Terrorismo e intelligenza artificiale

L'intelligenza artificiale

Non esiste un'unica definizione riconosciuta di intelligenza artificiale (nell'espressione originale in inglese *artificial intelligence*, AI). In questa sede, l'IA può essere sinteticamente definita come una disciplina dedicata all'elaborazione e allo sviluppo di sistemi informatici in grado di eseguire compiti che normalmente richiedono l'impiego dell'intelligenza umana (come la percezione visiva, il riconoscimento vocale, la traduzione in altre lingue, il processo di presa delle decisioni e la risoluzione di problemi). Questi sistemi intelligenti artificiali possono assumere la forma, per esempio, di applicazioni software (*apps*), di robot o di automobili autonome.

Sebbene il concetto e la disciplina accademica dell'intelligenza artificiale siano emersi già negli anni Cinquanta, le sue applicazioni più rilevanti sono apparse soltanto nel XXI secolo, con l'enorme incremento della potenza di calcolo dei computer e con la diffusione di internet, e hanno registrato un notevole avanzamento a partire dal 2022, con l'avvento di piattaforme online di IA di facili accesso e utilizzo, come ChatGPT (la celebre applicazione software progettata per simulare conversazioni simili a quelle degli esseri umani, lanciata nel novembre 2022).

Oggi l'IA comprende molti sottocampi diversi (tra gli altri, si veda UNOCT-UNICRI 2021a). Tra i più rilevanti vi è il cosiddetto *machine learning* («apprendimento automatico»), una branca dell'IA che riguarda gli algoritmi in grado di «imparare» dai dati, ovvero di migliorare progressivamente le prestazioni su un compito specifico.

Un sistema di intelligenza artificiale può includere, in aggiunta a un algoritmo di apprendimento automatico, sensori e dispositivi esterni necessari per eseguire tale compito; per esempio, un sistema di «visione artificiale» (*computer vision*) è composto da un *software* di riconoscimento delle immagini e da una o più telecamere per catturare l'immagine che l'algoritmo andrà a elaborare.

Un'area dell'IA che ha ricevuto molta attenzione è poi quella dell'«elaborazione del linguaggio naturale» (*natural language processing,* NLP). Facendo affidamento su una memoria interna che elabora sequenze di input, una macchina può, per esempio, eseguire il riconoscimento vocale, comprendendo sequenze di parole, la loro morfosintassi e funzione semantica. Oltre al riconoscimento vocale, la NLP può essere utilizzata anche per la generazione (*generation*) di testi. Questo processo costituisce la base dei cosiddetti «*chatbots*» (o *chatterbots*) – ossia quei programmi software online che simulano conversazioni di base, come la già ricordata ChatGPT.

Tra le molteplici applicazioni dell'IA si può contare anche il confezionamento dei cosiddetti *deepfakes*; il termine, coniato nel 2017 combinando le espressioni inglesi *deep learning* e *fake media*, si riferisce a un tipo di strumenti che usa l'intelligenza artificiale per manipolare o per generare contenuti visivi e audio falsi, ma difficilmente identificabili come tali nell'immediato. A causa della portata e velocità di internet e delle applicazioni di messaggistica online, i *deepfakes* possono raggiungere rapidamente grandi quantità di utenti. Questi contenuti *fake* vengono generalmente considerati come un'arma sempre più saliente nelle odierne campagne di disinformazione e si prestano a un utilizzo anche nei conflitti armati (Byman et al. 2023); potrebbero essere utilizzati anche da estremisti violenti e terroristi (cfr. Europol 2022). Anche se la natura fasulla di questi video o audio può essere verificata o rivelata nel corso del tempo, essi possono determinare conseguenze assai rilevanti nel breve periodo (per esempio, produrre sentimenti di terrore, panico, confusione oppure provocare immediate reazioni ostili).

Intelligenza artificiale e implicazioni per il terrorismo

L'IA è uno strumento assai potente al servizio di diversi attori e interessi (tra gli altri, si veda Kissinger et al. 2021). È già utilizzato nel settore privato e pubblico con l'ambizione di rendere gli individui, e la società in generale, più soddisfatti, più sani, più prosperi e più sicuri, oltre che per scopi di intrattenimento (UNOCT-UNICRI 2021a).

L'IA ha ulteriormente aumentato la propria popolarità nel corso del 2022 con la distribuzione di apprezzati prodotti online di pubblico accesso, come il *chatbot* ChatGPT; si stima che a gennaio 2023, appena poche settimane dopo il suo lancio, ChatGPT avesse già superato i 100 milioni di utenti e fosse diventata l'*app* con la crescita di utenti più veloce della storia (Hu 2023).

In aggiunta a meriti e benefici – incluse potenziali opportunità per le stesse attività di contrasto e di prevenzione del terrorismo (per esempio, UNOCT-UNICRI 2021b; Montasari 2022) – l'intelligenza artificiale, ovviamente, può anche avere un «lato oscuro»: in quanto tecnologia aperta a molteplici scopi, l'IA può essere utilizzata in modo improprio oppure deliberatamente malevolocon l'intenzione di recare danno ad altri attori.

In questo contesto, l'intelligenza artificiale potrebbe anche rappresentare uno strumento efficace in mano a estremisti violenti e terroristi. A questo proposito, è opportuno osservare che, ad oggi, non vi sono ancora prove o indicazioni chiare dell'effettivo utilizzo diretto dell'IA a scopi genuinamente terroristici (ovvero relativi all'esecuzione di atti di violenza fisica); nondimeno, la situazione potrebbe cambiare in futuro. Fin d'ora, inoltre, l'IA può giocare un ruolo rilevante in relazione ad altri ambiti dell'estremismo violento, come la propaganda.

A questo proposito, è importante sottolineare che la letteratura in materia è ancora limitata e al momento non riflette pienamente l'impatto degli ultimissimi sviluppi tecnologici, come il lancio della popolare ChatGPT.

Si possono distinguere schematicamente tre ambiti principali del potenziale impiego dell'IA da parte di terroristi ed estremisti violenti: 1) attacchi *cyber*; 2) attività di propaganda, manipolazione e influenza; 3) attacchi terroristici (fisici).

IA e terrorismo: attacchi nello spazio cibernetico

In generale, le minacce di natura informatica rappresentano, com'è noto, un'area di crescente preoccupazione, considerando le vulnerabilità intrinseche dello spazio cibernetico.

A titolo di esempio, tra le minacce cibernetiche oggi più comuni vi sono senz'altro gli attacchi di tipo DoS (*Denial of Service*) o DDoS (*Distributed Denial of Service*). Lo scopo ultimo di queste azioni ostili è rendere temporaneamente indisponibile per i propri utenti un sistema informatico connesso a internet, esaurendone la memoria attraverso molteplici richieste di connessione. Negli attacchi DDoS gli aggressori utilizzano più di una e spesso migliaia di macchine (le cosiddette *botnets*) per indirizzare le richieste contro il sistema informatico preso di mira, rendendo difficile rintracciare l'attaccante originario. Anche senza l'AI, gli attacchi DoS o DDoS possono essere oggi lanciati con un livello di sforzo relativamente ridotto.

Attraverso metodi di *machine learning*, essi possono diventare ancora più semplici e incisivi, grazie a un'automatizzazione dei processi tradizionalmente eseguiti dall'attaccante. Per esempio, gli algoritmi di apprendimento automatico possono essere utilizzati per controllare le *botnets* impiegate per l'attacco o possono consentire loro di identificare i sistemi vulnerabili attraverso una sofisticata ricognizione della rete (UNOCT-UNICRI 2021a).

IA e terrorismo: attività di propaganda e influenza

In generale, l'intelligenza artificiale potrebbe essere utilizzata per svolgere attività di propaganda, manipolazione e influenza a sostegno di cause estremistiche violente.

Per esempio, l'AI potrebbe essere utilizzata per rendere più accurata e rapida la ricerca di canali e materiali di propaganda estremistica già presenti su internet (Lakomy 2023).

Una funzione probabilmente ancora più saliente è quella di generare nuovi materiali e narrazioni radicalizzanti. L'avvento di modelli generativi (*generative*) di intelligenza artificiale, infatti, potrebbe consentire a una varietà di attori, compresi estremisti violenti e terroristi, di produrre una maggiore quantità di prodotti di propaganda, notevolmente più sofisticati, e con uno sforzo nettamente inferiore. Già oggi *chatbots* di facile accesso rischiano di aiutare l'utente interessato a sviluppare testi o narrazioni che espongano persino richiami all'estremismo violento o a miti complottisti radicali (cfr. Europol 2023), aggirando le restrizioni previste dalla relativa piattaforma. La generazione di testi potrebbe essere poi combinata alla creazione di *deepfakes* video o audio.

Secondo alcuni esperti, in futuro i modelli generativi di IA saranno in grado di trasformare profondamente il panorama della comunicazione e propaganda estremistica (Siegel e Bennett Doty 2023) con la produzione su misura di video, audio, videogiochi e così via.

Nuove tecniche avanzate nel *natural language processing* hanno sollevato, inoltre, preoccupazioni rispetto al potenziale utilizzo della tecnologia nella «micro-profilazione». L'IA potrebbe essere impiegata per individuare online singoli individui che siano più vulnerabili alla radicalizzazione, consentendo poi una distribuzione mirata di contenuti o messaggi estremistici, per mezzo di cosiddetti *algorithmic amplifiers*; potrebbero, per esempio, indirizzare messaggi mirati a individui che abbiano già cercato ripetutamente contenuti violenti online (UNOCT-UNICRI 2021a).

IA e terrorismo: attacchi terroristici

Non si può escludere, infine, che l'IA possa essere impiegata direttamente per preparare e portare a termine attacchi terroristici nel mondo reale.

Alcuni sviluppi dell'IA, specialmente nel campo della visione artificiale, potrebbero ipoteticamente rendere, in primo luogo, più semplice e veloce la fase di ricognizione e sorveglianza degli obiettivi da parte di terroristi. Con l'aiuto della tecnologia, gli attentatori sarebbero in grado, per esempio, di monitorare luoghi e seguire i movimenti delle persone, identificare individui ed edifici e valutare automaticamente, e a distanza, la natura delle misure di sicurezza fisica nel luogo preso di mira.

L'IA potrebbe essere utilizzata anche per consentire a simpatizzanti estremisti di ottenere informazioni utili per pianificare un attacco terroristico, fornendo per esempio istruzioni per fabbricare ordigni esplosivi (Lakomy 2023).

L'IA potrebbe trovare un impiego anche nell'esecuzione stessa di atti di violenza, per esempio attraverso il ricorso a veicoli autonomi. Considerando la lunga storia dell'uso di veicoli per atti di terrorismo (tra gli altri, Davis 2007), una maggiore autonomia delle automobili potrebbe costituire uno sviluppo favorevole anche per i terroristi, consentendo loro di effettuare efficacemente, e a distanza, un tipo di attacco che è diventato comune negli ultimi anni anche in Europa (si veda Marone 2021), senza la necessità che un membro del gruppo sacrifichi la propria vita o rischi di essere arrestato.

L'intelligenza artificiale potrebbe essere applicata anche all'uso di droni (cfr. Grossman 2018). Oltretutto, aeromobili a pilotaggio remoto potrebbero essere dotati di dispositivi per il riconoscimento facciale, attingendo a *databases* di immagini: in questo modo, un drone carico di esplosivo potrebbe identificare un bersaglio specifico e colpirlo. Fortunatamente, ad oggi, questo tipo di tecnologia presumibilmente non esiste in un formato «pronto all'uso»; nondimeno, rischi per il futuro non possono essere esclusi.

Conclusioni

L'intelligenza artificiale potrebbe diventare uno strumento potente e versatile in mano anche a estremisti violenti e terroristi. La manifestazione effettiva di questo rischio dipenderà dalla combinazione di intenzione e capacità di usare efficacemente questo tipo di tecnologia.

Da un lato, è presumibile che estremisti violenti e terroristi abbiano già maturato l'intenzione di utilizzare l'IA per perseguire i propri obiettivi (tra gli altri, UNOCT-UNICRI 2021a).

Dall'altro lato, ad oggi, appare invece poco probabile che essi abbiano le effettive capacità richieste per utilizzare efficacemente l'IA. Nonostante un processo generale di abbassamento della soglia di accesso alle e di uso delle tecnologie (in particolare, Kurth Cronin 2019), verosimilmente ben poche organizzazioni terroristiche possono vantare individui così qualificati nei propri ranghi, tanto più se esse non sono nelle condizioni di controllare un territorio e una popolazione (come era in grado di fare il cosiddetto Stato Islamico o *Dāʿish* ai tempi del «califfato» territoriale in Siria e Iraq). In linea di principio, esiste anche l'opzione di acquisire questo *expertise* da attori esterni, in cambio di una remunerazione o con altri metodi (per esempio, minacce o ricatti), ma al momento questa eventualità appare nel complesso poco probabile (Bazarkina 2023).

Nondimeno, questi apparenti vincoli e limiti odierni all'impiego dell'intelligenza artificiale nel campo dell'estremismo violento e del terrorismo potrebbero ridursi in futuro.

Bibliografia

- Bazarkina D. (2023). Current and Future Threats of the Malicious Use of Artificial Intelligence by Terrorists: Psychological Aspects. In E. Pashentsev, a cura di, *The Palgrave Handbook of Malicious Use of AI and Psychological Security*, Palgrave Macmillan – Springer: Cham, 251-272.
- Byman D.L., Gao C., Meserole C. e Subrahmanian V.S. (2023). *Deepfakes and international conflict*. Report. Brookings Institution, testo disponibile al sito: https://www.brookings.edu/articles/deepfakes-and-international-conflict/ (consultato il 4 ottobre 2023).
- Davis, M. (2007). *Buda's Wagon: A Brief History of the Car Bomb*, Verso Books: London.
- Europol (2022). *Facing reality? Law enforcement and the challenge of deepfakes*, Report, Europol, Europol Innovation Lab, testo disponibile al sito: https://www.europol.europa.eu/publications-events/publications/facing-reality-law-enforcement-and-challenge-of-deepfakes (consultato il 4 ottobre 2023).
- Europol (2023). *ChatGPT: The impact of Large Language Models on Law Enforcement*, Europol, testo disponibile al sito: https://www.europol.europa.eu/publications-events/publications/chatgpt-impact-of-large-language-models-law-enforcement (consultato il 4 ottobre 2023).
- Grossman N. (2018). *Drones and Terrorism: Asymmetric Warfare and the Threat to Global Security*, I.B. Tauris: London.
- Hu K. (2023). ChatGPT sets record for fastest-growing user base - analyst note. *Reuters*, 2 febbraio, testo disponibile al sito: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (consultato il 4 ottobre 2023).
- Kissinger H., Huttenlocher D. e Schmidt E. (2021). *The Age of AI: And Our Human Future*. John Murray: London.
- Kurth Cronin A. (2019). *Power to the people: How open technological innovation is arming tomorrow's terrorists*. Oxford University Press: Oxford.
- Lakomy, M. (2023). Artificial Intelligence as a Terrorism Enabler? Understanding the Potential Impact of Chatbots and Image Generators on Online Terrorist Activities. *Studies in Conflict & Terrorism*, online first.
- Marone F. (2021). A Farewell to Firearms? The logic of weapon selection in terrorism: the case of jihadist attacks in Europe. *Global Change, Peace & Security*, 3: 221-240.
- Montasari, R., a cura di (2022). *Artificial Intelligence and National Security*. Springer: Cham.
- Siegel D. e Bennett Doty M. (2023). Weapons of Mass Disruption: Artificial Intelligence and the Production of Extremist Propaganda. *Global Network on Extremism and technology (GNET)*, 17 febbraio, testo disponibile al sito: https://gnet-research.org/2023/02/17/weapons-of-mass-disruption-artificial-intelligence-and-the-production-of-extremist-propaganda/ (consultato il 4 ottobre 2023).
- UNOCT - UNOCT (2021a). *Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes*. Report. United Nations Office of Counter-Terrorism (UNOCT) and United Nations Interregional Crime and Justice Research Institute (UNICRI).
- UNOCT - UNOCT (2021b). *Countering Terrorism Online with Artificial Intelligence.* Report. United Nations Office of Counter-Terrorism (UNOCT) and United Nations Interregional Crime and Justice Research Institute (UNICRI).