ISSN 2612-047X

DOSSIER

MACHIAVELLI

Suadere atque agere

centro studi politici e strategici

n. 44 - gennaio 2024

IL RUOLO DELL'INTELLIGENZA ARTIFICIALE NELLE GUERRE IBRIDE

Emanuel Pietrobon

THE AVGVSTVS FOVNDATION

Con il contributo di Augustus Foundation

MachiavelliDossier

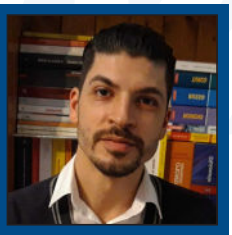
n. 44 - gennaio 2024

«**Il ruolo dell'intelligenza artificiale nelle guerre ibride**»
di Emanuele Pietrobon

ISSN 2612-047X

SOMMARIO ESECUTIVO

- **L'intelligenza artificiale è destinata a trasfigurare profondamente quotidianità, relazioni sociali, mondo del lavoro e guerre. L'impatto dell'intelligenza artificiale sulle guerre riguarderà sia la tecnologia militare impiegata nei conflitti convenzionali sia l'arsenale delle guerre ibride.**
- **Il perfezionamento di strumenti attualmente in fase embrionale, come *deepfake* e *chatbot*, aumenterà in maniera eccezionale la carica destabilizzativa di guerre informative e cognitive, mettendo a rischio la lucidità mentale di singoli e collettività, la stabilità sociale e i processi elettorali nei Paesi con meno difese.**
- **I *deepfake* sono creazioni audio, foto e video realizzate con l'intelligenza artificiale che si distinguono per l'elevata accuratezza. Sono dei potenti vettori di disinformazione già oggi e un domani il progresso tecnologico potrebbe renderli a prova di *fact-checking*.**
- **Gli assistenti virtuali intelligenti potrebbero diventare delle armi in grado di eseguire sabotaggi, operazioni cognitive e sorveglianza ai danni degli utilizzatori. Nel prossimo futuro aumenterà la commercializzazione di programmi, apparentemente innocui, progettati per veicolare narrazioni, propagare disinformazione e aggredire la mente dell'utenza.**
- **Il metaverso è una realtà che richiede una visione a lungo termine che sia in grado di coniugare profitto e sicurezza nazionale, perché sarà il teatro venturo di spionaggio, operazioni cognitive e campagne di radicalizzazione.**
- **Il piano nazionale di prevenzione e contrasto delle nuove minacce provenienti dall'intelligenza artificiale dovrebbe prevedere investimenti nella formazione di nuove figure professionali, legislazioni *ad hoc* contro la disinformazione avanzata e le operazioni cognitive, nonché un dibattito politico sulla necessità di creare un metaverso italiano.** ■



L'AUTORE

Emanuel Pietrobon è analista geopolitico, consulente di politica estera e scrittore. Laureato in *Area and Global Studies for International Cooperation* (Università di Torino), si è formato tra Polonia (Accademia di Umanistica e di Economia di Łódź), Portogallo (Lusiada di Lisbona) e Russia (*Higher School of Economics* di San Pietroburgo), ed è specializzato in guerre ibride, questioni latinoamericane e spazio postsovietico. ■

L'era dell'intelligenza artificiale (IA) è agli albori e la sua prole, dagli assistenti virtuali per casa e ufficio ai roboperai nelle fabbriche, promette di trasfigurare ogni aspetto della quotidianità umana.

Le opinioni degli esperti del settore divergono sul rapporto costi-benefici che accompagnerà la rivoluzione dell'IA, ma tutti concordano su due punti: a) non sarà priva di rischi, dall'inebetimento collettivo alla disoccupazione cronicizzata; b) l'entrata dell'umanità nell'era dell'ibridazione uomo-macchina e dei cervelli artificiali impatterà profondamente anche sulle guerre.

Mentre i conflitti convenzionali saranno dominati da droni, veicoli e velivoli computer-guidati capaci di condurre operazioni dalla precisione chirurgica, robo-soldati, sistemi d'arma offensivi e/o difensivi gestiti dall'IA e macchine decisionali capaci di pensare, calcolare e prevedere al posto dei generali[1], i conflitti ibridi vedranno l'impiego di *chatbot* maligni, di *deepfake* e la militarizzazione di dispositivi intelligenti e del metaverso.

L'impatto dell'IA sulle guerre ibride renderà più complesse le relazioni tra i domini della conflittualità, aumentandone peraltro il numero, rendendo obbligatoria l'introduzione di strumenti legislativi orientati al futuro, di nuove figure professionali, di organismi *ad hoc* e di investimenti multisettoriali per fronteggiare le applicazioni dell'IA nelle guerre ibride.

Intelligenza artificiale, presente e futuro dell'umanità (e delle guerre)

L'Europarlamento stima che gli investimenti e i progressi nell'IA porteranno, entro il 2035, alla creazione di una «nuova forza lavoro virtuale», basata sull'«automazione intelligente», votata all'autoapprendimento e capace di autoperfezionamento[2].

I *big data* raccolti, processati e analizzati dalle cellule dell'Internet delle Cose (IoT, *Internet of Things*) saranno la spina dorsale della prossima generazione di algoritmi, prodotti, servizi e tecnologie basate sull'IA, che avrà un impatto socioeconomico eccezionale a livello globale. Entro il 2030, secondo McKinsey e PricewaterhouseCoopers, il 70% delle compagnie di ogni dimensione potrebbe utilizzare «almeno un tipo di tecnologia basata su IA» e quasi il 50% delle grandi aziende potrebbe adottare «l'intera gamma delle tecnologie basate su IA», col risultato di un aumento del PIL globale trainato dall'IA del 14%[3].

Entro la metà del secolo, per esempio, quando sarà pienamente avvenuta la fusione di «tecnologie digitali, Internet e produzione manifatturiera in un unico sistema cyber-fisico», si stima che

1 Anthony King, *AI at War*, "War on the Rocks", 27 aprile 2023; Stephen Kelly, *Here's how a war between AI and humanity would actually end*, "BBC Science Focus", 29 settembre 2023.

2 *Economic impacts of artificial intelligence*, Parlamento Europeo, 2019.

3 Ethan Ilzetzki, Jain Suryaansh, *The impact of artificial intelligence on growth and employment*, CEPR, 20 giugno 2023.

nella sola UE il 54% del mercato del lavoro sarà stato divorato dalla computerizzazione[4].
La nascita di società a medio-alta disoccupazione fisiologica, causa l'automazione dei processi lavorativi, non è l'unico pericolo da considerare. Altre due questioni che i decisori politici dovranno trattare sono l'istupidimento delle masse alimentato dalla diffusione degli assistenti personali intelligenti e l'aggravamento delle guerre ibride. Il primo fenomeno sta accelerando la tendenza pregressa, innescata dall'avvento dell'Internet, del declino cognitivo degli esseri umani, incidendo negativamente su abilità di calcolo, creatività, propensione agli sforzi fisici e mentali, ragionamento e socialità[5]. Il secondo richiederà l'adozione di strategie di anticipazione e adattamento in tempi utili, pena conseguenze perverse a livello sociale.

L'IA porterà a una società a disoccupazione fisiologica medio alta. Inoltre vi sarà un aumento dell'inebetimento generale causato dalla diffusione degli assistenti in IA e l'aggravamento delle guerre ibride

La militarizzazione dell'assistenza personale intelligente

Una delle principali manifestazioni della seconda generazione dell'IA è l'assistenza intelligente, di cui i *chatbot* come ChatGPT e i dispositivi come Alexa rappresentano i prodotti più noti e utilizzati.
Le differenze intercorrenti tra *chatbot* e dispositivi di assistenza sono sostanziali: i primi sono dotati di un'intelligenza artificiale forte, i secondi di una debole. Le IA forti o generali possiedono la facoltà dell'apprendimento automatico e profondo, che permette loro di assorbire, comprendere ed emulare le abilità e le capacità cognitive degli esseri umani. Le IA deboli o ristrette sono progettate per interpretare il linguaggio naturale e rispondere ai comandi vocali ricevuti. Entrambe sono suscettibili alla militarizzazione.
Il punto di forza delle IA generali, l'apprendimento automatico (e profondo), sarà il punto debole della collettività. La loro intelligenza incrementale si deve all'immagazzinamento permanente in uno storico degli errori commessi e al metodo delle reti generative avversarie[6].
I programmi basati su IA forti dispongono di una «forma di intelligenza completamente differente, nuova e migliore» rispetto a quella degli esseri umani, che si esplica nell'elaborazione

4 Anand S. Rao, Gerard Verweij, *Sizing the prize What's the real value of AI for your business and how can you capitalise?*, *Pricewaterhouse-Coopers,* 2017; Jeremy Bowles, *Chart of the Week: 54% of EU jobs at risk of computerisation*, Bruegel, 24 luglio 2014.

5 Ahmad, S.F., Han, H., Alam, M.M. et al, *Impact of artificial intelligence on human loss in decision making, laziness and safety in education*, "Humanit Soc. Sci. Commun.", 10, 311 (2023); Nelson Grabados, *Human Borgs: How Artificial Intelligence Can Kill Creativity And Make Us Dumber*, "Forbes", 31 gennaio 2022.

6 La rete generativa avversaria è un metodo di apprendimento automatico sviluppato da Ian Goodfellow. Il modello prevede l'addestramento delle reti neurali artificiali in giochi a somma zero pensati per risolvere *bug* e *deficit*. Il risultato dello scontro è il miglioramento progressivo e incrementale dell'IA impegnata in suddetti giochi.

di miliardi di dati in breve tempo e nell'apprendimento quasi istantaneo delle conoscenze[7].
Possesso di una vasta gamma di conoscenze e propensione all'autoperfezionamento sono gli elementi che possono rendere le IA forti, nelle mani sbagliate, delle armi ad altissimo potenziale destabilizzativo. I *chatbot* semplici come ChatGPT, o complessi come Ion[8], vengono utilizzati in maniera crescente dagli utenti come se fossero degli insegnanti, dei consulenti e degli psicologi.
Il successo di ChatGPT, che ha superato quota cento milioni di utenti in soli sessanta giorni, è profetico circa il ruolo che i *chatbot* ricopriranno nelle società dell'informazione di domani[9]. ChatGPT e affini possono rispondere a quesiti storici e di attualità dei curiosi, elaborare tesine per gli studenti, proporre soluzioni ai problemi personali degli utilizzatori, scrivere articoli per i giornalisti e libri per gli scrittori. Sono dei contenitori di informazioni che, dietro l'apparenza di intelligenza superiore, neutralità e obiettività, operano (già oggi) come agenti delle guerre informative, psicologiche e cognitive.
Facili da utilizzare, iperproduttivi ed economici, ma soprattutto votati all'automiglioramento, i *chatbot* stanno trovando crescente impiego in rete per sfornare articoli acchiappa-clic su scala industriale e per veicolare disinformazione. Un'indagine di "NewsGuard", condotta nell'aprile-ottobre 2023, ha scoperto «516 siti di notizie e informazioni inaffidabili generate dall'IA […] che agli occhi di un lettore comune potrebbero sembrare legittimi […] ma operano con poca o nessuna supervisione umana e pubblicano articoli scritti in gran parte o interamente da *bot*», coprendo i temi più svariati, politica inclusa, e diffondendo, non di rado, bufale, semi-verità e notizie alterate[10]. Il problema riguarda anche l'Italia: un decimo di tutti i siti di cui sopra pubblica in italiano[11].
Ricerche sulla militarizzazione di ChatGPT, condotte a meno di un anno dal suo lancio, hanno confermato che può essere adoperato «per produrre testi chiari e convincenti che ripetono teorie del complotto e narrazioni fuorvianti» e che la tecnologia alla sua base ha il potenziale per dare vita ai «più potenti strumenti per la diffusione di misinformazione che siano mai esistiti nell'Internet»[12]. Gli esperimenti di "NewsGuard" sullo stesso tema hanno dimostrato che ChatGPT-3.5 e ChatGPT-4 possono prestarsi alla realizzazione di dettagliata e persuasiva disinformazione *on demand*, soddisfacendo rispettivamente l'80% e il 100% delle richieste ricevute[13].

7 David Hamilton, *The 'godfather of AI' says he's scared tech will get smarter than humans: 'How do we survive that?'*, Fortune, 4 maggio 2023.

8 Ion è il nome del roboconsigliere assunto dal governo Ciucă, in Romania, e ha rappresentato il primo caso di IA assunta da un esecutivo per scopi di policy e advising. Per sapere di più: Sophia Khatsenkova, *Romania's prime minister has hired the world's first AI government adviser. What will it do?,* "Euronews", 6 marzo 2023.

9 Kristal Hu, *ChatGPT sets record for fastest-growing user base - analyst note*, "Reuters", 2 febbraio 2023.

10 *Centro di monitoraggio sull'IA: i 516 siti inaffidabili di 'notizie generate dall'intelligenza artificiale' (in continua crescita) e le principali narrazioni false prodotte da strumenti basati sull'IA*, "NewsGuard", 16 ottobre 2023.

11 Marco Boscolo, *Plagio automatizzato: l'informazione online e le fake news generate dall'IA*, "Il BOLive – Università di Padova", 21 settembre 2023.

12 Tiffany Hsu, Stuart Thompson, *Disinformation Researchers Raise Alarms About A.I. Chatbots*, "New York Times", 20 giugno 2023.

13 *Lorenzo Arvanitis, McKenzie Sadeghi, Jack Brewster, GPT-4 produce più fake news e le rende più credibili, "La Repubblica", 21 marzo 2023.*

L'IA forte permetterà a chi ne sarà capace di militarizzare i *chatbot* anche per condurre operazioni cognitive raffinate. Per adesso è fantascienza, ma un domani sarà reale la presenza di *chatbot* maligni, o *killbot*, progettati e commercializzati, o hackerati e manipolati all'insaputa dei creatori, «per condurre operazioni psicologiche, su individui o campioni, di natura distruttiva – omicidi, stragi – o autodistruttiva – suicidi»[14].

Le operazioni cognitive del futuro impiegheranno *chatbot* che, protetti dalla parvenza dell'assistenza personale, saranno in grado di radicalizzare gli utenti più psicolabili, nonché di traviare i meno alfabetizzati, alimentando gli istinti violenti nei primi e la confusione nei secondi a beneficio di attori difficili da rintracciare. La comparsa di *chatbot* maligni potrebbe avvenire nel breve termine, perciò è necessario anticipare e prevenire il fenomeno. La distopia trasudante dall'immagine delle guerre cognitive IA-guidate non dovrebbe indurre i decisori nell'errore di sottovalutarne le probabilità di materializzazione: utenti psicolabili avrebbero già fatto del male a loro stessi e agli altri su istigazione di *chatbot*. Nel 2021, in Inghilterra, un *chatbot* chiamato Sarai avrebbe convinto Jaswant Singh Chail a intrufolarsi nel Castello di Windsor per assassinare la regina Elisabetta[15] e, nel 2023, in Belgio, un uomo si sarebbe suicidato su incitamento di un assistente psicologico virtuale[16].

Le operazioni cognitive del futuro impiegheranno *chatbot* che, protetti dalla parvenza dell'assistenza personale, saranno in grado di radicalizzare gli utenti più psicolabili

La militarizzazione delle IA deboli, come i dispositivi per la gestione delle case intelligenti, non sarà meno problematica per la sicurezza nazionale degli Stati. Su strumenti come Alexa, che necessitano una connessione alla rete per funzionare, sono stati effettuati degli esperimenti che ne hanno dimostrato l'elevata esposizione a infiltrazioni per fini di spionaggio e sabotaggio[17]. L'hackeraggio è la principale minaccia all'integrità e al ruolo amico delle IA deboli, in particolare dei programmi operanti nell'IoT (Internet delle Cose) e nell'Internet dei Corpi (IoB, *Internet of Bodies*), e la sua pericolosità aumenterà di pari passo con l'internetizzazione delle società e con l'ibridazione uomo-macchina.

Ricerche sulla militarizzazione delle IA deboli sono in corso negli Stati Uniti, in Russia, in Cina, sicuramente in altri Paesi, e convergono verso la stessa direzione: ogni dispositivo allacciato alla rete è hackerabile. Se il rischio con le IA forti è principalmente legato a una manipolazione algoritmica, con le IA deboli è dovuto agli hackeraggi: un dispositivo per la gestione di una *smart home* potrebbe essere utilizzato per spiare o per provocare un corto circuito, un

14 Emanuel Pietrobon, *ChatGPT, robot e deepfake: come saranno le guerre del futuro*, "InsideOver", 7 aprile 2023.

15 Il reo confesso è stato successivamente *condannato a nove anni di prigione per il piano omicida. Fonte:Tom Gerken, Liv McMahon, Tom Singleton, How a chatbot encouraged a man who wanted to kill the Queen, "BBC", 6 ottobre 2023.*

16 Chloe Xiang, *'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says*, "Vice", 30 marzo 2023.

17 Lily Hay Newman, *Turning an Echo Into a Spy Device Only Took Some Clever Coding*, "Wider," 25 aprile 2018.

macchinario ospedaliero potrebbe essere spento a distanza, un'autovettura intelligente potrebbe essere spinta a schiantarsi[18]. La militarizzazione delle IA deboli potrebbe verosimilmente spianare la strada ai delitti perfetti, provocando vittime eccellenti e determinando indagini inconcludenti. La quotidianizzazione e la capillarizzazione dell'IoT e dell'IoB dovranno avvenire con cautela e solo in seguito alla formulazione di adeguati strumenti capaci di minimizzare rischi e danni di una loro manipolazione.

La militarizzazione delle IA deboli potrebbe verosimilmente spianare la strada ai delitti perfetti

Falsi profondi, danni veri

I "falsi profondi", altresì noti come *deepfake*, sono una sorta di Photoshop sotto steroidi: audio, foto e video-montaggi di altissima qualità, generati da programmi e applicazioni basati su IA forti che, grazie all'apprendimento automatico e profondo e al metodo delle reti generative avversarie, sono in grado di produrre contenuti di volta in volta più verosimili.

Il loro ingresso nel panorama della disinformazione non sorprende, essendo un'evoluzione del falso fotografico, che esiste dall'invenzione della fotografia, ma preoccupa: la tendenza alla perfezione dei programmi che li producono è garanzia di crescita qualitativa e metterà a dura prova il *fact-checking*. La loro diffusione su scala industriale, invece, sarà la benzina del continuo disordine e delle sue manifestazioni, come le disinfodemie permanenti[19] e l'anarchia produttiva[20].

Ricostruire la breve storia dei falsi profondi può essere utile a comprendere la fondatezza della minaccia che costituiscono. *Deepfake* è un neologismo coniato nel 2017 e che letteralmente significa "falso dell'apprendimento profondo"[21], sebbene le origini del fenomeno risalgano al periodo, collocabile tra la fine degli anni Novanta e i primi anni Duemila, dello sviluppo delle IA con capacità immaginative e di clonazione digitale.

Dopo sedici anni di ricerca e sviluppo, periodicamente intervallati dall'irruzione nei mercati digitali di applicazioni come Face2Face, nel 2017 i *deepfake* raggiunsero la notorietà, in particolare negli Stati Uniti, complici l'abbattimento dei costi e lo sveltimento dei tempi per fabbricarli.

Inizialmente concentrati nei meandri della rete, come Reddit e 4Chan, dove vengono

18 Doug Irving, *Are We Ready for the Internet of Bodies?*, "Rand Corporation", 8 gennaio 2021.

19 È uno scenario fantapolitico in cui viene immaginato un ambiente permeato in maniera pervasiva dalla disinformazione, dunque estremamente diviso, fragile e conflittuale dal punto di vista sociale.

20 È una strategia della guerra ibrida che prevede di destabilizzare più livelli simultaneamente di una società, con l'obiettivo di condurre a un caos autoalimentante.

21 *Deepfake* è una parola-macedonia composta dai termini *deep learning* e *fake*.

prodotti da amatori per saziare preminentemente appetiti pornografici (nel 2019, secondo "Deeptrace", il 96% dei *deepfake* era a tema porno[22]), i super-falsi digitali subiscono un processo trasformativo a cavallo tra la pandemia di COVID19 e le presidenziali americane del 2020, quando diventano veicolo di disinformazione sanitaria e politica, per poi approdare nel resto del mondo.

La mutazione dei *deepfake* è stata repentina e radicale: gli anni Venti s'erano aperti con la loro incursione nell'intrattenimento per scopi pubblicitari, come rammentano i casi degli ologrammi di John Lennon ed Elvis Presley, ma sono proseguiti col loro divenire un potente strumento di disinformazione su guerre e relazioni internazionali.

La predisposizione al perfezionamento ha permesso alle IA di elevare straordinariamente la qualità dei falsi profondi in un lasso di tempo eccezionalmente breve. Il risultato è che i *deepfake*, oggi, possono essere fabbricati con facilità anche da amatori e sono passati dal replicare visi e movimenti all'imitare le voci. Esistono, infatti, programmi capaci di clonare una voce dopo averla ascoltata per cinque secondi. Diversamente dai *chatbot*, che richiedono competenze per essere creati e militarizzati, i *deepfake* sono un'atomica della disinformazione a portata di tutti. Il loro contenuto è fraudolento, talvolta palesemente, ma la qualità può essere così elevata da indurre in inganno porzioni più o meno ampie di una società e da richiedere un tempestivo *fact-checking* per evitare problemi di ordine interno, panico finanziario e persino incidenti internazionali – si pensi a un *deepfake* riguardante un capo di Stato intento a sparlare di una controparte o a rilasciare confessioni intime.

La predisposizione al perfezionamento ha permesso alle IA di elevare straordinariamente la qualità dei falsi profondi in un lasso di tempo eccezionalmente breve

Prova incontrovertibile che questa verrà ricordata come la decade dei *deepfake* è rappresentata dal fatto che sono stati protagonisti degli eventi più politicamente importanti del triennio 2020-23, dimostrando il loro enorme potenziale in termini di diffusione di disinformazione e di alimentazione della polarizzazione durante le presidenziali americane[23], la pandemia di COVID19[24], la guerra in Ucraina[25] e il confronto Israele-Hamas[26]. Hanno tratto in inganno i giornalisti per via della loro verosimiglianza, come il baciamano di Vladimir Putin a Xi Jinping viralizzato dalla macchina propagandistica ucraina nel marzo 2023[27], e hanno perfino cagionato dei crolli borsistici, come l'immagine del Pentagono in fiamme diffusa da attori russi due mesi

22 *The state of deepfakes. Landscape, threats and impact*. "DeepTrace", settembre 2019.

23 Eric Hofesmann, *Have Deepfakes influenced the 2020 Election?*, "Medium", 3 novembre 2020.

24 *Deepfake, propaganda and disinformation | How Russia and Ukraine battled the information war*, "WION", 26 febbraio 2023.

25 Kyle Matthews, *"Deepfake news: Artificial intelligence and Disinformation as a multilateral policy challenge"*, OSCE, 2021.

26 Matt Lebovic, *These Israelis are fighting Hamas on the war's emerging 'deepfake' cyberfront*, "The Times of Israel", 18 ottobre 2023.

27 Joscha Weber, *Fact check: No, Putin did not kneel before Xi Jinping*, "Deutsche Welle", 23 marzo 2023.

dopo[28]. I *deepfake* saranno uno degli strumenti prediletti degli agenti della destabilizzazione nell'era delle guerre ibride combattute con l'IA. La loro incomparabile capacità di alterare e distorcere la consapevolezza situazionale e il campo della realtà degli individui risponde agli obiettivi delle guerre cognitive e permette di produrre ondate di confusione collettiva. Le liberaldemocrazie, in ragione del possesso di reti aperte e pluralistiche, sono chiamate a normare il fenomeno e ad investire in figure e settori utili a minimizzare i danni delle operazioni cognitive condotte per loro tramite.

Metaverso, prossima frontiera delle guerre ibride

Il metaverso è lo spazio incorporeo e deterritorializzato in cui stanno trasferendosi gradualmente attività commerciali, grandi imprese, giornali, politici, *videogamer*, artisti dello spettacolo e *influencer*[29]. Sebbene non sia ancora chiara la forma definitiva che assumerà, certo è che Mark Zuckerberg e altri vorrebbero fare del metaverso il futuro delle società, delle economie e dell'intrattenimento, permettendo alle persone di vivere una vita a distanza grazie a realtà aumentata ed esperienze immersive.

Il mercato globale dei metaversi aveva un valore di 65-68 miliardi di dollari nel 2022 e si stima che potrebbe valere 1,5-5 mila miliardi di dollari nel 2030[30]. Entro quella data, a prescindere dal suo stato evolutivo e dalla sua demografia, il metaverso sarà diventato la nuova dimensione delle guerre ibride e dello spionaggio.

Esperimenti negli Stati Uniti e in Cina hanno dimostrato che il metaverso non soltanto è militarizzabile, ma che le sue caratteristiche intrinseche, come realtà aumentata e multisensorialità, amplificano eco ed impatto delle operazioni cognitive[31]. Ne consegue che la sicurezza nazionale del dopodomani, oltre che da una maggiore consapevolezza sull'IA, passerà inevitabilmente dalla comprensione e dalla gestione del metaverso, che andrebbe configurato, già da ora, come un nuovo dominio della conflittualità.

La Rand Corporation è dell'opinione che «gli ambienti di realtà virtuale consentiranno la manipolazione emotiva e psicologica dei loro utenti a un livello inimmaginabile rispetto ai media odierni», per una questione demografica e di sensazioni amplificate dall'immersione profonda nella realtà aumentata[32]. Il metaverso, invero, è un palcoscenico del paradosso per cui personaggi come Ariana Grande, che nel 2021 ha tenuto un concerto su "Fortnite" seguito

28 Luke Hurst, *How a fake image of a Pentagon explosion shared on Twitter caused a real dip on Wall Street*, "EuroNews", 23 maggio 2023.

29 Metaverso è un termine proveniente dalla letteratura fantascientifica che oggi è utilizzato per indicare i tentativi dei *Big Tech* di costruire dei grandi spazi a realtà virtuale e aumentata in grado di permettere agli utenti di condurre delle vere e proprie esistenze parallele, acquistando proprietà, socializzando e "viaggiando".

30 *Value creation in the metaverse*, McKinsey & Company, giugno 2022; *Metaverse Market*, Precedence Research.

31 Rand Waltzman, *Facebook Misinformation Is Bad Enough. The Metaverse Will Be Worse, Rand Corporation, 22 agosto 2022.*

32 Vedasi nota precedente.

da settantotto milioni di metanauti, hanno provato essere privo di barriere all'entrata. Una forza-debolezza che destabilizzatori e radicalizzatori potrebbero sfruttare, e sfrutteranno, per condurre operazioni psicologiche, informative e cognitive capaci di raggiungere milioni di metanauti in contemporanea.

Il metaverso sarà il futuro teatro di battaglia delle grandi potenze e degli attori maligni non-statuali, come mafie e terrorismi, che ivi agiranno per carpire informazioni sensibili, raccogliere *intelligence*, riciclare denaro illegale, attaccare meta-assetti, reclutare adepti e condurre vari tipi di operazioni ai danni degli ignari metanauti.

L'arrivo delle guerre ibride nel metaverso comporterà la proliferazione di *avatar* maligni gestiti da bot e/o da spie, lo spargimento di messaggi subliminali negli ambienti virtuali e la manipolazione impercettibile di luoghi ed eventi per mezzo dei *deepfake*. I risultati potrebbero essere delle super-operazioni cognitive, resistenti a future verifiche dei fatti, giacché le sperimentazioni sul mimetismo digitale «hanno dimostrato che cambiare leggermente i tratti di una figura politica poco familiare […] rende le persone più positivamente disposte verso di essa. E vale anche il contrario: si può fare in modo che i metanauti spengano il *computer* provando sentimenti negativi verso qualcosa, inspiegabilmente nervosi perché reduci da un'esperienza immersiva inzeppata di messaggi subliminali»[33].

Nei metaversi non regolamentati, e in una certa misura anche nelle controparti monitorate da Stati o da campioni nazionali, elevata sarà la probabilità di assistere a guerre cognitive a scarsa visibilità ma ad altissimo impatto. Le armi principali delle suddette saranno i messaggi subliminali introdotti in luoghi ed eventi per condizionare le esperienze dei metanauti e i *deepfake* in grado di «spaesare, potenzialmente, (decine di) milioni di persone in simultanea e instillare in loro dei dubbi tanto potenti da resistere a successive smentite», le cui capacità di influenzamento del subconscio saranno accresciute dalla «customizzazione [delle esperienze] e dall'amplificazione delle sensazioni nella realtà virtuale»[34].

Almeno quattro sono le ragioni che rendono necessaria l'inclusione del metaverso tra i domini della conflittualità:

a) le operazioni informative, psicologiche e cognitive ivi condotte saranno incredibilmente efficaci e capaci di raggiungere decine di milioni di persone contemporaneamente;

b) lo spostamento di assetti dalla realtà, da conti correnti a *database*, attirerà schiere di *hacker*;

c) il trasloco di imprenditori, politici e soldati sarà una calamita per spionaggio e *meta-kompromat*[35];

d) esiste un rischio medio-alto che le organizzazioni terroristiche creino dei meta-centri di radicalizzazione.

33 Emanuel Pietrobon, *L'arte della guerra ibrida. Teoria e prassi della destabilizzazione*, Castelvecchi, 2022, p. 300.

34 Ivi, p. 301.

35 *Kompromat* è una parola-macedonia di lingua russa che è entrata nel gergo dello spionaggio per indicare fascicoli che, in quanto contenenti documenti sensibili e compromettenti, possono essere utilizzati per estorcere informazioni e denaro e per ricattare qualcuno. Un *meta-kompromat* sarà un dossier costruito a partire dalle trappole disseminate nel metaverso.

Le guerre ibride del prossimo futuro avranno luogo anche nel metaverso, o meglio nei metaversi, di cui le grandi potenze stanno indagando fragilità e potenzialità da quando questo spazio futuristico è stato annunciato[36]. Si tratterà di conflitti invisibili alle masse, ma in grado di incidere sulla realtà fisica per via della loro capacità di dilatare impatto e durata delle operazioni cognitive, di indebolire i centri decisionali stabiliti nel metaverso e di colpire i meta-assetti strategici di Stati, entità e individui.

Conclusioni

L'internetizzazione della quotidianità, delle relazioni sociali e dei rapporti interstatali impone ai legislatori di comprendere, di accompagnare e di normare il fenomeno delle IA.
In Italia, che è un Paese dalla vulnerabilità cibernetica, informatica e cognitiva medio-alta[37], è necessaria l'adozione di un piano nazionale per l'IA che tenga conto sia delle opportunità sia dei rischi dell'ultima rivoluzione tecnologica.
L'Agenzia per la Cybersicurezza Nazionale è chiamata a estendere il proprio raggio d'azione, includendo le cyber-guerre combattute con l'IA tra i suoi scopi e investendo nella messa in sicurezza di metaversi e di luoghi e cose intelligenti (*smart cities* e *smart things*).
Occorre una Commissione sull'intelligenza artificiale che focalizzi le attività di indagine e di *policy making* su applicazioni opache come TikTok, *chatbot*, *content farm*, *deepfake* e fragilità dei dispositivi intelligenti che trovano crescente impiego nelle proprietà private, nell'industria e nella sanità.
L'IA è il crocevia in cui s'intersecano le vie delle guerre cibernetiche e cognitive, perciò è fondamentale la formulazione di un piano onnicomprensivo che preveda strumenti ad hoc, aggiornamenti legislativi e investimenti. Strumenti per drenare le *content farm* gestite da *bot*, facilitandone la chiusura in caso di responsabilità nella veicolazione voluta e ripetuta di disinformazione. Aggiornamenti legislativi per elevare la diffusione di *deepfake* destabilizzanti ad aggravante dell'articolo 656 del Codice Penale. Investimenti nella riforma del sistema educativo e in nuove figure professionali.
L'aumento quanti-qualitativo delle operazioni cognitive IA-trainate richiede sistemi scolastici che alimentino scetticismo attivo e spirito critico nella popolazione studentesca e nei quali le discipline Stem[38] e Alph[39] abbiano pari importanza, giacché le operazioni di destabilizzazione socio-psicologica incidono maggiormente laddove dominano povertà intellettuale e tecno-dipendenza.

36 Kiru Pillay, *Militarising the Metaverse, "Defence Web", 15 agosto 2022; Caitlin Dohrman, Jennifer McCardle, The Full Potential of a Military Metaverse, "War on the Rocks", 18 febbraio 2022.*

37 *Rapporto Clusit 2023 sulla sicurezza ICT in Italia*, Clusit, ottobre 2023; Emanuel Pietrobon, *Guerra cognitiva, la nuova minaccia ibrida*, "MachiavelliDossier", luglio 2023; Rita Plantera, *Perché l'Italia è ancora molto indietro sulla cybersicurezza*, "Pagella Politica", 7 marzo 2022.

38 Acronimo in lingua inglese che sta per S*cience, Technology, Engineering and Mathematics*, ossia scienza, tecnologia, ingegneria e matematica.

39 Acronimo in lingua inglese che sta per *Art, Literature, Philosophy and History*, ossia arte, letteratura, filosofia e storia.

È altresì imperativo che aumenti la sorveglianza sulle applicazioni a rischio, con possibilità di bandirle, e che si investa nella professionalizzazione del giornalismo, in tecnologie per individuare *chatbot* e *deepfake* maligni in tempi brevi e nell'incremento di figure come specialisti in *intelligence* delle fonti aperte e analisti dell'immagine e dei suoni. L'alternativa alla presenza di un organico esteso, attrezzato e aggiornato è un allungamento delle tempistiche del *fact-checking* causato dal miglioramento continuativo e incrementale dei *deepfake*.

Super-falsi e *chatbot* maligni sono la prossima frontiera della disinformazione, dell'interferenza elettorale e della guerra cognitiva, nonché dello spionaggio, e la loro presenza e la loro influenza vanno espandendosi progressivamente dalla rete al metaverso. Per questa ragione si raccomanda di ponderare l'inserimento del metaverso tra i domini della conflittualità e di costruire uno o più metaversi nazionali, previa creazione di agenzie deputate alla loro gestione e investimenti in tecnologie per individuare *avatar* maligni e realtà virtuali manipolate o manipolabili. ■

MACHIAVELLI

Suadere atque agere

centro studi
politici e strategici

Il Centro Studi Politici e Strategici Machiavelli dal 2017 si occupa di promuovere politiche improntate ai valori tradizionali e finalizzate a costruire un'Italia prospera e forte.

www.centromachiavelli.com

THE AVGVSTVS FOVNDATION

La presente pubblicazione è stata realizzata col contributo di *Augustus Foundation*