

## De wapenwedloop achter het beeldscherm

Jaus Müller

**E**rgens in het bergland van Noord-Jemen, tussen kale rotsellingen, zanderige vlaktes, diepe ravijnen en straatarme dorpen, klappt een groep mannen een laptop open. Zo zou het tafereel eruit kunnen hebben gezien. Via een internetverbinding krijgen ze de AI-assistent Claude aan de praat. De mannen, vermoedelijk verbonden aan de Houthi-beweging, hebben een probleem: hun ballistische raketten doen niet precies wat zij willen. Mogelijke doelwitten zijn schepen in de Rode Zee of doelen in Saudi-Arabië. Om de raketsoftware te verbeteren, schakelen ze Claude in. Na een aanpassing proberen ze het opnieuw. Kennelijk met onvoldoende resultaat, want enkele uren later vragen ze de AI-assistent waarom de raket niet naar behoren functioneerde en wat eraan moet worden verbeterd.

Zo probeerden vermoedelijke leden van de Houthi-beweging, een bondgenoot van Iran, de afgelopen maanden met westerse AI-software hun raketarsenaal te verbeteren. Op het hoofdkantoor van Anthropic, de maker van Claude in San Francisco, zo'n 14.000 kilometer verderop, kwamen alarmerende berichten binnen vanuit het eigen systeem. Wat gebeurde daar allemaal in Jemen met Claude? Anthropic greep in en blokkeerde de accounts. Maar een deel van de hulp middels het Large Language Model was toen al geleverd.

Deze casus van Jemenitische doe-het-zelf-raketgeleerden laat zien hoe kunstmatige intelligentie haar weg vindt naar het militaire domein – en hoe belangrijk regulering is binnen het militaire domein. Het valt enigszins te prijzen dat Anthropic ingreep en vervolgens een uitvoerig rapport over het incident publiceerde. Tegelijkertijd weten we niet in hoeveel gevallen ingrijpen achterwege bleef en welke incidenten niet zijn gepubliceerd.<sup>1</sup> Het wringt dat we voor het bewaken van zulke grenzen zo sterk vertrouwen op de zelf-regulering van Amerikaanse technologie-bedrijven met grote commerciële belangen. Wie bepaalt welke militaire toepassingen toelaatbaar zijn, en wie ziet erop toe dat die grenzen worden gehandhaafd?

Telkens weer laten we ons verrassen door het enorme tempo van de wapenwedloop achter het scherm. Onlangs pakte ik het uitstekende boek *De oorlog van morgen. Wat de conflicten van nu ons leren over toekomstige dreigingen* van NLDA-collega Martijn Kitzen en Tim Sweijs er weer eens bij. Het boek, waarvan de eerste druk in april 2025 verscheen, verkent met een vooruitziende blik de toekomst van oorlogvoering en staat daarbij stil bij de stormachtige ontwikkelingen rond AI. Op pagina 124 staat bijvoorbeeld: 'Later deze eeuw, zo voorspellen sommige technologen, kan het zelfs gevolgd worden door ASI, of artificial super intelligence, die slimmer is dan mensen en onze prestaties simpelweg overtreft'.<sup>2</sup> Met de snelheid waarop AI-systemen zich sindsdien al hebben ontwikkeld, lijkt het waarschijnlijk dat deze mijlpaal eind van dit decennium wordt bereikt; wat

1 Anthropic, *Detecting and countering misuse of AI: September 2026* (10 september 2026). Zie: <https://www-cdn.anthropic.com/e50be2e51e7695dc4b1366a37a245a597377d3b5/Anthropic-Detecting-and-countering-091026.pdf>.

2 Martijn Kitzen en Tim Sweijs, *De oorlog van morgen. Wat de conflicten van nu ons leren over toekomstige dreigingen* (Amsterdam, Uitgeverij Balans, 2025) 124.



gisteren nog verre toekomst leek, wordt vandaag als voldongen feit gepresenteerd.

Met de razendsnelle opmars van AI groeien ook de zorgen binnen militaire organisaties. Statelijke actoren, zoals China, kunnen kunstmatige intelligentie gebruiken om digitale aanvallen uit te voeren. Maar minstens zo zorgelijk is dat ook niet-statelijke actoren toegang krijgen tot uiterst krachtige kennis-systemen, zoals het voorbeeld van Claude in Jemen laat zien.

Het is zeker niet de eerste keer dat de wereld wordt geconfronteerd met een nieuwe technologie waarvan de mogelijk schadelijke gevolgen nauwelijks zijn te overzien. Daarbinnen zijn er ook voorbeelden van geslaagde zelfregulering. In de jaren zeventig beloofde nieuwe DNA-technologie nieuwe medicijnen en een snellere en tevens goedkopere productie van bestaande middelen, zoals antibiotica. Tegelijkertijd konden wetenschappers daarmee genen veranderen op manieren waarvan zij de gevolgen amper konden overzien. Bij vroege experimenten werd DNA gebruikt van een virus dat kanker kon veroorzaken; andere experimenten introduceerden resistentie tegen antibiotica. Wat als zulke organismen uit het laboratorium zouden ontsnappen? Een geschrokken wetenschapper waarschuwde dat het besmettingsgevaar de mensheid aan de rand van een soort biologisch Hiroshima bracht. Wetenschappers kwamen samen in conferentiecentrum Asilomar, in het Californische Monterey, om afspraken te maken over de veilige toepassing van deze nieuwe technologie.<sup>3</sup>

Een andere historische parallel is de ontwikkeling van het kernwapen in de jaren 40 en de regulering die daarop volgde. Tegenwoordig roepen leiders van AI-bedrijven op tot regulering en waarschuwen sommigen zelfs voor de mogelijke vernietiging van de mensheid als die uitblijft. Ook binnen het Manhattanproject klonken dergelijke waarschuwingen over de gevolgen van de technologie die wetenschappers zelf hielpen te ontwikkelen. In juni 1945, nog vóór de atoombom op Hiroshima viel, richtten

Als de wereld opnieuw een kwart eeuw nodig heeft om grenzen te stellen, hoeveel valt er dan nog te beteugelen?

zij zich in het grootste geheim met het *Franck Report* tot de Amerikaanse minister van Oorlog. Daarin waarschuwden ze voor een nucleaire wapenwedloop en de verwoestende gevolgen daarvan. Internationale controle was volgens hen noodzakelijk om dat gevaar te bezweren. Het was een oproep aan de politieke leiding om in actie te komen.<sup>4</sup>

Het grote probleem is de factor tijd. De nucleaire geschiedenis laat zien hoe traag internationale beheersing tot stand kan komen. In juni 1945 waarschuwden wetenschappers de Amerikaanse regering voor de gevaren van een nucleaire wapenwedloop. In 1957 werd het Internationaal Atoomenergieagentschap (IAEA) opgericht. Het Non-proliferatieverdrag volgde in 1968 en trad in 1970 in werking: een kwart eeuw na Hiroshima. Er zat vijftientig jaar tussen de eerste inzet van een kernwapen en het verdrag dat de hoeksteen van de internationale nucleaire beteugeling zou worden. Zelfs tijdens de Koude Oorlog bleken zulke afspraken mogelijk. Inmiddels staat het internationale stelsel dat ze zou moeten dragen zwaar onder druk. Bovendien werd de ontwikkeling van kernwapens geleid door staten, terwijl bij AI grote technologiebedrijven het tempo dicteren. Hun activiteiten overschrijden landsgrenzen en laten zich moeilijk onder één internationaal regime brengen. Als de wereld opnieuw een kwart eeuw nodig heeft om grenzen te stellen, hoeveel valt er dan nog te beteugelen? ■

3 Kate Zernike, 'How Scientists Contained a Threat That Could Have Destroyed Humanity', *The New York Times*, 23 september 2026.

4 James Franck e.a., 'The Franck Report. A Report to the Secretary of War', juni 1945, beschikbaar via de Federation of American Scientists. Zie: <https://sgp.fas.org/eprint/franck.html>.