Auteur: Dr. J. Kwik is onderzoeker internationaal recht bij het Asserinstituut in Den Haag.

Civiele sector in het vizier van de agressor

Google's recente besluit (1) om het verbod op AI-ontwikkeling voor militaire doeleinden te schrappen, stelt onderzoekers en ontwikkelaars in de technologiesector voor een uitdaging: hoe houden ze afstand van het militaire domein? Veel experts (2) geven de voorkeur aan die scheiding. Uit een recente peiling van ELSA Lab Defence (waar de auteur bij betrokken was) blijkt dat de meerderheid van AI-onderzoekers in Nederland persoonlijk bezwaar heeft tegen defensieve AI-toepassingen, zoals autonome (cyber)wapens en doelbestrijdingssoftware.

Het komt regelmatig voor dat AI-gerelateerd onderzoek een militair staartje krijgt. Waar in het verleden innovaties vaak ontstonden in een militaire context en vervolgens hun toepassingen vonden in de civiele sfeer (zoals de mobiele telefoon of de magnetron), is de situatie nu omgekeerd: defensie leert, leent en bouwt voort op producten uit de civiele sector. Denk hierbij aan openbaar beschikbare grote taalmodellen (LLM), die getraind worden tot militair-operationele planningsadviseurs (3).

Civiele sector doelwit

De civiele en militaire AI-ecosystemen raken steeds verder met elkaar verweven, en dat heeft consequenties voor dreigingen waar privébedrijven rekening mee zullen moeten houden. Momenteel zijn cyber- en informatieaanvallen in het kader van hybride conflicten zoals in Oekraïne (4) nog voornamelijk gericht tegen militaire of publieke installaties; digitale spionageoperaties tegen inlichtingendiensten of cyberaanvallen tegen energienetwerken. De toekomst zal leren dat ook de civiele sector een steeds aantrekkelijker doelwit wordt voor vijandelijke machten.

Om hier een voorbeeld van te noemen: stel dat partij X als doel heeft om de militaire planningsadviseur van tegenstander Y te saboteren. Eén optie is om Y's software te beschadigen met een virus. Dit heeft weinig kans van slagen, aangezien de software zich in een gesloten en beveiligd militair netwerk bevindt. Het alternatief viseert een gemakkelijker doelwit: de database van de LLM-ontwikkelaar waar de planningsadviseur op is gebaseerd. Door het injecteren van vergiftigde exemplaren in de dataset of het installeren van een backdoor in het LLM-model, kan X het eindproduct compromitteren (5) zonder ooit Y's beveiligde netwerk te hoeven hacken. Slachtoffer: de LLM-ontwikkelaar en de burgers die ook gebruikmaken van deze LLM.

Het tweede voorbeeld: AI is in staat hele verfijnde en doelgerichte desinformatie en cognitieve aanvallen (6) te genereren tegen personen, zolang er voldoende data beschikbaar is over de gedachtenpatronen en voorkeuren van deze persoon. Dit maakt het hacken van brain-computer interfaces (BCI) (7) zoals Neuralink, een aantrekkelijke strategie: dergelijke apparaten slaan immers continu de breinsignalen van de drager op. Slachtoffer: de drager van de BCI en het geassocieerde bedrijf.

Begrip voor nieuwe dreiging

Voor onderzoekers en ontwikkelaars wordt het steeds moeilijker om zich af te zonderen van de lange arm van het moderne slagveld. De verstrengeling tussen militair en civiel zal ervoor zorgen dat zelfs als een bedrijf het militaire domein niet opzoekt, laatstgenoemde hem toch zal vinden. Van belang is daarom dat deze nieuwe dreigingen begrepen worden en in acht worden gehouden, zelfs als de technologie op het eerste gezicht weinig te maken lijkt te hebben met defensie – zoals het voorbeeld van BCI's laat blijken.

De auteur is onderzoeker internationaal recht bij het Asser Instituut in Den Haag. In een latere uitgave verschijnt een uitgebreider artikel van zijn hand over de rol van drones, AI, desinformatie en andere digitale technieken op het slagveld, inclusief de bijbehorende ethische vraagstukken. Dit artikel dient alvast als smaakmaker.

Referenties

(1) https://www.nu.nl/tech/6344877/google-schrapt-verbod-op-het-gebruik-van-zijn-ai-voor-wapens.html
(2) https://nypost.com/2025/02/06/business/google-may-use-ai-for-weapons-surveillance-prompts-backlash/
(3) https://defensescoop.com/2024/11/04/scale-ai-unveils-defense-llama-large-language-model-llm-national-security-users/
(4) https://www.rodekruis.nl/nieuwsbericht/trollen-en-hackers-zo-werkt-vechten-in-een-cyberoorlog/
(5) https://lieber.westpoint.edu/anti-ai-countermeasures-warfare-terra-incognita-ihl/
(6) https://www.ie.edu/insights/articles/cognitive-war-turns-the-mind-into-battle-ground/
(7) https://www.aria.org.uk/opportunity-spaces/scalable-neural-interfaces