INFORMATIEBEVEILIGING MAGAZINE

Dossier: AI in cybersecurity

- Generatieve AI vergt meer dan code en data
- GenAI: een nieuwe destructieve technologie of een evolutionaire stap?
- Column: Vertrouwen in AI

JAARGANG 24 - 2024 - EDITIE 2

ISOPlanner

Eenvoudig Compliance Management in **Microsoft 365**

Waarom Microsoft 365?

Simpel: benut de kracht van Microsoft.

- Iedereen heeft Outlook en Teams. Naleving van compliance wordt als onderdeel van het werk ervaren.
- Vertrouwelijke gegevens zoals beleidsdocumenten en bewijsmateriaal blijven in jouw omgeving.
- Je lift vanzelf mee met innovaties. Bijvoorbeeld detectie van kwetsbaarheden en het automatisch ophalen van bewijs.

ISOPlanner is de enige integrale compliance oplossing in Microsoft 365. ISOPlanner brengt ISO naar jouw medewerkers toe en zorgt voor betere acceptatie en borging van informatiebeveiliging in jouw organisatie. Uiteraard ook voor de overheid (BIO), zorg (NEN7510) en de NIS2.

Snelle implementatie van je ISMS door meegeleverde voorbeeldmaatregelen, templates en voorbeelddocumenten.

SPIE NL IT heeft de ISO 27001 certificering met een positief resultaat doorlopen. Wat mede heeft bijgedragen aan dit mooie resultaat was de inzet van ISOPlanner als ISMS.

Leon van der Valk
SPIE Nederland B.V.

Kijk op **www.isoplanner.app** voor meer informatie en jouw gratis proefperiode.

ISO 27001 2023 COMPLIANT

GDPR 2023 COMPLIANT

PENTEST 2023 SECURE

O, zit dat zo!

Chris de Vries

Kunstmatige Intelligentie (K.I.), of zoals het gebruik eist Artificial Intelligence (AI), is het thema van dit nummer. Zoals zoveel hypebegrippen weten wij allemaal wat het is en staan wij er niet al te vaak bij stil wat het werkelijk betekent. In deze uitgave hebben wij de kans om onzekerheden weg te nemen en de 'O zit het zo!'-ervaring te beleven.

Een keur aan professionals dragen hun steen(tje) bij aan onze kennispoel, waaronder vanuit de ministeriële overheid (Dion Koeze & Ruben Faber). In hun artikel lees je welke risico's zij zien en hoe wij zouden kunnen handelen. Mr. Arnoud Engelfriet (IT-informaticus & jurist) licht de AI Act consequenties toe. En dan Vincent van Dijk (security scientist), die ons op zijn AI-ontdekkingsreis meeneemt.

KI/AI evolueert en zoals de ontwikkelingen laten zien, verrassend snel. Van de nauwe opvatting naar de generatieve tot de kunstmatige algemene intelligentie. Is dat enkel een woordplaats wisselspel of zegt het echt iets meer? Ervaar het in deze uitgave.
De nauwe opvatting gaat over machinaal – tot het diep-leren én van voorspellingsmodellen naar verwerking van natuurlijke taal. Generatieve intelligentie omvat de grote taalmodellen, generatie van afbeeldingen, audio/video en de multimodale, fijn afgestemde en grote actiemodellen (`Agentic Models'). De laatste groep is nog niet uitgekristalliseerd. Zij bestaat merendeels nog niet, maar vormt zich naar onze eigen menselijke denkprocessen. Enerzijds verwacht men een 'hockeystick' groei, anderzijds vrezen wij de existentiële dreiging. Waar kom jij uit na lezing van deze uitgave? Laat het ons op LinkedIn weten of reageer op de standpunten in Achter Het Nieuws van redactieleden Fook Hwa Tan en Leo van Koppen.

Menselijke intelligentie is meer dan alleen ratio. Gevoel, intuïtie, gewaarwording en zeker niet te vergeten humor maken er deel van uit. En dat vinden wij in de columns van Dimitri van Zantvliet volop terug. In deze uitgave 'vinden' en vanaf de volgende uitgave 'vonden', want na vele bijdragen van Dimitri geeft hij nu het stokje aan een ander over. Wij zullen zijn bijdragen missen. Ze daagden uit, prikkelden ons tot weerspraak en nadenken. Lieten ons de vaste waarden van ons ó zo rationeel vak ter discussie stellen. In zijn laatste bijdrage refereert ook hij aan KI/AI en wel met de woorden: *'... aangejaagd door A.I. niet alleen verder evolueert; het vereist daarboven ook een revolutie in hoe we denken en handelen ...'*. Om af te sluiten met: *'... Vaarwel, tot we wederkeren in deze bijzondere tijd, waar onze cyberpaden zich andermaal mogen kruisen onder het wakend oog van het lot ...'*.

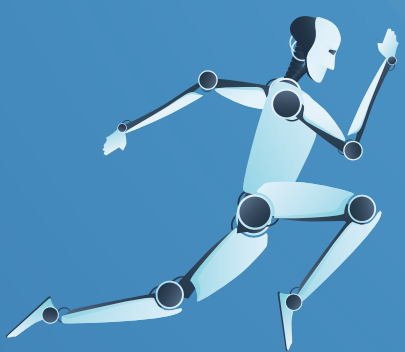
Wij, als redactie, zeggen tegen Dimitri: "Dank voor jouw mooie, kwalitatief hoogstaande bijdragen. Niet als een vaarwel, maar als een tot de volgende ontmoeting."

Chris

IN DIT NUMMER

Auteur: Fook Hwa Tan is redactielid van iB-Magazine en chief quality officer bij de Northwave Group.
Hij is bereikbaar via: fookhwatan@northwave-security.com.

cybersecurity



AI in cybersecurity: navigeren in een nieuw competentielandschap

Naarmate AI dieper integreert in cybersecurity, staan professionals voor ongekende besluiten: wanneer AI-adviezen op te volgen of terzijde te schuiven? Dit artikel belicht een nieuw soort competenties die nodig zijn om deze kritieke keuzes te maken. Het focust op de vaardigheden voor het beoordelen, interpreteren en ethisch inzetten van AI, en de impact daarvan op de toekomst van cybersecurity.

De opkomst van kunstmatige intelligentie (AI) transformeert de cybersecuritywereld, waarbij nieuwe soorten competenties en besluitvormingsprocessen centraal staan. Cybersecurityprofessionals moeten nu niet alleen technische, procedurele en menselijke kennis hebben, maar ook de vaardigheden om de betrouwbaarheid en relevantie van AI-gegenereerde adviezen te beoordelen op deze vlakken. Dit vereist een nieuwe benadering van training en praktijk, waarin ethiek en kritisch denken voorop staan.

Nieuwe competenties in het AI-tijdperk

Met nieuwe competenties moet men denken aan kritisch denken en beoordelingsvermogen, ethiek en verantwoordelijkheid en AI-geletterdheid. Een cybersecurityprofessional gebruikt een AI-systeem dat netwerkverkeer analyseert om verdachte activiteiten te identificeren. Wanneer het systeem een reeks transacties als risicovol markeert, vertrouwt de professional niet blindelings op deze beoordeling. In plaats daarvan onderzoekt hij de onderliggende data, evalueert de trainingsdataset van het AI-systeem, en overweegt of de markering beïnvloed kan zijn door recente, legitieme veranderingen in netwerkgedrag of mogelijke biases in de trainingsdata.

In de context van AI en cybersecurity vereist kritisch denken dat professionals niet blindelings vertrouwen op AI-adviezen, maar deze beoordelingen analyseren en vraagtekens plaatsen bij hun geldigheid. Het gaat om het begrijpen van hoe bepaalde conclusies zijn bereikt door:

- Te beoordelen of de data die gebruikt is voor het trainen van AI-systemen relevant, actueel en vrij van biases is;
- Inzicht te krijgen in de mechanismen achter AI-beslissingen, dat helpt bij het identificeren van eventuele zwakheden of foutmarges in de algoritmes;
- Te weten dat AI-systemen bestaande vooroordelen kunnen versterken als ze worden getraind op bevooroordeelde datasets. Het is essentieel dat professionals in staat zijn om deze biases te herkennen en hier kritisch op te reageren.

Stel dat een AI-gestuurd cybersecuritysysteem een potentieel gevaarlijke softwaretoepassing op een bedrijfsnetwerk identificeert. Voordat actie wordt ondernomen, overweegt de verantwoordelijke professional de mogelijke gevolgen van het isoleren of verwijderen van deze toepassing. Zij houdt rekening met vragen zoals: 'Kan dit legitieme bedrijfsprocessen verstoren?' en: 'Is er voldoende bewijs om deze actie te rechtvaardigen?' Hierbij houdt ze niet alleen rekening met de technische, maar ook met de ethische en bedrijfsmatige implicaties, waarbij ze de privacyrechten en het welzijn van de gebruikers waarborgt.

De ethische kant van AI in cybersecurity benadrukt dat technologische vooruitgang hand in hand moet gaan met morele

overwegingen. Professionals moeten:

- Kennis hebben van fundamentele ethische principes, zoals: rechtvaardigheid, eerlijkheid en respect voor privacy hebben, dat is cruciaal;
- Beslissingen over het gebruik van AI-adviezen zorgvuldig overwegen, waarbij de verantwoordelijkheid voor de gevolgen wordt erkend;
- Voorbereid zijn op het navigeren van complexe situaties waarin verschillende ethische waarden met elkaar in conflict kunnen zijn.

Een cybersecurityteam implementeert een nieuw AI-systeem dat is ontworpen om phishing-aanvallen te detecteren. Een lid van het team neemt de tijd om de documentatie van het systeem te bestuderen, begrijpt de aard van de algoritmen die worden gebruikt en de soorten data waarop het systeem is getraind. Dit begrip stelt hem in staat om de effectiviteit van het systeem beter te evalueren en te begrijpen in welke contexten het systeem het best presteert. Wanneer het systeem een false positive geeft, kan hij door zijn kennis ingrijpen, de fout corrigeren en feedback geven voor verdere verbetering van het systeem.

In een tijdperk waarin AI een steeds grotere rol speelt, is het essentieel dat cybersecurityprofessionals niet alleen gebruikers van AI zijn, maar ook inzicht hebben in de onderliggende technologie. Dit omvat:

- Een grondig begrip van hoe AI en machine learning werken, helpt professionals om de potentie en beperkingen van deze technologieën te begrijpen;
- Inzicht in hoe AI kan worden ingezet voor diverse cybersecuritydoeleinden, zoals dreigingsdetectie en respons, verbetert de effectiviteit en efficiëntie van beveiligingsstrategieën;
- Kennis van de beperkingen van AI, zoals de afhankelijkheid van de kwaliteit van de gebruikte data en het potentieel voor onvoorziene fouten, is essentieel voor het verantwoord inzetten van deze technologie.

Door zich deze competenties eigen te maken, kunnen cybersecurityprofessionals beter geïnformeerde beslissingen nemen over het gebruik van AI, de ethische implicaties van hun acties overwegen en effectief samenwerken met AI-technologieën om de cyberveiligheid te versterken.

Besluitvorming met AI

Bij het nemen van besluiten op basis van AI of AI autonoom besluiten te laten nemen rijzen de volgende vragen: wanneer is AI te vertrouwen? Wanneer is menselijke interventie noodzakelijk? Hoe werken mens en machine samen?

AI-systemen in cybersecurity kunnen buitengewoon effectief zijn bij het uitvoeren van repetitieve en data-intensieve taken, zoals het monitoren van netwerkverkeer, het identificeren van bekende malwarehandtekeningen of het detecteren van afwijkingen die wijzen op een datalek. Het vertrouwen in AI-gebaseerde aanbevelingen hangt echter sterk af van de context en de specifieke toepassing. Bijvoorbeeld: AI kan bijzonder betrouwbaar zijn in het snel identificeren en categoriseren van bekende dreigingen, vanwege het vermogen om enorme datasets met grote snelheid nauwkeurig te kunnen analyseren; dat kunnen mensen niet.

Een goed voorbeeld is een AI-systeem dat is getraind met uitgebreide datasets van phishing-e-mails. Zodra het systeem een hoge nauwkeurigheid bereikt in het herkennen van dergelijke e-mails, kunnen cybersecurityteams op AI vertrouwen om deze dreigingen automatisch te identificeren en te isoleren, waardoor de responstijd wordt geminimaliseerd en het potentieel voor menselijke fouten wordt verkleind.

Ondanks de kracht van AI zijn er situaties waarin de nuances van menselijke ervaring en beoordeling onmisbaar zijn. Dit is met name het geval in scenario's waar de context verandert of waar AI geconfronteerd wordt met nieuwe, onbekende dreigingstypen die niet in de trainingsdata voorkwamen. Menselijke experts hebben het vermogen om bredere contextuele aanwijzingen te interpreteren, creatief te denken en beslissingen te nemen op basis van onvolledige of tegenstrijdige informatie.

Een voorbeeld is de detectie van een zero-day exploit, een nieuwe kwetsbaarheid die nog niet bekend is bij de AI. In dit geval is de expertise van een menselijke analist vereist om ongebruikelijke systeemgedragingen te interpreteren, correlaties te leggen en de dreiging te bevestigen of te ontkennen op basis van een holistisch begrip van het IT-landschap en de heersende cyberdreigingen.

Om de samenwerking tussen mens en machine effectief te maken, is een duidelijk framework voor besluitvorming essentieel. Dit framework helpt te definiëren wanneer en hoe AI-adviezen kunnen worden geïntegreerd en wanneer menselijke interventie vereist is. Een dergelijk framework zou aspecten moeten omvatten zoals:

Ondanks de kracht van AI zijn er situaties waarin de nuances van menselijke ervaring en beoordeling onmisbaar zijn

- Duidelijke criteria voor de betrouwbaarheid van AI-adviezen, inclusief prestatiebenchmarks en foutmarges;
- Protocollen voor situaties waarin AI-adviezen moeten worden geëscaleerd naar menselijke beslissers;
- Trainingsprogramma's om de AI-geletterdheid van het personeel te vergroten, zodat zij beter geïnformeerde beslissingen kunnen nemen over de AI-adviezen;
- Regelmatige evaluatie van de AI-systemen, met aanpassingen gebaseerd op feedback van menselijke gebruikers en veranderingen in het dreigingslandschap.

Een praktische uitwerking van dit framework kan worden geïllustreerd aan de hand van een incident respons protocol dat specificeert hoe alerts van het AI-systeem worden behandeld, wie verantwoordelijk is voor de eindbeoordeling, en hoe beslissingen worden gedocumenteerd en geanalyseerd voor toekomstige verbeteringen.

Deze diepgaande benadering van besluitvorming met AI in cybersecurity zorgt voor een synergie tussen menselijke expertise en machine-efficiëntie, waarbij elk zijn rol speelt in het waarborgen van de organisatorische cyberveiligheid.

Ethiek voorop

Waar mensen het meest bang voor zijn, is het feit dat machines de controle overnemen van mensen. Daarom is het belangrijk om ethiek voorop te blijven stellen. Hierbij dient aandacht aan het volgende te worden besteed: ethische overwegingen bij gebruik AI, verantwoordelijkheid voor AI-beslissingen en praktische uitwerkingen van ethiek.

Een AI-systeem dat wordt ingezet voor het screenen van veiligheidsrisico's bij werknemers moet transparant zijn over de criteria die het gebruikt, de privacy van individuen respecteren en vrij zijn van elke vorm van discriminatoire bias. De organisatie moet deze aspecten duidelijk communiceren naar alle betrokkenen en regelmatig audits uitvoeren om de naleving van deze ethische normen te waarborgen.

De implementatie van AI in cybersecurity moet gepaard gaan met strikt ethische overwegingen. Transparantie is cruciaal; gebruikers en belanghebbenden moeten begrijpen hoe AI-systemen werken, op welke data ze zijn getraind, en hoe beslissingen worden genomen. Dit bevordert vertrouwen en acceptatie.

Privacybescherming is een ander essentieel aspect. AI-systemen die persoonlijke of gevoelige informatie verwerken, moeten dit doen met respect voor de privacy van individuen, conform de geldende wet- en regelgeving, zoals de Algemene Verordening Gegevensbescherming (AVG) in de EU.

Non-discriminatie is ook een belangrijk punt: AI-systemen moeten vrij zijn van vooroordelen die kunnen leiden tot discriminatie. Dit vereist zorgvuldige selectie en monitoring van de trainingsdata om te verzekeren dat deze representatief en onbevooroordeeld is.
In het geval van een AI-gestuurde detectie van cybersecurity-dreigingen, moet duidelijk zijn wie binnen de organisatie verantwoordelijk is voor de beoordeling en actie op basis van deze detecties. Als een AI-systeem een fout maakt, zoals het ten onrechte classificeren van legitiem netwerkverkeer als kwaadaardig, moet er een procedure zijn om deze beslissing te herzien en de verantwoordelijke personen of teams moeten de bevoegdheid hebben om corrigerende maatregelen te nemen.

Organisaties moeten verantwoordelijkheidsmechanismen instellen voor beslissingen die door AI worden beïnvloed of gemaakt. Dit betekent dat er altijd een duidelijke lijn moet zijn over wie verantwoordelijk is voor de uitkomsten van deze beslissingen, inclusief de mogelijkheid om in te grijpen wanneer een beslissing herzien moet worden.

Een cybersecurityteam zou regelmatige training moeten krijgen over ethische aspecten van AI, inclusief scenario's en oefeningen die specifiek gericht zijn op de ethische dilemma's die ze kunnen tegenkomen. Ethiekcommissies of adviesraden kunnen worden ingesteld om te overleggen over complexe gevallen en richtlijnen bieden voor ethisch verantwoorde beslissingen.
Het daadwerkelijk implementeren van ethische principes in de dagelijkse cybersecuritypraktijken en besluitvormingsprocessen vereist een cultuurverandering en voortdurende aandacht voor ethische training en bewustzijn.
Door deze stappen te nemen, kunnen organisaties ervoor zorgen dat hun gebruik van AI in cybersecurity niet alleen effectief is in het beschermen tegen bedreigingen, maar ook ethisch verantwoord, transparant en in overeenstemming met de hoogste standaarden van privacy en rechtvaardigheid.

Voorbereiding op de toekomst

AI dit gezegd hebbende, hoe kunnen we ons voorbereiden op de toekomst? Hierbij moeten we denken aan educatie en training, professionalisering en dialoog om AI op een juiste manier in te zetten ter bevordering van onze doelstellingen.
In een wereld waarin AI steeds meer verweven raakt met cybersecurity, is het cruciaal dat professionals uitgerust zijn met de juiste kennis en vaardigheden. Gespecialiseerde opleidingen die de kruising van AI, ethiek en cybersecurity aanpakken, zijn essentieel. Deze opleidingen moeten niet alleen technische vaardigheden bijbrengen, maar ook diepgaand inzicht geven in de ethische implicaties van AI-gebruik in cybersecurity.

Een cybersecurityprofessional volgt een cursus waarin hij leert over de nieuwste AI-technologieën die worden gebruikt voor dreigingsdetectie. De cursus behandelt ook hoe deze systemen worden getraind, welke ethische overwegingen erbij komen kijken, en hoe bias en privacykwesties kunnen worden aangepakt. Deze kennis stelt de professional in staat om niet alleen technisch effectiever te zijn, maar ook om ethische overwegingen een plaats te geven in zijn werk.

De technologiewereld evolueert razendsnel, en wat vandaag nieuw is, kan morgen verouderd zijn. Levenslang leren en het vermogen om zich snel aan te passen aan nieuwe ontwikkelingen zijn daarom cruciale competenties voor elke professional in dit veld. Dit houdt in dat men voortdurend op de hoogte moet blijven van de laatste trends, technologieën en best practices.
Een IT-beveiligingsteam organiseert maandelijkse bijeenkomsten om de laatste ontwikkelingen op het gebied van AI en cybersecurity te bespreken. Ze nodigen regelmatig externe experts uit om lezingen te geven en nemen deel aan online forums en communities. Door deze activiteiten blijven ze niet alleen geïnformeerd, maar worden ze ook gestimuleerd om hun kennis en vaardigheden voortdurend bij te werken.

AI, ethiek en cybersecurity zijn interdisciplinaire velden die baat hebben bij een brede aanpak. De samenwerking tussen techneuten, ethici, juristen en beleidsmakers is essentieel om holistische en duurzame oplossingen te ontwikkelen. Dit soort samenwerking kan helpen bij het identificeren van gemeenschappelijke uitdagingen en het uitwisselen van beste praktijken.
Een cybersecurityorganisatie richt een werkgroep op met AI-ontwikkelaars, ethische adviseurs, juridische experts en vertegenwoordigers van regelgevende instanties. Samen beoordelen ze nieuwe AI-tools, bespreken de implicaties ervan voor privacy en beveiliging, en werken aan beleidsaanbevelingen die zowel innovatie stimuleren als ethische en juridische normen respecteren. Door deze stappen te nemen, kunnen professionals en organisaties in de cybersecuritysector, maar ook daarbuiten zich effectief voorbereiden op de toekomst, waarbij ze niet alleen technologisch vooroplopen, maar ook ethisch en maatschappelijk verantwoord handelen.

Hoe nu verder?

Terwijl AI de grenzen van mogelijkheden in cybersecurity herdefinieert, staan we op een kruispunt van uitdaging en kans. De toekomst roept professionals op om verder te kijken dan de code en circuits; het is een uitnodiging om pioniers te zijn in een tijdperk waarin technologie en menselijkheid samensmelten. In deze dynamische arena moeten we niet alleen technische meesters zijn, maar ook ethische architecten, die met inzicht en integriteit de digitale werelden vormgeven.

De reis naar morgen vraagt om een nieuwe soort moed - de moed om voortdurend te leren, uit te dagen wat we kennen, en ethische principes te verankeren in elke beslissing en innovatie. Dit is niet alleen een roep om actie, maar ook een kans om te leiden, te inspireren en de toekomst van cybersecurity actief te vormen.

Laat ons samen deze kans grijpen om niet alleen de wachters van cyberspace te zijn, maar ook de bouwers van een veilige, rechtvaardige en florerende digitale toekomst. Sta op en omarm deze uitdaging, laat zien dat wij, de cybersecurityprofessionals van vandaag, klaar zijn om de architecten te zijn van morgen. Verenigd in onze toewijding en onze vaardigheden, laten we een toekomst smeden waarin AI en ethiek hand in hand gaan, waardoor we een wereld creëren die veiliger, eerlijker en vol mogelijkheden is voor iedereen.

COLUMN PRIVACY

Mr. Rachel Marbus
@RACHELMARBUS OP TWITTER

Het transparante, lerende kind

Eind januari bracht de Autoriteit Persoonsgegevens het Sectorbeeld Onderwijs uit. De reflecties die geschetst worden, zijn niet enorm verrassend – het kan vooral beter, maar dat geldt natuurlijk vaak ook binnen andere sectoren. Wel dwingt de AP ons met de neus op de feiten. Jongeren zijn een kwetsbare groep en hebben daarom extra bescherming nodig. In zoverre is de onderwijssector wel atypisch. Kinderen, schrijft de AP, hebben die extra bescherming nodig zodat zij zich in een vrije en veilige omgeving kunnen ontwikkelen. Als je dat in gedachten houdt, is het beeld van de huidige 'staat van zijn' wellicht verontrustend. Want die basis is nog niet overal op orde en nieuwe uitdagingen staan alweer voor de deur met de ontwikkelingen op het gebied van Artificial Intelligence en de inzet van algoritmes. Instellingen hebben ook steeds meer gevoelige informatie over jongeren tot hun beschikking over onder meer het gedrag dat zij vertonen. Maar ook worden zij beoordeeld met behulp van AI en algoritmes en daar schuilen gevaren van vooringenomenheid, een verkeerde interpretatie van de data en een verlies van controle over de persoonsgegevens.

Het is ook maar de vraag of het altijd rechtmatig is dat onderwijsinstellingen al te gevoelige gegevens verwerken. Je moet dan denken aan onder meer verwerkingen over psychisch welzijn van leerlingen en studenten, veiligheid, armoede, kansen(on)gelijkheid en polarisatie. AP signaleert dat dit vaak onvoldoende doordacht gebeurt, weliswaar met de juiste intenties, dat maakt het niet minder nijpend. De toezichthouder wijst hier ook met de vinger naar de wetgever, die daarin beter moet voorzien. Dat gaat denk ik wel volledig voorbij aan de vraag of we dit ook zouden moeten willen. Daar zou dus een inherent ethische vraag vooraf gesteld moeten worden. Je kunt je daarnaast afvragen in hoeverre schoolgaande jongeren überhaupt nog de ruimte hebben om onbevangen zichzelf te zijn. Zeker in het voortgezet onderwijs zie je dat de hele 'schoolreis' nauwgezet in de gaten gehouden kan worden en gedeeld wordt met ouders en verzorgers. Waar vroeger nog ruimte was om die pittige onvoldoende even onder de pet te houden en te compenseren, is daar nu geen plek meer voor. Ouders weten vaak eerder wat er gescoord is dan de jongere zelf (mobieltjes mogen immers niet meer mee de klas in, maar ouders hebben die continue in de hand). Een perspectief dat nog wel eens vergeten wordt en mij pijnlijk raakte, toen ik laatst een bericht las van een docent die door een bange jongere werd gevraagd de cijfers niet direct online te zetten "omdat mijn vader anders heel erg boos op mij gaat worden".

Dit raakte me gemeen hard. En het zou meer mensen moeten raken. Ooit eens eerder pleitte ik ervoor om het gesprek aan te gaan met jongeren (pubers), maar ik denk dat het eens te meer noodzakelijk is. Al was het maar om meer van dit soort signalen boven te krijgen zodat het mee kan wegen als wij volwassenen een ethische en privacy-impactassessment doen. Het is heel normaal om in assessments de belangen van de betrokkenen mee te wegen, maar hoe weeg je waar je zelf onvoldoende van op de hoogte bent? Door de jongeren zelf ook een duidelijker stem te geven.

Rachel

Auteurs: R.E.J. (Ruben) Faber, werkt bij het NCSC als strategisch adviseur en is bereikbaar via https://nl.linkedin/in/rubenfaber en dr. ir. D.J. (Dion) Koeze, eveneens werkzaam bij het NCSC als onderzoeker, hij is bereikbaar via research@ncsc.nl.

Generatieve AI vergt meer dan code en data

In het snel evoluerende landschap van de informatietechnologie staat generatieve kunstmatige intelligentie (GenAI) aan de vooravond van een transformatie. Deze technologie, die zich richt op het creëren van content – van tekst tot beelden en zelfs code – door het leren van enorme datasets, heeft zich snel ontwikkeld tot een kerncomponent binnen de innovatiestrategieën van organisaties. Met vooruitgangen zoals OpenAI GPT-4, Google's Gemini en Anthropic's Claude-3, die mensachtige teksten kunnen genereren op basis van een brede en ongestructureerde context, is de potentie van generatieve AI onmiskenbaar.

De interesse in GenAI overstijgt industrieën en grenzen, waarbij organisaties de mogelijkheden onderzoeken of al stappen zetten richting een concrete implementatie van de nieuwe technologie. De verschillende toepassingsvormen lijken door de brede inzetbaarheid van de technologie onbegrensd en bieden Nederlandse organisaties innovatiekansen die tot voor kort ondenkbaar waren. Bijna elke organisatie, van startup tot multinational, onderzoekt momenteel de mogelijkheden. Zo kan generatieve AI bijvoorbeeld worden ingezet om de communicatie met klanten in meerdere talen te vereenvoudigen, bij het genereren en stroomlijnen van juridische en financiële documenten, om complexe logistieke processen te optimaliseren of diagnostische ondersteuning van artsen binnen de gezondheidszorgsector.

Drie pijlers

Voor een effectieve integratie van generatieve AI in bedrijfsprocessen is het cruciaal om te begrijpen dat de ontwikkeling ervan niet alleen draait om code en data, maar dat het een nieuw paradigma vereist van de toeleveringsketen. Deze zienswijze bestaat uit drie pijlers, die hun weerslag kennen in alle dimensies van IT-management, -ontwikkeling en -operatie. Pijlers worden in dit geval breed gedefinieerd als: de route door het totale procuratie- en leveringsproces; de implementatie tot aan de oplevering in een productieomgeving en integratie in bestaande systemen. Binnen de IT-toeleveringsketen is deze driedeling van essentieel belang om GenAI op een toekomstbestendige manier aan te besteden, te ontwikkelen en te implementeren.

- **Code:** programmeren, ontwikkelen, testen, implementeren, kopen of adopteren van ICT oplossingen vallen onder deze pijler. Het zijn alle activiteiten die een ICT-organisatie vanuit haar bestaansrecht kent en doet – het ontwikkelen van een applicatie die een bepaald proces ondersteunt of in een bepaalde behoefte voorziet. Het gaat hier nadrukkelijk om de brede definitie, waaronder functioneel ontwerpen, testen, reviewen – deze processen kennen traditioneel gezien hun weerslag in de code en andere artefacten die uiteindelijk in een softwareoplossing bij elkaar worden gebracht.
- **Data:** de succesvolle werking van GenAI berust op de beschikbaarheid van voldoende data van goede kwaliteit. Zowel in de ontwikkeling van foundational models, de onderliggende modellen van bijvoorbeeld OpenAI, als bij het toespitsen (finetuning) van modellen voor de toepassing in de eigen organisatie. We kennen deze pijler ook uit toepassingen die voortkomen uit de Data Science en omvat alle

aspecten van de Data-pipeline: data-acquisitie, -verwerking en -beheer.
- **Machine Learning (ML):** deze pijler is de essentiële verrijking van de traditionele tweedeling tussen code en data. Het omvat trainen, valideren en implementeren van een AI-model. Ondanks dat Machine Learning gedaan wordt via een code en op data is het van belang om dit als separate pijler te onderkennen. Een getraind model wordt in moderne toepassingen niet meer losstaand gebruikt, maar bijvoorbeeld in een pipeline met *Retrieval Augmented Generation*, een techniek waarmee bestaande informatie van een systeem ontsloten kan worden door GenAI.

Het begrijpen en beheersen van deze drie pijlers is essentieel voor het veilig en effectief implementeren van GenAI. Dit vraagt om een verschuiving in denkwijze van traditionele ICT-benadering naar een meer geïntegreerde visie die de complexiteit en de onderlinge verbondenheid van code, data en ML erkent. Door deze benadering kunnen organisaties niet alleen de vruchten plukken van GenAI, maar ook zorgen voor robuuste cybersecurity en ethische inzet van deze krachtige technologie. De transitie naar dit nieuwe paradigma vereist leiderschap dat vooruitziet, bereid is om traditionele denkpatronen te doorbreken en de waarde van een multidisciplinaire benadering onderkent. Het succesvol navigeren door deze verandering zal niet alleen technische expertise vereisen, maar ook een diepgaand begrip van de strategische, ethische en maatschappelijke implicaties van GenAI. Zo kunnen we de belofte van deze revolutionaire technologie waarmaken, terwijl we de veiligheid en integriteit van ons digitale ecosysteem waarborgen.

Fundamenten van GenAI en cybersecurity

De samenvloeiing van GenAI en cybersecurity vormt een boeiend doch complex domein dat unieke uitdagingen en kansen biedt. Terwijl generatieve AI de grenzen van technologische innovatie verlegt, brengt het ook nieuwe risico's en kwetsbaarheden met zich mee die speciale aandacht vereisen binnen het cybersecuritylandschap. Het feit dat er een apart raamwerk voor aanvalstechnieken wordt bijgehouden door MITRE – het ATLAS raamwerk, naast het bekende ATT&CK raamwerk – geeft dit ook aan. Wat maakt cybersecurity bijzonder in de context van Generatieve AI? Het aanvalsoppervlak van applicaties die gebruik maken van GenAI kent nieuwe karakteristieken. Dit kunnen we beter begrijpen door een aantal typen aanvallen te bekijken.

- **Data Poisoning en Manipulatie:** bij GenAI is de integriteit van de trainingsdata cruciaal. Een aanval waarbij de data gemanipuleerd wordt (data poisoning) kan leiden tot het genereren van valse, misleidende, of ongewenste output. Dit vraagt om geavanceerde verificatie- en validatiemechanismen om de integriteit van data te waarborgen
- **Modeldiefstal en Reverse Engineering:** GenAI modellen vertegenwoordigen aanzienlijke intellectuele en financiële waarde. Aanvallers kunnen proberen deze modellen te stelen of via reverse engineering te dupliceren. Bescherming tegen dergelijke aanvallen vereist geavanceerde technieken, zoals modelversleuteling en -watermerking.
- **Injection Attacks:** ondanks dat injectieaanvallen – waarbij malafide code of instructies ingeschoten kunnen worden – ook in webapplicaties in het algemeen nog steeds een bekende aanvalstechniek zijn volgens de OWASP Top 10, zien deze aanvallen er fundamenteel anders uit wanneer met menselijke taal of andere media zoals foto's met een applicatie te interacteren is.

Om een adequate vertaalslag te maken naar de realiteit van het aanbesteden en implementeren van ICT-oplossingen kan de driedeling in de verschillende pijlers worden gebruikt.
De **codepijler** omvat de ontwikkeling van algoritmes en software die de ruggengraat vormen van AI-systemen. Alle beveiligingsrisico's die van toepassing zijn bij het ontwikkelen van veilige softwarecode, zijn daarom ook van toepassing wanneer binnen het domein van GenAI wordt ontwikkeld: secure coding, (geautomatiseerde) code review en het adequaat beveiligen van omgevingen waar de code in wordt ontwikkeld en gedistribueerd.

Binnen de **datapijler** is de grootste zorg de bescherming van de data die wordt gebruikt voor het trainen en valideren van AI-modellen. Dit omvat maatregelen tegen ongeautoriseerde toegang, datalekken, en de ongewilde manipulatie van trainingsdata. Belangrijk hierbij is het besef dat data een steeds grotere rol gaat spelen in de toeleveringsketen, doordat GenAI rust op veel data van goede kwaliteit. Net zoals code gecompromitteerd kan worden stroomopwaarts in de keten, kan datzelfde gebeuren met de onderliggende data van modellen die worden ingezet. Verder spelen op dit domein veel bestuurlijke en juridische overwegingen, zoals de herkomst van data en bescherming van persoonlijke gegevens. Deze vraagstukken kunnen gedeeltelijk technisch worden opgelost (pseudonimi-

Het begrijpen en beheersen van de pijlers code, data en Machine Learning is essentieel voor het veilig en effectief implementeren van GenAI

sering, anonimisering), maar vergen vooral een sterk gefundeerde visie die onderdeel is van de bedrijfscultuur: wat zien wij als veilig, rechtmatig en ethisch verantwoord gebruik van data?

In de **Machine Learning-pijler** gaat het om bescherming en doorlopende bijwerking van de modelarchitectuur, toetsing en validatie van uitkomsten en onbedoelde effecten, in kaart houden van modelspecifieke kwetsbaarheden, en ketenintegriteit van het model van training tot ingebruikname. Steeds meer zal het gebeuren dat het trainen van modellen niet meer gebeurt bij een organisatie zelf of een directe leverancier, maar dat er door een keten heen modellen worden aangepast, via finetuning of few-shot learning, voordat het tot ingebruikname komt.

Concrete productontwikkeling met Generatieve AI

Om de verdeling over de drie domeinen verder in te kleuren ontwikkelen we als voorbeeldcasus een chatbot die namens een organisatie communiceert met klanten of medewerkers. Dit is nog een betrekkelijk 'eenvoudige' toepassing in vergelijking met eerder genoemde kansen, maar daarmee een oplossing waar veel organisaties momenteel concreet mee bezig zijn. Daarnaast heeft elke consument (wisselende) ervaringen met dergelijke bots. Wanneer een organisatie een chatbot wil ontwikkelen kan de driedeling in pijlers wederom worden toegepast.

Code

Bij de ontwikkeling van de chatbot is het eerste aandachtspunt de codepijler. Dit behelst het ontwerp en de implementatie van de software die de basis vormt van de chatbot. Secure-by-design principes zijn hier cruciaal: vanaf het begin moet de chatbot worden ontwikkeld met een sterke focus op de beveiliging en privacy. Dit betekent dat ontwikkelaars rekening houden met potentiële beveiligingsrisico's en kwetsbaarheden, zoals cross-site scripting of fouten in access control, en deze proactief aanpakken door middel van onder andere moderne coderingsstandaarden, code reviews, voldoende monitoring en logging, en geautomatiseerde beveiligingstests. Wanneer de chatbot ook acties kan ondernemen naast het genereren van tekst of andere media, zogenaamd function calling, om bijvoorbeeld meer gegevens op te kunnen halen of om acties uit naam van de gebruiker uit te voeren, is het belangrijker dan ooit om zorgvuldige maatregelen te nemen. Daarnaast moet de chatbot worden geprogrammeerd om te voldoen aan de GDPR en andere relevante privacywetgeving, bijvoorbeeld door gebruikers duidelijk te informeren over het gebruik van hun gegevens en hen controle te geven over hun persoonlijke informatie.

Data

De datapijler betreft het verzamelen, verwerken en beheren van de data die de chatbot gebruikt om te leren en te functioneren. Dit omvat zowel de initiële trainingsdata, de data die gebruikt wordt bij fine-tuning of few-shot learning, alsook de voortdurende input van gebruikersinteracties. Het waarborgen van de kwaliteit en diversiteit van deze data is essentieel voor de effectiviteit van de chatbot. Om bias en onnauwkeurigheden te voorkomen, moeten data scientists en ontwikkelaars zorgvuldig selecteren welke data wordt gebruikt voor het trainen van de chatbot. Dit begint bij de selectie van het model waarop de applicatie wordt gebaseerd en loopt door tot de laatste prompt engineering die nodig is om de bot de gewenste interactie te geven. Tegelijkertijd moeten ze strikte privacyrichtlijnen volgen om te zorgen dat persoonsgegevens beschermd worden. Dat betekent dat alle verzamelde data geanonimiseerd of gepseudonimiseerd moet worden en dat data alleen voor specifieke, gerechtvaardigde doeleinden wordt gebruikt.

Machine Learning

Ten slotte omvat de Machine Learning-pijler het trainen, valideren en implementeren van de AI-modellen die de chatbot in staat stellen om te leren van interacties en in de loop van de tijd te verbeteren. Dit proces moet zorgvuldig worden beheerd om te

zorgen voor de betrouwbaarheid en veiligheid van de chatbot. Dit houdt in dat er maatregelen worden getroffen om overfitting te voorkomen, dat er validatiestappen worden ingebouwd om de accuraatheid van de chatbot te verzekeren en dat er voortdurend wordt gemonitord op mogelijke veiligheidsrisico's die kunnen ontstaan door manipulatie van de input, bijvoorbeeld via adversarial attacks. Daarnaast is het van belang dat het model regelmatig wordt geëvalueerd en bijgewerkt op basis van nieuwe data en feedback van gebruikers, om zo te zorgen voor een continue verbetering van de prestaties en gebruikerservaring.

Door de drie pijlers zorgvuldig toe te passen en te integreren in het ontwikkeltraject van een chatbot, kan een organisatie een krachtige tool ontwikkelen die niet alleen effectief communiceert met klanten, maar dit ook doet op een veilige, betrouwbare en ethisch verantwoorde manier. Het succes van dit traject hangt af van een multidisciplinaire aanpak waarbij ontwikkelaars, datawetenschappers, cybersecurity experts, en juridisch adviseurs samenwerken om de uitdagingen en mogelijkheden van elke pijler te adresseren.

De driedeling effectief implementeren

In het hart van de moderne digitale transformatie ligt een fundamentele verschuiving in de manier waarop we denken over en werken met technologie. Deze verschuiving, gedreven door de opkomst van GenAI en de immer aanwezige noodzaak voor robuuste cybersecurity, vraagt om een herziening van traditionele ICT-strategieën. Het is een uitnodiging aan managers en bestuurders om niet alleen technologische vernieuwers maar ook culturele architecten binnen hun organisatie te zijn. Het implementeren van de driedeling in pijlers vormt de kern van deze transformatie en stelt medewerkers in staat om op een andere manier te kijken naar de kansen en risico's van GenAI. Hoe kunnen leiders deze transitie zo effectief mogelijk begeleiden en een cultuur creëren die deze nieuwe benadering omarmt?
Het begint bij de erkenning dat de integratie van generatieve AI en cybersecurity verder gaat dan technologie alleen: het raakt aan de wijze waarop teams samenwerken, hoe projecten worden geleid en hoe succes wordt gemeten. Leiders moeten allereerst de unieke waarden en vereisten van elke pijler begrijpen en vaststellen hoe deze bijdraagt aan het grotere geheel.

Bij de codepijler is het van belang dat ontwikkelaars niet alleen schrijven wat functioneel, maar ook wat veilig en veerkrachtig is. Dit vereist een verschuiving naar secure-by-design principes, waarbij veiligheid vanaf het begin onderdeel is van de ontwikkelingscyclus. De datapijler vereist een cultuur die de waarde van data erkent en respecteert. Dit betekent dat teams werkwijzen en vaardigheden ontwikkelen op het vlak van het ethisch verzamelen, verwerken en gebruiken van data, met een duidelijk begrip van privacy- en beveiligingsrisico's. De Machine Learning-pijler introduceert complexiteit rondom het trainen en implementeren van modellen die zowel effectief als veilig zijn. Dit vereist nauwe samenwerking tussen datawetenschappers, cybersecurity experts en de rest van de IT-organisatie, om te zorgen dat modellen niet alleen nauwkeurig zijn, maar ook bestand tegen manipulatie en misbruik.

Een effectieve implementatie van deze driedeling vraagt om een cultuuromslag binnen de organisatie. Dit betekent dat leiderschap niet alleen moet komen vanuit het management, maar ook vanuit een breed gedragen visie en enthousiasme voor de nieuwe benadering van technologie en beveiliging volgens de voorgestelde driedeling.

Een dergelijke cultuuromslag begint met onderwijs en bewustwording. Bestuurders en managers moeten zorgen voor regelmatige trainingen en workshops die niet alleen de technische aspecten van AI en Cybersecurity behandelen, maar ook de ethische en strategische implicaties ervan. Het is essentieel dat alle medewerkers, ongeacht hun functie, begrijpen hoe hun werk aansluit op de bredere digitale strategie van de organisatie.
Organisaties moeten daarnaast investeren in hun capaciteiten, zowel in eigen specialistisch ML-personeel als het aan- of uitbesteden van ML-activiteiten en -componenten. Een geïntegreerde personeels- en leveranciersstrategie, die past in een visie gebaseerd op de drie pijlers, zal helpen om de juiste nieuwe collega's te vinden en ML-leveranciers te selecteren.
Als laatste is het belangrijk om een omgeving te creëren waarin experimenten, falen en leren worden aangemoedigd. Innovatie komt niet voort uit het strikt volgen van de regels, maar uit het durven verkennen van nieuwe ideeën en het accepteren dat niet elke poging succesvol zal zijn. Dit vereist een leiderschapsstijl die autonomie ondersteunt, initiatief aanmoedigt en waardeert, en de nadruk legt op continu verbeteren. Door het stimuleren van deze vrijheden, binnen de context van de drie pijlers en hun vereisten, is het mogelijk om de enorme snelheid van het AI-domein bij te houden en daar, op een cyberveilige manier, de vruchten van te plukken. Door het stimuleren van deze vrijheden, binnen de context van de drie pijlers en hun vereisten, is het mogelijk om de enorme snelheid van het AI-domein bij te houden en daar, op een cyberveilige manier, de vruchten van te plukken.

Dimitri van Zantvliet is Directeur Cybersecurity bij de Nederlandse Spoorwegen

COLUMN DIMITRI

Vaarwel

Het is inmiddels iets meer dan twee jaar geleden dat ik startte als columnist van dit mooie vakblad. Het was tijdens de staart van Log4J waar we behoorlijk druk mee waren met z'n allen. In de tussentijd passeerden enkele stevige zero-days (tijdens de vakanties), ontstonden verschrikkelijke hybride oorlogen in de Oekraïne en het Midden-Oosten en zagen we natuurlijk de opkomst van generatieve AI als aanvalstactiek.

AI en door AI ondersteunde automatisering zal een essentieel instrument worden voor cyberexperts om organisaties weerbaar te maken. Ik hoor sommige cybercollega's beweren dat het vertrouwen in AI een gok is, een sprong in het diepe waarvan de gevolgen onvoorspelbaar zijn. Maar laten we eerlijk zijn: de echte gok is het vasthouden aan achterhaalde methoden die keer op keer hebben bewezen inadequaat te zijn tegen de dynamische en onophoudelijke golf van cyberdreigingen. De weerstand tegen AI is dus niet alleen achterhaald, het is ronduit risicovol. Het is tijd om de gok te wagen, niet uit roekeloosheid, maar uit een overtuiging dat stagnatie in de digitale arena gelijk staat aan de volgende nederlaag.

Welkom dus in 2024, een jaar waarin het cyberlandschap, aangejaagd door AI niet alleen verder evolueert; het vereist daarboven ook een revolutie in hoe we denken en handelen. Die boodschap voor jullie is ondubbelzinnig: het is tijd om het persoonlijk op te vatten! Bruh? Welnu, dit jaar ben je niet alleen een bewaker van systemen, infrastructuur of algoritmen, opleider van management en gebruikers of toezichthouder op de door jou opgestelde voorschriften, je bent een hoeder van ethiek en integriteit. De rol van een cyberexpert gaat meer dan ooit over het beschermen van normen en waarden. Je staat dus op een kruispunt waar naleving niet alleen meer gaat over het volgen van regels, maar zeker ook over het uitdagen ervan wanneer ze in conflict zijn met ethische principes.

Je moet bereid zijn om te zeggen: "Je hebt besloten deze risico's niet te mitigeren, dus ik zal het senior management nu moeten informeren over de blootstelling van onze organisatie aan deze risico's," of zelfs: "Nee, ik weiger met alle respect te voldoen aan deze opdracht omdat het ethisch niet langer te verdedigen is. We zullen onze koers dus aanpassen, of je zult iemand anders moeten vinden." Deze gedurfde houding is geen insubordinatie; het is een demonstratie van de integriteit en vastberadenheid die elk van ons vanaf 2024 moet drijven.

Sta rechtop, zelfs als het voelt dat je alleen staat. Weet dat de cybergemeenschap achter je staat en je niet langer geïsoleerd bent. Cyberexpert zijn gaat niet langer over het drie lagen diep in stilte onder een CIO weggestopt zijn. Het gaat over het zijn van de duidelijke stem van de rede, de aanjager van verbetering. Vanaf dit jaar laat je je leiderschap niet alleen definiëren door hoe je (geholpen door AI) beschermt, maar ook door hoe je weigert compromissen te sluiten over wat juist is. Laat dit het jaar zijn waarin je het niet alleen persoonlijk opvat, maar het ook ethisch zuiver houdt.

En met deze woorden sluit ik, voordat ik in herhaling ga vallen, twee mooie jaren als columnist voor PvIB af en geef het stokje door aan weer een nieuwe auteur (wel een echte van vlees en bloed hoop ik :-). Vaarwel, tot we wederkeren in deze bijzondere tijd, waar onze cyberpaden zich andermaal mogen kruisen onder het wakend oog van het lot.

Auteur: Vincent van Dijk, eigenaar van Security Scientist. Hij is bereikbaar via: vincent@securityscientist.net

Wat ik leerde van het bouwen van AI-systemen

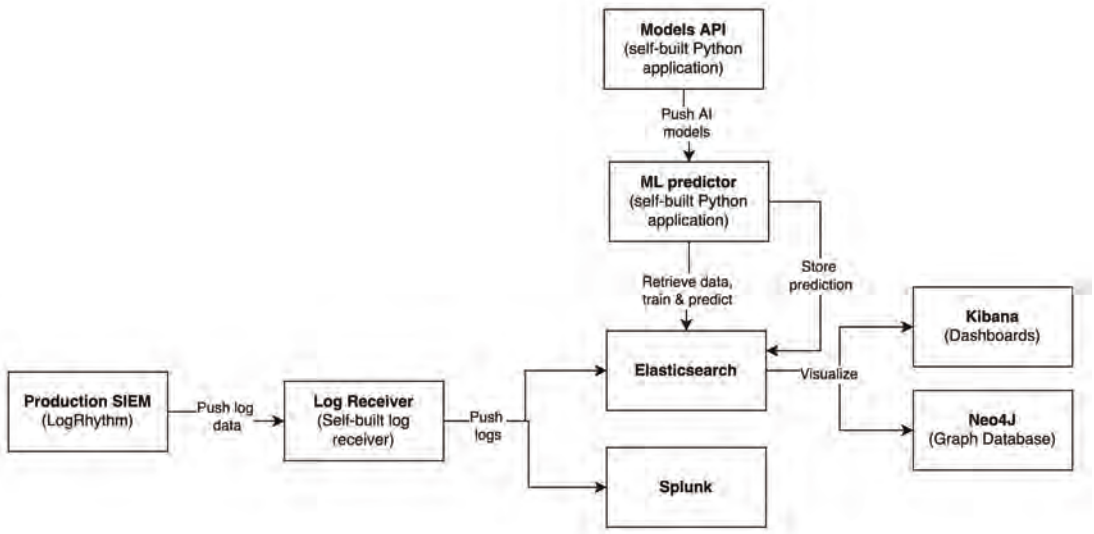
Jan, werkzaam in de verkoop, had problemen met zijn gezin. Hij kon werken, maar zijn gedachten waren elders. Zijn productiviteit daalde en AI classificeerde hem met 'snel productiviteitsverlies'. Kort daarna werd Jan ontslagen. Dit is niet echt gebeurd, met Jan gaat het prima. Het was een risico van één van de AI-modellen die ik in 2019 beschreef.

Voor mijn eerste werkgever, een cybersecurity bedrijf, maakte ik in 2019 een AI-systeem bestemd voor beveiligingsdoeleinden, maar - naar later bleek - met het risico van privacyschending als gevolg van classificering van de prestaties van werknemers. Ik hield het internetgedrag van werknemers bij door middel van een Excel-sheet met 40 rijen en één kolom per maand. Elke cel vertegenwoordigde het internetgedrag van een werknemer voor die maand, inclusief informatie over het meest dominante of afwijkende gedrag en hoe vaak dit zich voordeed.

Dit werd niet gerealiseerd met een nieuw Large Language Model. Het was pas 2019 en mijn oplossing werd al uitgedacht en ontwikkeld door Hugo Steinhaus in 1956. Met de juiste data en een paar creatieve ideeën, maak je zowel de slechtste en beste AI's met eenvoudige modellen. Zelfs kostenefficiënter en sneller (met een training in luttele seconden) dan elk groot taalmodel.

Onethische modellen

Ons idee over AI is verkeerd. Wij denken dat generatieve AI (GenAI) de veroorzaker van schade zal zijn, maar dat is niet



Figuur 1: De infrastructuur van het op maat gemaakte SIEM Machine Learningsysteem.

waar! Oudere modellen kunnen net zo slim zijn als de GenAI. Onder het mom van veiligheidsdenken creëren wij onethische modellen ogenschijnlijk met volkomen legitieme redenen.
Ook Amazon, met 1,5 miljoen werknemers wereldwijd (8,5% afgezet tegenover de Nederlandse bevolking), bouwde een dergelijk AI-systeem om beslissingen van werknemers te automatiseren. Jan kent een levensechte tegenhanger bij Amazon: Stephen. Hij werd door AI ontslagen.
Wij allen schatten in dat Algemene Kunstmatige Intelligentie een reëel gevaar betekent, misschien in de toekomst. Nu weerhoudt deze angst ons om te leren waar AI echt over gaat. En dat is dataverwerking.

Mijn ontdekking in 2019

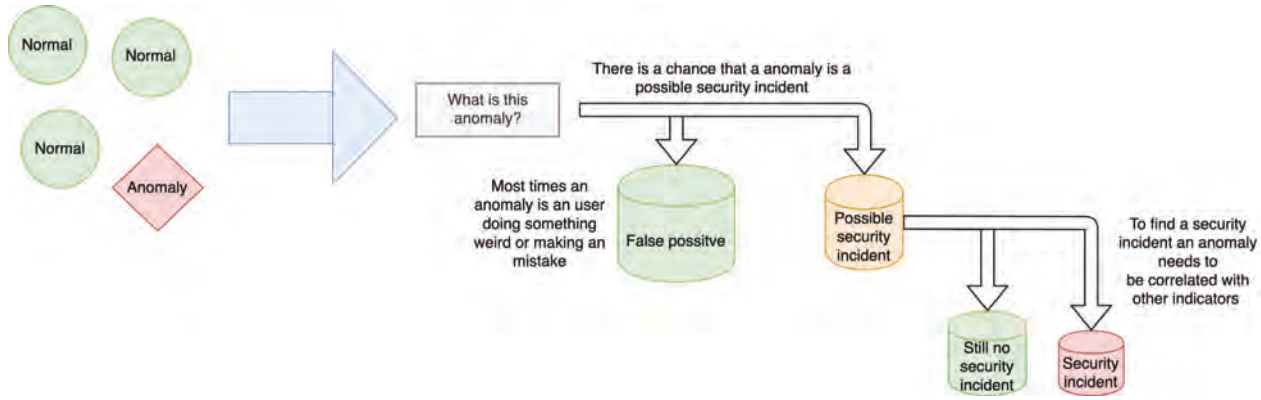
Ik bouwde elk Machine Learning-model dat ik wilde voor een Nederlands beveiligingsbedrijf. Ik had mijn eigen 'schaduw-SIEM' - van hun LogRhythm SIEM - in Elasticsearch. Met mijn schaduw was ik 'in control': ik verrijkte mijn logs, implementeerde geautomatiseerd modellen en realiseerde geautomatiseerde hertraining. Alles voor een complete AI-gerichte SIEM. De modellen konden het gedrag van gebruikers en systemen classificeren en vonden problemen die een compleet SOC-team vanuit Spanje niet kon detecteren: sneller, indringender en met betere resultaten dan Machine Learning-pakketten zoals: Splunk en LogRhythm. Dat project is helaas nooit afgerond.

Complexe modellen, slechte prestaties

Eind 2018 bezocht ik een conferentie over mijn bachelor school-project in Zwolle. Ik was vroeg en dat was mijn geluk want er was een sneeuwstorm die latere treinen blokkeerde. Op de conferentie was ik één van de vijf aanwezige scholieren; de anderen waren thuis gebleven. Ik ontmoette een CEO en als een van de weinige daar sprak hij mij aan. Ik had net voor mijn studie bij een technologisch gedreven verkeersinfrastructuurbedrijf het Machine Learning-project afgerond en zonder te weten wat hij wilde horen, bleef ik er maar over praten. Ik werd daarop aangenomen. Mijn volgende Machine Learning-project zou bij zijn beveiligingsbedrijf zijn. Mijn functie daar: datawetenschapper. AI was 'hot'. Voor hun Machine Learning-project kreeg ik mijn eigen servers. Ik bouwde mijn schaduw-SIEM, iedereen was geïnteresseerd. Nu begon de pret.

Het werd een op maat geschreven applicatie om logs te ontvangen (tot 20.000 per seconde!), een Machine Learning API, voor AI-modellen, opslag en publicatie. Een applicatie die deze modellen automatisch gebruikte en trainde op basis van de gegevens die waren opgeslagen. Ik had zelfs een Splunk-instance testlicentie om mijn tools te vergelijken. Alles bij elkaar zeven systemen, daarvan drie volledig op maat gemaakt en alle integraties daartussen eveneens.

Ik startte mijn onderzoek naar Machine Learning-modellen om beveiligingsproblemen te vinden. Ik deed veel modellenon-

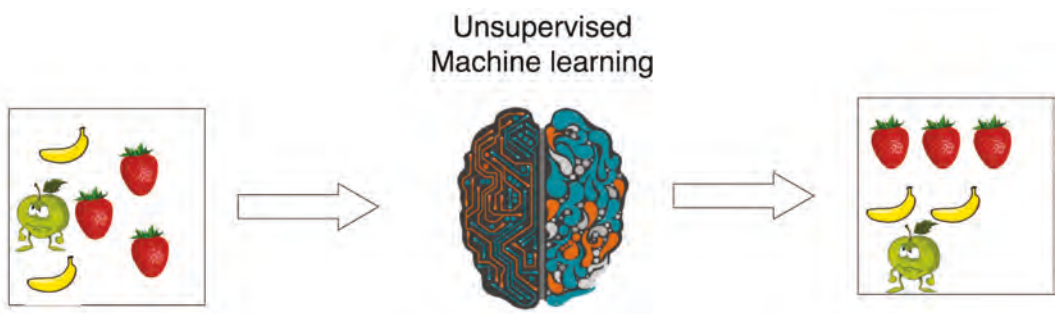


Figuur 2: Anomaliedetectie in cyberbeveiliging.

derzoek: van vroege Large Language Models en Deep Learning tot modellen zonder supervisie. Tot mijn verbazing ondervond ik dat hoe complexer het model, hoe slechter de prestatie. Deze modellen konden geen interessante conclusies trekken uit de gegevens. De reden: op welke locatie moesten deze modellen zoeken? Elk beveiligingsincident is uniek, beveiligingsincidenten komen niet zo vaak voor. Het Verizon Data Breach Investigation Report bevat slechts 1.000+ incidenten per jaar. Dat zijn maar een paar megabytes aan gegevens tegenover de dagelijkse petabytes-overdracht, dat is een speld in een hooiberg zoeken.

Ik wijzigde mijn aanpak liever dan beveiligingsincidenten te voorspellen. Ik zocht anomalieën met eenvoudig uit te leggen modellen. Anomaliedetectie kent echter ook problemen: is het een beveiligingsincident of een systeem dat, of een persoon die, zich afwijkend gedraagt? Elke dag handelt iemand ongewoon. Er logt iemand in op een applicatie wat hij niet eerder had gedaan, omdat hij als chef besloot dat dit zijn verantwoordelijkheid werd. Of een systeembeheerder zet terabytes aan data over wegens een nieuwe architectuur, zie figuur 2 voor de werkelijkheid anno 2019.

Ik wist dat niet elke anomalie een waarschuwing moest triggeren, omdat de meeste niet gerelateerd zijn aan beveiligingsvraagstukken. Tijd voor een compleet nieuwe aanpak. Ik verdiepte mij in clustermodellen. Clusteren, een techniek om te groeperen in datasets. Het algoritme voor machinaal leren probeert de dataset te begrijpen en vormt clusters van vergelijkbare datagroepen.



Figuur 3: Model voor toezichtloos machinaal leren en clusteren.

K-means clustering

Ik kwam uit bij K-Means clustering (1). Het verdeelt de gegevens in K-clusters (vectorkwantisatie - signaalanalyse), waarbij elk gegevenspunt behoort tot het cluster met het dichtst nabijgelegen gemiddelde. Een eenvoudig voorbeeld:

Voedsel	Gezondheidsscore
Appel	88
Banaan	90
Big Mac	10
Koekje	20

Appel en Banaan bezitten vergelijkbare scores, zij vormen een groep. Big Mac en Koekje hebben ook een vergelijkbare maar lagere score, zij zijn de tweede groep. K-Means maakt twee groepen. Dit is een voorbeeld waarbij een mens de groepen gemakkelijk overziet, maar in het echt zijn er honderden kolommen en soms miljoenen rijen. Dat is waar het K-Means-model uitblinkt.

K-Means is efficiënter en extreem snel geworden. De eenvoudigste versies van Deep Learning-modellen kunnen gemakkelijk uren kosten om te trainen. Dit K-Means-model traint in slechts enkele seconden. Dat maakt toegankelijke hertraining mogelijk, zodat het model zich gemakkelijk aanpast aan het dagelijkse, soms vreemde gedrag van mensen en technologie.

Met K-Means bij de hand vond ik een databron voor testen: de Palo Alto firewalls. Die informatie was geweldig omdat, met de juiste configuratie, je ziet welke applicaties gebruikers benutten, hoe vaak en hoeveel gegevens er werden overgedragen. Ik begon met interne tests, waarbij ik al het netwerkverkeer van de medewerkers had verzameld en gefilterd, zie voorbeeld onderaan pagina 19 Vereenvoudigde fictieve gegevens van de Palo Alto firewalls.

Stel een dataset voor van 10.000 records, 40 Azure AD-gebruikers en honderden applicaties. Alleen daaruit al trek je opmerkelijke conclusies. Je achterhaalt wie cloudapplicaties of schaduw-IT gebruikt, die ze niet mogen gebruiken. Toch zegt dit niets over de beveiliging. Ze gebruiken de cloudapplicatie of schaduw-IT wellicht om persoonlijke redenen. Misschien is het normaal gedrag? Dat riep de vraag op: 'Wat is echt abnormaal/ongewoon gedrag?'

Met K-Means zoeken wij beter uit wat normaal gedrag is en wat niet. Zoals eerder groepeer je data, maar in plaats van een appel en een Big Mac, groeperen wij specifieke gebruikers en hun applicatiegedrag.

Het resultaat:

Azure AD-gebruikersnaam	Toepassing
Jan	Groep 1
Peter	Groep 1
David	Groep 2
Kees	Groep 2

K-Means groepeerde op basis van toepassingengebruik. Jan en Peter delen een groep, zo ook David en Kees. Geweldig, maar je trekt hieruit niet echt een conclusie. Zijn David of Kees in groep 2 abnormaal? Of zijn beide groepen normaal? En waarom zijn ze bij elkaar ingedeeld? Je krijgt geen antwoord op

Voorbeeld:

Azure AD-gebruikersnaam	Toepassing	Gebruikte tijden	Datum
Jan	Google Drive	10	10-01-2019
Peter	OneDrive	20	10-01-2019
David	YouTube	12	10-01-2019
Kees	Amazon Web Services	19	10-01-2019

Vereenvoudigde, fictieve gegevens van de Palo Alto firewalls.

deze vragen, tenzij je verder onderzoekt. K-Means is een vrij eenvoudig instrument en interpreteert de resultaten niet voor je.

Naamgeving groepen

Deze vragen hielden mij wakker. Met Machine Learning krijg je spannende resultaten, maar de moeilijkheid is: wat doe je er mee? In cyberbeveiliging onderzoek je de onderliggende redenering om te zien of de gebeurtenis kwaadaardig is of dat het gaat om een willekeurige, legitieme afwijking.
Mijn oplossing: geef de groepen een naam als alternatief voor een nummering. Een naam op basis van de meest dominante applicatie. Groep 1 kan bestaan uit Google Drive-, OneDrive- en iCloud-gebruikers. Werd Google Drive het meest gebruikt, dan heet de groep 'Google Drive'. Als er meerdere groepen zijn met meerdere dominante applicaties, scheidde ik de groepen op basis van het totaal aantal activaties, bijvoorbeeld Google Drive-Hoog en Google Drive-Laag. Tot slot werd elke toepassing die niet binnen de gebruikte kaders viel, als 'Abnormale toepassing' gekenmerkt en kreeg hogere prioriteit.

Het model opnieuw uitgevoerd:

Azure AD-gebruikersnaam	Toepassing
Jan	Google Drive - Hoog
Peter	Google Drive - Hoog
David	YouTube - Hoog
Kees	Abnormale toepassing - Laag

Nu werden de resultaten resultaten interessant, zie tabel onderaan pagina 20. Wij zagen een trend in hoe mensen zich gedroegen en op basis daarvan werden geclassificeerd. Met verder onderzoek verbonden wij zelfs conclusies aan de verschillende indelingen:

- Het toepassen van Google Documenten was de primaire manier van werken. Wij ontdekten bij 'Google Drive – Hoog' meestal productieve werknemers. Bij 'Google Drive – Laag', betekende dit vaak dat een medewerker ziek was of op vakantie ging;
- Bij een groot IT-project werd de systeembeheerder vaak ingedeeld bij 'Abnormale toepassing';
- In geen geval mag iemand buiten de IT-afdeling in de groep 'Abnormale toepassing' vallen. Dit was waarschijnlijk een indicatie van niet-goedgekeurde schaduw-IT taken.

Enkele conclusies benadrukt, zie onderstaande tabel, cursief geschreven:

- Jan was niet op vakantie in februari, hij was onproductief. Waarschijnlijk door zijn verhuizing;
- Jan mag nooit worden ingedeeld bij 'Abnormale toepassingen', hij is geen IT-beheerder. Bij nader onderzoek bleek dat zijn computer spyware bevatte;
- Wij verwachtten geen hoge abnormale classificatie voor Kees in maart. Er waren geen grote projecten gepland. Het bleek dat Kees vanaf bedrijfssystemen toegang had tot zijn thuisserver via Amazon Web Services;
- David was niet productief. Zonder uitzondering stelden we vast dat alle gebruikers die met 'YouTube-Hoog' werden geclassificeerd, ondermaats presteerden.

Dit model was zeer sterk in het monitoren van medewerkers. Het was niet perfect omdat details ontbraken, maar gedrag kon op hoog niveau waargenomen en afgezet worden tegenover

Indien verrijkt met de code over de afgelopen 3 maanden:

Azure AD-gebruikersnaam	Januari	Februari	Maart
Jan	Google Drive - Hoog	*Google Drive - Laag*	*Abnormale toepassing - Laag*
Peter	Google Drive - Hoog	YouTube - Hoog	Google Drive - Laag
David	*YouTube - Hoog*	*YouTube - Hoog*	*YouTube - Hoog*
Kees	Abnormale toepassing - Laag	Google Drive - Hoog	*Abnormale toepassing - hoog*

Zelfs AI, zoals ChatGPT, heeft moeite met het logisch vinden van bedreigende actoren

verwachtingen. Door onderzoek konden we valideren of alle classificaties van gebruikers al dan niet overeenkwamen met verwachtingen en bepalen of daar een goede reden voor afwijking bestond.

Dit eenvoudige op K-Means gebaseerde model was... **griezelig**.

Conclusies

Zorgen rondom het model:

1. Hoe zit het met de privacy? Beveiliging verbeterde, maar ten koste van privacy;
2. Zonder begrip voor het model onderzochten onze junior analisten weinig succesvol de waarschuwingen. Dat vereiste een andere manier van denken;
3. Zonder organisatiebegrip waren deze modellen te complex. Doorgronden en inzicht creëren in de nuances van zo'n Machine Learning-project is nodig.

Gevolgen: de klassieke manier van beveiligingsmonitoring is voor klanten al complex, laat staan deze nieuwe modellen. Zonder training van veiligheidsanalisten zijn 'use cases' betekenisloos. Professionals met ervaring in Machine Learning en cyberbeveiliging zijn vereist, de organisatie moet zich aanpassen. De overstap naar nauwkeurige detecties op basis van AI was te complex. Er waren betere opties om de kwaliteit van onze beveiligingsdienstverlening te verhogen.

Essentiële vraag: 'Welk probleem proberen we op te lossen met AI? ' Proberen wij cyberbedreigingen efficiënter aan te pakken? Dan lijkt AI niet de juiste oplossing. Wij maken grotere sprongen door betere en gerichtere strategieën te definiëren.
Of proberen wij ons vermogen om cyberbedreigingen te detecteren te verbeteren? AI is dan ook niet de juiste keuze, omdat wij niet over grootschalige waarnemingen beschikken om gevaarlijke actoren te detecteren. De grootste dataset van incidenten bevat slechts een paar duizend per jaar, is te weinig gedetailleerd en onvoldoende voor een AI-model training. Wij moeten vertrouwen op anomaliedetectiemethoden zoals het K-Means, dat wel training en inspanning vereist.

Zelfs AI, zoals ChatGPT, heeft moeite met het logisch vinden van bedreigende actoren, omdat het vooral is getraind op internetgegevens en geen echte data heeft over bedreigende actoren.

Het ontwikkelen van AI-modellen leert ons veel. AI is niet het antwoord als je niet weet welke vraag je stelt. De meeste aangeboden AI-oplossingen schieten tekort; leveranciers maken slimme oplossingen, maar niet AI zorgt ervoor dat het werkt. AI in combinatie met bovengenoemde modellen is een leermodel, mits je tijd neemt voor het vinden en stellen van de juiste vragen, net als bij de griezelige K-Means AI, die ik maakte en onbedoeld mensen 'bespioneerde'.

Referentie

(1) https://en.wikipedia.org/wiki/K-means_clustering

Auteur: Arnoud Engelfriet is informaticus en IT-jurist. Hij verdiept zich graag in complexe uitdagingen op het snijvlak van ICT en recht. Arnoud staat aan het hoofd van de Academy waar hij diverse IT-gerelateerde cursussen voor juridische en zakelijke professionals heeft ontwikkeld. Arnoud nam het initiatief tot het creëren van de gecertificeerde CAICO®-cursus voor AI Compliance Officers. Hij blogt sinds 2007 elke werkdag over IT-recht en technologie en is bereikbaar via a.engelfriet@ictrecht.nl.

Risicobeheersing: de AI Act en de AVG

De AI Act doet iets aparts: deze reguleert AI niet in het algemeen, maar focust op de beheersing van risico's van deze innovatieve technologie. De wet kent daartoe diverse instrumenten, waarvan de conformiteitsbeoordeling (Conformity Assessment) de belangrijkste is. Deze lijkt op de Data Protection Impact Assessment (DPIA) die we al kennen uit de AVG, maar er zijn belangrijke verschillen. Hoe werkt de AI Act?

De AI Act kent drie risiconiveaus: verboden, hoogrisico en laagrisico. De meeste eisen gaan gelden voor AI-systemen die een hoog risico vormen voor de gezondheid, veiligheid, grondrechten of het milieu. Zo moet duidelijk zijn waar de data vandaan komt waarmee de AI is getraind, is menselijk toezicht vereist en moet de technische documentatie op orde zijn. Het afhandelen van verzekeringsclaims, bepaalde medische hulpmiddelen en algoritmes die sollicitanten beoordelen zijn voorbeelden van hoog risico-AI.

Bepalen of een AI hoogrisico of verboden is, is een kwestie van inschatten of de toepassing binnen een bepaalde lijst valt. Het is dus geen open norm waarbij je voors en tegens tegen elkaar afweegt, zoals bij de vraag of je een DPIA moet uitvoeren onder de AVG. Het omgekeerde is wel waar: als een AI hoog risico is en persoonsgegevens verwerkt, dan is een DPIA verplicht.

Doel van conformiteitsbeoordelingen

Het doel van conformiteitsbeoordelingen (conformity assessments of CA's) is het toetsen aan normen en beheersen van risico's. In de AI Act is het doel specifiek het toetsen aan specifieke wettelijke vereisten voordat deze op de Europese markt worden gebracht. De verantwoordelijkheid voor het uitvoeren van deze beoordelingen ligt primair bij de aanbieders van deze systemen, maar kan ook betrekking hebben op fabrikanten, distributeurs of importeurs. Dit proces benadrukt de noodzaak van transparantie en accountability in de ontwikkeling en implementatie van AI-systemen.
De opstellers van de AI Act zetten hierbij zwaar in op Europese normen waarmee het assessment kan worden uitgevoerd. Voldoet men aan deze normen, dan wordt de conformiteit verondersteld te bestaan. Zijn er geen relevante normen of standaarden te vinden, dan gelden de normen uit Annex VII van de AI Act.

Verantwoordelijke partij

Een CA moet worden uitgevoerd voordat een AI-systeem op de EU-markt beschikbaar komt. Dat kan simpelweg zijn doordat producten met de AI erin worden verkocht, maar ook door het aanbieden van een online dienst of app met daarin verweven de AI. Ook wanneer een bedrijf voor eigen toepassing een AI ontwikkelt, moet de CA worden uitgevoerd voordat deze in gebruik wordt genomen.
Een CA is niet een eenmalige exercitie. Als het AI-systeem aanzienlijk wordt gewijzigd, moet de CA opnieuw worden uitgevoerd. Het enkele feit dat het AI-systeem bijleert en daarmee nieuw gedrag vertoont, is echter niet genoeg om een nieuwe CA te hoeven doen. Wel moet in de CA natuurlijk zijn opgenomen dat het systeem kan bijleren.

De CA wordt primair uitgevoerd door de provider van het AI-systeem. Als deze dat nalaat, dan kunnen de importeur of deployer van het systeem deze taak op zich nemen om er zo voor te zorgen dat het AI-systeem (of product met de AI erin) alsnog de Europese markt op kan. Een partij is de provider als deze de AI op de markt brengt met een eigen merknaam. Dit hoeft dus niet de feitelijke ontwikkelaar te zijn.

Wijze van uitvoering

Er zijn twee manieren waarop een CA kan worden uitgevoerd: intern of door een derde partij. In het interne CA-proces is het de provider zelf (of de distributeur/importeur/deployer) die de CA uitvoert. De CA door een derde partij wordt uitgevoerd door een externe zogeheten 'notified body'. Deze 'aangemelde instanties' zijn conformiteitsbeoordelingsinstanties die aan specifieke vereisten voldoen en zijn aangewezen door de nationale notifying authorities.

Uitgangspunt is dat een provider werkt met een interne CA. De provider zal immers beter uitgerust zijn en de nodige expertise hebben om de naleving van AI-systemen te beoordelen. Alleen bij de inzet van realtime en post remote biometrische identificatie van personen is dit niet mogelijk. En wanneer het product op de wettenlijst van Bijlage II staat (dus hoog risico vanwege veiligheidscomponent) dan moet uit de toepasselijke wet volgen welke keuze van intern of extern wordt gemaakt.

Bij een interne CA moet de provider:

1. Verifiëren dat het vastgestelde kwaliteitsmanagementsysteem in overeenstemming is met de vereisten;
2. De informatie in de technische documentatie onderzoeken om te beoordelen of aan de vereisten is voldaan;
3. Verifiëren dat het ontwerp- en ontwikkelingsproces van het AI-systeem en de post-markt monitoring consistent is met de technische documentatie.

Na het uitvoeren van een interne CA stelt de verantwoordelijke entiteit een schriftelijke EU-conformiteitsverklaring op voor het AI-systeem. Bijlage V van de AI Act somt de informatie op die moet worden opgenomen in de EU-conformiteitsverklaring. Deze verklaring moet actueel worden gehouden tot tien jaar

De AVG kent geen sjabloon of voorbeeld van hoe een DPIA eruit moet zien

nadat het systeem op de markt is gebracht of in gebruik is genomen. De provider moet ook een zichtbare, leesbare en onuitwisbare CE-markering van conformiteit aanbrengen. En als laatste moet de provider een EU-verklaringsformulier opstellen met daarin onder meer een beschrijving van de uitgevoerde procedure. In het geval van een CA door een derde partij beoordeelt de notified body het kwaliteitsmanagementsysteem en de technische documentatie. De provider doet hiertoe een verzoek en moet de benodigde informatie aanleveren. Bijlage VII somt de informatie op die moet worden opgenomen in de aanvraag aan de aangemelde instantie. Zowel het kwaliteitsmanagementsysteem als de technische documentatie moeten in de aanvraag staan.
Als de aangemelde instantie vaststelt dat het hoog risico-AI-systeem in overeenstemming is met de vereisten, zal het een EU-certificaat van technische documentatiebeoordeling afgeven dat een beperkte geldigheid heeft. Net zoals bij de interne CA, moet de provider onder het CA-proces door een derde partij de EU-conformiteitsverklaring opstellen en de CE-markering van conformiteit aanbrengen. Om het proces af te ronden, moet de aanbieder een EU-verklaringsformulier opstellen met daarin onder meer een beschrijving van de uitgevoerde conformiteitsbeoordelingsprocedure.

In het geval de aangemelde instantie beoordeelt dat het hoog risico-AI-systeem niet in overeenstemming is met de vereisten voor hoog risico-AI-systemen, moet dit gedetailleerd worden gecommuniceerd en uitgelegd aan de aanbieder of andere verantwoordelijke entiteit. Artikel 45 geeft de provider het recht om beroep aan te tekenen tegen de beslissing van de aangemelde instantie. Als het oordeel overeind blijft, kan de provider worden bevolen het systeem aan te passen, terug te trekken of van de markt te halen.

Let op: de CA is geen eenmalige oefening. Providers moeten een post-markt monitoringssysteem opzetten en documenteren, dat tot doel heeft de voortdurende naleving van AI-systemen met de AIA-vereisten voor hoog risico-AI-systemen te evalueren. Het post-markt monitoringsplan kan deel uitmaken van de technische documentatie of het productplan. Daarnaast moet in het geval van een externe CA partij de notified body periodieke audits uitvoeren om ervoor te zorgen dat de provider het kwaliteitsmanagementsysteem handhaaft en toepast.

Data Protection Impact Assessments

Het instrument van de Data Protection Impact Assessment of DPIA (in het Nederlands: gegevensbeschermingseffectbeoordeling) is in de AVG opgenomen om bepaalde risico's en gevolgen van verwerken van persoonsgegevens te kunnen beheersen. Vanwege deze insteek lijken er overeenkomsten te zijn met een CA. Toch zijn er belangrijke verschillen.

Doel van de DPIA

De DPIA is een wettelijke verplichting onder de AVG die vereist dat de entiteit verantwoordelijk voor een verwerking van persoonsgegevens (de 'controller') een beoordeling uitvoert van de impact van de beoogde verwerking op de

bescherming van persoonsgegevens, in het bijzonder wanneer de betreffende verwerking waarschijnlijk een hoog risico vormt voor de rechten en vrijheden van individuen, voordat de verwerking plaatsvindt.

Doel van de DPIA is het beoordelen van noodzaak en evenredigheid van de verwerkingen en de risico's die daarmee samengaan, en het formuleren van maatregelen om deze te mitigeren. Deze dient voorafgaand aan het starten van de verwerking te worden uitgevoerd (net zoals de CA), en kan leiden tot de plicht om toestemming te vragen aan de toezichthouder voordat de verwerking wordt ingezet. Is die plicht er niet, en meent de controller dat de risico's adequaat bestreden zijn, dan mag men de keuze maken de markt op te gaan.

Dit is iets anders dan bij de CA: daar mag het gebruik van het AI-systeem pas beginnen nadat het systeem conform is gemaakt en het CE-logo aangebracht is. Er is geen optie om bijvoorbeeld een ontheffing te vragen als de conformiteitsbeoordeling niet slaagt.
De AVG kent geen sjabloon of voorbeeld van hoe een DPIA eruit moet zien. Ook zegt de AVG niet welke maatregelen zouden passen bij welke risico's. Er zijn ook geen normen waarmee men compliance aan de AVG kan vaststellen. Dit maakt een DPIA meer open-ended dan een CA.

Verantwoordelijke partij

Onder de AVG is de datacontroller of verwerkingsverantwoordelijke de partij die de doelen en middelen van de verwerking van persoonsgegevens bepaalt. De datacontroller is dan ook verantwoordelijk voor alle compliance aspecten van de AVG, en dus ook voor het beoordelen of een DPIA moet worden uitgevoerd en voor het feitelijk uitvoeren daarvan. Uiteraard mag de controller ook anderen inhuren, of informatie gebruiken die een verwerker hem aanlevert.
In ICT in het algemeen, en bij AI in het bijzonder, heeft die verwerker vaak een zeer sturende en prominente rol. Wie bijvoorbeeld een SaaS-dienst afneemt en daarmee persoonsgegevens verwerkt, is volgens de AVG de 'controller' ook al kan hij feitelijk niets veranderen aan hoe de dienst werkt en is de inspraak vooral theoretisch. Als die dienst AI gebruikt, dan is de aanbieder van de SaaS-dienst de 'provider' daarvan en de afnemer de 'deployer'. Daarmee ligt dus de plicht tot een DPIA bij de afnemer, en de plicht tot een CA bij de leverancier.

Het kan dus zijn dat relevante delen van de CA uitgevoerd door de provider van het AI-systeem (zoals die gerelateerd aan de naleving van de gegevenskwaliteit en cybersecurity vereisten) vervolgens gebruikt worden om de afnemer te informeren, waarna die deze gegevens overneemt in de DPIA.
Net als bij de CA is een DPIA geen eenmalige exercitie. In de Guidelines over DPIA's geeft de EDPB duidelijk aan dat dit een continu proces van bijwerken en bijsturen is.

Wijze van uitvoering

De AVG bepaalt geen formeel proces voor het uitvoeren van een DPIA. De controller is de verantwoordelijke partij. Deze kan ervoor kiezen een derde het werk te laten uitvoeren, en mag zoals gezegd afgaan op informatie van de verwerker, maar blijft eindverantwoordelijk voor de juistheid en actualiteit van de DPIA. Is er een functionaris gegevensbescherming, dan moet deze worden geraadpleegd. Ook moet de controller waar relevant de input van betrokken personen verkrijgen. De CA procedure kent dit laatste aspect niet, hoewel het ook niet verboden is om bij het proces input van personen van buitenaf te verkrijgen.
Zowel de DPIA als de CA focussen op risico's van buitenaf. De scope is wel iets anders: bij een DPIA gaat het om risico's die verband houden met gebruik van persoonsgegevens, terwijl bij een CA het gaat om alle risico's die het AI-systeem kan veroorzaken of verhogen. Een risico op het vernielen van vloerbedekking door een robotstofzuiger zal bijvoorbeeld wél in een CA aan de orde moeten komen, maar in een DPIA niet relevant zijn.

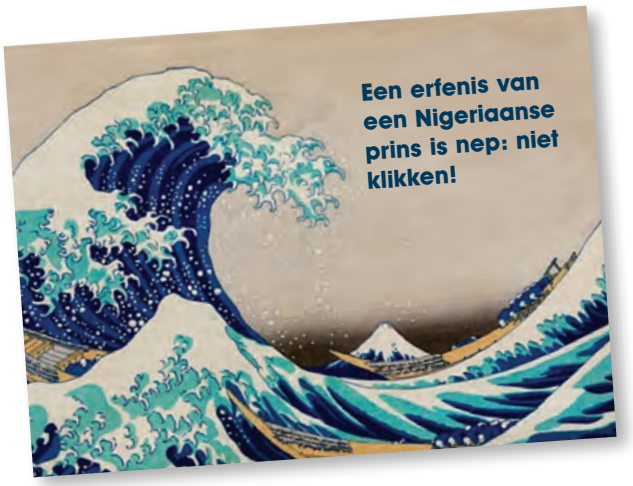
Bij een CA wordt in principe uitgegaan van normen waartegen men het assessment maakt. Voor DPIA's zijn dergelijke normen er niet. Diverse handreikingen en modellen voor DPIA's verwijzen naar ISO-norm 31000, de internationale norm voor risicomanagement. Het is echter geen gegeven dat dit de juiste norm is voor een specifieke organisatie.

De inhoud van een DPIA hoeft niet openbaar gemaakt te worden. Vaak zien organisaties deze ook als bedrijfsgeheim. Ook een CA hoeft niet gepubliceerd te worden. Bij beiden geldt wel dat de toezichthouder inzage mag eisen in deze documenten. Onder artikel 51 moet de provider van een hoog risico AI wel een samenvatting van de DPIA aanleveren voor publicatie in de openbare database van deze AI's.

Samenloop van CA en DPIA

Organisaties die AI op de markt brengen of inzetten, kunnen zowel met een CA als met een DPIA te maken krijgen. Een DPIA focust op beheersen van risico's rondom persoonsgegevens, terwijl een CA een proces biedt voor het mitigeren van risico's rondom AI. Hoewel er dus zeker overeenstemmingen zijn, is het zeker geen gegeven dat het uitvoeren van een DPIA ertoe leidt dat de CA overbodig is – of andersom. Let dus goed op de vereisten van beiden.

Auteur: Drs. Robert Metsemakers RA RE CISSP is als ervaren auditor en informatiebeveiligingsexpert beschikbaar voor securityadviesopdrachten en bereikbaar op robert.metsemakers@gmail.com.



BLOG

Security-awareness met hAIku's

Het kenmerkende van Artificiële Intelligentie is voor mij niet dat het met computers, software en grote taalmodellen werkt, maar dat het intelligentie is van buiten je eigen hoofd. Zo gebruik ik al jaren AI en wel in de vormen skill, scale en scope.

Mijn vader leerde mij om in een vreemde stad naar de juiste weg te gaan vragen bij een tankstation voor benzine. Dit was meer dan vijftig jaar geleden, dus zonder eigen navigatiesysteem in je auto. Er was meestal slechts één pompbediende die je benzinetank voor je vulde en door een lager aantal personenauto's nog tijd had voor een persoonlijk praatje, in elk geval bij het contant afrekenen. Nu kun je het beter vragen aan een pizzabakker die thuis bezorgt. Want naast de skill van het bakken van die van oorsprong Turkse platte koeken, hebben de bezorgers gedetailleerde kennis van straten, pleinen en wegen in de omtrek. En ze zijn in staat om overal snel te komen, dat wil zeggen binnen de – bij de bestelling aan klanten beloofde – bezorgtijd. In jouw eigen stad kun je zelfs de kosten voor een taxi uitsparen door telefonisch een pizza te bestellen en op het moment dat de bezorger de warme pizza ontvangt van de bakker, binnen te stappen en te vragen of je een lift kunt krijgen naar het opgegeven besteladres. Ze rijden er toch heen, dus dat kan meestal wel. Als securityprofessional kun je op dezelfde manier in jouw eigen organisatie bij het bedrijfsrestaurant vragen of ze nog tips hebben om een Denial of Service aanval af te slaan!? Namelijk vanuit hun ervaring met een massatoeloop op de uitgiftebalie op de dagen dat ze sushi of biefstuk met friet serveren voor vijf euro, het normale dinertarief voor personeelsleden. Bij de personeelsvereniging kun je nagaan welke aangeboden uitstapjes (zoals Efteling, skiën, Kerstmarkt, workshop jeu de boules voor gevorderden) veel belangstelling trekken – en daarmee in theorie ook interessante onderwerpen zijn voor phishingaanvallen op diezelfde medewerkers. Je gebruikt zo de skill van iemand anders voor security.

Scale

Een nieuwe medewerker bij de IT-servicedesk kun je inwerken door haar of hem 'alles' te leren over Windows, printers en de softwarepakketten die jouw organisatie standaard gebruikt. Meestal is dat Excel, Word en PowerPoint. En die 'nieuwe' dan te laten wachten op de binnenkomende vragen van medewerkers met IT-problemen. Je kunt hem of haar echter ook een paar jaargangen computerbladen (Nederlandse, maar ook Engelse en Duitse) geven en vragen om de 'lezersvragen' rubrieken goed te bestuderen. In dat soort rubrieken worden namelijk ingestuurde vragen van lezers behandeld. Het zijn vragen waarop überhaupt een antwoord mogelijk is en de redactie zoekt natuurlijk de vragen uit die (heel) vaak gesteld worden en waarvan men verwacht dat andere lezers er ook belang bij (zullen) hebben. Tijdschriften hebben een hogere informatiedichtheid per vierkante centimeter dan boeken. En precies in die vragenrubriek, soms ook 'tips' genoemd, is de (nuttige) dichtheid het hoogst. Je gebruikt zo externe intelligentie vanuit de scale (schaal) van een breed verspreid tijdschrift, met veel meer computers en printers gebruikende lezers dan er mensen werken in jouw organisatie. Security-informatie uitwisselen met andere organisaties in jouw branche is daarom ook een goed idee. Problemen, oplossingen en ervaringen bij de aanpak van problemen kun je prima over en weer delen. En elkaar even bellen bij een ransomwarebesmetting of eerder, al bij de waargenomen aanval.

Scope

De derde bron van externe intelligentie boor je aan door buiten je eigen denkraam te treden. Uit films of tv-programma's kun je namelijk ook dingen leren over security. Inspecteur Columbo is tegenwoordig vooral bekend van zijn zeer effectieve 'nog een laatste vraag' aan de hoofdverdachte aan het eind van zijn gesprekjes. Informatieverzameling lijkt bij deze, een beetje verfomfaaid ogende inspecteur nooit een verhoor, al is het dat natuurlijk wel. Ook zijn vasthoudendheid (je ziet de verdachte denken 'daar heb je hem wéér!'), grote aandacht voor in eerste instantie onbelangrijk lijkende details, bescheidenheid (zijn standpunt is: je kunt niet alles weten, maar je kunt wel alles vrágen) en onverwachte invalshoeken ('mijn vrouw vroeg gisteravond…') zijn wat mij betreft een inspiratie voor securityprofessionals.

Ook een boek over een ogenschijnlijk ander onderwerp (scope) kan je security-inzichten geven. Machiavelli schrijft in zijn beruchte boek 'De Prins' onder andere over het als vorst verstandig (of was het nou geslepen en uitgekiend?) omgaan met huurlingenlegers. Die ervaringen kun je prima vertalen naar het doordacht omgaan met tijdelijk ingehuurde externe krachten. Chinese krijgsheren zoals Sun Tzu schreven over oorlog en veldslagen, maar ook over strategie en zelfs over het vermijden van gewapende strijd. Waarom zou je dat als beveiligingsdeskundige niet kunnen gebruiken? Ik ontdekte in 2023 een interessant boek van vóór Sun Tzu met de titel '36 Chinese Stratagemes'. Deze 36 krijgslisten zijn gegroepeerd in zes hoofdstukken van ieder zes listen:

1. Listen in een gewonnen situatie;
2. Listen om met een vijand om te gaan;
3. Aanvalslisten
4. Listen voor chaotische situaties;
5. Listen voor terreinwinst;
6. Listen in een verloren situatie.

De eerste drie hoofdstukken behandelen listen in voordelige situaties (men is aan de winnende hand of heeft een voordeel behaald in de strijd); de laatste drie hoofdstukken listen in nadelige situaties. Red teams hebben dus meer aan Hoofdstuk 1, 2 en 3 en een SOC aan hoofdstuk 4 (nou vooruit, ook aan 5 en 6). In een later artikel zal ik dit boek uitgebreider behandelen.

De 'cloud' bestaat niet: het is puur de computer van iemand anders

Maak zelf hAIku's met een biertje, wijntje en sushi tijdens een brainstorm

Haiku's

Ook kun je scope letterlijk als een 'denkraam' toepassen in de vorm van een haiku, bijvoorbeeld voor een security-awareness campagne. Deze Japanse dichtvorm bestaat uit drie regels van respectievelijk 5, 7 en 5 lettergrepen. De eerste zin beschrijft vaak een situatie. Deze vormeisen dwingen je op speelse wijze de securityboodschap die je wilt overbrengen, kort en krachtig te formuleren. Het hoeft niet te rijmen, maar je moet wel spelen met taal en zo de securityslogans en uitspraken van directieleden opnieuw formuleren, totdat ze passen in die 5-7-5 vorm. Met de traditionele vormgeving van Japanse prenten eromheen, zijn deze teksten zeer geschikt als aandachttrekkende items op posters in gangen of liften binnen jouw organisatie (op alle locaties!), als screensavers op alle gelukkig met wachtwoord geblokkeerde werkstations of op een afsluitende slide in elke presentatie gegeven door een medewerker van de afdeling Security. Papieren koffiebekers met opdruk liggen moeilijk in Nederland sinds 1 januari 2024. Maar bierviltjes, muismatten en suikerzakjes kunnen nog steeds prima bedrukt worden met een haiku. Vermoedelijk ten overvloede, in de naam is de populaire afkorting AI al opgenomen.

De haiku's in de afbeeldingen in dit artikel kun je zo gebruiken, maar het lijkt me leuker om met het hele team, wijntje, biertje en/of sushi erbij twee uurtjes te brainstormen en zelf teksten te verzinnen. Goed voor het interne begrip, kennisdeling en de samenwerking met collega's in het team.

Kampai! (Japans voor 'droog glas' of 'bottoms up!' of proost).

COLUMN LEX

Vertrouwen in AI

De hoeveelheid werk dat met de hulp van AI uitgevoerd wordt neemt snel toe. Kunnen we wel vertrouwen op die inzet van AI? AI kan eentonig werk met veel informatieverwerking automatiseren. In informatiebeveiliging is er bij uitstek één functie waarop dit van toepassing is: SOC-analisten, zij moeten bijzonderheden zoeken in een stroom van informatie. Je ziet steeds meer hulpmiddelen die het saaie uitpluiswerk doen voor de analisten, zodat zij zich meer kunnen focussen op de anomalieën die gevonden worden.

Klassieke automatisering met SIEM-tools beperkt zich tot het vinden van use-cases, combinaties van logregels uit verschillende bronnen. De nieuwe generatie EDR- en NDR-tools kunnen nu al met AI realtime zoeken naar bijzonderheden en dat snel combineren met aanvullende informatie vanuit andere hoeken, zonder use-cases.

De volgende stappen in automatisering van securitywerk zijn waarschijnlijk het forensisch onderzoek en CERT-activiteiten. Daar is het belangrijk om volgens strakke protocollen te werken. Een AI zal dat vlekkeloos kunnen doen, onderzoekers rest het alleen nog simpele aanwijzingen te geven. Nog even een verificatie van een expert om de kwaliteit te controleren en klaar is Kees.

Ander security werk dat wel eens zou kunnen veranderen door AI is het inrichten en uitvoeren van een ISMS. De securitybeleidstukken hierin lijken toch steeds meer op elkaar. Over een paar jaar geef je ChatGPT 12 de karakteristieken van je bedrijf en rolt de hele set aan beleidstukken eruit. Natuurlijk moet je ook die nog controleren op gehallucineerde passages. Tot zover hebben we het zelf in de hand.
Vertrouwen in een AI kan op verschillende momenten beschaamd worden. Bij het trainen kan het fout gaan, bijvoorbeeld door een denkfout in het algoritme. Het algoritme is het recept waarmee de ingevoerde data-ingrediënten door het model gecombineerd worden om een doel te bereiken.

Het gebeurt ook vaak dat de data waarmee het algoritme getraind wordt geen neutraal overzicht geeft. Historische of empirische data kunnen eenzijdig zijn. Studenten met een donkere huidtint worden minder goed herkend door een surveillance-AI bij universiteitsexamens, omdat er veel minder trainingsdata was over mensen van kleur.

En dan kan input mogelijk ook nog gemanipuleerd worden, wanneer bijvoorbeeld helpdeskchats worden gebruikt om het algoritme te trainen. Door bewust een bepaald soort vragen veel te stellen krijg je in het model een onbalans in die richting. Als je bewust verkeerd antwoordt, train je het model fout te antwoorden.

Bij het gebruiken van het model is manipulatie ook mogelijk. Door zaken te veranderen in de gebruikscontext kan het model verzuimen een beslissing te nemen of zelfs een verkeerde beslissing nemen. Bijvoorbeeld door een sticker op een verkeersbord te plakken wordt deze door de AI niet meer herkend of door rijstrepen op de weg bij te schilderen kan het model die 'fake' markering gaan volgen, mogelijk met ernstige gevolgen. Een model dat getraind is met persoonsgegevens kan wellicht gemanipuleerd worden om die gegevens prijs te geven.

Maar ook gewoon gebruik van de AI kan resultaten produceren die voor criminele doeleinden gebruikt kunnen worden. Een AI kan gebruikt worden om beter gerichte phishing e-mails te schrijven met gemanipuleerde video's of een kiezer via social media te manipuleren om voor een ander te kiezen of niet te gaan stemmen.

AI met al genoeg reden om zeer kritisch te zijn op de resultaten die met AI behaald worden.

SIEM = Security Information and Event Management
EDR = Endpoint Detection and Response
NDR = Network Detection and Response
CERT = Computer Emergency Response Team
ISMS = Information Security Management System

Lex Borger is security consultant bij Tesorion en oud-hoofdredacteur van IB Magazine. Lex is bereikbaar via lex.borger@tesorion.nl

Geïnterviewde: Dasha Simons is Managing Consultant voor Trustworthy AI bij IBM Consulting en adviseert klanten over eerlijkere en transparantere AI-ontwikkeling. Ook is zij onderdeel van Team Europe Direct met de focus op Trustworthy AI bij de Europese Commissie. Zij studeerde als 'best graduate' af aan de Technische Universiteit in Delft (industrieel ontwerpen 2019). Dasha is bereikbaar via: https://www.linkedin.com/in/dasha-simons.

Op zoek naar een eerlijk AI-model

Wat is een eerlijk AI-model? Hoe beperk je risico's op desinformatie? De belangrijkste uitdaging ligt in het bepalen wat in elke specifieke toepassing als eerlijk wordt beschouwd en wat de meest geschikte manier is om de werking van het systeem uit te leggen. IBM introduceerde watsonx. Een AI en dataplatform met een set van AI-assistenten die bedrijven helpen om de impact van AI te schalen en te versnellen met betrouwbare data.

Over dit thema hebben wij met Dasha Simons van IBM gesproken en haar gevraagd naar haar inzichten, ervaringen en lessen. Onderstaand het verslag van dat interview.

Ik kwam bij IBM terecht door mijn afstudeerproject en ontdekte dat IBM een van de weinige, grote techbedrijven is bij wie het hen in het businessmodel niet gaat om de verkoop van data. IBM omarmde al vroeg de ethische principes en was daarmee één van de eerste die daarop inzette. De reden van mijn keuze voor het vak was tweeledig: aan de ene kant, bijna filosofisch nadenken over de vraagstukken waarmee je bezig bent. Jezelf vragen stellen als: wat vinden wij eerlijk? Wat omvat daadwerkelijke transparantie? Wat is voldoende transparantie zonder schending van het privacyrecht? Aan de andere kant betekent dit binnen het vakgebied van Kunstmatige Intelligentie het vertalen naar technische oplossingen die ervoor moeten zorgen dat keuzes op filosofisch, sociaal en politiek gebied hun weerslag vinden in de technische realiteit van alledag. Dat is een interessante en intellectuele uitdaging. Daarnaast draag ik als millennial bij aan een iets betere wereld – al is het een heel klein beetje.

Een kijkje bij IBM en de AI-wereld

Er zijn verschillende vormen van AI en de meest recente is die van generatieve AI (GenAI). De meer traditionele vormen zoals Machine Learning bestaan al langer, maar het grote publiek is pas recentelijk met de lancering van ChatGPT met GenAI in aanraking gekomen. Feitelijk bestaan er al GenAI modellen vanaf circa 2012. GenAI is gebaseerd op onder andere foundation models. Een voorbeeld van een Machine Learning-model is een model dat kredietrisicovoorspellingen kan doen. Ook daarvoor moet je veel data verzamelen en het model trainen. Het model beperkt zich enkel tot die taak en is zeer specifiek.

Het verschil met de Foundation Models is dat je met één model traint. Dat vereist heel veel data en veel rekenkracht. Dat model train je niet met slechts één doel voor ogen. Neem de Large Language Modellen (LLMs), de grote taalmodellen die gebruikmaken van foundational models. LLMs zijn heel goed in het voorspellen van het eerstvolgende woord in een tekst. Zo'n model is goed toepasbaar op veel laag risico taken en persoonlijke vragen uit de praktijk. Ze kunnen een liefdesbrief of gedicht schrijven of feedback geven op jouw CV, zonder of met beperkte training. Maatwerktraining is voor zoiets veel minder nodig. Dit betekent dat je het model één keer traint en op vele taken kunt toepassen. Deze modellen zijn handig voor generieke en laag risico taken. Maar bij toepassing binnen het bedrijfsleven is het wel verstandig ze toe te spitsen op de specifieke taken waarnaar je kijkt, en het extra in acht nemen van de veiligheid, de betrouwbaarheid en andere AI- en ethiekfacetten.

AI-systemen zijn uiteindelijk het verlengde van besluitvorming

Oorzaken van risico's

Het is interessant om te kijken naar de oorzaken van de risico's. Het kan helpen om te kijken naar risico's bij de input, de output en de governance. Als je naar de input kijkt, let je op vragen als: met welke data train je? En: welke risico's zijn eraan gerelateerd? Een evident voorbeeld is duidelijk te zien bij het genereren van afbeeldingen, zoals bij de Bloombergstudies (1), waarin geconstateerd wordt dat het gezicht van de gemiddelde advocaat of CEO relatief blanker en mannelijker is, terwijl het gezicht van de schoonmaakhulp relatief donkerder en vrouwelijker is. Het risico op vooringenomenheid wordt versterkt in de uitkomsten van het model. Bij tekst gebeurt hetzelfde, maar dat kan minder gemakkelijk te herkennen zijn. Dit komt doordat de trainingsdata deze vooringenomenheid bezat.

Als je naar de output kijkt, kun je ook kijken naar hoe deze misbruikt kan worden. Wat vooral te maken heeft met manipulatie en misbruik die leiden tot des(mis)informatie. Stel dat een financieel planner een simulatiemodel gebruikt om de kans te berekenen op het behalen van verschillende beleggingsdoelen in verschillende marktomstandigheden. Het model geeft echter een vertekend beeld van de werkelijkheid, doordat de uitvoerresultaten significant afwijken en positiever zijn dan de werkelijke uitkomsten. Dit laat een risico zien van de uitkomst van een model, dat leidt tot onjuiste conclusies en incorrecte besluitvorming.

Als we kijken naar governance is een voorbeeld van een risico het energieverbruik van generatieve modellen. Om één afbeelding te genereren is evenveel energie nodig als bij het opladen van een mobiele telefoon. Dat lijkt weinig, maar om de afbeelding tevredenstellend te genereren kan het je wel twintig tot dertig pogingen kosten. Het is belangrijk om je hiervan bewust te zijn zodat je daarop kunt mitigeren.

Risico op desinformatie

De risico's van GenAI zijn afhankelijk van het model dat je hebt en waar je het voor gebruikt. Sommige risico's zijn urgenter dan andere. Daarbij spelen onder andere de urgentie van de taak, de risicomanagementpraktijken, transparantie in datamanagement en robuuste modeltrainingprocedures een rol in. Stel dat je een model alleen intern gebruikt, door eigen medewerkers die vragen stellen met betrekking tot HR, dan is desinformatie minder waarschijnlijk. Maar gebruik je een ChatGPT of een andere vraag-consumentenbot voor het algemene publiek, dan kan iedereen een vraag formuleren die uiteindelijk qua output schadelijk zou kunnen zijn. Bijvoorbeeld, als je om een recept had gevraagd voor je avondeten, zou dit kunnen resulteren in het verkrijgen van recepten die uiteindelijk giftige stoffen bevatten. Een gebruiker of iemand die schade wil toebrengen aan het product, kan op tal van manieren proberen te achterhalen hoe hij het systeem kan omzeilen. Voor een intern systeem is dat risico lager, er is minder risico op desinformatie.

Watsonx platform van IBM

IBM presenteert het watsonx platform met: watsonx.ai, watsonx.data en watsonx.governance (2). Hoe kan je de samenhang daarin zien? IBM watsonx is een AI en dataplatform met een set van AI-assistenten die bedrijven helpen om de impact van AI te schalen en te versnellen met betrouwbare data. Je kunt alle drie platformen: watsonx.governance, watsonx.ai or watsonx.data, separaat of samen gebruiken voor een toolkit voor model maatwerk, governance en dataopslag.
Wat watsonx.governance interessant maakt is dat het je helpt te kunnen voldoen aan regels zoals de komende EU wet- en regelgeving, ook kun je met watsonx.governance je AI governance inregelen en tegelijkertijd de manuele taken voor de data scienceteams beperken voor bedrijven.
De aankomende EU AI Act hanteert een risicogebaseerde


NEGF
BANKCALL
GOMARKF
FIXLOC
INTPRET
GETMKS
CDULOGIC
SCAXIS

Er zijn nieuwe vaardigheden en kennis nodig bij zowel management als ontwikkelteams, die juridische, filosofisch-ethische en technische kennis omvatten

aanpak waarbij niet de technologie zelf wordt gereguleerd, maar de toepassing van de technologie. Een voorbeeld daarvan is het aanbieden van essentiële publieke of private diensten, die dan kunnen vallen onder de hoog risico categorie. Je moet dan voldoen aan modeldocumentatie en een post-market monitoringsysteem bezitten. Dit om te monitoren of er sprake kan zijn van vooringenomenheid en je dat moet kunnen uitleggen. Manueel gaat je dat veel tijd kosten om al die model documentatie bij alsook up-to-date te houden.

De watsonx.governance tooling helpt je in de automatisering en het toegankelijk maken van de relevante data voor je compliance publiek alsook voor de ontwikkelaars. Ook helpt de tooling met het automatiseren van de modeldocumentatie en het standaardiseren ervan. De tooling ondersteunt het monitoren van modellen tijdens ontwikkeling en productie op bijvoorbeeld vooringenomenheid, en zorgt dat de verantwoordelijkheden duidelijk en transparant belegd zijn binnen de organisatie.

Voor bedrijven blijft het vraagstuk: 'Wat vinden wij een eerlijk model?' Daarvoor moeten goede keuzes gemaakt worden. Hoe processen moeten worden ingericht, die niet alleen door technologie kunnen worden opgelost. Dat zijn strategische keuzes op C-niveau, die moeten bepalen welke rol ze willen spelen binnen het AI-landschap: voldoen aan minimale wet- en regelgeving of ook nog proactief handelen naar andere waardes?

De nodige uitdagingen

Toen ik vijf jaar geleden begon met dit werk was het onderwerp van betrouwbaardere AI-ontwikkeling meer een onderwerp van gesprek voor bij de vrijdagmiddagborrel. Mensen vinden het een leuk onderwerp om het er over te hebben, maar niet bij een maandagochtend budget alloceringsafspraak. En dan hebben wij het over gesprekken bij klanten. Eindelijk kreeg het onderwerp in 2023 meer prioriteit bij het grote publiek. Want AI-systemen zijn uiteindelijk het verlengde van besluitvorming, want het is het eerste systeem dat dit doet. Daarmee vormen zij het verlengde van elk bedrijf in de digitale wereld en dat met impact op de reële wereld!

De grote uitdaging zit in het definiëren wat per toepassing eerlijker is of wat de juiste manier van uitlegbaarheid voor een bepaalde toepassing blijkt te zijn. Uiteindelijk zijn het politieke vraagstukken. Die zijn lastig omdat bedrijven en technici niet gewoon zijn dat soort vraagstukken op te pakken. Er zijn nieuwe vaardigheden en kennis nodig bij zowel management als ontwikkelteams, die juridische, filosofisch-ethische en technische kennis omvatten. Dat samenspel is nieuw en in beweging. Gelukkig zijn er vele bestaande methoden of ontwikkelingen om deze risico's te mitigeren, al mag daar naar mijn mening meer concrete aandacht naar toe gaan en niet alleen door evenementen te organiseren. Ik verwacht dat (Gen)AI een met het onderwerp duurzaamheid vergelijkbaar traject ingaat. Dus geen 'ethics washing' voor (Gen)AI, maar daadwerkelijk uitvoeren. Doe je mee?

Referenties

(1) https://www.bloomberg.com/graphics/2023-generative-ai-bias

(2) watsonx.ai: training, validering, afstemming, uitrol grondbeginselen en ML-modellen; watsonx.data: schaalbaarheid van AI-werklast met betrekking tot data waar dan ook opgeslagen en watsonx.governance: verantwoordelijke versnelling, transparantie en verklaarbaarheid data en AI werkstromen

Martijn Hoogesteger is Head of Cyber bij S-RM en is bereikbaar via m.hoogesteger@s-rminform.com.

COLUMN MARTIJN

Hackers in hyperdrive: de wapenwedloop met AI

Elk groot softwarebedrijf is tegenwoordig druk bezig met het testen van software en het uitbrengen van patches. Ze rapporteren netjes wat voor kwetsbaarheden de patches oplossen, en zorgen voor de juiste informatie zodat IT-teams weten welke stappen ze moeten ondernemen. Maar, wat voor effect zou dit in de toekomst kunnen hebben? Laten we eens kijken naar een fictief futuristisch scenario bij een VPN-softwareleverancier. Deze publiceert in de toekomst een reguliere patch waar een buffer overflow kwetsbaarheid gepatched wordt (een mogelijkheid om willekeurige commando's uit te voeren).

Automatische tools zien deze patch door regelmatig de websites van dit soort leveranciers te scrapen. Met behulp van language models wordt de tekst geïnterpreteerd. Een AI besluit dat een exploit maken interessant kan zijn en de applicatie wordt automatisch gedownload. Afhankelijk van waar de kwetsbaarheid zit en wat de patch doet, genereert een AI een code die misbruik maakt van de kwetsbaarheid en toegang geeft tot het systeem. Door dit een aantal keer te doen en verschillende variaties te maken kan de AI de meest effectieve versie selecteren. De proof of concept is af, en dat in slechts enkele minuten.

Ondertussen is door hetzelfde systeem op het internet gescand waar deze VPN geïnstalleerd is. Er zijn duizenden targets over de wereld. De exploit wordt automatisch afgevuurd op alle targets. Dit is voor de cybercrimineel maar een paar keer klikken, technische kennis is niet meer nodig. Deze modus operandi kennen we op dit moment al en hebben we gezien met bijvoorbeeld Citrix Netscaler of Pulse Connect Secure kwetsbaarheden. Ondersteund door AI gaat het proces nog vele malen sneller. Dit geeft kwetsbare organisaties significant minder tijd om zich voor te bereiden op een aanval.

Door deze aanval hebben criminelen een eerste stap in de netwerken van honderden organisaties gezet. Deze toegang wordt automatisch op het darkweb verhandeld, waar een andere groep de toegang opkoopt en hun eigen gespecialiseerde tools inzet. Een AI getraind op het gebruik van verschillende hacking-tools kijkt naar het systeem en de omgeving. Met de verzamelde informatie selecteert een AI de beste manier om rechten te verhogen en meer toegang te krijgen. Ook dit gebeurt razendsnel en parallel. Binnen enkele dagen hebben de criminelen vergaande toegang bij honderden organisaties. De criminelen krijgen automatisch te zien welke bedrijven de AI heeft gehackt, hoeveel omzet deze bedrijven hebben, in welke sector ze zitten en hoe ze het beste afgeperst kunnen worden. De crimineel kan met een druk op de knop afpersing starten per e-mail, telefoon of zelfs video. Sterk gepersonaliseerd, in de juiste taal en gebruikmakend van de informatie die de AI gevonden heeft in het netwerk.

Gelukkig zijn we nog niet in deze dystopische toekomst beland. Cybercriminelen hebben de afgelopen jaren enorme hoeveelheden geld buit weten te maken. Geld dat ze vervolgens investeren in deze tools en technieken. Gelukkig is de aandacht van de wereld ook gevestigd op deze criminele groeperingen, en weten we hun operaties steeds beter te dwarsbomen. We moeten dit juist nu keihard aanpakken, en onszelf heel scherp houden, voordat we in een exponentiele groei van criminaliteit komen door het gebruik van AI.

Hopelijk zet dit fictieve scenario je aan het denken! Het moment om door te pakken is nu.

In deze rubriek geven enkele IB-redacteuren in een kort stukje hun reactie op recente nieuwsitems over informatiebeveiliging. Dit zijn persoonlijke reacties van de auteurs en deze geven niet noodzakelijkerwijs het officiële standpunt weer van hun werkgever of van PvIB. Vragen en/of opmerkingen kun je sturen naar ibmagazine@pvib.nl.

GenAI: een nieuwe destructieve technologie of een evolutionaire stap?

In een rapport van McKinsey (juni 2023) werd gesteld dat 75% van de generatieve AI-impact zal worden gerealiseerd in: 1a. Software engineering bedrijfs-IT, 1b. Software engineering productontwikkeling, 2. Cliëntactiviteiten, 3a. Marketing, 3b. Verkopen en 4. Onderzoek & ontwikkeling. Impact ten aanzien van functionele inkoop varieert van 4% tot 39% en de ermee gemoeide bedragen van 320 tot en met 490 miljard US dollar.

Uitgangspunt daarbij was onder andere de aanname dat er bij specifieke toepassingen een revolutie plaatsvindt over de interne kennis managementsystemen en dat ongeveer 20% van de tijd van kenniswerkers ingezet wordt voor het zoeken en verzamelen van informatie.

Daarnaast stelt McKinsey een lijst op met zes essentiële overwegingen, waarvan de uitvoering het succes tot stand laat komen, of niet.

Deze vragen zijn:

1. Zijn de bedrijfskansen significant beter met GenAI?
   Tip: samen laten gaan met andere opkomende trends!
2. Bezitten wij genoeg ambitie met GenAI?
   Tip: hefboomeffect en radicale kostencurve verschuiving nastreven.
3. Waar zit het vermogen in de waardeketen?
   Tip: differentiatie door eigen databezit.
4. Beschikken wij over voldoende getalenteerde medewerkers?
5. Wat moeten wij doen om bij opschaling uit de dodenvallei te komen?


Fook Hwa Tan


Leo van Koppen

Tip: betreft ontwerpproblematiek met zes dimensies, geef dat aandacht.

6. Wegen wij de risicofactoren op de juiste wijze?

Het voorgaande geplaatst naast de hype die AI meemaakt, en dat iedereen het nu over GenAI heeft, roept de vraag op of GenAI een destructieve technologie is of slechts een volgende evolutiestap. Overwaarderen wij de mogelijkheden van GenAI en overschatten wij de risico's? Deze vragen leggen wij voor aan de redacteuren.

Fook Hwa Tan - De toekomst heruitgevonden: mens en AI in synergie

Generatieve AI is de katalysator die de toekomst van werk en innovatie hervormt, een kracht die ons uitnodigt om ons potentieel opnieuw te definiëren en te benutten. Het belichaamt niet slechts een evolutionaire stap, maar een transformatie die ons naar nieuwe hoogtes van creativiteit en efficiëntie leidt. Stel je een wereld voor waar kenniswerkers bevrijd zijn van de monotonie van informatieverzameling, waar hun capaciteiten gericht zijn op het creëren en innoveren, het verleggen van grenzen van wat mogelijk is.

Deze technologische vooruitgang vereist echter nieuwe competenties. Het vraagt om een mentaliteit die even dynamisch is als de technologie zelf, waarbij professionals zich voortdurend aanpassen, leren en evolueren. Het gaat om het ontwikkelen van een symbiotische relatie met AI, waarbij menselijke intuïtie gecombineerd wordt met machinale precisie en snelheid, om zo samen ongekende resultaten te bereiken.

In deze nieuwe realiteit is het essentieel om niet alleen technische vaardigheden te ontwikkelen, maar ook de 'soft skills' die ons menselijk maken, zoals creatief denken, empathie, en aanpassingsvermogen, te cultiveren. Dit zijn de kwaliteiten die ons onderscheiden en die, in combinatie met GenAI, de basis vormen voor een toekomst die rijk is aan innovatie en menselijke vooruitgang.

Laten we daarom deze nieuwe horizon omarmen, onze vaardigheden ontwikkelen en ons voorbereiden op een toekomst waarin we, samen met generatieve AI, onze wereld opnieuw uitvinden en een tijdperk van ongekende mogelijkheden en verwezenlijkingen binnentreden.

Leo van Koppen - Risicogebaseerd GenAI

Ik zou, vrij naar Beatrice te Graaf, graag willen starten met 'de historie heeft ons geleerd' dat we moeten oppassen met de adoptie van nieuwe technologie. Ons eigen vakgebied staat bol van de leerpunten, sterker nog komt voort uit het blind toepassen van nieuwe technologieën. Informatiebeveiliging en later cybersecurity huldigen het gedachtegoed van een risicogebaseerde aanpak en ook ik ben een aanhanger geworden van een dergelijke benadering. Niet alleen omdat het nu eenmaal past binnen ons vakgebied, maar ook omdat het vanuit een holistische en ethische zienswijze passend is. Niet alleen maar blind gaan op de voordelen van de nieuwe technologie, maar ook de andere kant van de medaille durven zien. Wat zijn de mogelijke gevolgen, of beter de risico's als we hiermee aan de slag gaan? De hype rondom GenAI waart volop rond en we raken o zo gemakkelijk gebiased door de veelheid aan artikelen over de mogelijkheden (kansen) van GenAI. We zien de risico's gewoon niet meer. Het onderzoek van McKinsey dat wordt aangehaald ontbeert de benadering van de keerzijde. De kansen worden met name belicht en de risico's worden achterwege gelaten, het is tekenend voor de bias. Vooralsnog wacht ik op de kritische beschouwingen van één van de goeroes in dit domein, waarmee vervolgens de eerste wetenschapsfilosofen en wellicht vervolgens ook journalisten nadruk gaan leggen op de mogelijk negatieve gevolgen van GenAI. Dat debat is hard nodig om de politiek wakker te schudden en om vervolgens regelgeving (lees regie) in te voeren. Net zoals bij een goede ontwerpmethode zou ik ook nu graag naast de use case van GenAI ook een abuse case opgesteld willen zien. Wat zijn de mogelijkheden voor de kwaadwillende? ChatGPT lepelt er zo een aantal voor me op zoals: misbruik, desinformatie, auteursrechten, creatief eigendom, veiligheid, privacy, ethiek, etc. De eerste kwesties over schending van auteursrechten zijn inmiddels aangekaart en er gaan er nog velen volgen. Eerst jurisprudentie en dan heel veel later pas wetgeving. De sturing of beter de regie op GenAI zal wederom te lang op zich laten wachten. Wetgeving hobbelt er immers altijd jaren later achteraan. Dat is eigenlijk ook vaak mijn punt. Ik zou graag wat meer regie willen bij dit soort ontwikkelingen en weet ook dat die regie pas (heel/te) laat komt.

Terecht benoemt mijn collega Fook Hwa Tan de symbiose van mens en technologie, zijn we als mens voldoende capabel om op een goede wijze om te gaan met deze technologie? Ik verwacht dat we op dat punt nog wel het nodige leergeld zullen moeten gaan betalen. Het zit nu eenmaal in de mens om het experiment aan te gaan, uit te proberen en te kijken wat het oplevert. Ook al lijkt het er op, – ik ben echt geen pessimist – ik weet dat door schade en schande de mens wijs is geworden. De historie heeft ons dat geleerd, maar we moeten die verworven wijsheid niet zomaar te grabbel gooien.

PvIB
Platform voor InformatieBeveiliging

COLOFON

IB Magazine is het huisorgaan van het Platform voor InformatieBeveiliging (PvIB) en bevat ontwikkelingen en achtergronden over onderwerpen op het gebied van informatiebeveiliging.

HOOFDREDACTEUR
Chris de Vries

REDACTIE
Bianca Brooijmans
Alex Dingemanse
Maarten Hartsuijker
Fook Hwa Tan
Lilian Knippenberg
Leo van Koppen
Rachel Marbus
Chris de Vries

BLADMANAGEMENT
MOS bv
Caroline Knobbe
Sam Dekkers
E ibmagazine@pvib.nl

ADVERTENTIE-ACQUISITIE
MOS bv
Eric Noordam
E acquisitie@mos-net.nl
T 033 247 34 00

VORMGEVING
Neverseen Art & Design
Dimitri van den Berg

DRUK
Veldhuis Media, Meppel

UITGEVER
Platform voor InformatieBeveiliging (PvIB)
Postbus 1058
3860 BB NIJKERK
T (033) 247 34 92
E secretariaat@pvib.nl
W www.pvib.nl

ABONNEMENTEN
De abonnementsprijs in 2023 bedraagt € 118,50 (exclusief btw), prijswijzigingen voorbehouden.

ABONNEMENTENADMINISTRATIE
Platform voor InformatieBeveiliging (PvIB)
Postbus 1058
3860 BB NIJKERK
E secretariaat@pvib.nl

ISSN 1569-1063

VERSTERK UW ISO/IEC 27001 MANAGEMENTSYSTEEM!

Wilt u zeker weten of uw organisatie klaar is voor certificering volgens ISO/IEC 27001? Doe de online DNV Self-Assessment en ontvang een rapportage in uw inbox. Of volg de Normkennis ISO/IEC 27001

Kent u de DNV Self-Assessment Suite al? Deze tool stelt u in staat te testen hoe goed u ISO/IEC 27001 kent en te beoordelen in hoeverre uw managementsysteem klaar is voor certificering. De evaluatie is op basis van puntenscores en laat zien waar er tekortkomingen bestaan en waar u verbeteringen kunt doorvoeren.

Stel een nulmeting op, stel kwantitatieve doelen vast voor een specifiek aandachtsgebied en meet regelmatig de geboekte vooruitgang. De Self-Assessment Suite biedt u in alle gevallen een gedetailleerd inzicht in uw kennis of prestaties en uw mate van beheersing.

Wilt u graag uw kennis vergroten over ISO/IEC ISO 27001? Volg dan de Normkennis ISO/IEC 27001 training! De trainer geeft u handige tips en voorbeelden uit de praktijk, zo leert u te kijken naar de norm, zoals een auditor dit doet.

Kijk voor meer informatie over de Self-Assessment Suite op dnv.nl/self-assessment of scan de QR-code als u wilt deelnemen aan de training.

DNV

TSTC

ICT en Security Trainingen

Ransomware? Log4j?

ADVANCE YOUR CAREER WITH SECURITY IN 2024

AIGP - Certified AI Governance Professional
CND - Certified Network Defender v2
CEH - Certified Ethical Hacker v12
OSCP - Offensive Security PEN-200
BIO - Certified Bio Professional
NIS2 - NIS2 Lead Implementer

GET SKILLED
WWW.TSTC.NL

Want security start bij mensen!!

TECHNICAL SECURITY TRAININGEN

CEH - Certified Ethical Hacker
CHFI - Computer Hacking Forensic Investigator
CPENT - Certified Penetration Testing Professional
SSCP - Systems Security Certified Professional
OSCP - Offensive Security Certified Professional

SECURITY MANAGEMENT TRAININGEN

CISSP - Certified Information Systems Security Professional
CISM - Certified Information Security Manager
CISA - Certified Information Systems Auditor
CRISC - Certified In Risk And Information Systems Control
C|CISO - Certified Chief Information Security Officer

PRIVACY TRAININGEN

CIPP/E - Certified Information Privacy Professional / Europe
CIPM - Certified Information Privacy Manager
CIPT - Certified Information Privacy Technologist
CDPO - Certified Data Protection Officer

CLOUD SECURITY TRAININGEN

CCSP - Certified Cloud Security Professional

ISO TRAININGEN

ISO 27001 - Foundation
ISO 27001 - Lead Implementer
ISO 27001 - Lead Auditor
ISO 27005 - Risk Manager
ISO 27701 - Privacy Management

www.tstc.nl

Onze trainingen zijn klassikaal of Live Online te volgen