I PROFILI DI RESPONSABILITÀ PENALE NELL'IMPIEGO DELL'INTELLIGENZA ARTIFICIALE

QUALE FUTURO PER L'INTELLIGENZA ARTIFICIALE NEL SETTORE GIUDIZIARIO?

FIRENZE, 31 MARZO 2023

BEATRICE FRAGASSO

Dottoranda di ricerca in Diritto penale
Università degli Studi di Milano

2 Killed in Driverless Tesla Car Crash, Officials Say

"No one was driving the vehicle" when the car crashed and burst into flames, killing two men, a constable said.

Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam

Waymo's driverless cars were involved in 18 accidents over 20 months

U.S. identifies 12th Tesla Autopilot car crash involving emergency vehicle

Google's Self-Driving Car Caused Its First Crash

Google's self-driving car appears to have caused its first crash on February 14, when it changed lanes and put itself in the path of an oncoming bus.

Autonomous Drones Have Attacked Humans. This Is a Turning Point

Drone experts have long dreaded this moment.

Drones That Kill on Their Own: Will Artificial Intelligence Reach the Battlefield?

Aeronautics | Artificial Intelligence | Big Data | Future | Machine learning | Machine learning

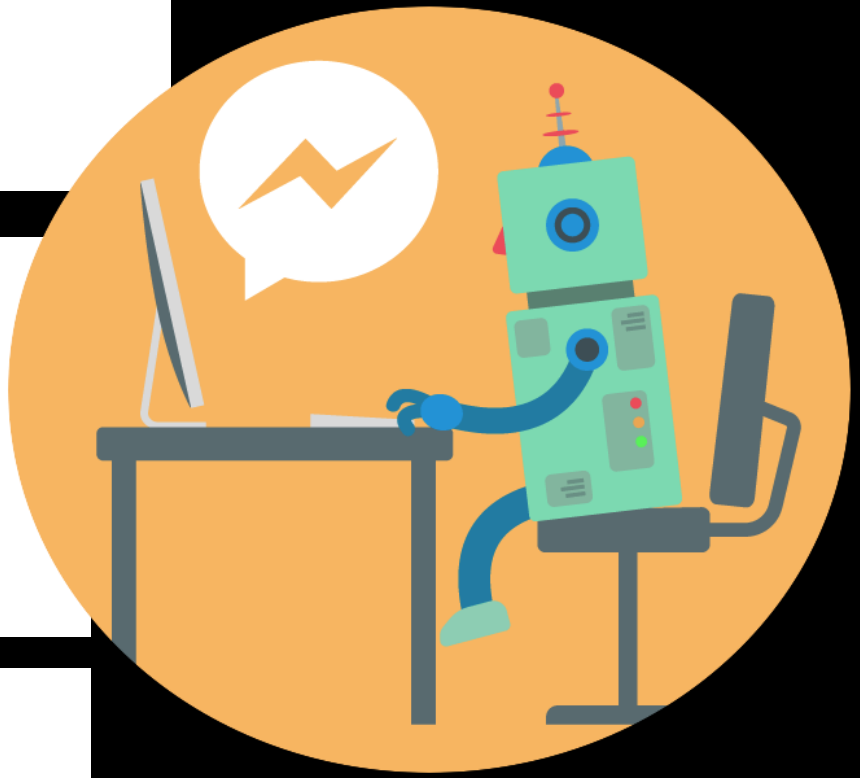
A.I. Drone May Have Acted on Its Own in Attacking Fighters, U.N. Says

A United Nations report suggested that a drone, used against militia fighters in Libya's civil war, may have selected a target autonomously.

Randomly generated tweet by bot prompts investigation by Dutch police

Police investigate after bot created by web developer Jeffry van der Goot tweets 'I seriously want to kill people'

Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk.



Swiss police release robot that bought ecstasy online

The robot - which goes by the name Random Darknet Shopper - was part of an art installation meant to explore the dark web

LA DEFINIZIONE DI INTELLIGENZA ARTIFICIALE FORNITA DALL'AI ACT

Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale, COM/2021/206 final, 21 aprile 2021 (cd. AI act) – art. 3, lett. a)

> Il sistema di i.a. è «un *software* sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare *output* quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono»

UNA NOZIONE «PENALISTICAMENTE ORIENTATA» DI INTELLIGENZA ARTIFICIALE

AUTONOMIA

Capacità di compensare l'incompletezza delle informazioni ricevute in partenza attraverso l'apprendimento; capacità di prendere decisioni in situazioni di incertezza (*Decision under Uncertainty*)

INTERATTIVITÀ

Capacità di interagire con l'ambiente fisico esterno (es. internet of things) e con altri dispositivi dotati di intelligenza artificiale.

OPACITÀ (C.D. BLACK BOX)

Imperscrutabilità dei meccanismi causali interni ai sistemi di i.a. È possibile individuare input e output, ma non è invece possibile ricostruire cosa accade all'interno della scatola nera, ovverosia la catena causale che ha determinato il passaggio dagli input agli output.

LA RESPONSABILITÀ CIVILE PER DANNO DA PRODOTTO

- Dir. 85/374/CEE (Cod. Consumo, artt. 114-127):

  Responsabilità oggettiva del produttore per il danno cagionato dal difetto del prodotto.

- Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi, COM (2022) 495 final, 28 settembre 2022:
  - Modifica delle nozioni di «prodotto» e «produttore»
  - Modifica della nozione di «difetto»
  - Meccanismi di *disclosure* obbligatoria in capo al produttore
  - Inversione dell'onere probatorio: presunzione di difettosità del prodotto e presunzione di causalità tra difetto e danno.

LA RESPONSABILITÀ PENALE DELL'UTILIZZATORE

Utilizzatore di sistema di i.a. TOTALMENTE autonomo

- Assoluta perdita di dominio sull'attività algoritmica.
- Eventuale predisposizione di posizioni di controllo: da responsabilità commissiva a responsabilità omissiva.

Utilizzatore di sistema di i.a. PARZIALMENTE autonomo

- La funzione di governo è condivisa tra operatore e algoritmo → profili di colpa relazionale.
- L'esigibilità della conformazione alle norme cautelari può venire meno a causa di:
  - opacità;
  - distorsione dell'automazione;
  - ridotta capacità di reazione

LA RESPONSABILITÀ PENALE DEL PRODUTTORE

Il danno da dispositivo intelligente ripropone alcuni dei profili più problematici già sorti in relazione alla responsabilità penale per danno da prodotto:

- sostanziale fungibilità tra struttura commissiva e omissiva del reato
- difficile individuazione dei soggetti personalmente responsabili all'interno delle organizzazioni complesse
- accertamento del nesso di causalità in relazione a prodotti dotati di *black box*
- accertamento dell'elemento soggettivo in contesti di incertezza scientifica

IL NESSO DI CAUSALITÀ

Attualmente, non sussistono leggi scientifiche consolidate che siano in grado di descrivere il dispiegarsi della catena causale nel funzionamento degli algoritmi di *machine learning*.

Discrasia tra approccio deterministico e modello probabilistico tipico del *machine learning*?

EVENT DATA RECORDER

Gli Event Data Recorder (c.d. EDR; le «scatole nere») non sono strumenti risolutivi: pur potendo fornire dati utili per la ricostruzione della dinamica dell'incidente, le scatole nere non sono in grado di rimediare al *deficit di comprensibilità* delle decisioni algoritmiche.

Ipotizzabile, in futuro, lo sviluppo di sistemi di ricostruzione *ex post* dell'evento lesivo (*reverse engineering*).

In ogni caso, sarà possibile individuare una regolarità causale?

ELEMENTO SOGGETTIVO: LA COLPA

Gli eventi lesivi scaturenti dai sistemi di i.a. possono considerarsi

→ prevedibili per quanto concerne l'*an*

→ imprevedibili con riferimento al *quantum* e al *quomodo*

IMPREVEDIBILITÀ GENERICAMENTE PREVEDIBILE

RISCHIO CONSENTITO

- «…area di condotte pericolose, ammesse dall'ordinamento nonostante che l'adozione di cautele idonee a contrastare i possibili svolgimenti lesivi sia destinata a residuare un certo grado di pericolosità» [G. Forti]

- C.d. "rischio residuale (*Restrisiko*): pericolo marginale che le misure preventive non sono in grado di disinnescare e che ricade sulla società.

- L'area di impermeabilità alla colpa generica è delimitata da questa ipotesi di ordinario fallimento della norma cautelare

LE REGOLE CAUTELARI SCRITTE: LA COLPA SPECIFICA

AI ACT (2021)

- adozione di un “sistema di gestione dei rischi” (art. 9);
- utilizzo di dataset “di qualità” (art. 10);
- predisposizione di idonea documentazione tecnica (art. 11);
- installazione di event data recorder sui dispositivi (art. 12);
- previsione di meccanismi che garantiscano il principio dell’*human-in-the-loop* (art. 14).

Standard ISO (International Organization for Standardization) e IEC (International Electrotechnical Commission)

- ISO/IEC TR 24027:2021 – Bias in AI systems and AI aided decision making
- ISO/IEC TR 24368:2022 – Overview of ethical and societal concerns
- ISO/IEC TR 24028:2020 – Overview of trustworthiness in artificial intelligence
- ISO/IEC TR 24029-1:2021 – Artificial Intelligence (AI) – Assessment of the robustness of neural networks

UNA TUTELA ANTICIPATA NEI CONFRONTI DEI SISTEMI DI I.A. PERICOLOSI?

PROSPETTIVE *DE JURE CONDENDO*

- Art. 71, § 1, AI Act: Gli Stati Membri dovranno introdurre sanzioni «effettive, proporzionate e dissuasive» per il caso di mancato rispetto della disciplina stabilita dall'AI Act

- Reati colposi di mera condotta:

(i) omessa predisposizione, da parte del programmatore, di presidi di sicurezza;

(ii) disattivazione, mancata attivazione o mancato aggiornamento, da parte dell'utilizzatore, dei presidi di sicurezza.

- Modello ingiunzionale

GRAZIE
PER L'ATTENZIONE!

Beatrice.fragasso@unimi.it