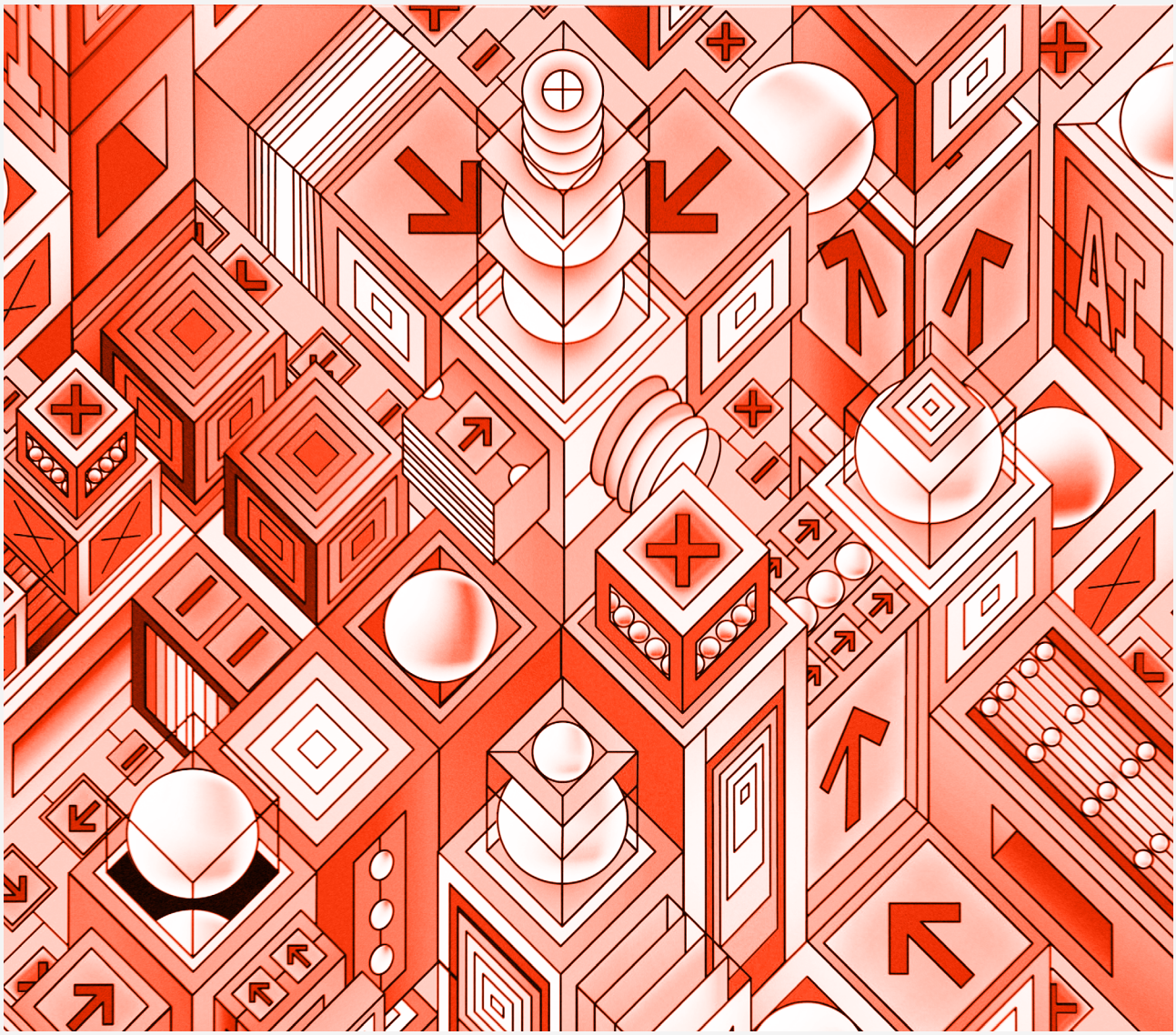
I RISCHI DELL'INTELLIGENZA ARTIFICIALE

Analisi e raccomandazioni per l'applicazione del regolamento europeo sull'intelligenza artificiale (AI Act)



A CURA DI: **PHILIP DI SALVO**
SUPERVISIONE E COORDINAMENTO: **ANTONELLA NAPOLITANO**
CON IL CONTRIBUTO DI: **CLAUDIO AGOSTI, LAURA CARRER, LAURA FERRARI, DAVIDE DEL MONTE, MARTINA TUROLA**

COLLABORATORI:
HERMES HACKING FOR HUMAN RIGHTS
THE good LOBBY

Questo policy paper è parte di un progetto di Hermes Center e The Good Lobby Italia ed è stato redatto da Philip Di Salvo con la supervisione delle due organizzazioni. Le raccomandazioni finali sono a cura di Hermes Center e The Good Lobby Italia. La sezione 5.3 del capitolo sull'AI ACT dedicata all'Autorità nazionale e le relative raccomandazioni finali sono state sviluppate a partire dal documento "Motivazioni e Caratteristiche dell'Autorità Nazionale per l'Intelligenza Artificiale", realizzato da Hermes Center, The Good Lobby e Privacy Network e pubblicato a marzo 2024.

Il policy paper è stato realizzato nel quadro del progetto "The CARE - Civil Actors for Rights and Empowerment" finanziato da ActionAid International Italia E.T.S e Fondazione Realizza il Cambiamento nell'ambito del progetto "The CARE - Civil Actors for Rights and Empowerment" cofinanziato dall'Unione Europea.

MAGGIO 2024

REALIZZATO NELL'AMBITO DI:

The care

FINANZIATORI

actionaid
REALIZZA IL CAMBIAMENTO

FONDAZIONE REALIZZA IL CAMBIAMENTO

Cofinanziato dall'Unione europea

INDEX

1. OBIETTIVI E PRESENTAZIONE

Il recente passaggio e l'entrata in vigore dell'Artificial Intelligence Act da parte dell'Unione Europea delineano un passaggio storico e politico cruciale a livello europeo.

Il primo tentativo al mondo di regolamentare l'intelligenza artificiale è il risultato di un complesso lavoro di mediazione, trattativa e bilanciamento tra diversi attori: come società civile, l'esperienza pregressa in campo di politiche tecnologiche insegna che spesso sono proprio gli spazi lasciati aperti dalla mediazione politica a rendere possibili potenziali abusi non facilmente riscontrabili immediatamente nei testi di legge[1].

Obiettivo di questo testo è quindi quello di analizzare questi potenziali vuoti normativi e portare nel dibattito nazionale una particolare attenzione sui rischi intrinsechi per i diritti umani e la data *justice*[2] nelle tecnologie di IA, per poter indirizzare le decisioni future in materia con la consapevolezza di quelle che sono le potenzialità, ma anche e soprattutto i limiti e i rischi, di questa dirompente tecnologia.

1. https://algorithmwatch.org/en/ai-act-deal-key-safeguards-and-dangerous-loopholes/
2. https://us.sagepub.com/en-us/nam/data-justice/book271599

L'AI ACT SI BASA SULLA DISTINZIONE DI LIVELLI DI RISCHIO. MA SE IL 2023 È STATO L'ANNO IN CUI IL CONCETTO DI "RISCHIO" È STATO SPESSO ASSOCIATO ALL'INTELLIGENZA ARTIFICIALE, SPESSO SE NE È PARLATO IN RELAZIONE AI COSIDDETTI "RISCHI ESISTENZIALI", LA CUI NATURA È AL MOMENTO ANCORA PER LO PIÙ SPECULATIVA[3].

Al contrario, siamo convinti che esistano già dei ben documentati[4] rischi espliciti, diretti e tangibili e dei pericoli immediati in termini di discriminazione, equità e disuguaglianze che i sistemi di IA potrebbero amplificare e replicare su larga scala, automatizzandoli[5]. Si tratta certamente di una questione politica e di regolamentazione, che l'applicazione dell'AI Act non potrà non tenere in considerazione, ma anche di narrazione mediatica e di costruzione sociale delle tecnologie[6]. È quindi fondamentale ragionare su come l'entrata in vigore dell'AI Act debba rappresentare un momento di attivazione di tutti i soggetti coinvolti, nei confronti di quelli che oggi sono i più evidenti ed immediati pericoli connessi a un futuro condiviso con macchine in grado di prendere decisioni con effetti significativi sulla vita delle persone.

Il testo si apre con una introduzione del tema, seguita da una sua contestualizzazione nel dibattito italiano. Affronterà quindi alcuni degli aspetti principali relativi alla sorveglianza di massa e come questi si colleghino ad alcuni degli aspetti chiave delle tecnologie regolamentate dall'AI Act.

3. https://www.technologyreview.com/2023/06/12/1074449/real-ai-risks/
4. https://yalebooks.yale.edu/book/9780300264630/atlas-of-ai/
5. https://us.macmillan.com/books/9781250074317/automatinginequality
6. https://thehumanerrorproject.ch/journalism-ai-futures-narratives/

L'analisi si concentrerà poi su alcuni aspetti dell'AI Act che consideriamo particolarmente problematici e dei rischi che pongono: ci occuperemo in particolare di riconoscimento biometrico negli spazi pubblici, utilizzo dell'intelligenza artificiale ai confini e in ambito di fenomeni migratori, autorità nazionale competente, General Purpose AI Models (GPAI) e sandbox. Concluderemo infine il documento con una serie di raccomandazioni per ciascuna delle aree analizzate.

2. INTRODUZIONE

L'accelerazione nello sviluppo di sistemi di intelligenza artificiale (IA) pone senza dubbio questioni urgenti nel dibattito contemporaneo sugli impatti sociali e politici delle tecnologie. In particolare, il lancio e la rapida diffusione su larga scala di sistemi di IA generativa come ChatGPT ha portato questi temi al centro dell'attenzione di soggetti appartenenti a diversi ambiti, non necessariamente tecnici, ora costretti a fare i conti con una tecnologia il cui impatto potenziale non ha eguali negli ultimi due decenni, o addirittura dalla nascita del World Wide Web a inizio anni '90.

Diverse analisi pubblicate da centri di ricerca e da società di consulenza ci permettono di avere un'idea, per quanto vaga e tutta da confermare, del possibile impatto economico e sociale a breve termine, derivato dalla diffusione dell'AI.

S&P Global Market Intelligence[7] prevede un'importante crescita del mercato del software di intelligenza artificiale generativa, stimando che il fatturato del settore possa arrivare a 36 miliardi di dollari entro il 2028.

7

7. "Generative AI Software Market Forecast to Expand Near 10 Times by 2028 to $36 Billion", https://press.spglobal.com/2023-06-08-Generative-AI-Software-Market-Forecast-to-Expand-Near-10-Times-by-2028-to-36-Billion,-S-P-Global-Market-Intelligence-Says

ANCHE MCKINSEY HA PUBBLICATO UN SUO PRIMO RAPPORTO SULL'IA[8]: SECONDO IL GIGANTE MONDIALE DELLA CONSULENZA, LA DIFFUSIONE DI AI GENERATIVA E ALTRE TECNOLOGIE ARRIVERANNO AD AUTOMATIZZARE LE ATTIVITÀ LAVORATIVE CHE ATTUALMENTE OCCUPANO IL 60-70% DEL TEMPO DEI DIPENDENTI.

Il rapporto evidenzia che saranno necessarie alcune significative riqualificazioni, che comporteranno la necessità di ristrutturazione delle risorse umane e, probabilmente, la perdita di posti di lavoro. Fattori con cui i governi dovranno confrontarsi e per cui dovranno presto trovare delle soluzioni.

Oltre ai rischi posti dall'AI agli individui, si cui ci occuperemo nei capitoli successivi, secondo lo Stockholm Environment Institute[9] lo sviluppo dell'AI amplificherà anche le disuguaglianze tra Stati, ovvero tra chi avrà le capacità di sviluppo, acquisto e applicazione della tecnologia e tra chi, al contrario, rimarrà indietro.

L'IA non è certamente esplosa all'improvviso negli ultimi anni, né è stata sviluppata ex novo nel contesto della AI generativa, ma è piuttosto un tema, almeno in termini strettamente tecnologici e di ricerca, che si potrebbe dire antico[10]. Eppure, la presenza della tecnologia nella realtà e nella sfera pubblica è diventata tangibile per la cittadinanza e su larga scala solo in tempi più recenti.

***8**. "The state of AI in 2023: Generative AI's breakout year", https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year*
***9**. "Technological disruption: Will artificial intelligence solve global problems or widen equity gaps?", https://www.sei.org/perspectives/technology-equity/*
***10**. https://www.einaudi.it/catalogo-libri/scienze-sociali/media/macchine-ingannevoli-simone-natale-9788806254063/*

Il 2024, in particolare, sarà certamente l'anno della conferma di questa tendenza: da un lato, il settore tecnologico continuerà a mettere a disposizione prodotti e applicazioni commerciali a qualche titolo gravitanti attorno all'IA; dall'altro saranno sempre di più anche le istituzioni e le amministrazioni pubbliche che adotteranno sistemi basati sull'intelligenza artificiale. Per questo motivo, l'approvazione dell'Artificial Intelligence Act (AI Act) dell'Unione Europea il 13 marzo 2024 è avvenuta in un momento cruciale[11]. Senza dubbio, infatti, l'AI Act è da considerarsi come il più avanzato e ampio – nonché il primo - regolamento esistente a livello mondiale per quanto riguarda questa tecnologia. Come tale sarà anche destinato a diventare il punto di riferimento per qualsiasi altro intervento normativo simile su scala globale, come già avvenuto in precedenza con la General Data Protection Regulation (GDPR), altra occasione in cui l'Unione Europea si pose come traino e punto di riferimento per la regolamentazione della tecnologia[12].

Dopo un processo legislativo lungo e tortuoso, l'AI Act troverà ora applicazione diretta negli Stati membri, andando a normare il rapporto tra IA e i quasi 450 milioni di cittadine e cittadini dell'Unione Europea e, con esso, anche il ruolo che la tecnologia andrà ad assumere nella società contemporanea e nel futuro prossimo. Non è una esagerazione: il testo, infatti, copre un ampio spettro di scenari di applicazione, che spaziano da quello commerciale a quello securitario e di controllo, toccando di fatto un novero molto elevato di ambiti sociali.

11. *https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai*
12. *https://www.cnbc.com/2021/04/08/from-california-to-brazil-gdpr-has-created-recipe-for-the-world.html*

3. CONTESTO ITALIANO

L'Italia, come tutti gli Stati Membri dell'Unione Europea, non può ovviamente considerarsi aliena ai temi sollevati dai recenti sviluppi dell'intelligenza artificiale, al suo utilizzo e alla sua regolamentazione, a cominciare dai punti critici sollevati dall'AI Act stesso. Al contrario, quello italiano è stato uno dei primi contesti a essersi trovato di fronte alla necessità di confrontarsi con queste questioni. Ciò è avvenuto in modo evidente per quanto riguarda l'uso delle tecnologie biometriche di sorveglianza, il riconoscimento facciale in particolare, e la verifica della conformità dei large language model (LLM) - come ChatGPT - con le leggi vigenti in termini di data justice, privacy e trasparenza.

Ovviamente, questi sono solo due tra gli scenari che si sono manifestati in modo più evidente ed esplicito, ma non certo gli unici. Come ogni altro Paese, anche l'Italia si troverà sempre più esposta alle pressioni dell'IA in diversi settori e, in quanto Stato Membro dell'UE, dovrà dare pieno adempimento anche alle nuove regole introdotte dall'AI Act e provvedere alla loro implementazione.

AD OGNI MODO, LO SCENARIO ITALIANO SEMBRA ESSERE PARTICOLARMENTE VIVACE, COME TESTIMONIATO ANCHE DALLA PRESENZA DI DIVERSE ORGANIZZAZIONI DELLA SOCIETÀ CIVILE CHE SI SONO IMPEGNATE IN CAMPAGNE E INIZIATIVE IN QUESTO AMBITO, ANCHE NEL CONTESTO DELLE NEGOZIAZIONI ATTORNO AL TESTO DELL'IA ACT[13].

Le questioni italiane inerenti l'uso dell'IA nel contesto della sorveglianza sono comunque emerse prima che l'AI Act prendesse forma.

Ad esempio, mentre a livello internazionale era in corso un dibattito serrato sulle opportunità e i rischi connessi all'uso degli strumenti di sorveglianza biometrica negli spazi pubblici, con numerose prove a sostegno degli aspetti problematici di queste tecnologie, diverse città italiane hanno comunque spinto per l'acquisto e l'utilizzo di questi strumenti[14].
Se all'estero il dibattito si concentrava sulla sorveglianza e sulle discriminazioni a cui il riconoscimento facciale può esporre la cittadinanza, in Italia si procedeva invece all'installazione di telecamere "intelligenti" in diversi contesti, e senza la dovuta attenzione né al dibattito in questione, né, in alcuni casi, ai requisiti di legge.

13. *Barassi Veronica, Scharenberg Antje and Di Salvo Philip (2024), 'Civil Society's Struggle Against Algorithmic Injustice in Europe', Research Report (II) The Human Error Project: AI, Human Rights and the Conflict over Algorithmic Profiling, School of Humanities and Social Sciences and MCM Institute University of St. Gallen, St. Gallen, Switzerland.*
https://thehumanerrorproject.ch/civil-society-struggle-algorithmic-injustice-ai-errors-europe-report/.
14. *https://www.ilpost.it/2021/09/16/riconoscimento-facciale-comuni-telecamere/*

Il caso più emblematico è stato certamente quello di Como dove, nell'estate del 2020, l'amministrazione comunale ha acquistato e installato in un parco pubblico un sistema di riconoscimento facciale, testandolo per alcune settimane. Un' inchiesta giornalistica[15] ha rivelato l'assenza di base giuridica per l'utilizzo del riconoscimento facciale, portando a un primo provvedimento del Garante della Privacy[16], a una interrogazione parlamentare sul tema[17] e all'approvazione da parte del Parlamento di una moratoria sull'uso di queste tecnologie[18], poi estesa fino alla fine del 2025[19]. Per quanto non escluda aspetti di criticità, come notato da diversi osservatori[20], la moratoria ha rappresentato uno dei primi interventi ufficiali per la regolamentazione dell'uso del riconoscimento facciale in Europa, in un momento storico in cui le discussioni e negoziazioni attorno al testo dell'AI Act erano ancora lontane dal giungere a un punto di conclusione.

Nonostante questi momenti di tensione, il riconoscimento facciale e la possibilità del suo utilizzo hanno continuato ad attirare la costante attenzione delle autorità italiane. È noto, ad esempio, che la Polizia di Stato abbia a disposizione il Sistema Automatico Riconoscimento Immagini (SARI) già dal 2017: utilizzato per il momento solamente nella sua

15. *https://www.wired.it/internet/regole/2020/06/09/riconoscimento-facciale-como/*
16. *https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9309458*
17. *https://www.wired.it/internet/regole/2020/06/10/riconoscimento-facciale-como-interrogazione/*
18. *https://www.ilsole24ore.com/art/italia-primo-paese-vietare-riconoscimento-facciale*
19. *https://www.ilpost.it/2023/06/22/moratoria-riconoscimento-facciale-2025/*
20. *https://privacy-network.it/iniziative/italia-moratoria-sul-riconoscimento-facciale/*

modalità standard, è stato oggetto dell'attenzione del Garante della Privacy, che nel 2021 ne ha limitato le funzionalità "Real Time" in quanto "privo di una base giuridica che legittimi il trattamento automatizzato dei dati biometrici per il riconoscimento facciale a fini di sicurezza. [Il sistema] realizzerebbe per come è progettato una forma di sorveglianza indiscriminata/ di massa"[21]. Prima dello stop imposto dal Garante, SARI era stato anche proposto come potenziale strumento di contrasto all'immigrazione[22]. Nonostante la moratoria sul riconoscimento facciale e la sua successiva estensione e nonostante i casi controversi qui discussi, nel 2023 il Ministro dell'Interno Matteo Piantedosi è comunque tornato a parlare di questa tecnologia, auspicandone l'introduzione nelle stazioni, negli ospedali e nelle aree commerciali di Roma, Milano e Napoli[23]. Allo stesso modo, in Italia si è discusso di riconoscimento facciale anche nel

***21**. https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9575842*
***22**. https://irpimedia.irpi.eu/viminale-garante-privacy-riconoscimento-facciale-in-tempo-reale/*
***23**. https://www.wired.it/article/piantedosi-riconoscimento-facciale-stazioni-luoghi-pubblici/*

contesto delle grandi manifestazioni sportive e del loro utilizzo negli stadi, a cominciare dall'Olimpico di Roma dove un sistema di questo tipo è attivo dal 2016[24]. Anche lo stadio Friuli di Udine, nel 2019, è stato interessato dal test di un sistema di questo tipo[25]. Più di recente, la possibilità di utilizzare la tecnologia in questo contesto è tornata di attualità dopo gli episodi di razzismo che hanno caratterizzato la partita Udinese-Milan nel gennaio del 2024[26]. Infine, sempre nel 2024, la città di Trento è stata a sua volta sanzionata per aver utilizzato tecnologie di sorveglianza biometrica basate sull'adozione dell'intelligenza artificiale[27].

Il successo del riconoscimento facciale e il suo ricorrente utilizzo sono elementi che confermano la centralità delle tecnologie biometriche negli scenari futuri dell'IA, in Italia come altrove.

24. *https://irpimedia.irpi.eu/sorveglianze-stadio-olimpico-roma-riconoscimento-facciale/*
25. *https://www.calcioefinanza.it/2019/11/05/udinese-alla-dacia-arena-sperimentato-sistema-per-il-riconoscimento-facciale/*
26. *https://www.rainews.it/articoli/2024/01/calcio-razzismo-discriminazione-violenza-gabriele-gravina-figc-viminale-riconoscimento-facciale-e-tecnologia-negli-stadi-03dbd9a2-f12d-4c2c-81f5-60d0c89db6c1.html*
27. *https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9977299*

Le questioni aperte dalla diffusione dell'intelligenza artificiale e le sue tensioni con gli aspetti normativi sono poi tornate visibili nella sfera pubblica italiana nel marzo del 2023 quando, in seguito alla diffusione virale di ChatGPT di OpenAI, il Garante della Privacy italiano ha imposto all'azienda di interrompere il trattamento di tutti i dati personali degli utenti italiani[28] fino all'effettiva verifica che questi fossero stati raccolti in ottemperanza con le normative del GDPR (e aprendo anche un'indagine nei confronti di un data breach avvenuto nel medesimo periodo[29]). All'intervento del Garante, il primo a livello mondiale, OpenAI aveva risposto con la momentanea messa offline di ChatGPT in Italia[30]. Il chatbot è poi tornato disponibile dopo alcune settimane, in seguito alle risposte dell'azienda e ad alcuni interventi volti a superare le criticità sollevate. Il contenzioso si è comunque riaperto nel gennaio del 2024, quando il Garante della Privacy è tornato a bussare agli uffici di Open AI, sollevando ulteriori problematicità in termini di riservatezza e trattamento dei dati[31].

28. *https://www.wired.it/article/chatgpt-blocco-italia-garante-privacy/*
29. *https://www.wired.com/story/italy-ban-chatgpt-privacy-gdpr/*
30. *https://www.nytimes.com/2023/03/31/technology/chatgpt-italy-ban.html*
31. *https://techcrunch.com/2024/01/29/chatgpt-italy-gdpr-notification/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAIYwKfB_kO8bpOyC5275CVkI6m2PlYfMiSMUgdplj8BLAsF3EWuFQHCVElqJyjSNvV-6xX57S3NQuacghRC-qbhepEyKlphKxGzqWN9Oxeg-Js0-4FonCexIm15BSjFwuOwy6omf6g00q-7_vSNyMi_ncQxkylx46BmBMnA1cCNn4*

4. LOTTA ALLA SORVEGLIANZA DI MASSA

I temi relativi alla lotta alla sorveglianza di massa sono al centro di ogni discussione relativa alle questioni connesse ai diritti digitali. Nell'ultimo decennio, e precisamente sulla scia del caso Snowden esploso nel 2013[32], la sorveglianza è diventata un tema politico di ampia portata, non più relegato esclusivamente agli ambienti degli addetti ai lavori.

Nel decennio trascorso dalle rivelazioni di Edward Snowden[33], il tema della sorveglianza è stato affrontato da diversi punti di vista e in ottiche diverse, connesse anche a specifiche lotte contro di essa. A una fase iniziale fortemente ancorata al tentativo di limitare gli spazi di azione e gli abusi dei governi nel monitoraggio delle comunicazioni digitali, ne è seguita una che, complice anche la concomitante esplosione del caso Cambridge Analytica[34], si è di nuovo incentrata maggiormente sugli aspetti di data justice insiti nello sfruttamento dei dati personali a fini commerciali. Progressivamente, si è reso evidente come questo tipo di abuso di dati potesse diventare il volano per scenari ancora più controversi, come ad esempio la strumentalizzazione (weaponization) delle piattaforme commerciali per scopi propagandistici e di influenza politica. Queste due fasi si sono avvicendate, influenzandosi a vicenda, mentre sullo sfondo avanzava spedito lo sviluppo dell'intelligenza artificiale e la diffusione degli strumenti di sorveglianza biometrica che su di essa si basano. Ciò ha ancorato di conseguenza i temi della lotta alla sorveglianza a quelli, più ampi, della data justice e della discriminazione algoritmica[35].

32. https://www.wired.it/article/10-anni-caso-snowden/
33. https://www.eff.org/deeplinks/2023/05/10-years-after-snowden-some-things-are-better-some-were-still-fighting
34. https://www.ilpost.it/2018/03/19/facebook-cambridge-analytica/
35. Per "discriminazione algoritmica" o "pregiudizio algoritmico" si intende un "contenuto etico o ideologico distorto o discriminatorio (per es. verso le fasce più fragili della popolazione) processato dall'algoritmo nella fase di raccolta massiva dei dati e poi generato automaticamente". Definizione di Treccani: https://www.treccani.it/vocabolario/neo-pregiudizio-algoritmico_(Neologismi)/

INEVITABILMENTE, QUESTO HA COSTRETTO A UNA RIFLESSIONE SU COME L'INTELLIGENZA ARTIFICIALE POSSA SOSTENERE, RAFFORZARE, ED ACCELERARE NUOVE FORME DI SORVEGLIANZA DI MASSA.

Come si è detto, il riconoscimento facciale è certamente la tecnologia che ha trovato più rapida applicazione e più ampia diffusione tra quelle che si basano sull'analisi dei dati biometrici, specialmente negli spazi pubblici, proprio perché offre la possibilità di sorveglianza non solo massiva, ma anche in tempo reale. Come riassumono Mark Andrejevic e Neil Selwyn nel loro libro dedicato al riconoscimento facciale[36], la storia di questa tecnologia va fatta risalire agli anni '60, ma il suo sviluppo nella forma attuale si è consolidato attorno agli anni dell'esplosione del web commerciale e in parallelo all'ascesa dell'economia digitale basata sullo sfruttamento dei dati. Una volta trovate diverse applicazioni commerciali in spazi pubblici come stadi, aeroporti e centri commerciali, la tecnologia ha presto mostrato anche le sue ampie e profonde ramificazioni in termini di controllo, tramite anche la spinta offerta dalle circostanze della pandemia da COVID-19 e le relative necessità di automazione[37].

36. *https://www.wiley.com/en-us/Facial+Recognition-p-9781509547333*
37. *https://algorithmwatch.org/en/automating-society-2020-covid19/*

La tecnologia, ad ogni modo, qualunque sia la sua applicazione, ha mostrato numerose falle e problematicità, certificate da una letteratura accademica estesa[38] e da una serie di casi di cronaca[39] che hanno messo in evidenza le numerose controversie relative all'utilizzo del riconoscimento facciale. Per quanto riguarda l'uso di questa tecnologia per scopi di controllo e sorveglianza, gli aspetti più controversi, si possono riassumere in questi ambiti:

1. la sua fallibilità;
2. la sua innata natura intrusiva;
3. la sua tendenza a favorire effetti di *function creep*[40] (il graduale ampliamento dell'uso di una tecnologia al di là dello scopo per cui era stato originariamente concepito),
4. le logiche oppressive e di controllo insite nel razionale nella tecnologia, indipendentemente dal contesto di utilizzo.

Non stupisce, quindi, che le attenzioni delle più recenti lotte contro la sorveglianza di massa che hanno interessato la società civile sia in Europa[41] che negli Stati Uniti[42] abbiano avuto come obiettivo primario il riconoscimento facciale. Allo stesso modo non stupisce che proprio attorno alla sorveglianza degli spazi pubblici tramite l'uso di riconoscimento facciale si siano viste le più accese discussioni anche attorno al testo dell'AI Act. Il testo finale in linea di principio vieta l'uso del riconoscimento facciale in tempo reale nei luoghi pubblici, ma lascia aperti diversi spiragli grazie ad

38. *Un ottimo riassunto è quello offerto da Kate Crawford nel suo "Atlas of AI": https://yalebooks.yale.edu/book/9780300264630/atlas-of-ai/. Traduzione italiana: https://www.mulino.it/isbn/9788815294197.*
39. *https://www.theguardian.com/us-news/2023/apr/27/california-police-facial-recognition-software*
40. *https://www.tandfonline.com/doi/full/10.1080/17579961.2021.1898299*
41. *https://reclaimyourface.eu/*
42. *https://www.aclu.org/news/privacy-technology/grassroots-activists-are-leading-the-fight-to-stop-face-recognition-its-time-for-congress-to-step-up-too*

alcune eccezioni importanti, tra cui quella concessa alle forze dell'ordine di utilizzare la tecnologia per indagini su crimini gravi o per la ricerca di persone scomparse, a condizione di ottenere l'autorizzazione di un giudice[43]. L'opinione prevalente dei soggetti della società civile che operano in questo settore, è che le maglie di queste eccezioni siano troppo ampie e le stesse troppo generiche, e che, di conseguenza, l'uso del riconoscimento facciale da parte delle forze dell'ordine per ragioni di sicurezza nazionale e nel contesto delle migrazioni[44] non sia stato regolamentato adeguatamente.

NEI PROSSIMI CAPITOLI ANALIZZEREMO PIÙ APPROFONDITAMENTE ALCUNE DELLE MAGGIORI CRITICITÀ PRESENTI NEL TESTO FINALE DELL'AI ACT, SPIEGANDO LA LOGICA ALLA BASE DELLE NOSTRE PREOCCUPAZIONI E PROPONENDO DELLE RACCOMANDAZIONI UTILI A ELIMINARE, O QUANTO MENO MITIGARE, I RISCHI CORRELATI.

43. https://www.politico.eu/article/eu-artificial-intelligence-act-ai-technology-risk-rules/
44. https://edri.org/our-work/civil-society-statement-regulate-police-tech-ai-act/

5. AI ACT

5.1 Riconoscimento biometrico negli spazi pubblici

L'approccio dell'AI Act in termini di regolamentazione dell'intelligenza artificiale è stato, sin dagli albori del processo di elaborazione del testo, ancorato a due diversi principi. Da un lato, un approccio regolatorio orizzontale, orientato alla regolamentazione complessiva della tecnologia, ma non dei suoi specifici utilizzi. Questo approccio si inserisce in un disegno politico di più ampio spettro, volto a sostenere e a favorire il ruolo dell'Unione Europea come leader mondiale nello sviluppo di una IA sicura ed etica. Dall'altro, l'AI Act è stato elaborato anche con l'obiettivo di introdurre un quadro normativo specifico per i sistemi di intelligenza artificiale giudicati come più a rischio[45]. In questo ambito, l'AI Act vieta esplicitamente l'immissione sul mercato e diversi ambiti di utilizzo delle tecnologie di intelligenza artificiale, come il predictive policing, il social scoring, o altre pratiche in grado di influenzare il comportamento umano in modi poco trasparenti, se non addirittura occulti[46].

45. https://www.mulino.it/isbn/9788815388353
46. https://www.wired.it/article/ai-act-testo-ultima-versione-gennaio-divieti-riconoscimento-facciale/#uno

LA DETERMINAZIONE DI COME L'AI ACT AVREBBE DOVUTO PORSI NEI CONFRONTI DEL RICONOSCIMENTO BIOMETRICO DELLE PERSONE NEGLI SPAZI PUBBLICI È STATO UNO DEI MAGGIORI TERRENI DI SCONTRO SIN DAI PRIMI PASSI DELL'ELABORAZIONE DEL TESTO DI LEGGE NEL 2021.

Il testo è stato infine approvato[47] dal Parlamento europeo il 13 marzo 2024[48]. Nella sua veste definitiva l'AI Act vieta, in linea di principio, i sistemi di categorizzazione biometrica basati su caratteristiche sensibili e la raccolta indiscriminata di immagini facciali da Internet (al fine di evitare il ripetersi di un nuovo caso ClearviewAI[49]) o da video di sorveglianza per creare database di riconoscimento facciale. Inoltre, vieta il riconoscimento delle emozioni sul posto di lavoro e nelle scuole.

È proprio su genericità, misura e tenuta di questi divieti che la società civile europea ha espresso i dubbi più forti, giudicando questi parametri eccessivamente vaghi e suscettibili di interpretazioni diverse e sottolineando come il numero di eccezioni e concessioni per il loro utilizzo da parte delle forze dell'ordine sia troppo esteso. L'organizzazione Access Now, ad esempio, ha definito il testo approvato dal Parlamento a marzo come "pieno di scappatoie, eccezioni e deroghe, il che significa che non proteggerà le persone, né i loro diritti umani, da alcuni degli utilizzi più pericolosi dell'IA"[50].

47. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_IT.pdf
48. https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law
49. https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html
50. https://www.accessnow.org/press-release/ai-act-failure-for-human-rights-victory-for-industry-and-law-enforcement/

Allo stesso modo la Ong Article 19 si è espressa criticamente nei confronti del testo approvato dal voto parlamentare, dichiarandosi delusa per il fatto che "l'AI Act non riesce a vietare completamente l'uso delle tecnologie di riconoscimento emotivo e di identificazione biometrica remota in tempo reale negli spazi accessibili al pubblico" e sottolineando come "vietare l'uso, negli spazi pubblici, dell'identificazione biometrica remota soltanto in tempo reale (e non anche "post") non è sufficiente per proteggere i diritti fondamentali; inoltre, le eccezioni previste a questo divieto già troppo limitato sono ampie e non includono sufficienti garanzie, creando pericolose scappatoie per gli utilizzi delle forze dell'ordine"[51].

Si può facilmente constatare come il testo approvato dal Parlamento europeo a marzo segni un passo indietro rispetto alla bozza risultante dalle precedenti negoziazioni tra Commissione, Parlamento e Consiglio europeo, chiusa nel dicembre del 2023. Nel testo di partenza il divieto era più esplicito, al netto di alcune eccezioni che venivano già allora introdotte. Già a fine 2023 e proprio in reazione alla bozza di accordo raggiunto a dicembre, l'European Digital Rights (EDRi) giudicava gli articoli del testo

51. https://www.article19.org/resources/eu-ai-act-passed-in-parliament-fails-to-ban-harmful-biometric-technologies/

relativi alla sorveglianza biometrica al massimo "tiepidi" nel prevedere concrete limitazioni alla possibile proliferazione della sorveglianza di massa tramite tecnologie biometriche[52]. Persino più negativo il giudizio di Amnesty International che, a dicembre 2023 denunciava come l'assenza di una messa al bando esplicita del riconoscimento facciale di fatto corrispondesse al dargli un via libera[53].

Un altro aspetto controverso del testo riguarda la non applicazione dei divieti ai settori militari e di sicurezza nazionale, dove il regolamento non si applica. L'eccezione in nome della "sicurezza nazionale", infatti, è giudicata dalla società civile come troppo ampia e discrezionale e non incentrata su contesti applicativi chiari, che rendono i potenziali abusi della tecnologia più facili, come anche la giustificazione di potenziali suoi utilizzi. Come ha fatto notare AccessNow questa esenzione generalizzata "in teoria potrebbe essere utilizzata per qualsiasi questione di migrazione, polizia e sicurezza"[54].

L'utilizzo della sorveglianza biometrica da remoto e in tempo reale è comunque consentito per ragioni di contrasto alla criminalità in circostanze che sono piuttosto ampie, dato che il testo autorizza l'uso di questi strumenti per l'indagine di almeno 16 reati, previa autorizzazione di un'autorità giudiziaria o amministrativa indipendente.
I reati in questione sono:
- terrorismo;
- tratta di esseri umani;
- sfruttamento sessuale di minori e materiale pedopornografico;
- traffico illecito di stupefacenti e sostanze psicotrope;
- traffico illecito di armi, munizioni ed esplosivi;
- omicidio volontario;
- lesioni personali gravi;

52. https://edri.org/our-work/eu-ai-act-deal-reached-but-too-soon-to-celebrate/
53. https://www.amnesty.org/en/latest/news/2023/12/eu-blocs-decision-to-not-ban-public-mass-surveillance-in-ai-act-sets-a-devastating-global-precedent/
54. https://www.accessnow.org/press-release/joint-statement-ai-act-fails-migrants-and-people-on-the-move/rights.pdf

- traffico illecito di organi e tessuti umani;
- traffico illecito di materie nucleari e radioattive, rapimento, sequestro e presa di ostaggi;
- reati che rientrano nella competenza giurisdizionale della Corte penale internazionale;
- dirottamento di un aeromobile o una nave;
- stupro;
- reati ambientali;
- furto organizzato o rapina a mano armata;
- sabotaggio, partecipazione a un'organizzazione criminale coinvolta in uno o più dei reati qui elencati sopra.

Inoltre, il riconoscimento biometrico da remoto in tempo reale è possibile anche nei casi di ricerca di specifiche vittime (in casi come rapimento, tratta e sfruttamento sessuale di esseri umani e persone scomparse) e per la prevenzione di minacce per la vita o l'incolumità fisica delle persone o in risposta a una minaccia attuale o prevedibile di attacco terroristico.

Come ha invece fatto notare AlgorithmWatch[55], anche per quanto riguarda il riconoscimento facciale "retroattivo", quindi applicato a immagini archiviate e non ottenute in tempo reale, il testo di legge approvato dal Parlamento ha ridotto le limitazioni all'utilizzo. Con l'eliminazione del riferimento ai soli reati internazionali, secondo l'organizzazione tedesca, basterebbe ora il sospetto di un generico potenziale reato penale a rendere possibile l'uso del riconoscimento facciale retroattivo in spazi pubblici. Complessivamente, come ha ribadito anche EDRi nella sua più recente analisi delle tutele dei diritti umani incluse nell'AI Act, nei confronti del riconoscimento facciale "retroattivo" queste salvaguardie sarebbero non sufficientemente significative e facilmente bypassabili

55. https://algorithmwatch.org/en/eu-parliament-votes-on-ai-act/

dalle forze di polizia[56].

Complessivamente, ha fatto notare EDRi, alcune lacune del testo - e in particolare l'assenza di divieti specifici - potrebbero addirittura essere interpretate come un segnale che alcune forme di sorveglianza biometrica di massa siano legittime nell'UE.

È QUESTO, AD ESEMPIO, IL CASO DEI SISTEMI DI RICONOSCIMENTO DELLE EMOZIONI E DI CATEGORIZZAZIONE BIOMETRICA, VIETATI SUL POSTO DI LAVORO E NEI CONTESTI EDUCATIVI, MA CONSENTITI PER LE FORZE DELL'ORDINE E PER LE AUTORITÀ COMPETENTI IN MATERIA DI IMMIGRAZIONE, APRENDO DI FATTO AL LORO POSSIBILE UTILIZZO IN AMBITI PARTICOLARMENTE SENSIBILI[57].

56. *https://edri.org/wp-content/uploads/2024/04/EUs-AI-Act-fails-to-set-gold-standard-for-human-rights.pdf*
57. *https://edri.org/wp-content/uploads/2024/04/EUs-AI-Act-fails-to-set-gold-standard-for-human-rights.pdf*

5.2 Migrazione e confini

Uno dei temi di discussione rimasti aperti lungo tutto il processo di avanzamento e di stesura dell'AI Act è stato quello relativo all'uso di queste tecnologie nel contesto delle migrazioni. In particolare, sin dagli albori del processo legislativo, diverse organizzazioni del settore o operative nell'ambito della tutela dei diritti umani hanno fatto notare come il testo di legge non tutelasse sufficientemente, se non per niente, i diritti delle persone migranti e richiedenti asilo da potenziali abusi derivanti dall'utilizzo di queste tecnologie. In seguito all'approvazione parlamentare del testo di legge avvenuta lo scorso 13 marzo, ad esempio, European Digital Rights (EDRi) ha preso posizione in modo netto su questo tema, sottolineando come il testo non fornisca sufficienti protezioni alle persone migranti[58].

In particolare, EDRi ha fatto notare come il testo risulti debole e addirittura permetta l'uso di sistemi IA giudicati "a rischio" quando si tratta di migrazione, creando un pericoloso precedente con la proposta di un quadro legale parallelo per l'uso dell'AI nel contesto delle migrazioni (come anche per la sicurezza nazionale, ad esempio) ed esentando così tali utilizzi dalle norme e dalle garanzie altrimenti previste all'interno dell'AI Act. Allo stesso tempo,

58. *https://edri.org/our-work/protect-not-surveil-eu-ai-act-fails-migrants-people-on-the-move/*

è stata sottolineata la mancata considerazione nel testo di molte delle tecnologie specifiche utilizzate in contesto migratorio: sempre EDRi,ha sottolineato ad esempio come numerosi database - tra cui Eurodac[59], Schengen Information System[60] e European Travel Information and Authorisation System (ETIAS)[61] - non saranno soggetti ai paletti dell'IA Act fino al 2030.
La recente approvazione del nuovo pacchetto di riforme alle politiche europee su migrazioni e diritto d'asilo, avvenuta il 10 aprile 2024, dimostra come le preoccupazioni fossero più che fondate: il cosiddetto pacchetto migrazione va infatti ad inasprire le condizioni dei migranti che raggiungono i confini europei, arrivando addirittura ad abbassare a 6 anni l'età minima per la raccolta dei dati biometrici che confluiranno in Eurodac.

La coalizione #ProtectNotSurveil, invece, ha sottolineato come le tecnologie basate sull'intelligenza artificiale e gli algoritmi utilizzati nel contesto migratorio non abbiano solamente una natura di sorveglianza, ma vadano a svolgere funzioni di controllo variegate e siano spesso testate sulle persone migranti per poi essere diffuse più ampiamente nella società, sfruttando proprio le opacità e debolezze dei contesti di legge cui sono soggette le persone migranti[62]. Nel concreto, Access Now ha evidenziato come la mancata introduzione di specifiche norme in questo ambito apra a potenziali utilizzi di queste tecnologie, come, ad esempio, nel caso dei cosiddetti "lie detector", nel progetto iBorderCtrl finanziato proprio dalla UE con 4,5 milioni di euro[63].

INFINE, LA SOCIETÀ CIVILE HA PUNTATO L'ATTENZIONE ANCHE SUL RUOLO DELL'UNIONE EUROPEA COME ESPORTATRICE DI TECNOLOGIE DI SORVEGLIANZA POTENZIALMENTE LESIVE DEI DIRITTI UMANI IN CONTESTI EXTRA-EUROPEI, DOVE LE GARANZIE DI LEGGE E DI TRASPARENZA SONO ANCORA MENO FORTI: L'AI ACT, IN QUESTO SENSO, NON VIETA L'UTILIZZO IN ALTRI CONTESTI DI SISTEMI INVECE VIETATI IN EUROPA.

59. https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/eurodac_en
60. https://home-affairs.ec.europa.eu/policies/schengen-borders-and-visa/schengen-information-system_en
61. https://home-affairs.ec.europa.eu/policies/schengen-borders-and-visa/smart-borders/european-travel-information-authorisation-system_en
62. https://edri.org/wp-content/uploads/2023/07/Civil-society-AI-Act-trilogues-statement.pdf
63. https://theintercept.com/2019/07/26/europe-border-control-ai-lie-detector/

5.3 Autorità nazionale

Oltre alle questioni più strettamente normative e all'introduzione di regole per l'uso dei sistemi di intelligenza artificiale, l'AI Act interviene direttamente anche in materia di governance della stessa, stabilendo che gli Stati membri dovranno dotarsi di un'authority nazionale competente responsabile dell'introduzione e dell'applicazione dell'AI Act. In parallelo, la Commissione europea ha anche istituito il suo AI Office, che sarà responsabile del coordinamento dell'applicazione dell'AI Act[64].

Il testo di legge stabilisce che tali autorità debbano essere indipendenti. Il ruolo di queste authority, ovviamente, sarà fondamentale nel contesto della reale applicazione dell'AI Act nei contesti nazionali, e come tale la loro costituzione e la loro natura meritano profonda attenzione. Gli stati membri avranno 12 mesi per creare queste autorità. Allo stato attuale, solo Spagna, Irlanda, Olanda e Lussemburgo si sono già dotate di una loro authority dedicata[65]: Agency for the Supervision of Artificial Intelligence (AESIA) in Spagna, un dipartimento dedicato all'interno dell'autorità di tutela dei dati personali in Olanda, il dipartimento per l'Impresa, il Commercio e il Lavoro in Irlanda e il dipartimento Media, Telecomunicazioni e Politiche Digitali in Lussemburgo.

64. *https://digital-strategy.ec.europa.eu/en/policies/ai-office*
65. *https://www.euronews.com/next/2024/03/12/as-ai-act-enters-into-force-focus-shifts-to-countries-oversight-appointments*

Nel 2021, sia lo European Data Protection Board (EDPB) che lo European Data Protection Supervisor (EDPS) si erano espressi[66] sul tema, sottolineando come questo ruolo dovesse essere ricoperto dalle autorità nazionali per la protezione dei dati, ovvero per quanto riguarda l'Italia dal Garante della Privacy. Nel nostro Paese, la questione è rimasta irrisolta fino a dopo l'approvazione del testo finale dell'AI Act da parte del Parlamento europeo. Oltre alla potenziale assegnazione di tale ruolo al Garante, la stampa ha in più occasioni riportato anche la concreta possibilità che potesse essere l'Agenzia per l'Italia digitale (AgId) as assumere il ruolo di authority per l'IA[67]. Questa scelta è stata infine ufficializzata a marzo 2024 dal Governo italiano. Al netto di modifiche dell'ultimo momento, da noi auspicate, saranno infatti proprio l'AgId insieme all'Agenzia per la cybersicurezza nazionale (ACN) a ricoprire quel ruolo, come annunciato dal Sottosegretario all'Innovazione Alessio Butti alla stampa il 20 marzo. Successivamente all'annuncio di Butti, l'iter decisionale che interessa l'autorità e la regolamentazione dell'IA ha visto come passo successivo l'approvazione del disegno di legge del Consiglio dei Ministri n.78 del 23 aprile 2024, che verrà presto discusso in parlamento insieme alle proposte di legge già in corso d'esame sulla stessa materia. In questa sede potrà essere messa in discussione la visione del governo di conferire ad Agid e ACN il ruolo di autorità, benché, considerata la compattezza della maggioranza, non si vede effettivo spazio di riuscita[68].

66. *https://www.edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_it*
67. *https://www.agendadigitale.eu/cultura-digitale/governance-dellai-la-strategia-ue-e-il-dilemma-italiano/*
68. *https://www.key4biz.it/intelligenza-artificiale-ora-e-ufficiale-la-vigilanza-in-italia-ad-agid-e-acn/484086/*

La scelta è stata contestata[69] dal Garante della Privacy che, tramite una segnalazione[70] ai Presidenti di Senato e Camera e al Presidente del Consiglio, ha sottolineato come il ruolo dovrebbe spettare proprio al Garante, alla luce della sua competenza tecnica specifica e lo status di autorità indipendente necessario. Sia AgID che ACN, al contrario, sono agenzie di nomina governativa.

A QUESTO PROPOSITO, HERMES CENTER THE GOOD LOBBY E PRIVACY NETWORK SONO TRA I FIRMATARI DI UNA LETTERA APERTA[71] AL GOVERNO ITALIANO PER CHIEDERE UN PASSO INDIETRO RISPETTO ALLA SCELTA DI ASSEGNARE AD AGID E ACN QUESTO RUOLO, IN FAVORE DI UN'AGENZIA INDIPENDENTE LE CUI CARATTERISTICHE AUSPICATE ERANO STATE INDICATE IN UN DOCUMENTO[72] AVANZATO DA HERMES CENTER, PRIVACY NETWORK E THE GOOD LOBBY A MARZO DEL 2024.

69. *https://www.corrierecomunicazioni.it/digital-economy/ai-act-tensione-governo-garante-privacy-stanzione-autorita-di-controllo-sia-indipendente/*
70. *https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9996508*
71. *https://www.hermescenter.org/intelligenza-artificiale-lettera-aperta-alle-istituzioni-per-una-governance-indipendente-e-inclusiva/*
72. *https://privacy-network.it/wp-content/uploads/2024/03/Autorita-AI-Ita-2024.pdf*

5.4 General Purpose AI Models (GPAI)

L'AI Act è anche la prima legislazione al mondo a regolamentare i cosiddetti "foundation models", o "General Purpose AI Models" (GPAI). Come indicato dall'Ada Lovelace Institute[73], con GPAI model si intendono quei sistemi in grado di svolgere una serie di compiti generali (come la sintesi di testi, la manipolazione di immagini e la generazione di audio). Esempi noti sono il GPT-3 e il GPT-4 di OpenAI, modelli di base per ChatGPT o LaMDA per Google Bard.

Nell'AI Act è indicato come tutti i fornitori di questo tipo di modelli dovranno fornire documentazioni tecniche, istruzioni per il loro uso, indicazione di compliance con la Direttiva sul copyright e impegnarsi alla pubblicazione di informazioni su quali contenuti siano stati utilizzati per il training del loro modello[74]. Per i modelli che vengono definiti come "ad alto impatto", perché la loro capacità computazionale è pari a 10^25 FLOPs (floating point operations per second) sono previste norme più stringenti[75]. I produttori di modelli GPAI di questa natura e portata dovranno svolgere valutazioni dirette del modello, valutazioni per la mitigazione di possibili rischi sistemici, monitoraggio e segnalazione di incidenti gravi alla Commissione Europea e fornire garanzie su processi adeguati di cybersecurity[76]. I GPAI model open-source e gratuiti che condividono dettagli su come sono stati costruiti, compresa la loro architettura, i parametri e i pesi, sono esentati da molti degli obblighi dell'AI Act[77].

73. https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/
74. https://artificialintelligenceact.eu/high-level-summary/
75. https://www.wired.it/article/ai-act-testo-ultima-versione-gennaio-divieti-riconoscimento-facciale/#tre
76. https://www.lexology.com/library/detail.aspx?g=617ae881-a79f-447c-9441-409eaa72d20f
77. https://www.technologyreview.com/2024/03/19/1089919/the-ai-act-is-done-heres-what-will-and-wont-change/

L'approvazione di queste indicazioni per i modelli GPAI era stata messa in discussione a fine 2023, quando alcuni Stati Membri - compresa l'Italia - si erano improvvisamente opposti a questo tipo di regolamentazione, ritenendola troppo stringente per i modelli in fase di sviluppo dalle aziende dei loro territori e temendo che questa avrebbe potuto creare un rallentamento nello sviluppo e quindi un danno alla loro competitività sul mercato[78]. Questi Stati Membri avevano cercato di promuovere l'adozione di un approccio basato sull'autoregolamentazione. L'accordo raggiunto, e poi confermato nel testo approvato dal Parlamento, è considerato un compromesso[79] nei confronti degli Stati Membri, guidati dalla Francia - dove non a caso ha sede la società Mistral, che sta sviluppando un suo GPAI - che avevano cercato di escludere i GPAI dal testo dell'AI Act.

La necessità di regolamentare in modo esplicito i GPAI model era stata avanzata anche da Reporters Without Borders (RSF)[80] che ha sottolineato come questi modelli dovessero rispettare standard di apertura, spiegabilità del funzionamento e trasparenza, nonché rispettare misure per la tutela del diritto all'informazione. In precedenza, la Ong AlgorithmWatch aveva chiesto a sua volta la regolamentazione esplicita dei GPAI model nel testo dell'IA Act, proprio alla luce del loro potenziale ampio impatto sociale, unendosi all'analisi tecnica di una serie di esperti internazionali del settore dell'intelligenza artificiale[81].

Il testo finale risulta, purtroppo, un compromesso al ribasso, soprattutto a causa delle restrizioni relative all'addestramento dei modelli, giudicate troppo lievi. Il risultato insoddisfacente è anche il prodotto delle pressanti attività di lobbying da parte delle maggiori società del settore, tra cui la già citata Mistral, che hanno spinto nella direzione di una completa deregolamentazione e autocertificazione.

78. *https://www.techpolicy.press/will-disagreement-over-foundation-models-put-the-eu-ai-act-at-risk/*
79. *https://www.euractiv.com/section/digital/opinion/lobbying-for-loopholes-the-battle-over-foundation-models-in-the-eu-ai-act/*
80. *https://rsf.org/en/ai-foundation-models-must-be-regulated-protect-right-information*
81. *https://algorithmwatch.org/en/algorithmwatch-demands-regulation-of-general-purpose-ai/*

Tra le maggiori criticità si evidenzia l'assenza di obblighi di documentazione e pubblicazione dei meccanismi di addestramento dei modelli di AI.
L'attitudine a trattare masse di dati disponibili online come fonti di conoscenza, senza tenere in considerazione licenze, affidabilità delle fonti, e consenso all'utilizzo, è criticata da anni da diversi soggetti, uniti nella critica sebbene di mondi eterogenei, come l'industria, la società civile e l'accademia[82], eppure gli attuali leader di mercato hanno difeso con successo questa pratica. L'AI Act non disciplina esplicitamente questa pratica, ma la riconosce come uno standard de facto e non definisce restrizioni significative.

IN DEFINITIVA, LA NARRATIVA SECONDO CUI UN REGOLAMENTO TROPPO STRINGENTE AVREBBE LIMITATO LA CAPACITÀ DI INNOVAZIONE NON È PER NOI CONDIVISIBILE, CI APPARE ANZI EVIDENTE ESPRESSIONE DELLE AZIENDE CHE, PROPRIO IN VIRTÙ DI UNA TOTALE DEREGOLAMENTAZIONE, HANNO POTUTO SFRUTTARE QUESTE PRATICHE E DIVENTARE COSÌ LEADER DEL MERCATO.

82. *https://iapp.org/news/a/garante-issues-notice-to-openai-over-alleged-gdpr-violations/ e https://www.aimyths.org/ai-can-be-objective-or-unbiased*

5.5 Sandbox

L'AI Act interviene anche in merito alle sandbox per i sistemi di intelligenza artificiale, definite come "regulatory sandbox" nel testo del regolamento. Il Parlamento europeo definisce le sandbox come "strumenti che consentono alle imprese di esplorare e sperimentare prodotti, servizi o attività nuove e innovative sotto la supervisione di un regolatore"[83].

Quello delle sandbox è un approccio già comune in diversi settori, come, tra gli altri, quelli dei trasporti, dell'energia e, più di recente, del settore fintech. La Norvegia è stato uno dei paesi pionieri nell'uso delle sandbox, introducendo una sperimentazione nazionale sull'AI Generativa già nel 2021[84], i cui risultati sono stati giudicati complessivamente in modo favorevole[85]. Allo stesso tempo, come notato anche in relazione all'esempio norvegese, le sandbox sembrano offrire le migliori possibilità quando utilizzate in concerto con diverse realtà sociali e in modo interdisciplinare. Al contrario, come metaforicamente scrivono Alex Moltzau del Norwegian Artificial Intelligence Research Consortium e Robindra Prabhu della Norwegian Labour and Welfare Administration, si correrebbe il rischio di costruire "castelli sulla sabbia"[86].

83. https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf
84. https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/
85. https://assets.pubpub.org/5efi1a4i/Moltzau%20&%20Prabhu%20(2024)_Just%20Accepted-21704238663422.pdf
86. https://assets.pubpub.org/5efi1a4i/Moltzau%20&%20Prabhu%20(2024)_Just%20Accepted-21704238663422.pdf

Con il fine di estendere l'adozione delle sandbox anche al settore dell'intelligenza artificiale, queste sono state incluse nel testo dell'AI Act già nel 2021. Per quanto orientate a fornire più trasparenza e accessibilità nei confronti di nuovi prodotti, servizi o applicazioni a rischio, è importante sottolineare come le sandbox possano anche essere abusate o utilizzate come lasciapassare meno stringenti della stessa regolamentazione. Le sandbox possono in prospettiva sostituirsi a una rigida regolamentazione "a monte", fornendo invece una più aleatoria revisione caso per caso.

Altra situazione di ambiguità è negli aggiornamenti progressivi. Nel mondo del software rilasciare aggiornamenti è parte del ciclo di sviluppo, e la produzione di nuove versioni è più rapida che in ogni altro settore industriale.

LA DINAMICA RICORRENTE È QUELLA PER CUI SPECIALISTI ESTERNI DI CONFORMITÀ SVOLGONO ANALISI E PRODUCONO DOCUMENTAZIONE CON INTERVENTI UNA TANTUM. QUESTA PRASSI DISINCENTIVA UNA CONSULTAZIONE FREQUENTE, RISCHIANDO DI FAR PASSARE SOTTO TRACCIA L'EFFETTIVO IMPATTO DEGLI AGGIORNAMENTI SUCCESSIVI ALLA FASE DI TEST.

6. RACCOMANDAZIONI

Riconoscimento biometrico negli spazi pubblici

Come espresso dal Garante per la protezione dei dati personali, l'installazione di sistemi di riconoscimento facciale in real time negli spazi pubblici, determinerebbe "una evoluzione della natura stessa dell'attività di sorveglianza, che segnerebbe un passaggio dalla sorveglianza mirata di alcuni individui alla possibilità di sorveglianza universale".

Inoltre, bisogna considerare le conseguenze impreviste (unintended consequences) dalla posa di un'infrastruttura tecnologica di tale portata: ciò che oggi ci appare come implausibile e distante dai valori comuni condivisi - sorveglianza di massa, sorveglianza di specifiche minoranze - non è detto che diventi in futuro una pratica invece accettata e, con l'infrastruttura già presente, immediatamente attuabile.

1. Raccomandiamo quindi che i sistemi di riconoscimento facciale siano completamente vietati, senza la previsione di alcuna eccezione.

Infrastrutture tecnologiche su cui si potrebbe posare il riconoscimento facciale esistono già: sono i sistemi di videosorveglianza cittadini, sparsi per tutta l'Italia. Sistemi che ogni anno vengono finanziati dal Ministero dell'Interno a seguito di appositi bandi con il coinvolgimento delle prefetture e dei sindaci.

2.Raccomandiamo di rafforzare le verifiche su come i progetti presentati dai Comuni alle prefetture e successivamente al Ministero dell'Interno vengono implementati, e di effettuare dei controlli sugli impianti di videosorveglianza installati, fornendo una reportistica pubblica sui risultati ottenuti.

Migrazione e confini

Nel caso si decida di utilizzare comunque i sistemi di riconoscimento facciale negli spazi pubblici, seguendo le eccezioni previste dal regolamento europeo:

3. Raccomandiamo che il Ministero dell'Interno, possibilmente in accordo con il Garante per la Privacy e/o l'Autorità nazionale per l'intelligenza artificiale, pubblichi nella maniera più accessibile e disaggregata possibile, i dati relativi all'utilizzo dei sistemi di riconoscimento facciale, riportando la percentuale di errori riscontrati e le statistiche relative al tipo di reato per cui è stato richiesto e autorizzato l'utilizzo, corredate da una valutazione d'impatto.

La sperimentazione di tecnologie digitali in materia di migrazione rivela un approccio per cui migranti, stranieri e richiedenti asilo sono rappresentati come gruppi di persone da dover controllare, tracciare e sorvegliare semplicemente in quanto "non cittadini europei".
Questo tipo di sorveglianza si verifica in zone grigie, in cui la responsabilità dello Stato e dei governi è fortemente ridimensionata per via dell'ambito in cui viene agita.

4. Raccomandiamo che la gestione dei database, come AFIS, che raccolgono dati biometrici di persone appartenenti a categorie vulnerabili come i migranti, i rifugiati e i richiedenti asilo, così come il funzionamento degli algoritmi utilizzati dal sistema SARI della Polizia di Stato, siano trasparenti e accessibili a giornalisti, organizzazioni non governative e osservatori nazionali e internazionali per i diritti umani.

Autorità Nazionale

Le responsabilità che la nuova Autorità per la gestione dell'Intelligenza Artificiale andrà ad assumere saranno cruciali per il futuro del Paese, sia per ciò che concerne la tutela dei diritti delle persone, sia per l'impatto di tali tecnologie sulla sfera economica, politica e sociale.

Per una versione estesa delle seguenti raccomandazioni si rimanda al documento "Motivazioni e Caratteristiche dell'Autorità Nazionale per l'Intelligenza Artificiale", realizzato da Hermes Center, The Good Lobby e Privacy Network e pubblicato a marzo 2024[87]:

5. Raccomandiamo pertanto che l'Autorità nazionale sia totalmente indipendente ed autonoma, in modo da poter prendere decisioni politicamente neutrali e nell'esclusivo interesse della collettività.

6. Raccomandiamo che i componenti del collegio vengano nominati attraverso un percorso parlamentare trasparente e possibilmente partecipativo e che sia prevista un'adeguata durata del mandato per i membri del Consiglio direttivo dell'Autorità, che consenta loro di evitare pressioni politiche.

In ogni caso, se come sembra probabile al momento della pubblicazione di questo report, si propenderà ad assegnare i compiti e le responsabilità previste ad un'Autorità nazionale governativa, contrariamente a quanto sopra auspicato,

7. accomandiamo che l'Autorità nazionale possa disporre di un numero sufficiente di risorse umane, con competenze, esperienze e una conoscenza approfondita e multidisciplinare delle tecnologie di intelligenza artificiale, dei dati e dell'informatica, dei diritti fondamentali, dei rischi per la salute e la sicurezza, nonché la conoscenza delle norme e dei requisiti giuridici esistenti e delle scienze sociali e umanistiche per assicurare un'analisi esaustiva degli impatti dell'IA sulla società.

8. Raccomandiamo inoltre che l'Autorità venga costituita garantendone le caratteristiche di trasparenza, apertura al pubblico scrutinio, rappresentatività ed inclusività del collegio, capacità finanziaria ed operativa.

87. https://www.hermescenter.org/wp-content/uploads/2024/03/Autorita-AI-Ita-2024.pdf

Sandbox

Sebbene le Sandbox possano rappresentare un momento di scambio e revisione tra sviluppatori ed istituzioni, il rischio è che il meccanismo possa essere puramente cosmetico e poco normativo. Il testo dell'AI Act approvato dal Parlamento offre degli interventi positivi, ad esempio quando propone metriche di analisi che riguardano l'identificazione dei possibili rischi, i diritti fondamentali, salute e sicurezza e parametri tecnologici specifici e misure di mitigazione.

9. Raccomandiamo di ampliare il più possibile, durante le fasi di testing, il coinvolgimento della società civile e delle associazioni che rappresentano gruppi particolarmente esposti ai rischi connessi ai sistemi di AI.

In riferimento al Recital 89 del regolamento europeo sull'intelligenza artificiale, chi sviluppa librerie, framework, e altri utilità open source riutilizzabili non sono tenute agli obblighi di conformità, purché tali componenti siano generici supporti tecnici, o agnostici rispetto al tipo di dato trattato ed ai casi d'uso. Va inoltre ricordato che dare visibilità sorgente non è sufficiente, nei sistemi di AI, perché si abbiano proprietà di trasparenza, scrutinio, ricompilazione.

10. accomandiamo che sia assicurata l'effettiva natura generalista dei componenti esterni usati dagli applicativi, che devono essere documentati all'inizio del processo di Sandbox. Se infatti dovesse contenere logiche appartenenti ad un dominio specifico, dovrebbe essere considerato parte del test indipendentemente dalla licenza.

11. Raccomandiamo inoltre di focalizzare i test e lo scrutinio sulle parti di codice originali, così da offrire feedback più pertinenti ai produttori software, una volta isolati i componenti standard come le librerie ed i framework - che sovente rappresentano la maggior parte del codice stesso.

12. Raccomandiamo infine che l'Autorità pubblichi in maniera trasparente ed accessibile i report relativi ai test effettuati, così da permettere anche ai produttori di verificare internamente la conformità durante i diversi aggiornamenti del software, senza dover necessariamente ricorrere ad assistenza specializzata esterna.

Gli Autori

Philip Di Salvo
è ricercatore e docente presso la School of Humanities and Social Sciences dell'Università di San Gallo, Svizzera. Si occupa di giornalismo investigativo, sorveglianza di Internet, tecnologie black box e dei rapporti tra hacking e informazione. Come giornalista, collabora con diverse testate.

Hermes Center
si batte per una società in cui la tecnologia sia uno strumento abilitante per la libertà, non per la sorveglianza. Proteggiamo le persone dalle attività di sorveglianza di governi e aziende, conducendo attività di ricerca, di advocacy, di formazione e di comunicazione. Contribuiamo inoltre allo sviluppo di strumenti informatici per tutelare la libertà delle persone online e offline.
Sito: hermescenter.org

The Good Lobby Italia
è un'organizzazione non profit impegnata a tutelare l'interesse pubblico. Promuoviamo leggi, regolamenti e prassi che incoraggino la partecipazione pubblica, che difendano i valori democratici e lo stato di diritto, che preservino lo spazio civico e promuovano l'integrità pubblica. Contribuiamo inoltre a rafforzare le capacità della società civile (terzo settore, movimenti sociali, reti territoriali, cittadinanza attiva) ad agire nella sfera politica, a inserirsi nei processi decisionali portando la voce degli interessi collettivi.
Sito: thegoodlobby.it

CONTATTI

THE GOOD LOBBY
Martina Turola
martina@thegoodlobby.it

HERMES CENTER
Davide Del Monte
davide@hermescenter.org

Questo report è stato realizzato nel quadro del progetto "The CARE - Civil Actors for Rights and Empowerment" finanziato da ActionAid International Italia E.T.S e Fondazione Realizza il Cambiamento nell'ambito del progetto "The CARE - Civil Actors for Rights and Empowerment" cofinanziato dall'Unione Europea.

The CARE mira a promuovere, proteggere e far rispettare i Diritti e i Valori dell'Unione Europea con un approccio fondato sulla partecipazione dei/delle portatori/trici di diritti e sull'empowerment degli/delle stessi/e nel rivendicare i propri diritti. Il progetto coinvolge 70 realtà attive in tutta Italia, creando così una rete del cambiamento in grado di ascoltare e rispondere ai bisogni specifici e concreti di ogni territorio e comunità.

Scopri di più sul progetto >
thecare.actionaid.it

SCAN ME

Per maggiori informazioni sull'intero progetto
The CARE - Ufficio Stampa ActionAid
ufficiostampaactionaid@actionaid.org

REALIZZATO NELL'AMBITO DI:

The Care

actionaid
REALIZZA IL CAMBIAMENTO

FINANZIATORI

FONDAZIONE REALIZZA IL CAMBIAMENTO

Cofinanziato dall'Unione europea