente inimmaginabili. I sistemi implementati non si limitano alla semplice automazione di processi esistenti, ma introducono capacità completamente nuove di analisi predittiva, riconoscimento di pattern comportamentali e *targeting* dinamico, trasformando radicalmente la natura stessa del conflitto armato.[155]

L'integrazione di tecnologie di intelligenza artificiale nei conflitti armati solleva questioni fondamentali riguardo alla natura stessa della guerra moderna e alle implicazioni per il diritto internazionale umanitario. La capacità di questi sistemi di operare con livelli di automazione precedentemente inimmaginabili ha creato nuove forme di *warfare* che sfidano i paradigmi tradizionali del controllo umano sulle operazioni militari. La velocità e la scala con cui questi sistemi possono processare informazioni e prendere decisioni operative rappresentano un salto qualitativo che va oltre la semplice evoluzione tecnologica, configurandosi come una vera e propria

[154] Amnesty International, "*Automated Apartheid: How Facial Recognition Fragments, Segregates and Controls Palestinians in the OPT*", 2022, pp. 6-10 del PDF. Il rapporto documenta come l'implementazione sistematica di tecnologie di riconoscimento facciale nei Territori Palestinesi Occupati abbia creato un sistema di controllo automatizzato che opera simultaneamente come strumento di *warfare* e meccanismo di controllo sociale, evidenziando le sfide operative uniche create dall'elevata densità demografica e dalla compresenza di combattenti e civili. PDF disponibile su: https://www.amnesty.org/en/documents/mde15/6701/2023/en/

[155] Gray, C.H., "*AI, Sacred Violence, and War—The Case of Gaza*", Crown College University of California Santa Cruz, CA, USA, 2024, pp. 12-15 del PDF. L'inchiesta rivela come l'operazione militare israeliana a Gaza abbia rappresentato un punto di svolta nell'utilizzo dell'intelligenza artificiale per scopi militari, con l'implementazione di sistemi automatizzati che hanno raggiunto livelli di sofisticazione precedentemente inimmaginabili durante l'operazione "Spade di Ferro", introducendo capacità completamente nuove di analisi predittiva e targeting dinamico. PDF disponibile su: https://link.springer.com/book/10.1007/978-3-031-81501-0

rivoluzione nel modo in cui vengono concepite e condotte le operazioni militari contemporanee.[156]

2.1. Red Wolf: il sistema di identificazione automatizzata nei checkpoint

Il sistema Red Wolf rappresenta una delle innovazioni più significative nell'ambito del controllo automatizzato dei territori palestinesi, configurandosi come un sistema di riconoscimento facciale progettato specificamente per l'identificazione automatica di individui nei checkpoint militari. Implementato nei principali punti di controllo tra Israele e i Territori Palestinesi Occupati, Red Wolf utilizza algoritmi di deep learning per analizzare in tempo reale i volti delle persone che attraversano i checkpoint, confrontandoli con un database contenente profili di presunti militanti, attivisti e individui considerati "persone di interesse" dalle autorità israeliane. Il sistema è particolarmente concentrato nel checkpoint 56 di Hebron, dove sono state installate 24 telecamere che effettuano acquisizione facciale automatica e controllano lo sblocco o il blocco dei tornelli di accesso.[157]

La sofisticazione tecnologica di questo sistema risiede nella sua capacità di operare in condizioni ambientali difficili, processando immagini di qualità variabile e mantenendo elevati livelli di accuratezza anche in presenza di tentativi di occultamento dell'identità. L'architettura tecnica di Red Wolf si basa su una rete neurale addestrata su un dataset di immagini facciali raccolte attraverso anni di sorveglianza sistematica nei territori palestinesi. Il sistema è in grado di processare identificazioni simultanee utilizzando algoritmi di matching biometrico che raggiungono elevati tassi di accuratezza, con la vera innovazione che risiede nella

[156] Seixas-Nunes, A., " *The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective* ", Cambridge University Press, 2024, pp. 103-107. L'autore analizza come l'integrazione di tecnologie di intelligenza artificiale nei conflitti armati sollevi questioni fondamentali riguardo alla natura della guerra moderna, evidenziando come la velocità e la scala con cui questi sistemi possono processare informazioni rappresentino un salto qualitativo che va oltre la semplice evoluzione tecnologica.

[157] Amnesty International, op. cit., pp. 44-46 del PDF. Il rapporto descrive dettagliatamente il sistema Red Wolf implementato nel checkpoint 56 di Hebron, dove sono state installate 24 telecamere che effettuano acquisizione facciale automatica e controllano lo sblocco o il blocco dei tornelli di accesso, utilizzando algoritmi di deep learning per l'identificazione automatica di individui considerati "persone di interesse".

sua integrazione con sistemi di intelligence più ampi che permettono di correlare l'identità degli individui con informazioni comportamentali, storiche e associative.[158]

Il database utilizzato da Red Wolf contiene profili biometrici di migliaia di palestinesi, inclusi attivisti politici, giornalisti, studenti universitari e membri di organizzazioni della società civile, molti dei quali non hanno mai commesso reati ma sono stati inclusi nel sistema sulla base di criteri opachi e potenzialmente discriminatori. La raccolta di questi dati biometrici è avvenuta attraverso controlli militari, perquisizioni domiciliari e operazioni di sorveglianza condotte nei Territori Palestinesi Occupati, sollevando questioni riguardo alla legalità della raccolta senza consenso informato. L'interazione di Red Wolf con i registri di Blue Wolf e Wolf Pack determina di fatto il pass o no pass ai tornelli, creando un sistema integrato di controllo automatizzato che opera secondo logiche algoritmiche predefinite.[159]

2.2. Blue Wolf: la sorveglianza biometrica diffusa

Blue Wolf rappresenta l'evoluzione del concetto di sorveglianza biometrica, estendendo le capacità di riconoscimento facciale oltre i checkpoint fissi per creare un sistema di monitoraggio diffuso che copre l'intero territorio dei Territori Palestinesi Occupati. A differenza di Red Wolf, che opera principalmente in punti di controllo specifici, Blue Wolf utilizza una rete distribuita di telecamere e sensori per creare una copertura di sorveglianza continua e pervasiva che permette di tracciare i movimenti degli individui in tempo reale attraverso tutto il territorio. Il sistema è progettato per identificare e tracciare individui mentre si muovono attraverso aree urbane, strade, mercati e spazi pubblici, creando un profilo dettagliato dei loro movimenti, associazioni e comportamenti che viene costantemente aggiornato e arricchito attraverso l'analisi automatizzata dei dati raccolti.[160]

---

[158] Ivi, pp. 42-45 del PDF. L'analisi tecnica evidenzia come l'architettura di Red Wolf si basa su algoritmi di riconoscimento facciale addestrati su immagini raccolte negli anni.

[159] Ivi, pp. 6–7 e 67 del PDF. A p. 6 il rapporto spiega che Red Wolf si basa su database composti esclusivamente da dati di persone palestinesi; a p. 7 descrive l'interoperabilità di Red Wolf, Blue Wolf e Wolf Pack, impiegati per determinare **la** possibilità o meno di attraversare quartieri e varchi (pass/no pass) in base alle autorizzazioni e alle voci presenti in archivio. A p. 67 si chiarisce che, quando un individuo non è riconosciuto dal sistema ai checkpoint, viene effettuata la cattura biometrica del volto e l'inserimento nel database senza fornire adeguate informazioni su uso e trattamento dei dati, con evidenti criticità rispetto al consenso informato.

[160] Ivi, pp. 42-43 del PDF. Blue Wolf rappresenta l'evoluzione del concetto di sorveglianza biometrica, estendendo le capacità oltre i checkpoint fissi per creare un sistema di monitoraggio diffuso che

L'implementazione operativa di Blue Wolf si basa su un'applicazione per smartphone che fornisce accesso immediato ai dati di Wolf Pack, permettendo ai soldati di acquisire foto di palestinesi e di accedere istantaneamente alle loro informazioni personali. Il sistema utilizza una dinamica gamificata che incentiva l'acquisizione massiva di volti attraverso un sistema di punteggi e ranking, trasformando la sorveglianza in una competizione tra soldati. Questa gamificazione ha portato a un aumento esponenziale della raccolta di dati biometrici, con soldati che competono per acquisire il maggior numero di foto e identificazioni possibili. Il sistema viene utilizzato anche per il mapping e la gestione di attraversamenti e arresti, permettendo di coordinare operazioni di controllo su larga scala.[161]

La capacità di *Blue Wolf* di operare come piattaforma di *watchlisting* in tempo reale costituisce uno degli aspetti più controversi del sistema, tramite l'app i soldati acquisiscono volti e interrogano immediatamente il database *Wolf Pack*, ottenendo indicazioni operative (ad es. segnalazioni di persone ricercate o da trattenere) e alert automatici al verificarsi di determinate corrispondenze.[162]

2.3. Wolf Pack: l'integrazione sistemica della sorveglianza

Wolf Pack rappresenta il culmine dell'evoluzione dei sistemi di sorveglianza biometrica israeliani, configurandosi come un vasto database contenente informazioni esclusivamente sui palestinesi, inclusi permessi, familiari, targhe automobilistiche e altre informazioni personali. Il sistema è accessibile attraverso war room centralizzate e oggi anche attraverso l'applicazione Blue Wolf, creando un ecosistema integrato di sorveglianza che permette l'accesso istantaneo a informazioni dettagliate su ogni palestinese nei Territori Occupati. Wolf Pack non è semplicemente

---

utilizza una rete distribuita di telecamere e sensori per tracciare movimenti degli individui in tempo reale attraverso tutto il territorio.

[161] Ivi, pp. 42-45 del PDF. L'implementazione operativa si basa su un'applicazione per smartphone che fornisce accesso immediato ai dati di Wolf Pack, utilizzando una dinamica "gamificata" che incentiva l'acquisizione massiva di volti attraverso un sistema di punteggi e ranking, trasformando la sorveglianza in una competizione tra soldati per il mapping e la gestione di attraversamenti e arresti.

[162] Ivi, pp. 42–46 del PDF. Amnesty descrive *Blue Wolf* come un'app in dotazione ai soldati per scattare foto ai palestinesi, associare dati identificativi e interrogarli in tempo reale nel database centrale *Wolf Pack*; l'app restituisce esiti operativi immediati (ad esempio matching con *watchlist* e istruzioni su fermo/ arresto) e prevede meccanismi "gamificati" che incentivano la raccolta massiva di volti. Nelle pagine indicate non vengono descritte funzioni "predittive" in senso stretto, ma piuttosto segnalazioni/alert basati su corrispondenze di banca dati.

la somma dei suoi componenti, ma rappresenta un salto qualitativo nella capacità di monitoraggio e controllo, fungendo da repository centrale per tutti i dati di sorveglianza raccolti attraverso diverse piattaforme e sistemi.[163]

L'architettura di Wolf Pack si basa su un grande database progettato specificamente per contenere profili completi della popolazione palestinese, con informazioni che spaziano dai dati biometrici alle associazioni familiari, dai permessi di movimento alle attività lavorative. Il sistema permette di creare profili multidimensionali degli individui monitorati che includono non solo dati identificativi, ma anche analisi comportamentali e predittive sui movimenti futuri. L'accesso a Wolf Pack attraverso dispositivi mobili ha rivoluzionato le operazioni di controllo sul territorio, permettendo ai soldati di accedere istantaneamente a informazioni dettagliate su qualsiasi palestinese incontrato durante le operazioni.[164]

3. I sistemi d'arma autonomi: Lavender, l'algoritmo della morte automatizzata

Il sistema Lavender rappresenta probabilmente l'innovazione più controversa e significativa nell'ambito dell'automazione del *targeting* militare, configurandosi come un algoritmo di intelligenza artificiale progettato specificamente per identificare e classificare automaticamente potenziali bersagli umani attraverso l'analisi di enormi quantità di dati di intelligence. Sviluppato dall'Unità 8200 dell'intelligence militare israeliana, Lavender utilizza tecniche avanzate di machine learning per processare e correlare informazioni provenienti da comunicazioni intercettate, dati di geolocalizzazione, associazioni sociali, attività online, pattern comportamentali e altre fonti di intelligence per creare profili di rischio utilizzati per decisioni di vita o di morte. Il sistema rappresenta un'evoluzione radicale nel *targeting* automatizzato, trasformando il processo di identificazione dei bersagli da un'attività manuale e selettiva basata su intelligence umana tradizionale a un sistema

[163] Ivi, pp. 42-43 del PDF. Wolf Pack si configura come un vasto database contenente informazioni esclusivamente sui palestinesi, inclusi permessi, familiari, targhe automobilistiche e "*wanted*", accessibile attraverso war room centralizzate e oggi anche attraverso l'applicazione Blue Wolf, fungendo da repository centrale per tutti i dati di sorveglianza raccolti.

[164] Ivi, pp. 42–43. Si precisa che Wolf Pack è un database che raccoglie immagini e dati identificativi dei palestinesi della Cisgiordania con lo scopo di conservare il profilo di ogni palestinese e che l'accesso, prima mediato dalla "war room", è oggi consultabile istantaneamente sul campo tramite l'app Blue Wolf.

automatizzato di produzione di massa di bersagli basato su correlazioni algoritmiche e pattern comportamentali.[165]

L'implementazione operativa di Lavender durante l'operazione "Spade di Ferro" ha segnato un punto di svolta nell'automazione del *warfare*, permettendo di generare liste di bersagli a una velocità e scala precedentemente inimmaginabili. Il sistema è stato progettato per operare con un livello di automazione che riduce drasticamente il ruolo dell'intervento umano nel processo di selezione dei bersagli, con operatori umani che spesso si limitano a confermare le raccomandazioni generate dal sistema senza condurre verifiche approfondite. Secondo le testimonianze disponibili, il tempo dedicato alla verifica umana delle raccomandazioni di Lavender è stato ridotto a pochi secondi per bersaglio, trasformando il ruolo degli operatori umani da decisori attivi a validatori passivi delle decisioni algoritmiche.[166]

La capacità di Lavender di processare simultaneamente migliaia di potenziali bersagli rappresenta un'innovazione significativa nel *targeting* automatizzato, ma solleva al contempo questioni fondamentali riguardo all'accuratezza e all'affidabilità delle classificazioni effettuate dal sistema. Secondo le testimonianze disponibili, il sistema opera con un margine di errore che solleva gravi preoccupazioni riguardo all'accuratezza delle classificazioni effettuate, con la possibilità che civili innocenti vengano erroneamente identificati come bersagli militari legittimi. Il processo operativo di Lavender prevede che gli algoritmi di machine learning analizzino continuamente flussi di dati provenienti da multiple fonti di intelligence per identificare individui che corrispondono a profili comportamentali considerati indicativi di appartenenza a organizzazioni militari.[167]

---

[165] Gray, C.H., op. cit., pp. 95–96 del PDF. L'autore descrive Lavender come il "*database smart*" programmato dall'Unità 8200 e impiegato con Gospel: raccoglie e correla flussi eterogenei di sorveglianza e comunicazioni per generare rapidamente bersagli, lavorando in sinergia con Where's Daddy per il tracciamento domestico.

[166] Amoroso D. S*istemi di supporto alle decisioni basati sull'IA e crimini di guerra: alcune riflessioni alla luce di una recente inchiesta giornalistica*, rivista di diritto internazionale, 2024, pp. 356-357. Sulla scorta di +972 e *The Guardian* ricostruisce l'impiego di *Lavender* con supervisione umana minima nelle prime settimane: agli operatori sarebbe stato chiesto di trattare le raccomandazioni del sistema come ordini, limitandosi a controlli formali rapidissimi (ad es. esclusione di donne), con frequenti attacchi eseguiti in abitazioni e impiego di ordigni non guidati ai livelli operativi più bassi.

[167] Ibid, p. 356. Evidenzia i rischi di errore su scala: *Lavender* avrebbe "marcato" circa 37.000 palestinesi come militanti, con conseguenti criticità per accuratezza e tutela dei civili.

Una volta identificato un potenziale bersaglio, il sistema genera automaticamente un "file bersaglio" che include informazioni dettagliate sull'individuo, inclusi dati personali, pattern comportamentali, associazioni sociali e una valutazione del rischio basata su algoritmi predittivi. Questi file vengono poi utilizzati per pianificare ed eseguire operazioni di eliminazione mirata, con il sistema che fornisce anche raccomandazioni sui tempi e le modalità ottimali per l'esecuzione degli attacchi. La velocità con cui Lavender è in grado di generare questi profili ha permesso un'escalation significativa nel numero di operazioni di *targeting* condotte durante l'operazione a Gaza.[168]

3.1. Gospel: la predizione comportamentale per il *targeting*

Il sistema Gospel rappresenta un'evoluzione significativa nel *targeting* predittivo, configurandosi come un algoritmo di intelligenza artificiale progettato per ottimizzare il timing e la modalità degli attacchi attraverso l'analisi predittiva dei comportamenti e dei movimenti dei bersagli identificati da Lavender. A differenza di Lavender, che si concentra principalmente sull'identificazione e classificazione dei potenziali bersagli, Gospel utilizza algoritmi avanzati di analisi comportamentale e modelli predittivi per determinare quando e dove i bersagli sono più vulnerabili e teoricamente isolati da potenziali vittime civili. L'architettura tecnologica di Gospel si basa su algoritmi di machine learning che analizzano pattern comportamentali storici, movimenti geospaziali, abitudini quotidiane e altri indicatori per creare modelli predittivi sui comportamenti futuri dei bersagli.[169]

Il sistema utilizza tecniche di analisi temporale e spaziale per identificare finestre di opportunità ottimali per l'esecuzione degli attacchi, considerando fattori come la probabilità di presenza del bersaglio in una determinata location, la densità di popolazione civile nell'area, le condizioni meteorologiche e altri parametri operativi che possono influenzare l'efficacia e la precisione dell'attacco. Gospel opera attraverso l'analisi di enormi quantità di dati comportamentali raccolti attraverso anni

[168] Gray, C. H., op. cit., p. 88 del PDF. Sull'integrazione *Lavender–Gospel* nella produzione/aggiornamento massivo di target e nella pianificazione operativa ("quando" e "come" colpire) in coordinamento con *Where's Daddy*.
[169] Ivi, p. 87 del PDF. Presenta Gospel come Decision Data System che aggrega indicatori operativi e tende a privilegiare bersagli "di edificio" rispetto a singole persone.

di sorveglianza sistematica della popolazione palestinese, utilizzando questi dati per creare modelli predittivi che vengono poi utilizzati per decisioni operative che possono avere conseguenze letali.[170]

La capacità predittiva di Gospel rappresenta uno degli aspetti più innovativi e controversi del sistema, poiché si basa su correlazioni statistiche e pattern algoritmici che potrebbero non corrispondere necessariamente a intenzioni reali o che potrebbero essere influenzati da bias discriminatori incorporati nei dati di addestramento. Il sistema è progettato per identificare pattern ricorrenti nei comportamenti dei bersagli, come orari di movimento, luoghi frequentati, associazioni sociali e altre variabili comportamentali che vengono utilizzate per predire i movimenti futuri e identificare momenti di vulnerabilità. Questa capacità predittiva permette di pianificare attacchi con maggiore precisione temporale, ma solleva questioni etiche fondamentali riguardo all'utilizzo di algoritmi probabilistici per decisioni di vita o di morte.[171]

L'implementazione operativa di Gospel ha permesso di ottimizzare significativamente l'efficacia degli attacchi mirati, riducendo i tempi di pianificazione e aumentando la probabilità di successo delle operazioni. Tuttavia, l'utilizzo di algoritmi predittivi per decisioni di vita o di morte solleva questioni etiche e legali fondamentali, poiché tali decisioni vengono basate su probabilità e correlazioni statistiche piuttosto che su prove concrete di attività militari o di minacce immediate alla sicurezza. Questa delega di decisioni critiche a sistemi algoritmici che operano secondo logiche probabilistiche crea situazioni in cui individui possono essere eliminati sulla base di previsioni comportamentali che potrebbero essere errate, incomplete o discriminatorie.[172]

3.2. Where's Daddy: il *targeting* familiare automatizzato

Il sistema Where's Daddy rappresenta forse l'aspetto più controverso e moralmente problematico dell'automazione del *targeting* militare israeliano,

[170] Ivi, p. 88 del PDF. Spiega la componente predittiva e il ruolo di Where's Daddy nel localizzare i soggetti nelle abitazioni per sfruttare finestre temporali di vulnerabilità.
[171] Ivi, p. 87 del PDF. Richiama le criticità strutturali del *targeting* algoritmico (automation bias, scenario fulfillment, "misplaced concreteness", "AI alibi" rispetto al diritto internazionale).
[172] Ivi, pp. 95–96 del PDF. L'autore evidenzia che il ricorso a modelli predittivi in Gospel fornisce una copertura politico-operativa alle scelte di *targeting,* un vero "alibi di IA"; segnala i rischi strutturali del metodo probabilistico e gli effetti di automation bias, scenario fulfillment e "misplaced concreteness", che mettono in crisi le pretese di precisione e sollevano rilevanti profili etici e giuridici.

configurandosi come un algoritmo progettato specificamente per tracciare i movimenti dei bersagli identificati da Lavender e determinare quando si trovano nelle loro abitazioni familiari. Il nome stesso del sistema, "Where's Daddy", rivela esplicitamente la sua funzione: identificare quando i "papà", termine utilizzato dall'esercito israeliano per riferirsi ai bersagli maschili, si trovano a casa con le loro famiglie, momento considerato ottimale per gli attacchi nonostante la presenza di civili, ma proprio per garantire l'eliminazione certa del bersaglio utilizzando la presenza di familiari come garanzia che il bersaglio non possa sfuggire all'attacco.[173]

La logica operativa sottostante a Where's Daddy evidenzia una strategia militare che prioritizza deliberatamente l'eliminazione certa del bersaglio rispetto alla protezione dei civili, utilizzando la presenza di familiari come garanzia operativa piuttosto che come fattore deterrente. Questo approccio rappresenta una deviazione significativa dai principi tradizionali del *targeting* mirato, che storicamente hanno enfatizzato la necessità di minimizzare i danni collaterali e di proteggere la popolazione civile, sostituendo questi principi con una logica operativa che considera accettabile e persino desiderabile la presenza di civili durante gli attacchi per garantire l'efficacia dell'eliminazione del bersaglio.[174]

L'implementazione operativa di Where's Daddy si basa su algoritmi di tracking che utilizzano dati di geolocalizzazione, pattern di movimento, comunicazioni intercettate e altre fonti di intelligence per monitorare continuamente i movimenti dei bersagli e identificare quando si trovano nelle loro abitazioni. Il sistema è progettato per generare automaticamente alert quando i bersagli vengono identificati nelle loro case, attivando procedure operative che portano alla pianificazione ed esecuzione di attacchi che coinvolgono deliberatamente l'intera struttura abitativa, garantendo così l'eliminazione del bersaglio ma causando inevitabilmente vittime civili tra i familiari.[175]

---

[173] Ivi, p. 95 del PDF. È precisato che Where's Daddy è concepito per seguire i combattenti fino alle abitazioni, individuando quando si trovano a casa per colpirli insieme ai familiari e ai vicini; il sistema opera in combinazione con Gospel e Lavender.

[174] Ivi, pp. 95–96 del PDF. Si sottolinea una prassi operativa che privilegia l'eliminazione certa del bersaglio rispetto alla protezione dei civili; si ricordano attacchi condotti nelle case e l'uso di ordigni poco selettivi, con considerazioni critiche sul significato stesso del nome "Where's Daddy".

[175] Ivi, p. 95 del PDF. È descritta l'integrazione tra Gospel, Lavender e Where's Daddy: i primi due sistemi generano e gestiscono i dati di intelligence, il terzo consente di identificare le finestre temporali di presenza domestica del bersaglio, funzionali alla pianificazione dell'attacco.

L'utilizzo di Where's Daddy ha portato a un aumento significativo del numero di vittime civili durante l'operazione a Gaza, con particolare impatto su donne e bambini che si trovavano nelle abitazioni al momento degli attacchi. Questa strategia operativa solleva gravi preoccupazioni riguardo al rispetto del principio di distinzione, che richiede alle parti in conflitto di distinguere sempre tra combattenti e civili, e del principio di proporzionalità, che proibisce attacchi che possano causare danni civili eccessivi rispetto al vantaggio militare concreto e diretto previsto. La documentazione disponibile indica che Where's Daddy è stato utilizzato sistematicamente durante l'operazione "Spade di Ferro", con migliaia di attacchi condotti contro abitazioni familiari basandosi sulle raccomandazioni generate dal sistema.[176]

4. Implicazioni giuridiche e prospettive future: la sfida al diritto internazionale umanitario

L'implementazione massiva di sistemi di intelligenza artificiale nel conflitto israelo-palestinese, ha posto sfide inedite al diritto internazionale umanitario, mettendo alla prova la capacità delle norme esistenti, sviluppate in un'epoca pre-digitale, di regolare forme di warfare caratterizzate da automazione e delega decisionale a sistemi algoritmici. La natura automatizzata di sistemi come Lavender, Gospel e Where's Daddy solleva questioni fondamentali riguardo all'applicabilità dei principi tradizionali del diritto dei conflitti armati, poiché tali principi furono concepiti assumendo che le decisioni critiche riguardo alla conduzione delle ostilità sarebbero sempre state prese da esseri umani capaci di valutazioni etiche, contestuali e proporzionate.[177]

Il principio di distinzione, pilastro fondamentale del diritto internazionale umanitario, viene messo alla prova da algoritmi che operano sulla base di

[176] Ivi, p. 94 del PDF. Viene richiamata la nozione di "target factory" e l'incremento su scala industriale del numero di obiettivi e siti colpiti nei primi giorni dell'operazione; l'autore parla di "mass assassination factory", con effetti dirompenti sulle famiglie, a riscontro del marcato aumento delle vittime civili.

[177] Amoroso, D., "Sistemi di supporto alle decisioni basati sull'IA e crimini di guerra", Rivista di Diritto Internazionale, 2024, pp. 1-3 del PDF. L'autore analizza come l'implementazione massiva di sistemi di intelligenza artificiale abbia posto sfide inedite al diritto internazionale umanitario, mettendo alla prova norme sviluppate in epoca pre-digitale per regolare forme di warfare caratterizzate da automazione e delega decisionale a sistemi algoritmici.

correlazioni statistiche e pattern comportamentali che non necessariamente corrispondono a una partecipazione diretta alle ostilità. La questione della responsabilità penale individuale per crimini di guerra commessi attraverso sistemi automatizzati rappresenta una delle sfide più complesse per il diritto internazionale penale, poiché la catena causale tra decisione umana e risultato letale viene mediata da algoritmi che possono operare in modi imprevisti o non intenzionali. Il principio di proporzionalità, che richiede che i danni civili attesi non siano eccessivi rispetto al vantaggio militare concreto e diretto previsto, diventa particolarmente problematico quando applicato a sistemi automatizzati che effettuano calcoli di proporzionalità basati su parametri predefiniti e modelli matematici.[178]

4.1. Il problema del controllo umano significativo

La questione del controllo umano significativo rappresenta uno dei nodi centrali del dibattito giuridico sui sistemi d'arma autonomi. Il concetto di "meaningful human control" è emerso come punto di convergenza nelle discussioni internazionali, tuttavia i dettagli specifici del concetto sono stati lasciati così aperti da favorire conversazioni e accordi senza raggiungere una definizione operativa chiara. La diversità di interpretazioni del concetto di controllo umano significativo da parte di stati, sviluppatori, produttori, avvocati militari, attivisti per i diritti umani e studiosi rende cruciale raggiungere una comprensione condivisa del concetto.[179]

Il dibattito sul controllo umano significativo non fornisce una soluzione alle questioni tecniche, legali, morali e regolatorie che i sistemi d'arma autonomi pongono. La sfida principale consiste nello stabilire cosa costituisca precisamente un "controllo umano significativo" e come questo possa essere implementato in sistemi che operano a velocità e scale che superano le capacità umane di supervisione diretta. La questione diventa particolarmente complessa quando si considerano sistemi come

[178] Ivi, pp. 20-22 del PDF. Il principio di distinzione viene messo alla prova da algoritmi che operano su correlazioni statistiche che non necessariamente corrispondono a partecipazione diretta alle ostilità, con la questione della responsabilità penale che rappresenta una delle sfide più complesse, mentre il principio di proporzionalità diventa problematico quando applicato a sistemi automatizzati.

[179] Seixas-Nunes, A., "Do AWS Require a Distinct Level of Human Control?", Cambridge University Press, 2024, pp. 192-195. L'autore evidenzia come il concetto di "meaningful human control" sia emerso come punto di convergenza nelle discussioni internazionali, tuttavia i dettagli specifici del concetto sono stati lasciati così aperti da favorire conversazioni senza raggiungere una definizione operativa chiara.

Lavender, Gospel e Where's Daddy, che operano con livelli di automazione che riducono drasticamente il ruolo dell'intervento umano nel processo decisionale.[180]

4.2. Responsabilità e *accountability*

La questione della responsabilità per l'impiego di sistemi d'arma autonomi rappresenta una delle preoccupazioni principali espresse dalla maggioranza degli stati parte alla Convention on Certain Conventional Weapons.[181] Il rischio di un possibile "gap di responsabilità" è uno degli argomenti principali di coloro che sostengono la necessità di un "freno di emergenza" sullo sviluppo e l'acquisizione di sistemi d'arma autonomi. La preoccupazione è che situazioni in cui nessuno possa essere ritenuto responsabile se sistemi completamente autonomi vengono utilizzati in violazione del diritto internazionale possano avere conseguenze molto serie ed erodere i progressi sostanziali che sono stati raggiunti in questo settore negli ultimi anni.[182]

Tuttavia, l'argomento che con l'introduzione di sistemi d'arma autonomi non ci sia controllo umano è semplicemente inaccurato. I sistemi d'arma autonomi saranno progettati da operatori umani e programmati per compiere una missione militare prestabilita senza richiedere intervento umano. Questa innovazione, questa

[180] Ivi, pp. 196-198. Il dibattito sul controllo umano significativo non fornisce una soluzione alle questioni tecniche, legali, morali e regolatorie che i sistemi d'arma autonomi pongono, con la sfida principale che consiste nello stabilire cosa costituisca precisamente un "controllo umano significativo" in sistemi che operano a velocità e scale che superano le capacità umane.

[181] Assemblea degli Stati Parte alla Convenzione sulle proibizioni o restrizioni all'uso di alcune armi convenzionali (CCW), *Testo della Convenzione* come emendato il 21 dicembre 2001. Preambolo (richiamo ai principi per cui il diritto delle parti a scegliere mezzi e metodi di guerra non è illimitato; divieto di armi di natura tale da causare lesioni superflue o sofferenze inutili; protezione della popolazione civile; rinvio ai principi del diritto internazionale umanitario inclusa la clausola di Martens); art. 1 (*campo di applicazione*, esteso anche ai conflitti armati non internazionali in seguito all'emendamento del 2001); art. 2 (*rapporti con gli altri obblighi di DIU*, che rimangono salvi); art. 6 (*diffusione e istruzione militare*, rilevante per la catena di comando e i doveri interni di conformità). Nel quadro della medesima Convenzione, vedi anche GGE su AWS, Rapporto 2019, Allegato IV, Guiding Principles: lett. (b) (*la responsabilità umana per le decisioni sull'uso di armi deve essere mantenuta, l'accountability non si trasferisce alle macchine*), (d) (*catena di comando umana responsabile*), (e) (*determinazione preventiva sulla liceità dei nuovi mezzi/metodi di guerra*). Disponibili su https://geneva-s3.unoda.org/static-unoda-site/pages/templates/the-convention-on-certain-conventional-weapons/CCW%2Btext.pdf; https://treaties.un.org/doc/source/RecentTexts/26_2_cEng.pdf; https://documents.unoda.org/wp-content/uploads/2020/09/CCW_GGE.1_2019_3_E.pdf

[182] Ivi pp. 191-192. La questione della responsabilità per l'impiego di sistemi d'arma autonomi rappresenta una delle preoccupazioni principali espresse dalla maggioranza degli stati parte alla CCW, con il rischio di un possibile "gap di responsabilità" che è uno degli argomenti principali di coloro che sostengono la necessità di un "freno di emergenza" sullo sviluppo di tali sistemi.

possibilità di schierare sistemi d'arma capaci di operare senza intervento umano, è alla radice di tutti i problemi emergenti che assillano la ricerca e l'ulteriore sviluppo di sistemi d'arma autonomi. L'avvento di sistemi d'arma in grado di operare senza supervisione umana è possibile perché gli algoritmi di machine learning sono ora una realtà.[183]

4.3. Prospettive future e regolamentazione

L'utilizzo di tecnologie satellitari avanzate combinate con algoritmi di intelligenza artificiale per il monitoraggio rappresenta un'evoluzione significativa nelle capacità di intelligence e assessment, permettendo di valutare in tempo reale l'impatto degli attacchi e di adattare le strategie operative basandosi su dati oggettivi piuttosto che su stime approssimative. L'analisi satellitare combinata con l'intelligenza artificiale offre capacità senza precedenti di monitoraggio delle crisi umanitarie, permettendo una valutazione più accurata dell'impatto dei conflitti sulla popolazione civile e sulle infrastrutture critiche.[184]

La realtà operativa dei sistemi di intelligenza artificiale implementati nel conflitto israelo-palestinese evidenzia l'urgente necessità di sviluppare nuovi framework giuridici internazionali specificamente progettati per regolare l'utilizzo di tali tecnologie nei conflitti armati. La comunità internazionale deve affrontare la questione della regolamentazione delle armi autonome e dei sistemi di *targeting* automatizzato per garantire il rispetto del diritto internazionale umanitario e la protezione dei civili nei conflitti armati. L'assenza di un quadro normativo chiaro rischia di portare a un'escalation nell'utilizzo di queste tecnologie, con conseguenze

[183] Ivi, pp. 265-266. L'autore conclude che l'argomento secondo cui con l'introduzione di sistemi d'arma autonomi non ci sia controllo umano è inaccurato, poiché i sistemi saranno progettati da operatori umani e programmati per compiere missioni militari prestabilite, con l'avvento di tali sistemi reso possibile dal fatto che gli algoritmi di machine learning sono ora una realtà.

[184] Zhao, Q. et al., "Satellite and AI monitoring humanitarian crises in the Gaza Strip during the early stage of Israeli–Palestinian conflict", International Journal of Digital Earth, vol. 17, n. 1, 2024, pp. 2-16 del PDF. Lo studio documenta come l'utilizzo di tecnologie satellitari avanzate combinate con algoritmi di intelligenza artificiale rappresenti un'evoluzione significativa nelle capacità di intelligence e assessment, offrendo capacità senza precedenti di monitoraggio delle crisi umanitarie. PDF disponibile su: https://www.tandfonline.com/doi/pdf/10.1080/17538947.2024.2430678

potenzialmente devastanti per la popolazione civile e per la stabilità internazionale.[185]

Il fatto che sia ora noto che “roboticisti” e ingegneri avranno presto la capacità di progettare e costruire sistemi d'arma autonomi che possono "pensare per se stessi" ha preparato il terreno per una massa travolgente di speculazioni sullo status, le capacità e la legittimità dei sistemi d'arma autonomi sotto il framework del diritto internazionale umanitario. È impossibile escludere dall'orizzonte situazioni in cui il sistema potrebbe non riuscire a operare correttamente, risultando in esiti che violano le regolamentazioni del diritto internazionale umanitario. Tuttavia, è imperativo che qualsiasi soluzione dipenda sempre dalla buona volontà delle parti intervenenti in un conflitto armato, ma in nessun modo dovrebbero essere compromesse le regole di trasparenza.[186]

La tecnologia bellica non è infatti “neutrale”, incorpora scelte su dati, metriche di performance, soglie d’ingaggio e regole d’uso definite da progettisti e comandanti. Per questo l’architettura dei sistemi andrebbe vincolata, fin dalla fase di progettazione, a obblighi coerenti con i principi DIU (distinzione, proporzionalità, precauzione) tramite requisiti verificabili di *data governance* (provenienza, qualità, tracciabilità), test di bias *ex ante*, *red-teaming* periodico e clausole di *human-in/on-the-loop* tradotte nelle *rules of engagement*. In assenza di questi ancoraggi, l’automazione rischia di cristallizzare scelte opache difficili da correggere in tempo reale.[187]

---

[185] Seixas-Nunes, A., "AWS and the IHL Requirements", Cambridge University Press, 2024, pp. 140-142. L'analisi evidenzia come la realtà operativa dei sistemi di intelligenza artificiale dimostri l'urgente necessità di sviluppare nuovi framework giuridici internazionali per regolare l'utilizzo di tali tecnologie nei conflitti armati, con la comunità internazionale che deve affrontare la questione della regolamentazione per garantire il rispetto del diritto internazionale umanitario.

[186] Seixas-Nunes, A., "Final Conclusion", op. cit., pp. 266-267. L'autore conclude che è impossibile escludere situazioni in cui i sistemi potrebbero non riuscire a operare correttamente, risultando in esiti che violano le regolamentazioni del diritto internazionale umanitario, ma è imperativo che qualsiasi soluzione dipenda dalla buona volontà delle parti in conflitto senza compromettere le regole di trasparenza.

[187] Seixas-Nunes, A., *The Legality and Accountability of Autonomous Weapon Systems. A Humanitarian Law Perspective*, Cambridge–New York, Cambridge University Press, 2022, cap. 4 (“AWS and the IHL Requirements”), spec. pp. 140–141, 159 e 167: sull’obbligo di *weapons review* ex art. 36 API, sulla necessità di piani di verifica e validazione tecnico-operativa prima dell’impiego e sui doveri di precauzione del comando, inclusi meccanismi di controllo umano e possibilità di intervento/kill-switch lungo la condotta dell’attacco.

Nel contesto di Gaza, la rapida convergenza fra IA, sorveglianza e *targeting* automatizzato, dentro un ambiente urbano densissimo e altamente conflittuale, accresce il rischio di *lock-in* tecnologico e di normalizzazione di pratiche eccezionali. Questo richiede una cornice di governance sostanziale (limiti chiari, supervisione indipendente, audit esterni) capace di prevenire effetti strutturali sui civili e sulla stabilità regionale, evitando che l'eccezione operativa diventi regola.[188]

La questione centrale non è se il *machine learning* "rompa" la catena del controllo umano, ma come assicurare uno standard di giudizio umano tracciabile sulle decisioni d'ingaggio. Ciò implica strumenti concreti di *accountability*: registri decisionali e *event data recorder* per ogni attacco, soglie minime di intervento umano, *fail-safe/kill-switch* procedurali, audit tecnici forensi e obblighi di *after-action review*. Così si riducono i *responsibility gaps* quando i sistemi operano a velocità o scala che eccedono la supervisione ordinaria; in breve, non deve essere la tecnologia a dettare le regole, ma regole chiare a vincolare la tecnologia.[189]

CONCLUSIONI

Prospettive e riflessioni

Nel presente elaborato si è mostrato come l'Europa abbia scelto una via di regolamentazione che lega i principi a obblighi verificabili, lungo tutto il ciclo di vita dei sistemi di IA. La logica per livelli di rischio, la richiesta di trasparenza sostanziale, la tracciabilità dei passaggi tecnici e decisionali e la presenza di un controllo umano effettivo, sono gli strumenti da utilizzare per riportare la tecnologia entro il perimetro della legalità. Così operando non si intralcerebbe l'innovazione, bensì le si farebbe assumere la forma di un impegno pubblico alla spiegazione e alla

[188] Grundy-Warr, C., Sidaway, J.D., "*Gaza: The first full-scale AI war?*", *Political Geography*, 118 (2025), 103289, *Viewpoint*, spec. pp. 1–2 del PDF: quadro interpretativo di Gaza come laboratorio di guerra con IA, con integrazione su larga scala di sistemi di targeting automatizzato (*Lavender*, *Gospel*, *Where's Daddy?*), accelerazione del ciclo colpire-valutare e ampliamento dei bersagli in ambienti urbani densissimi, con rilevanti ricadute su distinzione civile/combattente, legalità e diritti.

[189] Seixas-Nunes, A., *The Legality and Accountability of Autonomous Weapon Systems. A Humanitarian Law Perspective*, cit., cap. 5 ("Accountability and Liability for the Deployment of AWS"), spec. pp. 202–203 e 221–222: distinzione tra responsabilità individuale e responsabilità dello Stato, discussione del "*responsibility gap*" connesso a esiti imprevedibili del *machine learning* e centralità di ruoli, tracciabilità e doveri di diligenza per assicurare accountability effettiva lungo il ciclo decisionale.

responsabilità. È il filo che tiene insieme i capitoli dedicati alla disciplina dell'Unione e alla cornice convenzionale europea, e che consente di leggere i casi concreti, senza smarrire il nesso tra norma e organizzazione.

Su queste basi, una parte del lavoro ha riguardato l'uso dell'intelligenza artificiale nel contesto bellico e nei territori occupati. L'integrazione tra sensori diffusi, raccolte di dati eterogenee, piattaforme di fusione informativa e interfacce che suggeriscono priorità e sequenze operative, hanno compresso i tempi dell'azione e spostato il baricentro del giudizio. Come è stato analizzato, nei teatri urbani mediorientali la letteratura giornalistica e d'inchiesta ha documentato sistemi capaci di classificare persone e luoghi, generare liste di obiettivi e mantenere un flusso costante di ingaggi. È un cambiamento che incide sulla distinzione tra sorveglianza mirata e sorveglianza generalizzata e che mette alla prova i principi di legalità, necessità e proporzionalità, soprattutto quando i processi diventano opachi, e il ruolo umano si riduce a una validazione rapida a valle del suggerimento della macchina.[190]

Il parallelismo con altri scenari conferma la tendenza di fondo. Anche nel conflitto ucraino si è osservata la saldatura tra capacità militari e infrastrutture digitali civili, con piattaforme di analisi e servizi di rete che entrano nella catena informativa e operativa. Questo non significa replicare meccanicamente contesti diversi. Significa riconoscere che l'innesto tra industria tecnologica e apparati della difesa produce dipendenze infrastrutturali, privatizza funzioni critiche e sposta porzioni di discrezionalità in ambienti che non sempre sono trasparenti per il decisore pubblico. Se la legittimazione dell'uso della forza si fonda sulla ricostruibilità delle scelte, la dispersione della responsabilità lungo filiere proprietarie diventa un problema giuridico e non soltanto etico.[191]

Il cuore della questione è il rapporto tra la velocità di esecuzione della macchina e la responsabilità umana delle conseguenze che ne derivano. Una delle promesse che si fanno sui sistemi di IA per il supporto alle decisioni è quello di avere

[190] VALORI.IT, redazione, L'intelligenza artificiale va al fronte. Un dossier di Valori.it, s.l., s.d., pp. 14–17: su Gaza, sistemi "Lavender" e "The Gospel", differenza tra marcatura di persone e di edifici, uso di "Where is daddy/Where's Daddy" e riduzione del ruolo umano in convalida, con soglie di danno collaterale riportate dalle inchieste citate. PDF disponibile su: https://valori.it/wp-content/uploads/2024/04/Lintelligenza-artificiale-va-al-fronte.pdf

[191] Ivi, pp. 9–12; 24–27: su "Ucraina laboratorio", ruolo di piattaforme civili e fornitori privati nella catena informativa e operativa, dipendenze infrastrutturali e privatizzazione di capacità critiche, con riferimenti a Palantir e alle iniziative Nat

maggiore consapevolezza situazionale e azioni più rapidi sul campo. Il diritto umanitario chiede invece che le decisioni con effetti sulla vita e sull'integrità di persone protette, siano prese sulla base di tutto ciò che è ragionevolmente esigibile in termini di verifica e di precauzione. La riflessione umanitaria più attenta, suggerisce che il tempo è parte della protezione e non un suo antagonista. Ridurre il ritmo quando la situazione lo impone, consente di valutare più approfonditamente le conseguenze, di confrontare fonti, di costruire alternative e di preservare un margine reale per il giudizio umano. Il controllo umano significativo ha contenuto solo se attraversa il progetto e l'impiego del sistema e se attribuisce poteri concreti di arresto, in presenza di incertezza o di discrepanze informative.[192]

Un secondo tema riguarda la qualità della conoscenza prodotta da sistemi complessi. Quando più livelli di analisi si accumulano su dati incompleti o sbilanciati, l'errore tende a propagarsi lungo la catena e a trasformarsi in decisione. In condizioni di tempo ridotto, cresce l'affidamento all'output e diminuisce la ricerca di problematiche nell'input, questo può portare alla normalità operativa nel commettere errori di classificazione di persone o di luoghi. L'unico "antidoto" giuridicamente praticabile sarebbe l'obbligo di controverifica tra fonti indipendenti, prima della traduzione di una raccomandazione in azione. La documentazione non è burocrazia. È la condizione minima di *accountability* e di rimedio effettivo.[193]

Nei contesti di occupazione militare la convergenza tra sorveglianza biometrica, analisi predittiva e *targeting* esercita una pressione continua sui diritti civili. L'identificazione diffusa, il monitoraggio dei movimenti e l'attribuzione di profili di rischio in base al comportamento umano predittivo ridisegnano la soglia della libertà. La guerra contemporanea spesso trasforma i luoghi di conflitto in veri e propri laboratori all'aperto, dove si sperimentano nuove tecnologie e strategie. Le soluzioni testate in questi contesti vengono poi migliorate e riutilizzate in altri ambiti,

[192] STEWART, Ruben; HINDS, Georgia, Algorithms of war. The use of artificial intelligence in decision making in armed conflict, ICRC Humanitarian Law & Policy Blog, pp. 3–4: su "tempo" operativo, "tactical patience" e necessità che il controllo umano resti effettivo lungo l'intera catena decisionale, con obblighi di vedere comprendere e sviluppare alternative prima dell'impiego della forza. PDF disponibile su: https://blogs.icrc.org/law-and-policy/wp-content/uploads/sites/102/2023/10/algorithms-of-war-use-of-artificial-intelligence-decision-making-armed-conflict.pdf

[193] Ivi, p. 2: su propagazione degli errori tra livelli di analisi, difficoltà di verifica dell'output, "*automation bias*" e necessità di controverifica fra fonti prima della traduzione in azione.

anche al di fuori del settore militare. Questo processo porta alla crescente presenza di attori privati nello sviluppo di capacità strategiche e rende sempre più difficile distinguere tra applicazioni civili e militari. Occorrerebbe quindi sviluppare un monitoraggio pubblico e privato con una valenza reale che possa condizionare l'utilizzo di questi sistemi anche da parte degli Stati. Senza questo monitoraggio, il controllo umano scivolerebbe verso una dimensione prettamente formale e la responsabilità si disperderebbe lungo catene di decisioni automatiche difficili da governare. Solo traducendo i principi di legalità che conosciamo con l'utilizzo di questi nuovi sistemi potremo salvaguardare la sicurezza e la difesa per la salvaguardia di tutti.

Tre condizioni risultano decisive perché l'impiego dell'intelligenza artificiale in guerra e nei contesti di occupazione resti governabile sul piano giuridico e concreto. La prima è un controllo umano progettato e misurabile, che consenta di stabilire in anticipo i compiti del sistema, le soglie di intervento e i poteri di arresto realmente esercitabili nel tempo operativo. La seconda è la qualità del dato intesa come dovere di eguaglianza, con origine verificabile, rappresentatività adeguata e confronto sistematico tra fonti prima che una raccomandazione diventi azione. La terza è la piena tracciabilità della decisione, con registri che rendano ricostruibile la logica dell'azione e aprano la strada a responsabilità effettive e a rimedi per i danni ai civili.

Il quadro europeo offre un metodo utile per consolidare queste tre condizioni. Le regole poste a tutela dei diritti fondamentali e della sicurezza informativa non sostituiscono le norme del diritto dei conflitti armati, ma innalzano gli standard di liceità nella raccolta dei dati, nella progettazione dei moduli intelligenti e nella documentazione delle scelte. In questo modo si riduce lo spazio per pratiche opache, si favorisce la verificabilità delle componenti critiche e si rende più chiara la catena delle responsabilità anche quando tecnologie nate in ambito civile confluiscono in sistemi destinati a impieghi militari.

Resta necessario un impegno pratico coerente con questi obiettivi. La progettazione deve prevedere meccanismi di arresto in sicurezza quando emergono errori o incertezze rilevanti. Le verifiche tecniche devono essere periodiche e, quando opportuno, indipendenti, con prove mirate a mettere in evidenza vulnerabilità

e comportamenti imprevisti. Occorrono registri di tracciabilità completi, conservati e accessibili alle autorità competenti, e valutazioni giuridiche preventive sull'uso di nuovi mezzi e metodi di guerra secondo le migliori prassi internazionali. Anche le procedure di approvvigionamento pubblico possono contribuire, imponendo clausole su dati leciti, registrazione degli eventi e possibilità di controllo.

In conclusione, l'evoluzione tecnologica non va respinta ma incanalata entro confini verificabili. Quando il controllo umano è realmente esigibile, quando i dati sono corretti e non discriminano e quando ogni passaggio lascia traccia, la tecnologia resta strumento e il diritto conserva la propria funzione ordinante anche quando la velocità delle operazioni tende a comprimere il tempo e il giudizio.

BIBLIOGRAFIA

Monografie e tesi

ALIC, Dalia, *The Role of Data Protection and Cybersecurity Regulations in Artificial Intelligence Global Governance: A Comparative Analysis of the European Union, the United States, and China Regulatory Framework*, tesi di laurea magistrale, Central European University, Vienna, 2021. Disponibile su: https://www.etd.ceu.edu/2021/alic_dalia.pdf.

FORGÓ, Nikolaus; PEHLIVAN, Ceyhun Necati; VALCKE, Peggy (a cura di), *The EU Artificial Intelligence (AI) Act: Commentary*, 2024.

GRAY, Chris H., *AI, Sacred Violence, and War—The Case of Gaza*, Cham: Springer, 2025. Disponibile su: https://link.springer.com/book/10.1007/978-3-031-81501-0.

McFARLAND, Tim, *Autonomous Weapon Systems and the Law of Armed Conflict: Compatibility with International Humanitarian Law*, Cambridge: Cambridge University Press, 2020. Disponibile su: https://www.cambridge.org/core/books/autonomous-weapon-systems-and-the-law-of-armed-conflict/09BFF6BB5B88E34935678B5A0606A8A7.

SEIXAS-NUNES, António, *The Legality and Accountability of Autonomous Weapon Systems: A Humanitarian Law Perspective*, Cambridge: Cambridge University Press, 2022. Disponibile su: https://www.cambridge.org/core/books/legality-and-

accountability-of-autonomous-weapon-systems/FE880FD3F459B29A495D79D0C8347D79.

Contributi in opere collettive

JASSERAND, Céline, «Facial Recognition in Public Spaces and the Principle of Necessity», in: N. Menéndez González; G. Mobilio (a cura di), *Next Democratic Frontiers for Facial Recognition Technology (FRT)*, Cham: Springer, 2025, pp. 49–69. Disponibile su: https://link.springer.com/content/pdf/10.1007/978-3-031-89794-8_4.

LEVANTINO, Francesco Paolo, «From Identity to Emotional Dominance? "Early Warnings" on Emotion Recognition Uses by Police Forces», in: N. Menéndez González; G. Mobilio (a cura di), *Next Democratic Frontiers for Facial Recognition Technology (FRT)*, Cham: Springer, 2025, pp. 129–157. Disponibile su: https://link.springer.com/content/pdf/10.1007/978-3-031-89794-8_8.

MORAWSKA, Ewa H., «Council of Europe standards and activities related to AI: towards a Framework Convention on AI and human rights?», in: *Artificial Intelligence and International Human Rights Law*, Cheltenham: Edward Elgar, 2024, pp. 24–44. Disponibile su: https://www.elgaronline.com/edcollchap-oa/book/9781035337934/book-part-9781035337934-9.xml.

PAOLUCCI, Francesca, «Enhancing Oversight and Addressing Gaps: Assessing the Impact of the AI Act on Biometric Identification Systems», in: N. Menéndez González; G. Mobilio (a cura di), *Next Democratic Frontiers for Facial Recognition Technology (FRT)*, Cham: Springer, 2025, pp. 71–89. Disponibile su: https://link.springer.com/chapter/10.1007/978-3-031-89794-8_5.

SELWYN, Neil; ANDREJEVIC, Mark; O'NEILL, Catherine; GU, Xiaoqing; SMITH, Greg, «Facial Recognition Technology: Key Issues and Emerging Concerns», in: R. Matulionyte; M. Zalnieriute (a cura di), *The Cambridge Handbook of Facial Recognition in the Modern State*, Cambridge: Cambridge University Press, 2024, pp. 11–28. PDF OA: https://researchmgt.monash.edu/ws/portalfiles/portal/592017706/583251749_oa.pdf.

Articoli di periodici

AMOROSO, Daniele, «Sistemi di supporto alle decisioni basati sull'IA e crimini di guerra: alcune riflessioni alla luce di una recente inchiesta giornalistica», in: *Diritti umani e diritto internazionale*, 18(2), 2024, pp. 347–368.

ATABEKOV, Arseny, «Artificial Intelligence in Contemporary Societies: Legal Status and Definition, Implementation in Public Sector across Various Countries», in: *Social Sciences*, 12(3), 2023, Art. 178. Disponibile su: https://www.mdpi.com/2076-0760/12/3/178.

DEVANY, Brenna E., «Clearview AI's First Amendment: A Dangerous Reality?», in: *Texas Law Review*, 101(2), 2022, pp. 473–507. Disponibile su: https://texaslawreview.org/wp-content/uploads/2023/01/Devany.Printer2-8.pdf.

DUL, Cristina, «Facial Recognition Technology vs Privacy: The Case of Clearview AI», in: *Queen Mary Law Journal*, 3, 2022, pp. 1–24. Disponibile su: https://www.qmul.ac.uk/law/research/journals/the-queen-mary-law-journal/media/law/docs/research/2022QMLJ1.pdf.

Ebers M. Regolamentazione dell'intelligenza artificiale veramente basata sul rischio: come attuare la legge sull'intelligenza artificiale dell'UE. *Giornale europeo della regolamentazione del rischio*. 2025; 16(2):684-703. DOI:10.1017/err.2024.78. Disponibile su: https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/truly-riskbased-regulation-of-artificial-intelligence-how-to-implement-the-eus-ai-act/E526C1D0D7368F9691082220609D60F4.

GABRIELLI, Giulia, «The Use of Facial Recognition Technologies in the Context of Peaceful Protest: The Risk of Mass Surveillance Practices and the Implications for the Protection of Human Rights», in: *European Journal of Risk Regulation*, 16, 2025, pp. 514–541. Disponibile su: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/A4B2FABA8F32DDBC0217C86837CDBAC6/S1867299X25000261a.pdf.

GRUNDY-WARR, Carl; SIDAWAY, James D., «Gaza: The first full-scale AI war?», in: *Political Geography*, 118, 2025, suppl. (online first). Disponibile su: https://www.sciencedirect.com/journal/political-geography/vol/118/suppl/C.

KOHN, Matthew, «Clearview AI, TikTok, and the Collection of Facial Images in

International Law», in: *Chicago Journal of International Law*, 23(1), 2022, pp. 195–234. Disponibile su: https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1832&context=cjil .

MARCHANT, Gary E.; GUTIERREZ, Carlos Ignacio, «Soft Law 2.0: An Agile and Effective Governance Approach for Artificial Intelligence», in: *Minnesota Journal of Law, Science & Technology*, 24(2), 2023, pp. 376–424. Disponibile su: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4473812.

NGUYEN, Quoc Viet; LAFRANCE, Stéphane; VU, Cong Tri, «China's Social Credit System. A Challenge to Human Rights», in: *The Law, State and Telecommunications Review*, 15(2), 2023, pp. 99–116. Disponibile su: https://pdfs.semanticscholar.org/b4ed/781a2f58e2a115197b7a0cef60d25c41c814.pdf.

PALMIOTTO, Francesca; MENÉNDEZ GONZÁLEZ, Nicolás, «Facial Recognition Technology, Democracy and Human Rights», in: *Computer Law & Security Review*, 56, 2025, Art. 105000. Disponibile su: https://www.sciencedirect.com/science/article/pii/S0267364923000675.

PRESNO, Manuel; MEUWESE, Anne, «Regulating AI from Europe: a joint analysis of the AI Act and the Framework Convention on AI», in: *Theory and Practice of Legislation*, 13(1), 2025, pp. 1–20. Disponibile su: https://www.tandfonline.com/doi/pdf/10.1080/20508840.2025.2492524.

Rezende, I. N. (2020). Riconoscimento facciale nelle mani della polizia: valutazione del "caso Clearview" da una prospettiva europea. *Nuova rivista di diritto penale europeo*, *11*(3), 375-389. https://doi.org/10.1177/2032284420948161 (Opera originale pubblicata nel 2020). Disponibile su: https://journals.sagepub.com/doi/epub/10.1177/2032284420948161.

ROSENZWEIG, Ido; PACHOLSKA, Marta, «The use of facial recognition for *targeting* under international law», in: *International Review of the Red Cross*, 107(928), 2025, pp. 238–255. Disponibile su: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/E8F47BBC8987E7E505C9152A3AADBCDE/S1816383124000705a.pdf.

SAKUBU, Deus, «Challenges of Artificial Intelligence today and future implications

for society and the world», in: *World Journal of Advanced Research and Reviews*, 26(1), 2025, pp. 3045–3054. DOI: 10.30574/wjarr.2025.26.1.1380. Disponibile su: https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-2025-1380.pdf.
SIMMLER, Manuel; CANOVA, Giulia, «Facial recognition technology in law enforcement: Regulating data analysis of another kind», in: *Computer Law & Security Review*, 56, 2025, Art. 106092. Disponibile su: https://www.sciencedirect.com/science/article/pii/S0267364924001572.
Veale, Michael e Zuiderveen Borgesius, Frederik, Demistificare il progetto di legge sull'intelligenza artificiale dell'UE (31 luglio 2021). Computer Law Review International (2021) 22(4) 97-112. Disponibile su: https://ssrn.com/abstract=3896852
ZALNIERIUTE, Monika, «International Decisions: Glukhin v. Russia», in: *American Journal of International Law*, 117(4), 2023, pp. 695–701. Disponibile su: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/AF9C0AAFA6E44CE9F1881B0542177D3C/S0002930023000520a.pdf.
ZHAO, Qian; et al., «Satellite and AI monitoring humanitarian crises in the Gaza Strip during the early stage of Israeli–Palestinian conflict», in: *International Journal of Digital Earth*, 17(1), 2024, pp. 1–15. Disponibile su: https://www.tandfonline.com/doi/pdf/10.1080/17538947.2024.2430678.

Documenti ufficiali e atti normativi

Agenzia dell'Unione europea per i diritti fondamentali (FRA), *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, 2019. PDF: https://staging.fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf.
Assemblea generale delle Nazioni Unite, *Lethal autonomous weapons systems*, Ris. A/RES/78/241, 22 dicembre 2023. Disponibile su: https://docs.un.org/en/A/RES/78/241.
Commissione europea, *European approach to artificial intelligence* (sintesi). Disponibile su: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.
Commissione europea, *White Paper on Artificial Intelligence – A European*

approach to excellence and trust, 2020. PDF: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065.

EDPB; EDPS, *Parere congiunto 5/2021 sulla proposta di regolamento sull'intelligenza artificiale*, 18 giugno 2021. PDF: https://www.edpb.europa.eu/system/files/2021-10/edpb-edps_joint_opinion_ai_regulation_it.pdf.

Consiglio d'Europa, *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law* (CETS n. 225), Vilnius, 5 settembre 2024. Disponibile su: https://rm.coe.int/1680afae3c.

Consiglio d'Europa, *Explanatory Report* alla Framework Convention on AI, 2024. Disponibile su: https://rm.coe.int/1680afae67.

Consiglio d'Europa, *Council of Europe opens first ever global treaty on AI for signature* (comunicato stampa), 2024. Disponibile su: https://www.coe.int/en/web/portal/-/council-of-europe-opens-first-ever-global-treaty-on-ai-for-signature.

German Federal Government, *Artificial Intelligence Strategy (update 2020)*, 2020. PDF: https://www.ki-strategie-deutschland.de/files/downloads/Nationale_KI-Strategie_engl.pdf.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST), *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, 2023. PDF: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

OCSE, *Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449)*, 2019 (aggiornata 2024). Scheda in italiano: https://www.biodiritto.org/AI-Legal-Atlas/AI-Docs/OCSE-Raccomandazione-sull-intelligenza-artificiale-principi-per-la-gestione-responsabile-di-una-AI-affidabile-e-raccomandazioni-agli-Stati-aderenti.

Parlamento europeo; Consiglio dell'Unione europea, *Regolamento (UE) 2024/1689 del 13 giugno 2024 (Artificial Intelligence Act)*. Disponibile su: https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

Parlamento europeo; Consiglio dell'Unione europea, *Regolamento (UE) 2016/679 (GDPR)*. Disponibile su: https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32016R0679.

STATE COUNCIL OF THE PEOPLE'S REPUBLIC OF CHINA, *New Generation Artificial Intelligence Development Plan*, 2017. (zh) http://www.gov.cn/zhengce/content/2017-07/20/content_5211990.htm; (en) https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf.

National Governance Committee for the New Generation Artificial Intelligence (Cina), *New Generation Artificial Intelligence Ethics Norms*, 2021. (zh) http://www.gov.cn/zhengce/content/2021-09/26/content_5639459.htm; (en) https://cset.georgetown.edu/publication/ethical-norms-for-new-generation-artificial-intelligence-released/.

UNITED NATIONS – CCW, *Convention on Certain Conventional Weapons* (testo consolidato). PDF: https://geneva-s3.unoda.org/static-unoda-site/pages/templates/the-convention-on-certain-conventional-weapons/CCW%2Btext.pdf — CCW GGE su AWS, *Rapporto 2019*, Allegato IV, *Guiding Principles*. PDF: https://documents.unoda.org/wp-content/uploads/2020/09/CCW_GGE.1_2019_3_E.pdf.

THE WHITE HOUSE, *Executive Order 14110 – Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, *Federal Register*, 88(211), 1 novembre 2023, p. 75192. Disponibile su: https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

Giurisprudenza

Corte di giustizia dell'Unione europea, 4 luglio 2000, *Laboratoires Pharmaceutiques Bergaderm SA e Goupil c. Commissione*, causa C-352/98 P. Disponibile su: https://curia.europa.eu/juris/document/document.jsf?docid=45097&doclang=IT.

Corte di giustizia dell'Unione europea, 16 luglio 2020, *Data Protection Commissioner c. Facebook Ireland Ltd e Maximillian Schrems (Schrems II)*, causa C-311/18, ECLI:EU:C:2020:559. Disponibile su: https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:62018CJ0311.

Corte europea dei diritti dell'uomo, *Use of facial-recognition technology breached rights of Moscow underground protestor (Glukhin v. Russia)*, Comunicato stampa

ECHR 207, 4 luglio 2023. PDF: https://hudoc.echr.coe.int/app/conversion/pdf/?library=ECHR&id=003-7694109-10618091.

Rapporti, blog e altre risorse online

Amnesty International, «A critical opportunity to ban killer robots – while we still can», 2 novembre 2021. Disponibile su: https://www.amnesty.org/en/latest/news/2021/11/global-a-critical-opportunity-to-ban-killer-robots-while-we-still-can/.

Amnesty International, *Automated Apartheid: How Facial Recognition Fragments, Segregates and Controls Palestinians in the OPT*, 2022. Disponibile su: https://www.amnesty.org/en/documents/mde15/6701/2023/en/.

AI4People – FLORIDI, Luciano et al., *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, 2018. PDF: https://ai4people.org/PDF/AI4People_Ethical_Framework_For_A_Good_AI_Society.pdf.

Forbes (Kate O'Flaherty), «Clearview AI, The Company Whose Database Has Amassed 3 Billion Photos, Hacked», 26 febbraio 2020. Disponibile su: https://www.forbes.com/sites/kateoflahertyuk/2020/02/26/clearview-ai-the-company-whose-database-has-amassed-3-billion-photos-hacked/.

Human Rights Watch, «Killer Robots: Military Powers Stymie Ban», 19 dicembre 2021. Disponibile su: https://www.hrw.org/news/2021/12/19/killer-robots-military-powers-stymie-ban.

IRPA – Istituto di Ricerche sulla Pubblica Amministrazione, «La via italiana all'intelligenza artificiale per scopi militari», 16 marzo 2021. Disponibile su: https://www.irpa.eu/la-via-italiana-allintelligenza-artificiale-per-scopi-militari/.

Micromega / Rete Italiana Pace e Disarmo, «Storica risoluzione Onu contro i sistemi d'arma autonomi», 9 novembre 2023. Disponibile su: https://www.micromega.net/storica-risoluzione-onu-contro-i-sistemi-darma-autonomi.

Politecnico di Torino Magazine, «Nuove regole per l'Intelligenza Artificiale: una

sfida europea», 16 dicembre 2021.
STEWART, Ruben; HINDS, Georgia, «Algorithms of war. The use of artificial intelligence in decision making in armed conflict», *ICRC Humanitarian Law & Policy Blog*, 2023. PDF: https://blogs.icrc.org/law-and-policy/wp-content/uploads/sites/102/2023/10/algorithms-of-war-use-of-artificial-intelligence-decision-making-armed-conflict.pdf.
VALORI.it (Redazione), *L'intelligenza artificiale va al fronte. Un dossier di Valori.it*, s.l., s.d. PDF: https://valori.it/wp-content/uploads/2024/04/Lintelligenza-artificiale-va-al-fronte.pdf.
ZOU, Mimi; ZHANG, Linxi, «Navigating China's regulatory approach to generative artificial intelligence and large language models», *Cambridge Forum on AI: Law & Governance*, 2025. Disponibile su: https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/navigating-chinas-regulatory-approach-to-generative-artificial-intelligence-and-large-language-models/969B2055997BF42DE693B7A1A1B4E8BA.