



Pentagono, l'IA sotto esame: dubbi sulla capacità di individuare chi mente, ridimensionato il programma

La DCSA esclude alcune tecnologie dal progetto attuale. Documenti precedenti descrivevano modelli per analizzare linguaggio ed emozioni.

Pubblicato il 08/10/2026

Il Pentagono sembra aver ridimensionato alcuni progetti che prevedevano l'impiego dell'intelligenza artificiale per analizzare il comportamento, la voce e le espressioni facciali del personale sottoposto a controlli di sicurezza. Le nuove dichiarazioni del Dipartimento della Difesa statunitense contrastano infatti con documenti e presentazioni precedenti, che descrivevano lo sviluppo di sistemi automatizzati per individuare possibili segnali di inganno durante gli interrogatori.

Secondo quanto ricostruito da Defense News, l'iniziativa rientra in un programma pluriennale di modernizzazione del poligrafo, lo strumento utilizzato anche nelle verifiche di controspionaggio e nelle procedure per il rilascio delle autorizzazioni di sicurezza.

Il Pentagono esclude alcune tecnologie dal programma

La Defense Counterintelligence and Security Agency (DCSA), responsabile dei controlli di affidabilità del personale che accede a informazioni classificate, ha precisato che l'attuale attività di ricerca e sviluppo del programma Modernizing Polygraph non comprende sistemi di intelligenza artificiale generativa, modelli linguistici di grandi dimensioni (LLM), analisi delle microespressioni facciali o sistemi di valutazione della voce.

L'agenzia ha tuttavia dichiarato che continuerà a monitorare gli sviluppi scientifici per valutarne un eventuale impiego futuro.

La precisazione non equivale a un abbandono completo dell'intelligenza artificiale. La DCSA non ha infatti chiarito se rimangano attivi altri algoritmi destinati all'analisi del linguaggio parlato e scritto, precedentemente sviluppati per riconoscere modelli comunicativi associati a comportamenti ingannevoli.

Il nuovo orientamento rappresenta comunque una differenza significativa rispetto alle indicazioni fornite dal Pentagono nei mesi precedenti.

Un programma da 31 milioni di dollari

La modernizzazione degli strumenti di valutazione dell'affidabilità personale è oggetto di studi da diversi anni.

Documenti ottenuti da Defense News attraverso richieste di accesso agli atti mostrano che, almeno dal 2019, l'US Air Force, la DCSA e diversi laboratori universitari hanno lavorato allo sviluppo di modelli di intelligenza artificiale per l'analisi dei sentimenti e il riconoscimento di possibili comportamenti ingannevoli.

Il programma, indicato nei documenti con le denominazioni Credibility Assessment Modernization, Polygraph+ e Polygraph Next, dispone di una previsione finanziaria complessiva di circa 31 milioni di dollari.

Le tecnologie considerate comprendono algoritmi per l'elaborazione automatizzata dei risultati, strumenti di supporto alle decisioni degli investigatori, sensori senza contatto e sistemi di imaging termico in grado di rilevare variazioni fisiologiche potenzialmente associate allo stress.

L'obiettivo è superare alcuni limiti del poligrafo tradizionale, che registra parametri come respirazione, pressione sanguigna e sudorazione, lasciando a un esaminatore umano il compito di interpretarli.

A luglio, un funzionario del Pentagono aveva spiegato che l'intelligenza artificiale avrebbe consentito di elaborare i dati in tempo reale, fornendo analisi più standardizzate a supporto degli investigatori, senza sostituirne il giudizio.

Algoritmi addestrati anche con dati provenienti da Reddit e Twitter

Tra gli elementi emersi dai documenti figura l'utilizzo di contenuti provenienti dai social network per addestrare alcuni modelli sperimentali.

Una presentazione della DCSA descrive infatti un progetto dell'US Air Force per lo sviluppo di algoritmi di deep learning destinati all'analisi del linguaggio.

Nelle prime fasi, i ricercatori hanno impiegato testi raccolti da piattaforme come Twitter e Reddit, insieme a risorse linguistiche come WordNet, per insegnare ai sistemi a riconoscere strutture grammaticali, espressioni colloquiali e caratteristiche della comunicazione quotidiana.

Un'altra presentazione, datata maggio 2025, illustrava un sistema sperimentale nel quale un avatar digitale avrebbe condotto parte del colloquio con un candidato all'autorizzazione di sicurezza.

Telecamere e microfoni avrebbero raccolto immagini, voce e contenuti verbali, successivamente elaborati da modelli di intelligenza artificiale per stimare lo stato emotivo dell'intervistato. Un investigatore umano avrebbe supervisionato la procedura, inserendo le domande da rivolgere al candidato.

Secondo un ulteriore documento, uno degli obiettivi di ricerca era raggiungere un'accuratezza del 75% nell'interpretazione degli stati emotivi attraverso tecnologie di elaborazione del linguaggio naturale.

La DCSA ha però precisato che l'università coinvolta nella progettazione di questo sistema non partecipa all'attuale fase di sviluppo.

La comunità scientifica mette in dubbio l'affidabilità dei risultati

Il principale ostacolo riguarda la capacità effettiva di un algoritmo di distinguere una dichiarazione falsa da una vera.

Numerosi studi scientifici hanno messo in discussione la possibilità di identificare una menzogna attraverso specifiche caratteristiche della voce, del linguaggio o delle espressioni facciali.

Lo stress, il nervosismo o una variazione nel comportamento non costituiscono necessariamente una prova di inganno.

Secondo David Markowitz, professore della Michigan State University specializzato nell'analisi della comunicazione attraverso l'intelligenza artificiale, non esistono indicatori diagnostici universalmente affidabili per riconoscere una menzogna.

In un esperimento condotto utilizzando il modello Gemini di Google, Markowitz ha confrontato le capacità dell'intelligenza artificiale con quelle degli esseri umani nell'identificare dichiarazioni false durante interrogatori simulati.

I risultati hanno mostrato prestazioni sostanzialmente paragonabili al caso, con un'accuratezza vicina al 50%.

L'esperimento ha inoltre evidenziato una particolare criticità: il sistema di intelligenza artificiale tendeva più frequentemente degli osservatori umani ad attribuire erroneamente comportamenti ingannevoli alle persone interrogate.

Il rischio è che gli algoritmi associno il semplice contesto di un interrogatorio alla probabilità che il soggetto stia nascondendo qualcosa.

Il rischio per le carriere militari e le autorizzazioni di sicurezza

La questione non riguarda soltanto l'efficacia tecnologica, ma anche le possibili conseguenze professionali di valutazioni errate.

Negli Stati Uniti, il risultato di un controllo di affidabilità può influire sull'assegnazione degli incarichi, sulle possibilità di avanzamento e sull'accesso alle informazioni classificate.

L'avvocato Mark Zaid, specializzato in questioni di sicurezza nazionale, ha evidenziato che una valutazione negativa può comportare lunghi periodi di sospensione o limitazione delle opportunità professionali.

L'introduzione dell'intelligenza artificiale potrebbe aggiungere un ulteriore livello di incertezza se gli investigatori attribuissero ai risultati automatizzati un'affidabilità superiore a quella effettivamente dimostrata.

Questo fenomeno, noto come automation bias, consiste nella tendenza ad accettare le indicazioni di un sistema automatizzato anche quando esistono elementi che potrebbero metterle in discussione.

Il problema è particolarmente delicato nei procedimenti di sicurezza nazionale, dove le possibilità di contestare una decisione possono essere limitate.

Già nel 1998, nella sentenza *United States v. Scheffer*, la Corte Suprema degli Stati Uniti aveva evidenziato l'assenza di un consenso scientifico sull'affidabilità del poligrafo, confermando la legittimità dell'esclusione dei suoi risultati dai tribunali militari.

Tra innovazione tecnologica e controllo umano

La revisione del programma del Pentagono evidenzia una questione destinata ad assumere crescente importanza nelle attività di intelligence e controspionaggio.

L'intelligenza artificiale può contribuire all'elaborazione di grandi quantità di informazioni, all'individuazione di anomalie e all'analisi di parametri fisiologici e comportamentali. Queste capacità, tuttavia, non dimostrano automaticamente che un sistema sia in grado di stabilire se una persona stia mentendo.

La distinzione assume particolare importanza quando le valutazioni riguardano militari, funzionari dell'intelligence e personale autorizzato a trattare informazioni classificate.

Per il momento, il Pentagono non ha chiarito completamente quali componenti di intelligenza artificiale rimangano nel programma di modernizzazione del poligrafo, né se l'esclusione delle tecnologie di analisi vocale e facciale rappresenti una decisione definitiva.

Resta quindi aperta la questione centrale: quale livello di affidabilità scientifica dovrà essere dimostrato prima che questi strumenti possano contribuire a decisioni capaci di incidere sulla carriera e sulle autorizzazioni di sicurezza del personale militare?

Link notizia: <https://www.defensenews.com/news/pentagon-congress/2026/10/05/has-the-pentagon-given-up-on-ai-polygraph-analysis-of-trustworthiness/>