FOTO: 123RF

Een digitaal **kat-en-muisspel**

Het was een schokgolf eerder dit jaar: het Amerikaanse AI-bedrijf Anthropic, gevestigd in San Francisco, zegt een generatief taalmodel, Mythos, te hebben ontwikkeld dat een veelvoud aan kwetsbaarheden, d.w.z. lekken in software, blootlegt in besturingsprogramma's van computers, browsers en servers. 'Je moet dit vergelijken met de situatie in de fysieke wereld dat je niks meer op slot kunt doen', reageert Aline Klingenberg. Haar collega Evgeni Moyakine vult aan: 'Zo'n AI-model kan namelijk óók misbruikt worden.'

KIRSTEN OTTEN WWW.RUG.NL/STAFF/A.M.KLINGENBERG WWW.RUG.NL/STAFF/E.V.MOYAKINE

ONDERZOEK

Klingenberg, hoogleraar Datadelen en Communicatierecht, bevindt zich in het Rölinggebouw in het hart van de stad Groningen. Naast haar zit Evgeni Moyakine, universitair hoofddocent IT-recht en Cyberveiligheid. Moyakine doet onderzoek naar de juridische mogelijkheid om landen verantwoordelijk te houden voor cyberaanvallen die door overheden en hun organisaties worden uitgevoerd. Het is voor hem en Klingenberg uiterst relevant dat door het oprukken van kunstmatige intelligentie nieuwe gaten worden geslagen in de beveiliging van computersystemen.

Cyberweerbaarheid verhogen

'AI is een mes dat aan twee kanten snijdt, vertelt Evgeni Moyakine. 'Terwijl zo'n AI-model als van Anthropic er eigenlijk voor bedoeld is om kwetsbaarheden op te sporen en bedrijven te helpen tijdig te reageren op cyberdreigingen', kan het ook misbruikt worden. Daar heb ik het vaak over in mijn onderwijs.' In een van de belangrijkste Europese richtlijnen, de NIS2-richtlijn, staat dat cybersecurity-experts AI moéten gebruiken om kwetsbaarheden op te sporen. De NIS2-richtlijn is bedoeld om de cyberweerbaarheid van bepaalde cruciale organisaties in de Europese Unie te verhogen. Denk aan sectoren als de gezondheidszorg, energie, transport en onderwijs.

'Anthropic heeft er nu voor gekozen dit

AI-model, met de naam Mythos Preview, voorlopig slechts vrij te geven aan een select aantal bedrijven zoals Google, Apple, Amazon, Cisco, Microsoft en NVIDIA. Zo kunnen ze onderzoeken wat hiervan de gevolgen zijn voor hun systemen. Maar een groepje gebruikers van het chatplatform Discord beweert ook toegang tot Mythos te hebben gekregen. Ze zeggen geen kwaadaardige bedoelingen te hebben, maar je weet nooit wie het precies in handen heeft en wat ze er mee gaan doen.

Gatenkaas in beveiliging

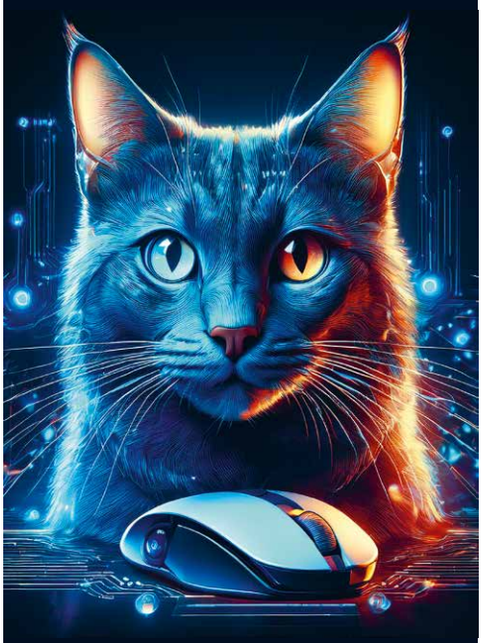
Wat kunnen de gevolgen zijn? Mythos wordt als gevaarlijker beschouwd dan andere modellen vanwege zijn uitzonderlijke vermogen om onbekende cyberkwetsbaarheden te ontdekken en te exploiteren, waardoor het risico op grootschalige cyberaanvallen en misbruik toeneemt. Klingenberg denkt terug aan de recente hack van onderwijssysteem Canvas, opgeëist door hackersgroep ShinyHunters. Deze groep zat ook achter de eerdere hack bij telecombedrijf Odido. ‘Wij hebben Canvas gelukkig niet op de universiteit’, vertelt de hoogleraar. ‘Maar ik hoorde van opleidingen die nu geen diploma’s kunnen uitgeven, omdat ze de tentamens of proefwerken niet kunnen nakijken. Dat heeft allerlei consequenties voor vervolgopleidingen en voor solliciteren.’

In theorie heeft ook een bedrijf als Anthropic er geen baat bij het internet te ontregelen met een disruptief AI-model dat gatenkaas maakt van digitale beveiligingen. Daarom is het bedrijf de samenwerking aangegaan met de grote Amerikaanse techbedrijven, onder de noemer van ‘Project Glasswing’. ‘Maar je bent wel opeens afhankelijk van de moraliteit van zo’n bedrijf wie allemaal dat systeem mag gaan gebruiken’, werpt Klingenberg tegen. ‘Terwijl dit vergelijkbaar is met een bedrijf dat gespecialiseerde inbrekersapparatuur op de markt brengt. Je mag toch ook niet zomaar wapens verkopen?”

Amerika vs Europa

Hieronder schuilt een fundamenteel verschil tussen de Amerikaanse en Europese marktordening. ‘In Amerika werkt het systeem anders dan in Europa: daar mag alles, tenzij het expliciet niet mag. In Europa denken we al eerder: is dit wel gewenst, en wat voor samenleving willen we eigenlijk hebben? Die manier van denken vloeit voort uit de periode na de Tweede Wereldoorlog, waarin we ons meer op mensenrechten zijn gaan baseren. In Amerika heerst veel meer de vrije ondernemersgeest, die pas later in het proces begrensd wordt door rechterlijke uitspraken over grote schadevergoedingen bijvoorbeeld bij Facebook. In Europa worden nieuwe technologieën doorgaans eerder beoordeeld op privacy, veiligheid en maatschappelijke gevolgen.

Aline Klingenberg: ‘In Amerika werkt het systeem anders dan in Europa: daar mag alles, tenzij het expliciet niet mag’



Evgeni Moyakine: ‘AI is een mes dat aan twee kanten snijdt'

Digitale wetgeving is moeilijk mondiaal te reguleren. ‘Wij kunnen wel zeggen: we willen niet van die chloorkippen op de markt hebben, of bepaald vlees met hormonen. Maar met digitale producten is zoiets moeilijker. Dat is een taaie strijd. De Europese Commissie probeert wel van alles, met toezicht houden en het opleggen van boetes. Ik denk dan bijvoorbeeld aan de AVG (de Algemene verordening gegevensbescherming van de EU, *red*). De Europese Commissie heeft aan WhatsApp boetes opgelegd vanwege het verzamelen en gebruiken van gegevens. Zo’n bedrijf wordt in Europa met boetes geconfronteerd die het in de VS niet krijgt opgelegd.’

Hackers voor zijn

Het is duidelijk dat taalmodellen als Mythos een grote kluif zijn voor de wereld van cybersecurity. Evgeni Moyakine: ‘Bedrijven en instellingen willen hun systemen en netwerken zo goed mogelijk beschermen. Maar hoe doe je dat als je weet dat hackers toegang hebben tot AI-systemen en effectief kunnen toeslaan? Dan moet je je bedrijfsprocessen aanpassen. Het Nationaal Cyber Security Centrum heeft ook gereageerd op het nieuws over Mythos van Anthropic en zegt: houd rekening met dit soort AI-ontwikkelingen en neem ze mee in de technische maatregelen die je treft, bijvoorbeeld bij het *patch management* (het pleisters plakken op gaten in de bescherming, *red*). Want je wilt hackers voor zijn. Als er een kwetsbaarheid is, kunnen deze kwaadwillende aanvallers meteen toeslaan, omdat ze veel sneller en efficiënter zijn dan organisaties met traag verlopende besluitvormingsprocessen.’

‘Het blijft een kat-en-muisspel’, zegt Klingenberg tot slot. ‘Regelgeving loopt in haar essentie achter op maatschappelijke ontwikkelingen. Anders zou je een Sovjetplaneconomie krijgen. De wetgever kan niet van tevoren bedenken: over twee jaar gebeurt in de digitalisering dít, dus laat ik alvast een wet maken.’