IA sul campo di battaglia: l’incubo dei dati avvelenati

scenarieconomici.it/ia-sul-campo-di-battaglia-lincubo-dei-dati-avvelenati

Jean Valjean 22 aprile 2024

La AI è al centro anche delle tecnologie militari per migliorarne l’efficacia sul campo di battaglia. Il suo ampio uso però la espone ad un attacco molto particolare, cioè “**l’avvelenamento del pozzo” dei dati** che alimentano la sua intelligenza.

“Non credo che i nostri dati siano avvelenati ora”, ha sottolineato il vice segretario alla difesa USA , Jennifer Swanson, mercoledì alla conferenza del Potomac Officers Club, “ma quando combatteremo contro un avversario quasi alla pari, dovremo sapere esattamente quali sono i vettori di minaccia”.

Ogni algoritmo di apprendimento automatico deve essere addestrato su dati – molti e molti dati. Il Pentagono sta facendo uno sforzo enorme per raccogliere, collazionare, curare e pulire i suoi dati, in modo che gli algoritmi analitici e le IA neonate possano dargli un senso. In particolare, il team di preparazione deve eliminare tutti i punti di dati errati, prima che l’algoritmo possa imparare la cosa sbagliata.

La battaglia dei dai

I chatbot commerciali, da Microsoft Tay del 2016 a ChatGPT del 2023, hanno mostrato come la qualità dei dati possa influenzare lo sviluppo della AI e il suo funzionamento. Nel caso di Ai utilizzate nel settore militare i dati potrebbero essere “Avvelenati”, cioè falsificati e distorti, in mdo deliberato e guidato dagli avversari potenziali delle forze armati di un paese, e questa tecnica è proprio definita “avvelenamento dei dati”.

“Qualsiasi LLM [Large Language Model] commerciale che è in circolazione e che apprende da Internet, oggi è avvelenato”, ha detto Swanson senza mezzi termini. “Ma sono onestamente più preoccupato di ciò che si chiama, come dire, l’IA ‘normale’, perché questi sono gli algoritmi che saranno realmente utilizzati dai nostri soldati per prendere decisioni sul campo di battaglia”.

Nel caso del Pentagono non si tratta di addestrare dei chatbot con i dati presi dalla rete. L’esercito dovrebbe addestrarlo su un set di dati militari affidabili e verificati all’interno di un ambiente sicuro e protetto. In particolare, ha raccomandato un sistema al livello di impatto 5 o 6 del DoD, adatto ai dati sensibili (5) o classificati (6).

Entro l’estate deovrebbe essere il primo campione di intelligenza artficiale IL-5 LLM, cioè basato sui dati di livello 5. Questo può essere utile per tutti i tipi di funzioni di back-office, riassumendo le risme di informazioni per rendere più efficienti i processi burocratici. “Ma la nostra preoccupazione principale sono gli algoritmi che informeranno le decisioni sul campo di battaglia”.

Avvelenare i dati della IA che decidono le attività sul campo di battaglai può essere invece un problmea molto più profondo e di difficile soluzione.

CJADC2, test AI ecome sconfiggere i dati avvelenati

Ottenere i giusti dati di addestramento specifici per le forze armate è particolarmente critico per il Pentagono, che mira a utilizzare l’AI per coordinare le future operazioni di combattimento attraverso la terra, l’aria, il mare, lo spazio e il cyberspazio. Il concetto si

chiama Combined Joint All-Domain Command and Control (CJADC2) e a febbraio il Pentagono ha annunciato che una "capacità minima di fattibilità" funzionante è già stata messa in campo in alcuni quartieri generali selezionati in tutto il mondo.

Le versioni future aggiungeranno i dati di targeting e la pianificazione degli attacchi, collegandosi ai progetti di comando di battaglia AI esistenti a livello di servizio: ABMS dell'Aeronautica, Project Overmatch della Marina e Project Convergence dell'Esercito.

Il Project Convergence, a sua volta, utilizzerà la tecnologia sviluppata dal neonato Project Linchpin, che sarà la Ai di punto nella guida delle decisioni strategiche.

In altre parole, l'Esercito sta cercando di applicare all'apprendimento automatico il ciclo di feedback "agile" tra sviluppo, cybersicurezza e operazioni correnti (DevSecOps) utilizzato dai principali sviluppatori di software per lanciare rapidamente nuove tecnologie e aggiornarle continuamente.

Il problema è che, in realtà, non si sa come guidare questi processi e le società che commercialmente gestiscono gli algoritmi non hanno bene idea di come questi funzionino. Normalmente si è abituati a vedere il funzionamento dei programmi come deterministico, come la risposta fissa a una certa situazione. Questo non accade con la AI, dove la risposta non sempre predeterminata, come dimostrano le applicaazioni di OpenAI.

C'è poi anche un problema ulteriore: ogni implementazione della Ai provoca un flusso di dati che viene a sua volta ad essere integrato nel database della AI e quindi ne viene a definire i processi futuri. Tutto questo viene a rendere il problema dell'avvelenament dei dati e della purezza delle informazioni che vengono fornite ai programmi Ai importantitssimo, essenziale, nella definizione delle decisioni strategiche e nella loro trasmissione.

I prossimi anni vedranno lo sviluppo di un nuovo campo di battaglia: quello dei dati, in cui ciascuna parte cercherà di corrompere e falsificare la abse dei dati su cui si basano le decisioni degli altri.