

Il contributo dell'intelligenza artificiale simbiotica nella protezione delle vittime vulnerabili e nel contrasto della violenza di genere

Lorenzo Pulito*

THE CONTRIBUTION OF SYMBIOTIC ARTIFICIAL INTELLIGENCE TO THE PROTECTION OF VULNERABLE VICTIMS AND IN THE FIGHT AGAINST GENDER-BASED VIOLENCE

ABSTRACT: Given the importance of assessing the risk of recurrence and escalation of violence for the prevention and fight against gender-based violence, the present paper emphasises the need to introduce specific assessment procedures, targeted and calibrated to the vulnerability of victims. Furthermore, the essay reconstructs the framework of algorithmic systems currently used in this field, the challenges they pose on a technical as well as ethical and legal level, and their potential developments. Finally, using the symbiotic paradigm, the paper provides suggestions for the acceptance of artificial intelligence in this sensitive area.

KEYWORDS: Risk assessment; gender-based violence; vulnerability; symbiotic artificial intelligence; significant human control.

ABSTRACT: Premessa l'importanza che la valutazione del rischio di ripetizione ed escalation della violenza assume per la prevenzione e il contrasto della violenza di genere, il contributo sottolinea la necessità di adottare procedure valutative specifiche, calibrate e mirate alla particolare vulnerabilità delle vittime. Inoltre, il saggio ricostruisce il quadro dei sistemi algoritmici attualmente in uso in tale dominio, delle sfide che sollevano, sia sul piano tecnico, che su quello etico e giuridico, e dei loro potenziali sviluppi. Infine, ricorrendo al paradigma simbiotico, fornisce suggerimenti per l'accettabilità dell'intelligenza artificiale in questo delicato ambito.

PAROLE CHIAVE: Valutazione del rischio; violenza di genere; vulnerabilità; intelligenza artificiale simbiotica; controllo umano significativo.

SOMMARIO: 1. Introduzione – 2. Vittime vulnerabili e obblighi di protezione nella cornice convenzionale – 3. L'ultimissimo tassello nel percorso legislativo interno di contrasto alla violenza domestica e di genere e gli spazi di valutazione del rischio di ripetizione ed *escalation* della violenza – 4. Lo *Spousal Assault Risk Assessment* (SARA) – 5. Simbolismo vs. effettività: *risk assessment* e ausilio algoritmico. Il sistema VioGén – 6. Intelligenza artificiale simbiotica e nuovi assetti paradigmatici – 7. Osservazioni conclusive (a partire da *ChatGPT*)

* Ricercatore di diritto processuale penale, Università di Bari. Mail: lorenzo.pulito@uniba.it. Questo lavoro è stato parzialmente sostenuto dal progetto FAIR – Future AI Research (PE00000013), nell'ambito del programma MUR del PNRR finanziato dal NextGenerationEU. Contributo sottoposto a doppio referaggio anonimo.

Special Issue



1. Introduzione

La violenza di genere rappresenta una delle violazioni dei diritti umani più sistematiche a livello mondiale ed impegna i legislatori ad individuare soluzioni per combatterla più efficacemente, tanto in ambito internazionale (dove la Convenzione di Istanbul¹, ratificata dall'Italia con l. 27 giugno 2013, n. 77², ha gettato i pilastri sinteticamente riassumibili nelle “tre P” – prevenzione, protezione e punizione) ed europeo (è di recentissima emanazione la Direttiva 2024/1385³, atto normativo che segue l'adesione dell'Ue alla stessa Convenzione di Istanbul), quanto a livello nazionale (dove è stato da ultimo “rafforzato” il “Codice Rosso”⁴, la cui disciplina – “ispirata” dal noto caso Talpis⁵ – è già stata *medio tempore* incisa dalle norme immediatamente precettive della c.d. “Riforma Cartabia”⁶ e dal *restyling* della l. 8 settembre 2023 n. 122⁷).

La Convenzione di Istanbul obbliga gli Stati aderenti a predisporre le necessarie azioni legislative o altre misure per proteggere tutte le vittime da ogni ulteriore atto di violenza (art. 18), esigendo indagini e procedimenti efficaci (art. 49), all'interno dei quali deve essere offerta una protezione adeguata ed immediata (art. 50) dalla vittimizzazione reiterata. In base a quanto disciplinato dal suo art. 51, la corretta valutazione del rischio di reiterazione dei comportamenti violenti (*risk assessment*) si colloca a monte della sua efficiente gestione (*risk management*), attuabile attraverso la messa in atto delle misure di protezione, che potranno consistere in interventi che offrono rifugio immediato alla vittima o in provvedimenti che determinano un allontanamento dell'autore della violenza ed una inibizione ad avvicinare la vittima (artt. 23, 52 e 53).

¹ Convenzione del Consiglio d'Europa sulla prevenzione e la lotta contro la violenza nei confronti delle donne e la violenza domestica, approvata il 7 aprile 2011 ed aperta alla firma l'11 maggio 2011 a Istanbul. Sulla Convenzione di Istanbul, v., tra gli altri, S. DE VIDO, M. FRULLI (a cura di), *Preventing and Combating Violence Against Women and Domestic Violence. A Commentary on the Istanbul Convention*, Cheltenham, 2023.

² Su cui cfr. G. BATTARINO, *Note sulla attuazione in ambito penale e processuale penale della Convenzione di Istanbul sulla prevenzione e la lotta contro la violenza nei confronti delle donne e la violenza domestica*, in *Dir. pen. cont.*, 2 ottobre 2013, 1 ss.

³ Direttiva (UE) 2024/1385 del Parlamento europeo e del Consiglio del 14 maggio 2024 sulla lotta alla violenza contro le donne e alla violenza domestica, in G.U.U.E., 24 maggio 2024, L. 1385. Per un primo inquadramento, M. FERRARI, *Violenza contro le donne: l'Unione europea adotta finalmente la direttiva (UE) 2024/1385*, in *Eurojus*, 17 giugno 2024. Sull'originaria Proposta di direttiva cfr. A. PITRONE, *Il lungo (ed incidentato) percorso della lotta alla violenza contro le donne nell'Unione europea. Dalla questione dell'adesione alla Convenzione di Istanbul alla proposta di una direttiva “ad hoc”*, in *Ordine internazionale e diritti umani*, 3, 2022, 692 ss.

⁴ L. 19 luglio 2019 n. 69, c.d. “Codice Rosso”. A riguardo, N. TRIGGIANI, *L'ultimo tassello nel percorso legislativo di contrasto alla violenza domestica e di genere: la legge ‘Codice Rosso’*, in *Proc. pen. giust.*, 2, 2020, 451 ss.

⁵ Corte EDU, 2 marzo 2017, *Talpis c. Italia*, ric. n. 41237/14. Per un commento, v. M. CASTELLANETA, *I ritardi e le misure inadeguate per combattere il fenomeno della violenza contro le donne rappresentano delle violazioni*, in *Guida dir.*, 14, 2017, 102 ss.

⁶ D. BIANCHI, *Le modifiche al codice penale immediatamente precettive: prescrizione del reato e sospensione condizionale*, in *Dir. pen. proc.*, 11, 2021, 1468 ss.

⁷ G. AMATO, *Intervento sicuramente apprezzabile ma non centra l'obiettivo perseguito*, in *Guida dir.*, 36, 2023, 17 ss.; ID., *Il supporto dell'informatizzazione necessario per garantire il controllo*, *ivi*, 23 ss.; A. MARANDOLA., *Codice Rosso rafforzato*, in *Dir. pen. proc.*, 11, 2023, 1420 ss.



L'apprezzamento del rischio rappresentato dall'autore del reato è richiesto, sin dal primo contatto tra vittima e autorità, dalla citata Direttiva 2024/1385, in via aggiuntiva rispetto agli obblighi della valutazione individuale a norma dell'art. 22 della Direttiva 2012/29/UE⁸.

L'uso di sistemi di intelligenza artificiale predittiva – e in particolare di approcci di *machine learning* – potrebbe ottimizzare tali processi valutativi, ma solleva diverse sfide, sia sul piano tecnico, che su quello etico⁹ e giuridico, con ampie implicazioni per gli individui e la società. Il contributo mira a comprendere queste ultime, in modo da mitigare efficacemente i rischi associati, nonché ad individuare le condizioni per l'accettabilità di tali sistemi intelligenti, al fine di assicurare valutazioni affidabili e rispettose dei diritti umani.

2. Vittime vulnerabili e obblighi di protezione nella cornice convenzionale

Sebbene Fondamentale in materia risulta anche l'esegesi della Corte europea dei diritti dell'uomo che, spesso richiamando la Convenzione di Istanbul, ha «da tempo proceduto ad una progressiva messa a fuoco delle centralità della procedura di valutazione del rischio»¹⁰.

Dalle disposizioni contenute nella Convenzione europea dei diritti umani (in particolare, dagli artt. 2, 3 e 4) la Corte di Strasburgo fa discendere in capo agli Stati ben precisi obblighi sostanziali e procedurali, individuati come strumento per contrastare la criminalità di genere e volti a garantire una tutela effettiva dei diritti convenzionali.

Per la Corte europea, in ossequio al principio di uguaglianza sostanziale, quando ad essere titolari dei diritti sono i soggetti più vulnerabili, gli Stati contraenti sono obbligati a fornire un livello di protezione più elevato, adeguato ai loro bisogni e alle loro particolari condizioni¹¹.

Ne deriva che, proprio «ai fini dell'adempimento degli obblighi di tutela delle vittime vulnerabili», quali sono quelle di violenza di genere, sessuale e domestica, specialmente se straniera¹², «appare cruciale l'adozione di efficaci procedure di valutazione e gestione del rischio»¹³, sia da parte della polizia giudiziaria che dei giudici.

⁸ Direttiva 2012/29/UE del Parlamento europeo e del Consiglio, che istituisce norme minime in materia di diritti, assistenza e protezione delle vittime di reato e che sostituisce la decisione quadro 2001/220/GAI, in G.U.U.E., 14 novembre 2012, L 315/57. In dottrina, v. L. LUPÁRIA (a cura di), *Victims and criminal justice. European standards and national good practices*, Milano, 2015.

⁹ Sulle sfide etiche e sulle opportunità offerte dalle tecnologie digitali e intelligenti nell'affrontare la violenza domestica cfr. P. NOVITZKY, J. JANSSEN, B. KOKKELER, *A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools*, in *Heliyon*, 6, 2023, 1 ss.

¹⁰ V. BONINI, *Protezione della vittima e valutazione del rischio nei procedimenti per violenza domestica tra indicazioni sovranazionali e deficit interni*, in *Sist. pen.*, 3, 2023, 53.

¹¹ Corte EDU, 24 giugno 2008, *Chapman c. Regno Unito*, ric. n. 27970/02, ha rappresentato il primo caso in cui la Corte ha avvertito l'esigenza che dei soggetti potessero godere di una tutela specifica in quanto "vulnerabili". Sul punto, cfr. E. LICATA, *Vulnerabilità e tutela dei "core rights": gli obblighi di protezione in materia di criminalità di genere derivanti dalla Cedu*, in *Sist. pen.*, 2, 2024, 63 s.

¹² Si parla in tal caso di doppia vulnerabilità. Su tale profilo e sulle interconnessioni problematiche tra violenza di genere e migrazione, si v. l'opera di A. DI STASI, R. CARDIN, A. IERMANO, V. ZAMBRANO (a cura di), *Donne migranti e violenza di genere nel contesto giuridico internazionale ed europeo*, Napoli, 2023.

¹³ E. LICATA, *op. cit.*, 65.



La Corte alsaziana individua le caratteristiche che la valutazione del rischio deve presentare nell'ambito della violenza di genere e domestica, ampliandola – rispetto al c.d. “Osman test”¹⁴ – con una serie di parametri valutativi volti a intercettare più adeguatamente l'esposizione di queste vittime al pericolo di subire nuove aggressioni, destinato altrimenti a rimanere nell'ombra se si adoperassero i meccanismi valutativi tradizionali, “tarati” esclusivamente sulla persona dell'accusato e sulle caratteristiche oggettive dell'azione criminosa già commessa¹⁵.

L'operazione valutativa, oltre ad essere improntata ai canoni della prontezza e della tempestività, dovrebbe giovare di strumenti standardizzati riconosciuti dalla comunità internazionale¹⁶ ed essere condotta in modo «*autonomous, proactive and comprehensive*»¹⁷.

3. L'ultimissimo tassello nel percorso legislativo interno di contrasto alla violenza domestica e di genere e gli spazi di valutazione del rischio di ripetizione ed *escalation* della violenza

La l. 24 novembre 2023, n. 168, recante «Disposizioni per il contrasto della violenza sulle donne e della violenza domestica», ribattezzata legge “Codice Rosso Rafforzato” o “legge Roccella”¹⁸, rappresenta l'ultimissimo tassello di un lungo e articolato percorso legislativo nazionale diretto alla tutela delle

¹⁴ Corte EDU, Grande Camera, 28 ottobre 1998, *Osman c. Regno Unito*, ric. n. 23452/94.

¹⁵ Corte EDU, Grande Camera, 15 giugno 2021, *Kurt c. Austria*, ric. n. 62903/15, par. 164.

¹⁶ Per una ricognizione degli strumenti si rimanda alla pubblicazione ufficiale dell'Unione europea elaborata dall'European Institute for Gender Equality (EIGE), *A guide to risk assessment and risk management of intimate partner violence against women for police*, Lussemburgo, 2019, 13 s., reperibile in <https://eige.europa.eu/publications-resources/publications/guide-risk-assessment-and-risk-management-intimate-partner-violence-against-women-police#> (ultima consultazione 10/07/2024).

¹⁷ Corte EDU, Grande Camera, 15 giugno 2021, *Kurt c. Austria*, par. 168. L'autonomia e la proattività della valutazione richiedono che la stessa non si basi esclusivamente sulla percezione del rischio riportata dalla vittima; la globalità implica che la valutazione riguardi una pluralità di fattori, quali quelli individuali, relazionali, statici e dinamici.

¹⁸ Sulla quale v., tra i primi volumi pubblicati, C. CECHELLA, C. PARODI, *Il nuovo codice rosso. L. n. 168/2023 contro la violenza sulle donne e la violenza domestica. Profili penali e civili*, Milano, 2023; V. DE GIOIA, G. MOLFESE, *Il nuovo codice rosso. Commento alla L. 24 novembre 2023, n. 168 recante le nuove disposizioni per il contrasto della violenza sulle donne e della violenza domestica (Riforma Roccella)*, Piacenza, 2024; J.F. LOREFICE, M. STORZINI, *Il nuovo codice rosso. Fondamenti in materia di misure cautelari e misure di prevenzione*, Milano, 2024; F. PICCIONI, *Abusi e violenza domestica. Il nuovo Codice rosso e le opportunità di difesa*, Rimini, 2024. Nell'impossibilità di tratteggiare in questa sede i molteplici aspetti dell'intervento normativo, si rinvia, tra i primi commenti apparsi in rivista, a G. AMATO, *Una corsia accelerata e preferenziale per definire le fattispecie più gravi*, in *Guida dir.*, 46, 2023, 59 ss.; ID., *Sul tema delle priorità nelle Procure serve una legge per i criteri generali*, *ivi*, 63 ss.; ID., *Violazione atti cautelari “attenuati”: rimediata un'incoerenza normativa*, *ivi*, 76 s.; ID., *Strumenti cautelari d'urgenza, arresto differito e via dalla casa*, *ivi*, 78 ss.; L. BIARELLA, *La gestione dei rischi aumenta l'efficacia*, *ivi*, 43 ss.; V. CANNAS, *Violenza di genere e misure di prevenzione: continuano le modifiche al codice rosso*, in *Dir. pen. proc.*, 2, 2024, 158 ss.; A. FAMIGLIETTI, *Disposizioni per il contrasto della violenza sulle donne e della violenza domestica*, in *Proc. pen. giust.*, 1, 2024, 9 ss.; A. MARANDOLA, *Violenza di genere: accelerazione del rito e formazione per l'adeguamento della legislazione domestica agli standard europei*, in *Dir. pen. proc.*, 2, 2024, 167 ss.; EAD., *L'accelerazione delle richieste cautelari e specializzazione dei magistrati*, in *Giur. it.*, 2024, 973 ss.; A. MARANDOLA, L. RISICATO, *Pregi e limiti del “Codice Rosso”*, *ivi*, 959 s.; M. PIERDONATI, *Il rafforzamento della tutela delle persone vulnerabili e l'eterogenesi dei fini*, *ivi*, 989.



vittime di violenza domestica e di genere: tale complesso di disposizioni costituisce, ormai, un vero e proprio sottosistema normativo sia nell'ambito del codice penale, ove è possibile individuare un ampio catalogo dei reati che si caratterizzano quale manifestazione di violenza domestica e di genere, sia nell'ambito del codice di procedura penale, ove si è prevista una disciplina differenziata ed un *iter* sempre più preferenziale¹⁹.

La stratificazione degli interventi normativi ha investito numerose disposizioni del codice di rito, incidendo particolarmente sul piano delle misure cautelari (applicabili in deroga ai criteri generali e da adottarsi, in maniera inedita, secondo precise cadenze temporali acceleratorie²⁰) e delle misure precautelari (ora "affrancate" dalla necessità che il soggetto sia colto in stato di flagranza e, in taluni casi, eseguibili anche quando questa è "differita"²¹).

Nonostante la moltitudine degli interventi normativi, continua a mancare una disciplina dedicata specificatamente alla valutazione del rischio di ripetizione ed *escalation* della violenza relazionale, che, attingendo dai criteri forniti dalla giurisprudenza europea, delinea una cornice organica che individui i soggetti tenuti a compierla, ne regolamenti i tempi, i modi e le caratteristiche.

L'adempimento «finisce per iscriversi all'interno di schemi normativi pensati e costruiti per altri processi valutativi» che, per limitarsi agli strumenti a caratura processuale, sono quelli concernenti: 1) l'apprezzamento dei presupposti di applicazione delle misure cautelari; 2) l'adozione di pre-cautele che prevedono spazi di discrezionalità, come nei casi di cui all'art. 381 c.p.p., all'art. 382-*bis* c.p.p., da ultimo introdotto e che consente l'arresto in flagranza differita²², e all'art. 384-*bis* c.p.p., che – così come novellato - consente al pubblico ministero, anche fuori dei casi di flagranza, di disporre l'allontanamento urgente dalla casa coniugale senza attendere il provvedimento del giudice²³; 3) in fase di esecuzione della pena, l'accesso a misure alternative alla detenzione o a benefici penitenziari che comportino una restituzione di spazi di libertà al soggetto *in vinculis*.

4. Lo Spousal Assault Risk Assessment (SARA)

A fronte del proliferare degli spazi normativi che richiedono la valutazione del rischio a cui la vittima vulnerabile è esposta e la prognosi sulla reiterazione delle condotte violente, il *risk assessment* resta affidato a prassi virtuose, che si avvalgono della consolidata esistenza di protocolli valutativi, sviluppati nell'ambito degli studi e delle ricerche sulla violenza di genere.

Tra questi vi è il SARA (*Spousal Assault Risk Assessment*)²⁴, un metodo messo a punto in Canada e ritenuto valido dalla comunità scientifica. Il modello, nella versione semplificata SARA-S, prende in

¹⁹ Così, in relazione al c.d. "Codice Rosso", N. TRIGGIANI, *op. cit.*, 2, 2020, 451 ss.

²⁰ L. ALGERI, *Tempi rapidi per l'adozione delle misure cautelari a protezione delle vittime*, in *Dir. pen. proc.*, 2, 2024, 172 ss.

²¹ N. ROMBI, *Novità in materia precautelare e cautelare e nuovi obblighi informativi*, in *Giur. it.*, 2024, 967 ss.

²² L. ALGERI, *L'arresto in flagranza differita per reati di violenza di genere e domestica*, in *Dir. pen. proc.*, 2, 2024, 181 ss.

²³ A. MARANDOLA, *I nuovi presidi a tutela della vittima: rimedi pre-cautelari, cautelari e obblighi informativi*, in *Dir. pen. proc.*, 2, 2024, 186 ss.

²⁴ Per la trattazione tecnica del quale si v. A.C. BALDRY, *Dai maltrattamenti all'omicidio. La valutazione del rischio di recidiva e dell'uxoricidio*, Milano, 2016.



considerazione dieci fattori di rischio (raggruppati in due sezioni, che tengono conto rispettivamente le condotte violente del partner e l'adattamento psico-sociale dell'accusato) e cinque fattori di vulnerabilità della vittima; infine, la *check-list* richiede anche di verificare se ricorrono ulteriori circostanze significative (quali *child abuse* e, come peraltro richiesto espressamente dall'art. 51, par. 2, della Convenzione di Istanbul, disponibilità di armi da fuoco). Si procede a stabilire il livello di presenza o meno di ogni singolo fattore, collocandolo allo stato attuale (ultime quattro settimane) e nel passato (prima di un mese); successivamente, si riporta il livello di rischio di recidiva, che può risultare basso, medio o elevato, sia nell'immediato (entro due mesi) che nel lungo termine (oltre i due mesi). Al valutatore viene anche chiesto di verificare se esiste un rischio di violenza letale e di un'*escalation* dell'atto violento.

È rimarcata la necessità di effettuare una nuova valutazione del rischio quando si verificano le cosiddette "circostanze critiche", in cui la vittima necessita di maggiori misure protettive (come nel caso di separazione, di contrasti circa l'affidamento o la visita dei figli, di scarcerazione del maltrattante).

L'efficacia operativa di questo sistema, riconosciuta dagli stessi organi di governo della magistratura²⁵, dipende dal fatto che la valutazione finale del rischio non viene effettuata soltanto sulla base della quantità di fattori presenti, ma soprattutto «sulla tipologia degli stessi e sulla loro interazione ed evoluzione nel tempo»²⁶.

Oltre a quello appena illustrato, tra i più noti strumenti valutativi riconosciuti a livello internazionale si annoverano B-SAFER (*brief spousal assault form for the evaluation of risk*), DASH (*domestic abuse, stalking and harassment and honour-based violence*)²⁷, MARAC (*multi-agency risk assessment conference*), DVSI (*domestic violence screening instrument*), ODARA (*Ontario domestic assault risk assessment*)²⁸, nonché VioGén (*Sistema de seguimiento integral en los casos de violencia de género*)²⁹, che è stato costruito proprio a partire dal modello SARA³⁰.

²⁵ Si allude, in particolare, alla risoluzione sulle linee guida in tema di organizzazione e buone prassi per la trattazione dei procedimenti relativi a reati di violenza di genere e domestica (delibera CSM 9 maggio 2018).

²⁶ L. ALGERI, *Tempi rapidi per l'adozione delle misure cautelari a protezione delle vittime*, cit., 177.

²⁷ A riguardo, v. E. TURNER, G. BROWN, J. MEDINA-ARIZA, *Predicting Domestic Abuse (Fairly) and Police Risk Assessment*, in *Psychosocial Intervention*, 3, 2022, 145 ss. La validità predittiva del Dash, strumento che pone l'accento sul giudizio professionale strutturato, è messa in dubbio da alcuni ricercatori, come può leggersi in M. BLAND, *Algorithms Can Predict Domestic Abuse, But Should We Let Them?*, in H. JAHANKHANI, B. AKHGAR, P. COCHRANE, M. DASTBAZ (a cura di), *Policing in the Era of AI and Smart Societies. Advanced Sciences and Technologies for Security Applications*, Cham, 2020, 147.

²⁸ Su questo strumento si v. J. HEGEL, K.D. PELLETIER, M.E. OLVER, *Predictive Properties of the Ontario Domestic Assault Risk Assessment (ODARA) in a Northern Canadian Prairie Sample*, in *Criminal Justice and Behavior*, 3, 2022, 411 ss., i quali richiamano anche l'interessante caso giurisprudenziale canadese *Ewert c. Canada*.

²⁹ Sullo strumento e sul suo *background* normativo cfr. J.L. GONZÁLEZ-ÁLVAREZ, J.J. LÓPEZ-OSSORIO, C. URRUELA, M. RODRÍGUEZ-DÍAZ, *Integral Monitoring System in Cases of Gender Violence. VioGén System*, in *Behavior & Law Journal*, 1, 2018, 29 ss. Sulla sua validazione, J.J. LÓPEZ-OSSORIO, J.L. GONZÁLEZ-ÁLVAREZ, J.M. MUÑOZ VICENTE, C. URRUELA CORTES, A. ANDRÉS PUEYO, *Validation and Calibration of the Spanish Police Intimate Partner Violence Risk Assessment System (VioGén)*, in *Journal of Police and Criminal Psychology*, 4, 2019, 439 ss.

³⁰ Lo rilevano A. VALDIVIA, C. HYDE-VAAMONDE, J. GARCÍA-MARCOS, *Judging the algorithm: A case study on the risk assessment tool for gender-based violence implemented in the Basque country*, in *arXiv*, 2022 (arXiv:2203.03723), 3.



5. Simbolismo vs. effettività: *risk assessment* e ausilio algoritmico. Il sistema VioGén

Sebbene la considerazione dell'esistenza di prassi virtuose sia valsa ad evitare (anche) all'Italia censure da parte del GREVIO (*Group of Expert on Action against Violence against Women and Domestic Violence*), l'organismo indipendente di monitoraggio dell'implementazione della Convenzione di Istanbul non ha mancato di sollecitare il nostro Paese ad intraprendere azioni migliorative³¹.

Del resto, l'arsenale di strumenti introdotto dal legislatore rischia di entrare in contrasto con l'ideale di un diritto effettivo se tali arnesi sono lasciati nella teca e non vengono adoperati per la mancanza o inadeguatezza della valutazione del rischio di esposizione a nuovi atti di violenza.

Non è casuale che, nonostante gli sforzi legislativi, l'Italia sia stata condannata più volte (ed anche in tempi recenti) dalla Corte di Strasburgo, ma le censure non sono dipese da inadeguatezze normative, bensì ruotano sempre intorno a deficit nel *risk assessment*, vuoi perché non eseguito dall'autorità giudiziaria³², inerte anche nell'avviare un'indagine effettiva nella quale condurre un'adeguata valutazione del rischio³³, vuoi perché quest'ultima non era stata immediata e proattiva come avrebbe dovuto³⁴.

È incidentalmente interessante notare come, in ben due casi, la Corte abbia evidenziato la scarsa conoscenza delle caratteristiche strutturali della violenza relazionale da parte dell'autorità giudiziaria³⁵, che invece rappresenta una «precondizione generale [per] riconoscere i tratti di una realtà ad elevata complessità [e] approcciarsi ad essa in modo libero da pregiudizi e *bias* ancora molto radicati sul piano culturale»³⁶.

Alla luce di tali decisioni risulta pertinente domandarsi se «le difficoltà che la complessità e la globalità della valutazione del rischio comportano» non possano «essere superate impiegando risorse tecnologiche che consentono, attraverso l'ausilio della macchina, una più veloce analisi di quantità anche assai imponenti di dati»³⁷.

Il quesito richiede di ritornare al summenzionato sistema attuariale VioGén, un algoritmo adoperato in Spagna e che trova base giuridica nella *Ley Orgánica 1/2004*.

Il funzionamento del sistema VioGén è basato su due questionari (*Protocolo Dual*): *Valoración Policial del Riesgo* (VPR) e *Valoración Policial de la Evolución del Riesgo* (VPER). Questi protocolli di va-

³¹ Si v. il par. 233 del documento del GREVIO intitolato *Baseline Evaluation Report Italy, 2020*, disponibile al seguente url: <https://rm.coe.int/grevio-report-italy-first-baselineevaluation/168099724e> (ultima consultazione 10/07/2024).

³² Corte EDU, 7 aprile 2022, *Landi c. Italia*, ric. n. 10929/19. A riguardo, si rinvia al commento di A.A. DEI CAS, *La Corte europea condanna ancora l'Italia per violazione degli obblighi positivi derivanti dall'art. 2 nei confronti di vittime di violenze domestiche*, in *Arch. pen. (web)*, 2, 2022, 1 ss.

³³ Corte EDU, 16 giugno 2022, *De Giorgi c. Italia*, ric. n. 23735/19.

³⁴ Corte EDU, 7 luglio 2022, *M.S. c. Italia*, ric. n. 32715/19.

³⁵ Nel caso *Landi* la Corte ha rilevato come non fosse stata effettuata una valutazione del rischio di letalità capace di considerare lo specifico contesto della violenza domestica, così censurando l'ignoranza da parte dell'autorità giudiziaria delle specificità strutturali e delle dinamiche di tale forma di violenza domestica; nel caso *De Giorgi* la violenza domestica era stata erroneamente sminuita ed interpretata alla stregua di una mera conflittualità coniugale.

³⁶ V. BONINI, *op. cit.*, 65.

³⁷ V. BONINI, *op. cit.*, 66.



lutazione vengono esaminati e revisionati da un *team* di esperti multidisciplinari; dal 2019 la valutazione del rischio è stata effettuata tramite VPR5.0-H e VPER4.1.

Quando una donna sporge denuncia contro il suo aggressore, gli agenti di polizia compilano il modulo VPR5.0-H. Questo comprende cinque domini con trentacinque indicatori di rischio: ogni elemento viene valutato come “presente” e “non presente”. In questo modo la raccolta delle informazioni è standardizzata su tutto il territorio nazionale. Una volta compilato il modulo, il sistema assegna un punteggio di rischio violenza di genere, i cui livelli sono “non apprezzato” (*no apreciado*), “basso” (*bajo*), “medio” (*medio*), “alto” (*alto*) ed “estremo” (*extremo*). Gli agenti di polizia possono solo modificare il punteggio portandolo a un livello di rischio più elevato. A seguito dello studio dell’*Equipo Nacional de Revisión Pormenorizada de Homicidios en el contexto de la Violencia de Género* (EHVdG), VioGén è stato adattato per rilevare i casi con rischio di aggressione letale³⁸. Questi vengono segnalati come di “particolare rilevanza”; in tale evenienza (così come nel caso di bambini a rischio o in situazione di vulnerabilità) una *Diligencia Automatizada* è allegata alla VPR al fine di richiamare l’attenzione di pubblici ministeri e giudici e raccomandare ulteriori approfondimenti.

Lo sviluppo delle metodologie computazionali e la disponibilità di *big data* digitali non solo rendono possibile applicare tecniche di *machine learning* alla previsione dei crimini di genere³⁹, ma – si sostiene – un approccio di apprendimento automatico riesce a valutare meglio i rischi rispetto a quello tradizionale⁴⁰. Infatti, i risultati empirici relativi alla valutazione di un modello creato applicando tecniche di *machine learning* ed alimentato con dati estratti dal sistema VioGén dimostrerebbero una netta *outperformance* dell’approccio automatizzato, con un sensibile miglioramento rispetto al sistema di valutazione del rischio preesistente⁴¹.

Si tratta di un ambito di impiego molto promettente, ma che – come condivisibilmente sostenuto – potrà avere successo solo a condizione di considerare le numerose sfide che si pongono⁴² e che riguardano, tra le altre, la trasparenza⁴³, la *accountability*⁴⁴ e la (eccessiva) dipendenza dall’algoritmo.

³⁸ Sul funzionamento di VioGén e sul suo duplice meccanismo algoritmico, si rinvia a A. GIRALDI, *Intelligenza artificiale e predictive policing nella rinnovata fase d’indagine*, in A. MASSARO (a cura di), *Intelligenza artificiale e giustizia penale*, Roma, 2020, 83.

³⁹ Si v. il documento *The External Audit of the VioGén System*, elaborato da Eticas foundation, disponibile al seguente url: <https://eticasfoundation.org/wp-content/uploads/2024/07/ETICAS-FND-The-External-Audit-of-the-VioGen-System-1.pdf> (ultima consultazione 10/07/2024), 2022, 9.

⁴⁰ J. GROGGER, S. GUPTA, R. IVANDIC, T. KIRCHMAIER, *Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases*, in *Journal of Empirical Legal Studies*, 1, 2021, 114.

⁴¹ Á. GONZÁLEZ-PRieto, A. BRÚ, J.C. NUÑO, J.L. GONZÁLEZ-ÁLVAREZ, *Hybrid machine learning methods for risk assessment in gender-based crime*, in *Knowledge-Based Systems*, 2, 2023, 1 ss.

⁴² M. BLAND, *op. cit.*, 148.

⁴³ Con il noto problema della *black-box*. Utile punto di partenza sul nodo della trasparenza è il quadro ALGO-CARE per il processo decisionale di M. OSWALDA, J. GRACEB, S. URWINC, G.C. BARNE, *Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality*, in *Information & Communications Technology Law*, 2, 2018, 223 ss.

⁴⁴ Nel senso che, in relazione a tutti gli usi della tecnologia decisionale algoritmica, l’obiettivo deve essere quello di «*to augment human legal intelligence, not to replace it*», M. HILDEBRANDT, *Law as computation in the era of artificial legal intelligence. Speaking law to the power of statistics*, in *The University of Toronto Law Journal*, vol. 68, suppl. 1 (*Artificial Intelligence, Technology, and the Law*), 2018, 33.

Stessi aspetti problematici che, insieme ad altri non meno delicati – come i riflessi negativi sulla presunzione di innocenza, il pericolo di cristallizzazione degli individui nel loro passato⁴⁵ nonché la supervisione umana minima e incoerente sul risultato del sistema automatizzato⁴⁶ –, sono già emersi in relazione al più “tradizionale” VioGén e che è facilmente pronosticabile assumeranno maggior enfasi critica con la transizione di questo strumento verso modelli di *machine learning*.

Tanto più che l'utilizzo di questi *tools* – in questa sede circoscritto al presupposto di una previa denuncia e, dunque, senza tenere conto delle ulteriori questioni che sollevano gli strumenti di *predictive policing*⁴⁷ e il loro impiego in chiave pre-investigativa⁴⁸ – non appare essere appannaggio delle sole autorità di *law enforcement*.

Strumenti algoritmici, come EPV-R, assistono infatti il processo decisionale del giudice: nella fattispecie, lo strumento, anch'esso ispirato al SARA, viene utilizzato nei tribunali baschi, dove i giudici decidono il livello di protezione da accordare alla vittima di violenza di genere sulla scorta dell'*output* algoritmico. Anche questo strumento di valutazione solleva questioni legali con riferimento ad almeno «*three factors: (1) opaque implementation, (2) efficiency's paradox and (3) feedback loop*»⁴⁹, per mitigare i quali si è ritenuto «*essential that judicial users understand their important overseeing role, that an algorithm cannot replace them and that its automatic assessment can be wrong*»⁵⁰, coltivando una visione antropocentrica dell'interazione tra intelligenza umana e artificiale⁵¹.

Il criterio di non esclusività del dato algoritmico per la decisione giudiziaria, con la correlata necessità che lo stesso sia corroborato da ulteriori e diversi elementi di prova, secondo l'avvertimento della Suprema Corte del Wisconsin nel caso *Loomis*, è stato già eletto da fonti di *soft law*⁵² e sostenuto da

⁴⁵ E. FALLETTI, *Human rights protection and high-risk AI systems: the Spanish model in gender-based violence prevention*, in *SSRN*, 16 aprile 2022, 6 s.

⁴⁶ Così *The External Audit of the VioGén System*, cit., 32 s., dove si raccomanda di «*accompanying the VioGén score with the justification of the police officers*», in quanto ciò «*would support the accountability of the Police for the risk assessment*».

⁴⁷ Con tale locuzione si indicano «le attività rivolte allo studio e all'applicazione di metodi statistici con l'obiettivo di “predire” chi potrà commettere un reato o dove e quando potrà essere commesso un reato, al fine di prevenirne la commissione»: così L. CAMALDO, *Intelligenza artificiale e investigazione penale predittiva*, in *Riv. it. dir. proc. pen.*, 1, 2024, 234. Sulla polizia predittiva, v. L. ALGERI, *Intelligenza artificiale e polizia predittiva*, in *Dir. pen. proc.*, 6, 2021, 724 ss.; K. BLOUNT, *Using artificial intelligence to prevent crime: implications for due process and criminal justice*, in *AI & Soc*, 1, 2024, 359 ss., J.L.M. MCDANIEL, K. G. PEASE (eds.), *Predictive Policing and Artificial Intelligence*, Abingdon, 2021; R. ORLANDI, *Uso poliziesco dell'intelligenza artificiale. L'insegnamento del Bundesverfassungsgericht*, in *Cass. pen.*, 7/8, 2023, 2167 ss.; B. PEREGO, *Predictive policing: trasparenza degli algoritmi, impatto sulla privacy e risvolti discriminatori*, in *BioLaw Journal – Rivista di BioDiritto*, 2, 2020, 447 ss.; E. PIETROCARLO, *Predictive policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali*, in *Sist. pen.*, 28 settembre 2023, 1 ss.

⁴⁸ Diversamente, M. BLAND, *op. cit.*, 142 ss., tratta la valutazione del rischio di reiterazione del reato accostandola alla nozione di “pre-crime”.

⁴⁹ A. VALDIVIA, C. HYDE-VAAMONDE, J. GARCÍA-MARCOS, *op. cit.*, 11 s.

⁵⁰ A. VALDIVIA, C. HYDE-VAAMONDE, J. GARCÍA-MARCOS, *op. cit.*, 14.

⁵¹ K. LA REGINA, *I.A. e ragionamento giuridico: la giustizia prevedibile*, in G.M. BACCARI, P. FELICIONI (a cura di), *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, Milano, 2023, 181.

⁵² Cfr. Commissione europea per l'efficienza della giustizia (CEPEJ), *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, Strasburgo, 3-4 dicembre 2018. In argomento, v. S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea, gli spunti per*



quelle unionali⁵³; inoltre, risulta accolto in una certa misura anche dall'*AI Act*⁵⁴, che chiarisce sin dall'esordio (sia al *considerando* 1, che all'art. 1) il suo obiettivo primario di «promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica».

Stando al regolamento sull'intelligenza artificiale, caratterizzato da un approccio «della individuazione e della gestione del “rischio” di violazione dei diritti fondamentali delle persone fisiche»⁵⁵, i software come quelli in esame sono riconducibili ai sistemi ad “alto rischio”⁵⁶ e devono sottostare a regole precise per la valutazione d'impatto, utilizzare dati ad alto standard qualitativo, avere risultati tracciabili, essere sottoposti a costante supervisione umana durante il loro utilizzo, possedere un elevato livello di robustezza, sicurezza e precisione⁵⁷. Inoltre, il Regolamento 2024/1689 non pregiudica l'esercizio dei diritti fondamentali riconosciuti dagli Stati membri, oltre che a livello di Unione, sicché operano i divieti vigenti sul piano interno, come ad esempio quello di perizia psicologica e criminologica di cui all'art. 220, comma 2 c.p.p.⁵⁸. Sempre curandosi di garantire i diritti della difesa e la presunzione di innocenza (come l'*AI Act* prevede ai *considerando* 48 e 59) e, laddove i dati da cui lo strumento predittivo provengano da risposte o informazioni fornite dallo stesso indagato o imputato, di non sopraffare la garanzia del *nemo tenetur se detegere*⁵⁹.

6. Intelligenza artificiale simbiotica e nuovi assetti paradigmatici

un'urgente discussione tra scienze penali e informatiche, in *Leg. pen. (web)*, 18 dicembre 2018, 4.

⁵³ Cfr. Direttiva 2016/680/UE del Parlamento europeo e del Consiglio del 27 aprile 2016 relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio, in G.U.U.E., 4 maggio 2016, L 119/89. A riguardo, v. G. BACCARI, *Il trattamento (anche elettronico) dei dati personali per finalità di accertamento dei reati*, in A. CADOPPI, S. CANESTRARI, A. MANNA, M. PAPA (diretto da), *Cybercrime*, Torino, 2019, 1611 ss. In particolare, l'art. 11, par. 1, della citata direttiva impone agli Stati membri di introdurre il divieto di decisioni basate «unicamente» su un trattamento automatizzato di dati, compresa la profilazione, che producano effetti giuridici negativi o incidano significativamente sui diritti fondamentali dell'interessato.

⁵⁴ Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 in G.U.U.E., 12 luglio 2024, L 1689.

⁵⁵ S. QUATTROCOLO, *Prova e intelligenza artificiale*, in C. CONTI, A. MARANDOLA (a cura di), *Prova scientifica*, Milano, 2023, 471.

⁵⁶ Cfr. l'art. 5, lett. d) nonché l'art. 6, par 2, che rinvia all'allegato III, di cui si vedano i punti 6. e 8. Secondo E. FALLETTI, *op. cit.*, VioGén è inquadrabile come “sistema IA ad alto rischio”.

⁵⁷ L'esemplificazione di tali requisiti è di S. QUATTROCOLO, *Prova e intelligenza artificiale*, cit., 472.

⁵⁸ Sulla possibilità di ricondurre gli *outputs* di software di *risk assessment* al novero delle prove generate automaticamente e, segnatamente, al paradigma della prova peritale, cfr. S. QUATTROCOLO, *Prova e intelligenza artificiale*, cit., 480. In senso contrario, ritiene che «i risk assessment tools non hanno una funzionalità tesa ad analizzare la personalità dell'autore del fatto» e che, pertanto, sarebbe «improprio ricondurli nell'ambito di operatività del divieto di cui all'art. 220 comma 2 c.p.p.», M. MONTAGNA, *Prognosi personologica, commisurazione della pena e applicazione di misure di sicurezza*, in G.M. BACCARI, P. FELICIONI, *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, cit., 241, la quale reputa pur sempre necessario «verificare, di volta in volta, la loro compatibilità con i principi fondamentali del sistema penale italiano», dovendosi impedire ogni forma di determinismo penale e che dal diritto penale del fatto si giunga a un diritto penale del profilo d'autore.

⁵⁹ M. MONTAGNA, *op. cit.*, 245.

La necessità di preservare la centralità dell'uomo ha indotto la dottrina a sostenere come sarebbe preferibile assegnare agli strumenti predittivi del rischio in esame compiti di mero *screening* iniziale, di guisa che tali *tools* verrebbero ad eseguire una «prima attività di rilevazione potenziale del rischio in modo generalizzato e standardizzato, senza creare un ingolfamento delle attività procedurali», mentre soltanto laddove tale *step* «si concluda con la rilevazione di un numero predeterminato di *red flags*, si impo[rebbe] una valutazione del rischio più accurata che può involgere accertamenti più approfonditi e il ricorso a competenze valutative specifiche», in modo da «nulla togli[re] alla (anzi, semmai restituendo) pienezza della valutazione umana»⁶⁰.

Tuttavia, siffatta tranquillante proposta pare limitativa delle potenzialità di questi strumenti, fruibili non solo per far scattare una *red flag*, ma anche per assistere l'autorità (tanto di *law enforcement* come pure quella giudiziaria) nel processo decisionale sulle misure da adottare, connettendo il singolo caso agli *n*-esimi altri "interiorizzati" dalla macchina e ricercando tra loro le interrelazioni più nascoste. Sono attività che la mente umana, computazionalmente limitata e condizionata da strategie cognitive definite euristiche, non potrebbe parimenti eseguire.

Ma finché si ragiona assumendo gli artefatti artificiali come strumenti, anziché come agenti integrati in una squadra e dotati di autonomia, adattabilità e capacità collaborative, si da creare un sistema multi-agente più ampio e più capace delle singole entità⁶¹, sarà difficile coniugare l'efficienza computazionale dei sistemi algoritmici con l'approccio incentrato sull'uomo.

Occorre, piuttosto, individuare strategie di partnership collaborativa tra esseri umani e macchine⁶², da concepirsi come entità che si aiutano a vicenda per un obiettivo comune.

Il potenziamento della collaborazione uomo-macchina e del *teaming* sociotecnico, con relazioni reciprocamente vantaggiose, che aumentino (e valorizzino) le capacità cognitive umane anziché sostituirle, è l'obiettivo dell'intelligenza artificiale simbiotica (SAI)⁶³.

Il paradigma simbiotico, che si ispira all'idea preconizzata dall'informatico e psicologo statunitense Joseph Carl Robnett Licklider nel saggio dal titolo *Man-Computer Symbiosis* (1960)⁶⁴, individua nel processo cooperativo uomo-macchina la soluzione antagonista «[a]lla sostituzione e [a]lla delega in bianco alle macchine»⁶⁵.

La sua adozione consente di superare l'interrogativo se collocare l'intelligenza artificiale alla stregua di un aiutante del giudice o di un suo sostituto. D'altro canto, la risposta fornita a tale quesito dal re-

⁶⁰ V. BONINI, *op. cit.*, 66 s.

⁶¹ T. O'NEILL, N. MCNEESE, A. BARRON, B. SCHELBLE, *Human-autonomy teaming: a review and analysis of the empirical literature*, in *Human Factors*, 5, 2022, 904 ss.

⁶² N. LETTIERI, A. GUARINO, R. ZACCAGNINO, D. MALANDRINO, *Keeping judges in the loop: a human-machine collaboration strategy against the blind spots of AI in criminal justice*, in *Soft Computing*, 16, 2023, 11275 ss.

⁶³ A. CARNEVALE, A. LOMBARDI, F.A. LISI, *Exploring Ethical and Conceptual Foundations of Human-Centred Symbiosis with Artificial Intelligence*, in G. BOELLA, F.A. D'ASARO, A. DYOUB, L. GORRIERI, F.A. LISI, C. MANGANINI, G. PRIMIERO (a cura di), *Proceedings of the 2nd Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming co-located with the 22nd International Conference of the Italian Association for Artificial Intelligence (AI*IA 2023)*, CEUR, 3615, 2023, 30.

⁶⁴ J.C.R. LICKLIDER, *Man-computer symbiosis*, in *IRE Transactions on Human Factors in Electronics*, 1, 1960, 4 ss.

⁶⁵ P. MARRA, *Giustizia digitale simbiotica e sue prospettive procedurali*, in *L'Ircocervo - First Italian digital journal of Legal Methodology, General Theory of Law and Doctrine of the State*, M.N. CAMPAGNOLI, P. MARRA (a cura di), *Artificial Intelligence and Neuro-cognitive Sciences in Law: From Symbiosis to Substitution*, 1, 2024, 26.



cente regolamento europeo che, imponendo al «processo decisionale finale d[i] rimanere un'attività a guida umana» (così il *considerando* 61), promette «la superiorità dell'intelletto umano su quello automatizzato», ha lasciato insoddisfatti, in quanto «nell'articolato normativo [si] dimentica la promessa», non individuandosi nel testo «le condizioni necessarie perché ciò accada»⁶⁶.

Lo sforzo si dovrebbe concentrare, invece, sul passaggio dalla supervisione al *teaming*, vera cifra che consente di rendere realmente significativo il controllo umano sull'algoritmo⁶⁷.

Quest'ultimo risulta concretamente possibile in quanto diviene il prodotto del contesto associativo, nel quale si registra un allineamento di valori tra gli esseri umani e l'intelligenza artificiale; l'integrazione tra il comportamento dei primi e quello dei sistemi automatizzati consente agli agenti umani non solo di sfruttare le capacità di quest'ultimi per eseguire compiti, ma anche di garantirne il controllo, preservando la «consapevolezza situazionale» da parte dell'uomo⁶⁸.

Poiché la valutazione del rischio nell'ambito della violenza di genere richiede di essere ripetuta ciclicamente, con crescente grado di complessità, sono immediatamente percepibili i vantaggi di questo approccio, superandosi la dinamica dell'«*automation bias*», la sfida dell'«*automation surprise*» per l'operatore umano e il pericolo che costui abdichi alla propria responsabilità decisionale⁶⁹, problemi tutti in varia misura registrati nell'esperienza del sistema VioGén.

La cooperazione, per essere effettiva, «non può che fondarsi sulle capacità comunicative, che non si riducono alla semplice trasmissione di contenuti, del tutto inutile se questi non vengono processati ed effettivamente compresi»⁷⁰; pertanto, la collaborazione uomo-computer richiede che gli *outputs* siano trasparenti e spiegabili⁷¹: simbioticamente, l'utente umano dovrebbe avere la possibilità d'intervenire sulla valutazione del sistema di intelligenza artificiale, di esaminare gli argomenti e, trovandosi in disaccordo con il «ragionamento» della macchina, anche di modificarli⁷². Conseguentemente, sarà anche più facile impugnare le decisioni che abbiano utilizzato l'*output* algoritmico, rispondendo più adeguatamente al problema della contestabilità, una delle nuove sfide democratiche poste

⁶⁶ G. DE MINICO, *Giustizia e intelligenza artificiale: un equilibrio mutevole*, in *Rivista AIC*, 2, 2024, 107.

⁶⁷ A. TSAMADOS, L. FLORIDI, M. TADDEO, *Human control of AI systems: from supervision to teaming*, in *AI Ethics*, 2024, p. 1 ss. G. UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sist. pen.*, 4, 2020, 83 s., chiarisce quali siano le condizioni alle quali assoggettare l'impiego della macchina in sede giurisdizionale, affinché sia mantenuto un controllo umano significativo.

⁶⁸ Sul problema della «loss of situational awareness» v. A. TSAMADOS, L. FLORIDI, M. TADDEO, *op. cit.*, 3 ss.

⁶⁹ A. TSAMADOS, L. FLORIDI, M. TADDEO, *op. cit.*, 5.

⁷⁰ P. MARRA, *op. cit.*, 24.

⁷¹ Nel senso che un approccio antropocentrico è fondamentale per l'intelligenza artificiale simbiotica e che, pertanto, la trasparenza e la *explainability* sono requisiti fondamentali per stabilire la fiducia nei sistemi SAI e renderli accettabili eticamente e legalmente, v. F. LISI, A. CARNEVALE, A. DYOUB, A. LOMBARDI, P. MARRA, L. PULITO, *Acceptability of symbiotic artificial intelligence: Highlights from the FAIR project*, in S. DI MARTINO, C. SANSONE, E. MASCIARI, S. ROSSI, M. GRAVINA (a cura di), *Proceedings of the Ital-IA Intelligenza Artificiale - Thematic Workshops co-located with the 4th CINI National Lab AIIS Conference on Artificial Intelligence (Ital-IA 2024)*, CEUR, vol. 3762, 2024, 115.

⁷² Sui modelli linguistici argomentativi di grandi dimensioni per processi decisionali spiegabili e contestabili, pensati anche per scenari complessi legali, v. G. FREEDMAN, A. DEJL, D. GORUR, X. YIN, A. RAGO, F. TONI, *Argumentative Large Language Models for Explainable and Contestable Decision-Making*, in *arXiv*, 2024 (arXiv:2405.02079), 1 ss.

dall'automazione del processo decisionale⁷³. Del resto, lo stesso *AI Act* non manca di sottolineare più volte come occorra garantire che «le previsioni, le raccomandazioni o le decisioni del sistema di IA possano essere efficacemente ribaltate e ignorate»⁷⁴.

7. Osservazioni conclusive (a partire da ChatGPT)

Interrogando *ChatGPT*, a domanda se «l'intelligenza artificiale simbiotica può aiutare le forze dell'ordine e i giudici nella valutazione del rischio nei casi di violenza di genere e nel contrasto a tale fenomeno», la *chatbot* ha fornito una risposta positiva e – almeno apparentemente – rassicurante, sostenendo che «l'intelligenza artificiale simbiotica può essere un potente alleato nella lotta contro la violenza di genere, fornendo strumenti avanzati per la valutazione del rischio, il monitoraggio, la prevenzione e l'azione legale»; ha enunciato diversi esempi; ha suggerito di considerare alcune sfide e implicazioni etiche, e ha concluso affermando che «è fondamentale implementare queste tecnologie in modo etico e responsabile, garantendo la protezione dei diritti delle vittime e la giustizia equa».

L'artefatto mostra di possedere una “consapevolezza” e una sensibilità sulle questioni di genere, terreno sfidante per la piena garanzia dei diritti e l'effettività delle tutele, che a tante volte sembra smarrirsi nell'essere umano.

Se, in teoria, le macchine offrono potenza di calcolo e capacità di automazione in grado di gestire in modo efficiente attività ripetitive e ad alta intensità di dati, mentre gli esseri umani forniscono le capacità cognitive ed emotive necessarie per la creatività e l'empatia⁷⁵, proprio quest'ultima – fondamentale per rispondere ai bisogni di chi è vulnerabile – risulta troppo spesso mancare nei casi di violenza di genere, dove si registrano da parte degli operatori comportamenti giudicanti, che possono persino indurre le vittime a rinunciare di denunciare⁷⁶.

Guardare alla simbiosi come costruito socio-tecnico e come modello procedurale del processo decisionale⁷⁷ significa altresì combinare gli approcci quantitativi con quelli qualitativi.

Il processo di operazionalizzazione cui si è accennato, necessario per allineare i valori etici e le regole giuridiche tra l'intelligenza artificiale e gli esseri umani, dovrebbe rappresentare anche l'occasione per ripensare e migliorare l'assetto esistente, allo scopo di alimentare il costante processo di rafforzamento delle tutele, di cui la Direttiva 1385/2024 rappresenta l'ultimo epilogo a livello europeo: in

⁷³ A.A. TUBELLA, A. THEODOROU, V. DIGNUM, L. MICHAEL, *Contestable Black Boxes*, in *arXiv*, 2020 (arXiv:2006.05133), 2.

⁷⁴ Cfr. il *Considerando* 141, nonché artt. 60, par. 4, lett. k) e 61, par. 1, lett. d), Regolamento (UE) 2024/1689.

⁷⁵ F. LISI, A. CARNEVALE, A. DYOUB, A. LOMBARDI, P. MARRA, L. PULITO, *Acceptability of symbiotic artificial intelligence: Highlights from the FAIR project*, *op. cit.*, 1.

⁷⁶ Cfr. *The External Audit of the VioGén System*, cit. 23. Un interessante studio sull'utilizzo di modelli di apprendimento automatico per prevedere il *disengagement* dai procedimenti legali nei casi di violenza domestica è stato portato avanti da E. ESCOBAR-LINERO, M. GARCÍA-JIMÉNEZ, M.E. TRIGO-SÁNCHEZ, M.J. CALA-CARRILLO, J.L. SEVILLANO, M. DOMÍNGUEZ-MORALES, *Using machine learning-based systems to help predict disengagement from the legal proceedings by women victims of intimate partner violence in Spain*, in *PLoS ONE*, 7 giugno 2023, 1 ss.

⁷⁷ P. MARRA, L. PULITO, A. CARNEVALE, A. LOMBARDI, A. DYOUB, F.A. LISI, *A Procedural Idea of Decision-making in the Context of Symbiotic AI*, in A. DIX, M. ROACH, T. TURCHI, A. MALIZIA, B. WILSON (a cura di), *Proceedings of the 1st International Workshop on Designing and Building Hybrid Human-AI Systems co-located with 17th International Conference on Advanced Visual Interfaces (AVI 2024)*, CEUR, vol. 3701, 2024, 4. s.



particolare, questo atto normativo si sforza di eliminare le barriere che impediscono l'accesso delle donne alla giustizia, propedeutico alla stessa valutazione del rischio, e si preoccupa della dimensione culturale del fenomeno della violenza di genere.

Così come per le macchine, è indispensabile allora che anche gli operatori siano sottoposti ad adeguato *training*, in modo da creare le opportune condizioni di supporto per la vittima⁷⁸.

Oltre alla necessaria specializzazione, andrà garantito il supporto legale e psicologico già al momento in cui vengono fornite le risposte ai questionari, delle quali i *tools* si nutrono, e che il "punteggio" sia spiegato dalla macchina e giustificato dagli operatori, anche per evitare effetti di vittimizzazione secondaria⁷⁹. Infine, occorrerà implementare i sistemi algoritmici adoperati in tale dominio mediante metodologie di co-progettazione e meccanismi di feedback con le organizzazioni femminili come un modo per esplorare le vulnerabilità attuali ed emergenti. La valutazione delle prestazioni di tali sistemi non potrà limitarsi alla loro analisi tecnica, ma è necessario che tenga conto delle esperienze, dell'applicazione pratica e delle percezioni degli *stakeholders* che si imbattono negli stessi o che li utilizzano nella loro attività lavorativa⁸⁰.

⁷⁸ Cfr. *The External Audit of the VioGén System*, cit. 25.

⁷⁹ Cfr. *The External Audit of the VioGén System*, cit., 30.

⁸⁰ Cfr. *The External Audit of the VioGén System*, cit., 30. Un primo quadro per comprendere e valutare i sistemi SAI è presentato in A. CARNEVALE, A. LOMBARDI, F.A. LISI, *A human-centred approach to symbiotic AI: Questioning the ethical and conceptual foundation*, in *Intelligenza artificiale*, 1, 2024, 9 ss.

