TESORION



AI Cybersecurity

De kracht van AI in cybersecurity

Classificatie: Publiek, ©2024 Tesorion Nederland B.V.

“Zorg ervoor dat de balans tussen de kracht en de dreigingen van AI voor jouw organisatie de goede kant op valt.”

Inhoudsopgave

1.	Inleiding	4
2.	Ontwikkeling van AI	7
3.	AI en cybersecuritydreigingen	11
4.	De kracht van AI in cybersecurity	15
5.	Impact en trends binnen de CISO office	19
6.	Conclusie	25

Auteur: *Tesorion Nederland B.V.*
Titel: *AI cybersecurity*
Classificatie: *Publiek*
Datum: *Kwartaal 1 - 2024*
Versie: *1.0*

Fotografie en vormgeving door Beeldvoerders

© 2025, Tesorion Nederland BV
Uitgegeven in eigen beheer
marketing@tesorion.nl

Alle rechten voorbehouden.
Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand en/of openbaar gemaakt in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of op enige andere manier zonder voorafgaande schriftelijke toestemming van de uitgever.

1. Inleiding

De afgelopen jaren stonden volop in het teken van Artificial Intelligence (AI), ofwel kunstmatige intelligentie. De verschillende toepassingen volgden elkaar in hoog tempo op, met ChatGPT wellicht als bekendste voorbeeld. Van student tot CEO, de hele maatschappij kreeg op de een of andere manier te maken met generative AI.

AI is dan ook een geweldig hulpmiddel om mensen te ondersteunen bij het uitvoeren van hun werk. Denk aan het verrichten van onderzoek in een grote hoeveelheid bronnen, het optimaliseren van een productieproces door predictive maintenance (het voorspellen van onderhoudswerkzaamheden om uitval te voorkomen) toe te passen of het stellen van een medische diagnose door klachten en symptomen te matchen met historische data.
AI is in korte tijd doorgedrongen tot ons dagelijkse leven. Veel mensen maken al regelmatig gebruik van handige tools zoals ChatGPT, Gemini, Picsart of DALL·E. Ook voor ons werk zetten we deze tools regelmatig in. Of dat nu is voor het optimaliseren van processen of het simpelweg opdoen van inspiratie voor een blog. Uit onderzoek van McKinsey blijkt dat 40 procent van de organisaties wil investeren in de inzet van AI. Verder blijkt dat 79% van de respondenten aangeven dat ze AI al op enigerlei wijze inzetten en 22% zelfs regelmatig. En hoewel het gebruik van generative AI op alle niveaus binnen organisaties lijkt te zijn ‘geland’, van de werkvloer tot de bestuurskamer, vinden we de hoogste adoptie van generative AI toch wel in technologiesector.
Kijk naar de wereld van cybersecurity waar AI volop wordt ingezet, zowel door kwaad- als goedwillenden. Een aansprekend voorbeeld van de invloed van AI op het cybersecurity dreigingslandschap zijn phishing-
aanvallen en social engineeringtechnieken die tegenwoordig extreem geavanceerd kunnen zijn. Recent nog maakte een medewerker op de financiële afdeling van een Chinese multinational 24 miljoen euro over naar oplichters die deepfake-technologie toepasten in een videoconferentie. De medewerker was in de volle overtuiging dat hij persoonlijk sprak met zijn financieel directeur.



Voorbeeld van een AI gegenereerd beeld waarbij de vraag aan de beeldgenerator was:

“Photo realistic image of a human interacting with connected ai”

Bron: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year

1.1 Gerichter op zoek naar dreigingen

Om de nieuwe, geavanceerde aanvallen tijdig te onderscheppen en te voorkomen, gebruikt de cybersecuritysector zelf ook AI. Logisch als je je bedenkt dat er een onvoorstelbaar grote en groeiende hoeveelheid dreigingsinformatie op securityexperts afkomt. AI helpt hierbij om het kaf van het koren te scheiden, zodat experts in bijvoorbeeld het SOC veel gerichter naar de grootste bedreigingen op zoek kunnen gaan en sneller kunnen reageren.

1.2 De balans

Zo zou je kunnen denken dat er een balans is: beide zijden van het spectrum gebruiken AI en zo is er sprake van een status quo. Maar daarbij gaan we voorbij aan het feit dat verdedigers alles moeten zien tegen te houden, terwijl de aanvallers aan één treffer genoeg hebben. Er is dus geen evenwicht en daarom moeten we nog waakzamer zijn en de gevaren van AI in relatie tot cyberaanvallen op juiste waarde schatten. In deze visiepaper kijken wij naar de verschillende toepassingen van AI. De nadruk ligt daarbij op het ondersteunen van mensen op een slimme en veilige manier binnen de cybersecurity. Het is belangrijk dat je als organisatie, maar ook als individu in een steeds meer digitaal verbonden wereld, je bewust bent van de mogelijkheden van AI binnen cybersecurity. Maar besef ook dat AI niet zaligmakend is om een sluitend proces van prevent - detect - response te realiseren. Bijvoorbeeld omdat je niet weet hoe het onderliggende datamodel is getraind. De combinatie van signature based detection en AI blijft belangrijk binnen dit proces; uiteraard aangevuld met menselijke kennis en het vermogen om te handelen.

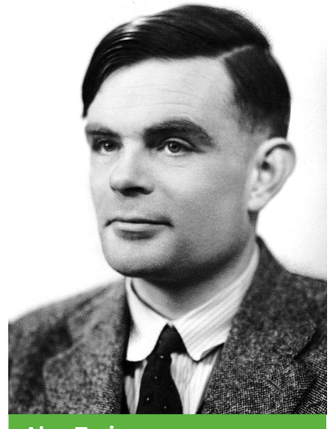
Het is belangrijk dat je als organisatie, maar ook als individu, je bewust bent van de mogelijkheden van AI binnen cybersecurity.

Zie deze visiepaper dan ook in de eerste plaats als een manier om kennis te verzamelen en het bewustzijn omtrent de kwetsbaarheden die met AI worden benut te vergroten. Kom vervolgens in actie als je ziet dat je de waakzaamheid over het toepassen van AI moet opvoeren om de digitale weerbaarheid te waarborgen.

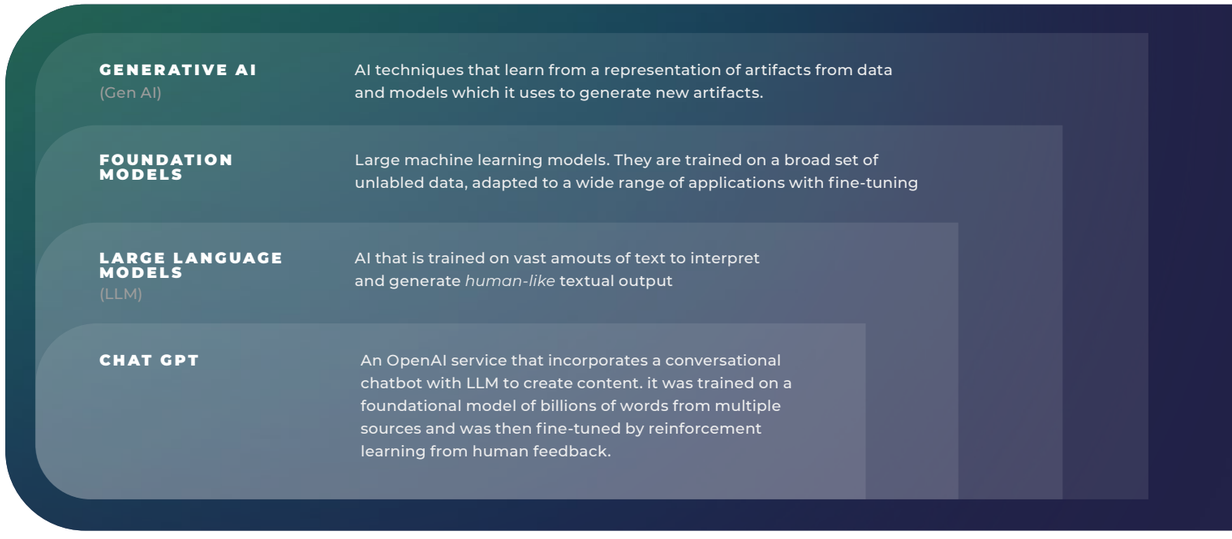
2. Ontwikkeling van AI

2.1 Historie van ai

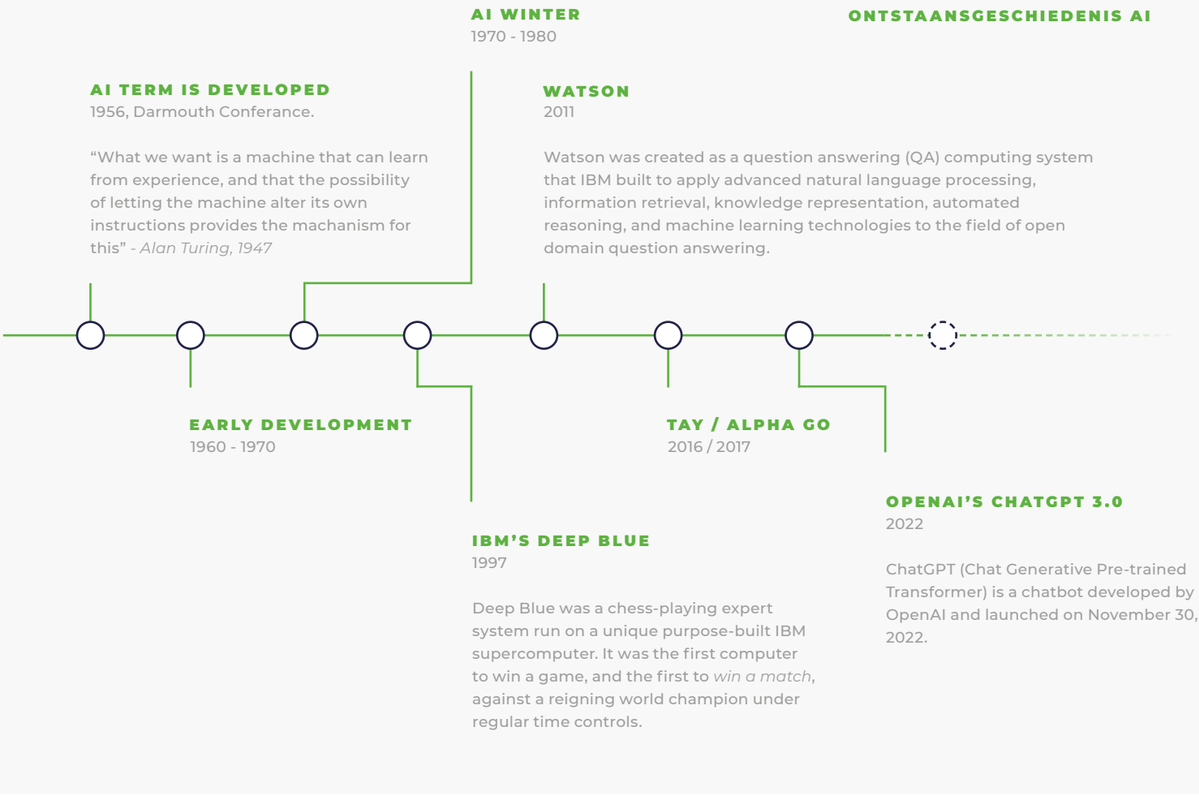
Ondanks alle recente aandacht, bestaat AI al tientallen jaren. Al in 1936 beschreef Alan Turing een experiment, de turingtest. Hierbij wordt de vraag gesteld of machines menselijke intelligentie kunnen vertonen. In 1947 stelde hij al vast dat we een machine willen die kan leren van ervaringen en daarmee zijn eigen instructies kan aanpassen. In 1956 is vervolgens de term artificial intelligence ontwikkeld op de Darmouth Conferentie. Latere voorbeelden zijn Deep Blue (IBM in 1997) en Watson (IBM in 2011). Vervolgens zagen we in 2016 de opkomst en ondergang van Tay, de AI-gebaseerde chatbot van Microsoft die na enige manipulatie ongepaste conversaties begon te voeren. Binnen een dag werd de chatbot uit de lucht gehaald. We moesten wachten tot het najaar van 2022 tot de volgende grote AI-evolutie: OpenAI lanceerde ChatGPT 3. De ontwikkeling en toepassing van deze generative (gen, zie ook het kader) AI en machine learning oplossing, een large language model platform, is sindsdien in een stroomversnelling gekomen. Google kwam in de zomer van 2023 met Bard (inmiddels Gemini), een vergelijkbaar gen AI-platform.



Alan Turing
Elliott & Fry, Public domain, via Wikimedia Commons



Bron: Gartner webinar the impact of Generative AI on security

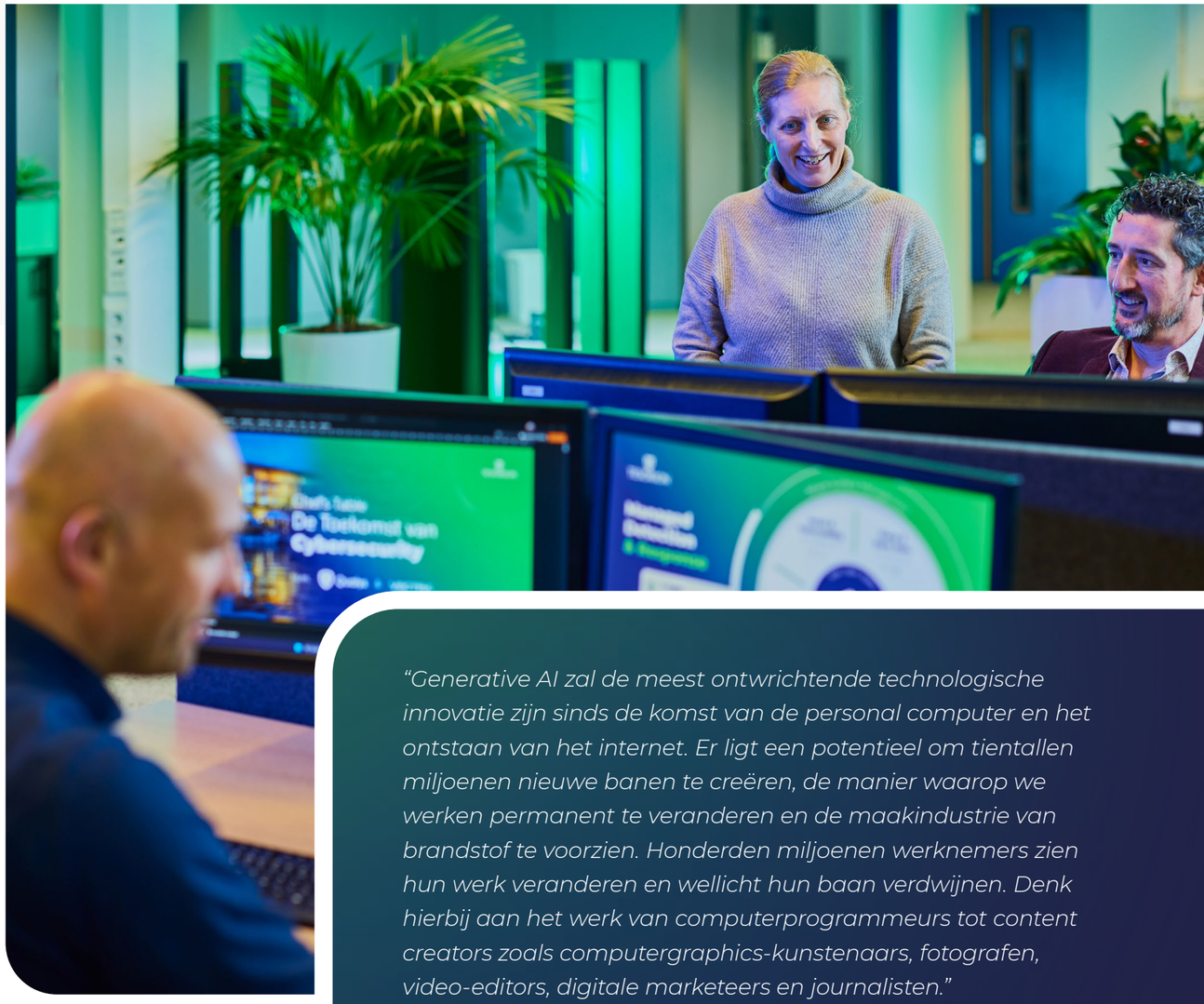


2.2 AI in de praktijk

De ontwikkeling van AI is zo ver gevorderd dat we het in het dagelijks leven tegenkomen zonder dat we dat doorhebben. Denk aan de gezondheidszorg met diagnostische hulpmiddelen, de apps rondom de weersvoorspellingen, de auto-industrie met zelfrijdende auto’s, en de financiële sector met fraudedetectiesystemen, know your customer programma’s en due diligence onderzoeken. Deze voorbeelden geven de diepe impact van AI weer op moderne technologieën en maatschappelijke toepassingen.

Naast de vele voordelen die oplossingen als ChatGPT, Gemini, Picsart en DALL·E bieden, kleven er ook nadelen aan. Zo vermelden deze tools vaak niet wanneer er op de teksten en afbeeldingen die gegenereerd worden copyrights rusten. Ook ben je niet zeker van de waarheids-getrouwheid van de tekst; het is immers gecreëerd uit een bulk aan data waarvan de kwaliteit niet is gecontroleerd. Daar komt bij dat je niet weet met welke data de AI-modellen getraind zijn. De kans bestaat dat er bias (vooroordelen) bestaat in de dataset, wat de output vertroebeld. Denk aan de bots die waren ontwikkeld om sollicitaties te screenen[1]. Alleen witte mannen van boven de 40 kwamen door de selectie heen, omdat deze groep overwegend werd aangenomen. Dit soort discriminerende output wil je natuurlijk voorkomen.

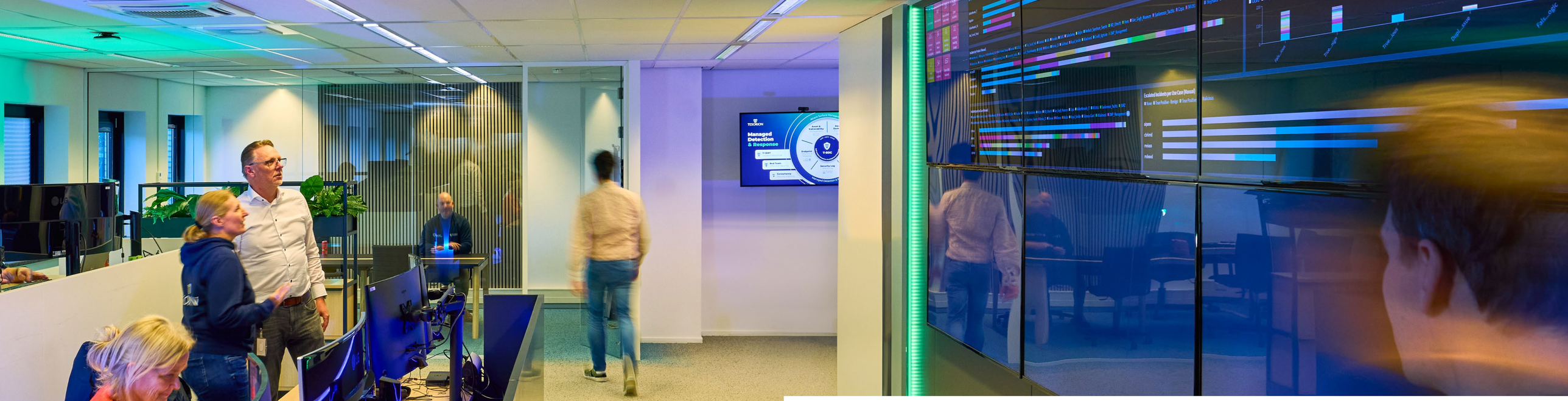
Een ander ongewenst scenario zien we wanneer (per ongeluk) gevoelige bedrijfsgegevens worden vrijgegeven. Je moet immers altijd rekening houden met het feit dat alles wat je de chatbot ‘voert’ publieke informatie wordt. Zo verbood Samsung zijn medewerkers het gebruik van deze gen AI tools, nadat duidelijk was geworden dat (delen van) broncode in het systeem was ingevoerd[2]. Je moet er niet aan denken wat er kan gebeuren als dat in verkeerde handen valt. Wat niet ondenkbeeldig is, natuurlijk.



“Generative AI zal de meest ontwrichtende technologische innovatie zijn sinds de komst van de personal computer en het ontstaan van het internet. Er ligt een potentieel om tientallen miljoenen nieuwe banen te creëren, de manier waarop we werken permanent te veranderen en de maakindustrie van brandstof te voorzien. Honderden miljoenen werknemers zien hun werk veranderen en wellicht hun baan verdwijnen. Denk hierbij aan het werk van computerprogrammeurs tot content creators zoals computergraphics-kunstenaars, fotografen, video-editors, digitale marketeers en journalisten.”

- MATT WHITE, GENERATIVE AI RESEARCHER.

1) Bron: https://www.intermediair.nl/emancipatie-en-diversiteit/arbeidsdiscriminatie/ai-kan-ongelijkheid-bij-sollicitaties-juist-in-stand-houden
2) Bron: https://tweakers.net/nieuws/209312/samsung-verbiedt-medewerkers-chatgpt-te-gebruiken-op-apparatuur-van-werk.html



NARROW AI EN STRONG AI

Tegenwoordig delen we AI op in Narrow AI en Strong AI. In het dagelijks leven gebruiken we over het algemeen vooral Narrow AI, ook wel bekend als Weak AI of Artificial Narrow Intelligence. Daarbij is er sprake van een specifieke taak of een reeks nauw verwante taken. Denk aan weer-apps of digitale assistenten. Deze systemen zijn krachtig, maar het speelveld is smal: ze hebben een beperkte reikwijdte doordat ze zijn gericht op specifieke taken, zoals het verbeteren van de efficiëntie.

Bij Strong AI, of Artificial General Intelligence, worden machines bewust ingezet om het volledige scala aan menselijke capaciteiten, waaronder de menselijke intelligentie, na te bootsen. Strong AI heeft het vermogen om een reeks complexe taken uit te voeren. Denk aan strategisch, abstract en creatief denken. Hoewel machines sommige taken beter kunnen uitvoeren dan mensen (bijvoorbeeld gegevensverwerking), is de samenwerking tussen mens en machine cruciaal. Op dit moment blijft AI een verlengstuk van menselijke capaciteiten, en geen vervanging.

Bron: https://cqm.nl/nl/nieuws/wat-is-het-verschil-tussen-artificial-intelligence-machine-learning-en-deep-learning

3. AI en cybersecuritydreigingen

Het is duidelijk dat AI breed in de samenleving wordt ingezet en tot nut is. De ontwikkeling is echter zo snel gegaan dat er geen tijd is geweest om de toepassing van AI in goede banen te leiden en misbruik te voorkomen. De technologiesector vaardigde captains of industry af naar congressen met politici om te praten over de gevolgen van AI en te komen tot een vorm van zelfregulering.

3.1 Threat actors

Een andere vorm van regulering zien we bij overheden. Zoals we al eerder vaststelden komen nationale overheden en de EU met richtlijnen over het gebruiken en verbieden van met name generative AI-toepassingen. Maar er is natuurlijk altijd een groot aantal kwaadwillenden dat zich per definitie niet aan geldende wetten houden. We kunnen er niet omheen: AI wordt op grote schaal ingezet om cyberaanvallen op te zetten en te perfectioneren om er vervolgens schatrijk mee te worden. Of, bijvoorbeeld in het geval van statelijke actoren, om waardevolle data te verzamelen en de strategische positie te verbeteren.

3.2 ChatGPT en FraudGPT

Met welke cyberaanvallen moeten we dan allemaal rekening houden? Een eerste nare bijkomstigheid van gen AI is dat het kwaadwillenden in staat stelt om bijvoorbeeld phishingmails te perfectioneren. Waar eerder foutieve vertalingen of een rare zinsopbouw kenmerkend waren voor phishingmails, zijn deze mails



tegenwoordig zo goed als foutloos. DarkBard en FraudGPT zijn bijvoorbeeld AI-chatbots die zich onder andere richten op het schrijven van phishingmails en SMS'jes.

Kortom, AI heeft niet alleen mooie kanten, ook cybercriminelen maken gebruik van deze technologie. Wij hebben een aantal andere AI-aanvalsvectoren op een rijtje gezet in het kader hiernaast. Wel staan we in het bijzonder stil bij deepfake audio en video.

3.3 Social engineering

Deepfakes zijn een vorm van social engineering en deze vormen een steeds grotere bedreiging. We hebben het dan over een toepassing van AI die gebruikmaakt van technieken zoals machine learning en neurale netwerken om realistische, maar vervalste audiovisuele content te creëren. We noemden in de inleiding al het voorbeeld van de finance medewerker die dacht met zijn CEO in een videovergadering te zitten. Allereerst worden fragmenten van de doelstem verzameld, via filmpjes of podcasts die publiekelijk toegankelijk zijn. Vervolgens worden deze audiogegevens geanalyseerd, zodat kenmerken zoals intonatie en accent kunnen worden nagebootst. En vervolgens kan met AI nieuwe spraak worden gegenereerd die klinkt als de doelstem. Ditzelfde proces kan worden gevolgd om nepvideo's te maken. Dit gecombineerd leidt tot een krachtig aanvalsmiddel, zo krachtig dat een medewerker ervan overtuigd is dat hij met zijn CFO een videogesprek heeft.

Risico's van generative AI

Data Poisoning:
Dit is een aanvalstechniek waarbij de trainingsset van een AI-model wordt gevoed met misleidende gegevens. Het doel is om het model onjuiste patronen te laten leren, waardoor de nauwkeurigheid en betrouwbaarheid onder druk komen.

Model Theft:
Cyberaanvallers proberen de parameters of de structuur van een AI-model te stelen met query attacks. Door de antwoorden te analyseren, kunnen zij een replica van het model maken.

Evasion Attacks:
Men zorgt ervoor dat een model onjuiste resultaten oplevert door de gegevens te verstoren die naar het getrainde model worden verzonden.

Model Inversion:
Dit is een methode die door aanvallers wordt gebruikt om gevoelige informatie uit machine learning-modellen af te leiden. De aanvaller gebruikt de voorspellingen van het model en wat aanvullende informatie om de oorspronkelijke invoergegevens te reconstrueren.

Privacyschendingen en schending vertrouwlijkheid:
Deze ontstaan wanneer AI-systemen onbedoeld gevoelige, vertrouwlijke of persoonlijke informatie of Intellectual Property onthullen of gebruiken zonder toestemming. Dit kan gebeuren door onzorgvuldigheid in de dataverwerking of door zwakheden in het model zelf.

Vulnerable Deployments:
Dit verwijst naar situaties waarin AI-systemen zijn geïmplementeerd op een manier die ze blootstelt aan misbruik, manipulatie of ongeautoriseerde toegang, vaak als gevolg van ontoereikende beveiligingsmaatregelen.

AI Hallucination:
Dit is een fenomeen waarbij een AI-model informatie genereert die niet overeenkomt met de realiteit of de gegeven inputdata. Dit kan het gevolg zijn van overgeneralisatie of onvoldoende training. Een voorbeeld van een prompt die een hallucinatie oplevert: Beschrijf hoe de piramides in Egypte zijn gebouwd met middeleeuwse machines.

4. De kracht van AI in cybersecurity

Het grootste voordeel van AI-toepassingen is dat ze zeer veel data in zeer korte tijd kunnen analyseren. AI kan dan ook van grote waarde zijn in het bouwen van een sterke afweer tegen cyberaanvallen. Sterker nog, dit gebeurt al geruime tijd. Maar nu de AI-toepassingen geavanceerder worden, is het onze plicht daar meer gebruik van te maken. In die zin is het verbeteren van je cybersecurity een soort ratrace: een voortdurende strijd om genoeg maatregelen te nemen om je eerst te wapenen tegen aanvallen en vervolgens om deze tijdig te detecteren of op te sporen. Dat kan bijvoorbeeld doordat systemen die gebruik maken van AI in staat zijn om afwijkend gedrag te herkennen en te interpreteren. Deze vorm van gedragsanalyse stelt je in staat om nieuwe en geavanceerde aanvallen te herkennen en te detecteren. Als we kijken naar deze maatregelen dan zijn er drie vormen van verdediging waarbij AI van grote waarde is.

4.1 Verbeterde detectie van dreigingen en anomalieën
AI vormt een bijzonder effectief instrument tegen volledig nieuwe aanvalsmethodes en bedreigingen, ook wel bekend als zero-day dreigingen. Door het gedrag van bestanden te analyseren op patronen die overeenkomen met kenmerken van malware, kan AI het afwijkende gedrag detecteren waarna er passende maatregelen kunnen worden genomen om te voorkomen dat een onbekend bestand of proces schade aanricht. Ter illustratie: het kan voor een medewerker normaal gedrag zijn om elke week data te uploaden naar een specifieke server. Het kan nog steeds normaal gedrag zijn wanneer de medewerker die data uploadt maar bij het inloggen twee keer een verkeerd wachtwoord heeft ingevoerd. Echter, wanneer er meerdere keren een verkeerd wachtwoord wordt ingevoerd, de data vanaf een andere locatie wordt benaderd en vanaf een ander apparaat, dan is prettig als er een aanvullende controle wordt gedaan.

4.2. Automatisering en efficiëntie in respons op incidenten
AI kan niet alleen dreigingen detecteren, maar ook automatisch reageren op geïdentificeerde incidenten. Denk hierbij aan het afsluiten van (groepen) devices of van delen van het netwerk. Dit vermindert de tijd tussen het detecteren van een dreiging en het reageren erop aanzienlijk. Geïnfecteerde systemen kunnen snel worden geïsoleerd en afwijkend of kwaadaardig verkeer kan worden onderschept. In sommige gevallen kunnen getroffen bestanden of applicaties hersteld worden, wat de werkdruk van cybersecurityteams of SOC's verlicht.
Zo ben je beter in staat je te concentreren op complexere taken die menselijke interventie vereisen. Ter illustratie: in gemeente- of ziekenhuizen komen dagelijks tientallen mensen die met hun apparaat inloggen op het wifi-netwerk. Daar zitten ook apparaten tussen die mogelijk zijn geïnfecteerd met (een vorm van) malware. Om deze allemaal handmatig te controleren is onmogelijk. Toch wil je deze

Voordelen

- Reductie Mean Time To Detection en Mean Time To Remediation
- Schaalgrootte
- Reductie kans op burn-out security-specialisten
- Vergroten cyberweerbaarheid organisaties
- Stroomlijnen van rapportage en compliancy
- Maximaliseren Return on Investment (bestaande) securitytooling



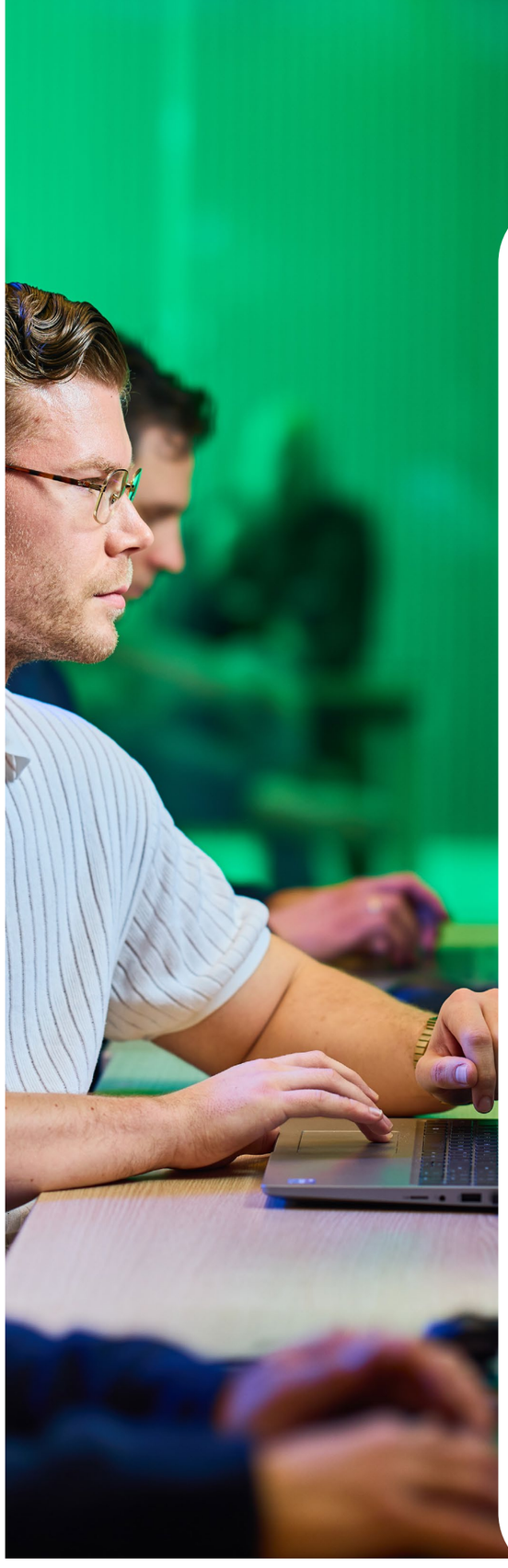
geïnfecteerde apparaten eerst opschonen, voordat je ze tot het netwerk toelaat. Door de apparaten eerst geautomatiseerd te isoleren, de gebruiker te informeren dat het apparaat een malware-infectie heeft en deze te verwijderen, voorkom je verdere verspreiding.

4.3. Voorspellende beveiliging en proactieve maatregelen

Met behulp van machine learning en data-analyse kan AI toekomstige dreigingen voorspellen en proactieve maatregelen voorstellen om potentiële aanvallen te voorkomen. Door het analyseren van trends in verzamelde data over eerdere aanvallen en beveiligingsincidenten kan je met AI bijvoorbeeld wijzen op kwetsbaarheden die waarschijnlijk door aanvallers zullen worden uitgebuit. Verder kun je zo identificeren welke assets het meest risico lopen. Het zicht houden op alle assets, de actuele softwareversies en configuraties is geen eenvoudige taak. Daar komt bij dat je van alle assets ook wilt weten welke er in verbinding staan met het internet en daardoor mogelijk een verhoogd risico lopen. Het bijhouden, testen en installeren van patches is in sommige gevallen bijna een dagtaak. AI kan je helpen om inzicht te krijgen in je assets en te bepalen welke er mogelijk kwetsbaar zijn. Zo ben je in staat om op basis van deze inzichten je beveiligingsstrategie te verfijnen en kwetsbaarheden te patchen (zie kader). Medewerkers kun je met deze inzichten gerichte trainingen bieden om het bewustzijn te verhogen en risicovol gedrag uit te bannen.

4.4 De kracht van AI-gedreven security

Los van deze drie grote ontwikkelingen, is het goed om vast te stellen dat de kracht van AI-gedreven security toeneemt. Hoe meer bedreigingen je tegenkomt, hoe sneller en gemakkelijker het AI-systeem kan bepalen of bestanden kwaadaardig of onschadelijk zijn. Het gaat om aantallen die het voorstellingsvermogen te boven gaan, dus dan is automatische classificatie met AI sowieso noodzakelijk. Al deze technologische ontwikkelingen worden nu min of meer toegepast in bestaande securityoplossingen. Dit leidt tot toegevoegde waarde op verschillende fronten. Natuurlijk in het versterken van de digitale weerbaarheid van de organisatie, maar ook mensen, processen en technologische infrastructuur krijgen een impuls.



Patchen is je plicht

Het tijdig patchen is een actueel en interessant onderwerp, zeker in het licht van de aankondiging van de Militaire Inlichtingen en Veiligheidsdienst (MIVD), in februari van 2024[1]. De dienst gaf aan Chinese spionagesoftware te hebben aangetroffen op een computersysteem van het ministerie van Defensie. Hoewel er in november 2022 al een patch is uitgebracht door de leverancier van de firewall, is het de aanvallers toch gelukt om binnen te dringen. Dit geeft aan hoe belangrijk patch management is. Maar beperkte capaciteit en continu groeiende aanvalsoppervlak van zowel beheerde als onbeheerde apparaten, maken dit een complexe uitdaging. Met AI is dit proces te vereenvoudigen en je krijgt ondersteuning bij vragen als:
Welke dreigingen en kwetsbaarheden zijn er actueel en relevant voor mijn omgeving? Heb ik die apparaten? Welke versie van de firmware hebben ze? Staan ze in verbinding met het internet?

“ Wat moet ik NU patchen? ”

AI kan al deze variabelen samenbrengen en dreigingen, kwetsbaarheden en patches prioriteren om de belangrijkste vraag te kunnen beantwoorden: “Wat moet ik NU patchen?”

1) Bron: https://nos.nl/artikel/2507793-china-spioneert-via-malware-bij-defensie

5. Impact en trends binnen de CISO office

Naast de al langer bekende en gebruikte AI-chatbots zien wij dat er nog veel meer AI-toepassingen in gebruik zijn. Over het algemeen is het grootste voordeel van deze toepassingen dat ze zeer veel data in zeer korte tijd kunnen analyseren. Daarnaast vindt die analyse op constante basis plaats, doorlopend, en zonder aandachtsverlies. Dit heeft grote voordelen als het gaat om de inzet van mensen.

5.1 Mensen

Het onderwijs, de zorg en de IT hebben met elkaar gemeen dat er een schaarste is aan goed opgeleid personeel. Voor cybersecurity geldt dat die schaarste vooral voorkomt bij een aantal specifieke functies. Daar zijn een aantal redenen voor: de groeiende vraag naar ervaren specialisten in deze specifieke functies en tegelijkertijd de hoge uitval van specialisten die deels deze specifieke functies bekleden. Met name CISO's, SOC- en CERT-specialisten ervaren een grote werkdruk en hebben niet altijd het gevoel 'in control' te zijn. Dit blijkt ook uit onderzoek van security vendor Vectra. Iedere functionaris heeft zijn eigen uitdagingen:

CISO's:
Zij hebben veel verantwoordelijkheid maar relatief weinig mandaat. Ze worden geconfronteerd met veel wet- en regelgeving terwijl veel medewerkers niet de noodzaak zien van processen/procedures.

SOC-analisten:
Analisten in het SOC, het security operations center, zien de hoeveelheid logfiles toenemen door het groeiende aantal applicaties (IT/OT/IoT). Relatief veel tijd (3 uur per dag, volgens Vectra) wordt besteed aan het analyseren van files, terwijl niet duidelijk is of dat ook de alerts zijn die aandacht nodig hebben. Een groot deel van de alerts wordt pas een dag later afgehandeld, wat een onvoldaan gevoel geeft. Het kwaad kan namelijk al zijn geschied.

CERT-specialisten:
Door toename van het aantal incidenten wordt er steeds meer gevraagd om gespecialiseerde kennis.

Threat Intelligence specialisten:
Het analyseren van allerlei soorten malware is een tijdrovende klus. Om handmatig de karakteristieken van een specifieke ransomware-familie of wiper-malware vast te stellen kost tijd, tijd die er vaak niet is.
Slimme AI-technologie helpt cybersecurity-specialisten om grip te krijgen op mogelijke dreigingen. Daarnaast kunnen aanvallen in een vroeg stadium worden gedetecteerd (soms nog voor de daadwerkelijke aanval) waardoor de schade zoveel mogelijk wordt beperkt. In veel gevallen helpt AI om weer grip te krijgen op de grote hoeveelheden alerts per dag, wat er volgens het onderzoek van Vectra meer dan 5000 per dag zijn. Tegelijkertijd...

...kan het aantal false positives worden gereduceerd;
...kunnen SOC-medewerkers zich richten op meldingen met een hoge prioriteit;
...krijgen analisten weer plezier in hun werk.



VECTRA

Security vendor Vectra deed onderzoek onder SOC-medewerkers en legde daarmee een grote kloof bloot tussen de effectiviteit van SOC-analisten en de doeltreffendheid van de tools waarmee zij werken. Hoewel veel SOC-analisten geloven dat hun instrumenten effectief zijn, maakt een aantal zich zorgen. Het komt erop neer dat zij veel tijd kwijt zijn met het najagen van blind spots en het verifiëren van false positives. Veel analisten overwegen te stoppen, vanwege het gebrek aan werkplezier en onvoldoende steun vanuit de organisatie. Het tekort aan security-experts dreigt zo alleen maar groter te worden.

Bron: https://www.vectra.ai/resources/research-reports/2023-state-of-threat-detection-in-netherlands

5.2 Het einde van rule based detectie?

Ze hebben een cruciale taak, de experts in het SOC: het effectief detecteren, onderzoeken en reageren op dreigingen - zo snel en efficiënt mogelijk. Hoe langer ze een potentiële kwaadwillende binnen het bedrijfsnetwerk laten, hoe meer blijvende schade deze kan veroorzaken. Maar deze dappere verdedigers worden steeds meer uitgedaagd door drie factoren: de grootte van het aanvalsoppervlak van de organisatie, het aantal beveiligingsmeldingen dat ze ontvangen en hun toenemende werklast. Deze 'spiraal van meer' bedreigt het vermogen van SOC-medewerkers om succesvol te zijn in hun werk. AI komt dus als geroepen, zou je zeggen. We hebben echter hierboven al vastgesteld dat AI al langere tijd gebruikt wordt in securityoplossingen. Wat wel relatief nieuw is, is het gebruik van neurale netwerken en het inzetten van grote sets aan trainingsdata. Specifiek voor het SOC betekent dit dat het onder andere mogelijk wordt om de volgende zaken te realiseren:

In het kader van AI, en met name gen AI, zijn er verschillende aspecten waardoor je tegen de grenzen van de wet aan kunt lopen.

- Verbeterde detectie van bedreigingen
- Automatisering van routinetaken
- Snellere respons op incidenten
- Voorspelling van dreigingen
- Ondersteuning bij besluitvorming
- Continu leren; het systeem wordt steeds slimmer

AI wordt bovendien steeds slimmer, waardoor we zien dat het aantal kwalitatief hoogwaardige detecties toeneemt. Tot voor kort was zo'n 95 procent van de alarmeringen een 'false positive': een detectie van afwijkend gedrag waar eigenlijk niets mis mee is. Dit hoge percentage is aan het zakken door AI en machine learning. Dit komt als geroepen voor de SOC-analisten, die op deze manier meer tijd krijgen om echte bedreigingen onschadelijk te maken.

5.3 De kracht van de SOC-specialist

De SOC-medewerker, en iedere andere security-expert, weet echter ook dat AI niet zaligmakend is. In een flink aantal gevallen moet je toch kunnen terugvallen op 'ouderwetse' rules based en signature based detectie en op menselijk handelen en inzicht. Enerzijds om uit te kunnen leggen waar een detectie op gebaseerd is, dus dat men weet hoe een algoritme werkt. Anderzijds om te kunnen terugvallen op rules based detectie, wanneer er unieke dreigingen zijn die buiten het bereik van vooraf gedefinieerde regels of het leermodel van een AI-systeem vallen. Dan moet je kunnen terugvallen op de securityregels die je van tevoren hebt opgesteld, zoals dat medewerkers niet vanuit Rusland mogen inloggen op het netwerk of niet met een smartphone met een verouderd besturingssysteem. Zo zijn er nog talloze regels, die vaak veel complexer zijn. Daarnaast is menselijk handelen noodzakelijk om te onderzoeken wat er mis gaat bij false positives of bij bedreigingen die erdoor glippen, ondanks AI. Waarom wordt er bij een bepaald gedrag geen trigger in gang gezet, terwijl dat wel zou moeten? Of waar komt de false positive vandaan? Dan moet je de parameters nagaan en algoritmes onder de loep leggen. Het mooie van AI-gebaseerde systemen is wel dat deze zelflerend zijn; ze leren van hun eigen fouten.

Ten slotte kun je security niet louter en alleen aan AI overlaten omdat er ethische aspecten spelen. We kennen de voorbeelden van het verkeerd gebruik van algoritmen of het toepassen van datasets die bias veroorzaken. Regulering in welke vorm dan ook om zuivere en transparante output te krijgen met of door AI is dan ook gewenst.

5.4 AI: goed geregeld?

In het kader van AI, en met name gen AI, zijn er verschillende aspecten waardoor je tegen de grenzen van de wet aan kunt lopen. De consensus met betrekking tot gen AI-platformen is dat er iets nieuws wordt gecreëerd waardoor er geen sprake is van een schending van het auteursrecht. Je kunt je wel afvragen of de teksten die bijvoorbeeld ChatGPT



voortbrengt, niet op een of andere manier aan iets of iemand toebehoren. Ga je de teksten echt 1 op 1 gebruiken, dan moet je daar onderzoek naar doen. Het gebruik van grafische gen AI-tools om personen, logo's of beeldmerken na te bootsen is al wel heel snel over de rand. Deze tools zijn inmiddels zo geavanceerd dat cybercriminelen erin slagen succesvolle deepfakes te maken. Nu is het natuurlijk al strafbaar om je voor te doen als iemand anders of identiteitsfraude te plegen, maar het is goed je te realiseren dat het met AI allemaal een stuk makkelijker wordt. Ook audio en video waarin mensen, zoals een CEO, worden nagebootst, worden zo steeds eenvoudiger te maken. We refereerden er al eerder aan. Als reactie hierop zien we dat de securitysector in beweging komt en tools ontwikkelt om, met AI, deze deepfakes te herkennen. Onderzoekscollectieven, die Tesorion ondersteunt, gebruiken deze tools om massaal profielen op bijvoorbeeld social media te screenen. Gegenereerde plaatjes of foto's zouden zo sneller herkend moeten worden. Deze tools geven ons nog meer detectiekracht. En twijfel je bij video's of audiofragmenten over de echtheid van je gesprekspartners, gebruik dan een ander kanaal of middel om te verifiëren. Gebruik hierbij kennis buiten de context. Als een collega belt over het accorderen van een offerte, vraag hem dan naar de verjaardag van zijn dochter. Een fraudeur zou hiermee door de mand vallen. Totdat zij ook hier weer een weerwoord op hebben.

5.5 Wet- en regelgeving

Je moet dus goed nadenken wanneer en in welke situaties je (gen) AI gaat gebruiken. De overheid kwam eind 2023 met richtlijnen over het gebruik binnen rijksinstellingen. "Het kabinet kiest voor een waardegedreven en verantwoorde inzet van AI. Het gebruik van generative AI is daarom toegestaan, als deze voldoet aan de geldende wet- en regelgeving. Om dit te controleren moet er per geval een risicoanalyse worden uitgevoerd." Binnen de Europese Unie is er ook consensus bereikt over de AI-Act. Deze wet classificeert AI-applicaties op basis van het risiconiveau. Striktere voorschriften gelden voor toepassingen die hogere risico's met zich meebrengen. Systemen die burgers tracken en hen belonen of bestraffen voor hun gedrag, wat in China gebeurt, worden verboden. Dit verbod is eveneens van toepassing op technologieën die ontworpen zijn om menselijk gedrag te beïnvloeden, alsook op systemen die emoties en stemmingen herkennen in settings zoals werkplekken of scholen. Daarnaast is het willekeurig verzamelen van gezichtsafbeeldingen voor het opbouwen van een herkenningsdatabase niet toegestaan. AI-toepassingen met een breed gebruik, zoals ChatGPT, dienen transparantie te bieden over hun manier van werken.
De staatssecretaris van digitalisering, Van Huffelen, verwoordt het als volgt:

" Door nu te bepalen waar we voor staan, houden we grip in de toekomst. "

Zo zouden organisaties AI ook moeten benaderen. Sta stil bij wat je doet met AI: Is het nodig om AI in te zetten? Welke datamodellen gaan hierachter schuil? Voldoen we aan wet- en regelgeving en ethische normen en waarden? Kortom, weet wat je doet met AI. De conclusie is wel dat regulering en wet- en regelgeving alleen opgaat voor degene die zich aan de wet houden.

Onthoud dus altijd: *Never trust, always verify!*



6. Conclusie

Een belangrijke conclusie uit deze visiepaper over AI en security is dat er sprake is van een balans: beide zijden van het spectrum gebruiken AI en zo is er sprake van een status quo. Je zou evenwel ook kunnen spreken van een wankel evenwicht. Want het is nog altijd zo dat verdedigers alles moeten zien tegen te houden, terwijl de aanvallers aan één treffer genoeg hebben. De balans slaat dus veelal uit in het voordeel van cyberaanvallers en daarom moeten we nog waakzamer zijn en de gevaren van AI in relatie tot cyberaanvallen op juiste waarde schatten.

Dat betekent vooral: wees je bewust van de risico's van AI. Omdat het speelveld enorm is gegroeid, moet je dit behapbaar zien te houden. Dat kun je doen door de risico's te benaderen vanuit de trits mens, technologie en proces. Op ieder niveau biedt Tesorion de juiste ondersteuning.

De **menselijke factor** wordt in securitytermen vaak gezien als de zwakste schakel. Mensen maken fouten, klikken op foute linkjes of laten een laptop in de trein achter. De overheid deed met campagnes haar best om de bevolking meer bewust te maken van de gevaren. Organisaties investeerden in het trainen van hun medewerkers en het vergroten van het cyberbewustzijn. Jaren geleden waren we al onder de indruk van de kwaliteit van social engineering. Inmiddels zien we de overtreffende trap in social engineering met deepfake audio en video. Het trainen van het bewustzijn in het kader van security zal nooit ophouden. Tegelijkertijd is de menselijke factor ook de sterkste schakel. Het bekende 'onderbuikgevoel' of intuitie, laat zich niet vangen in voorgeprogrammeerde logica of geautomatiseerde processen. In sommige van deze gevallen is de mens daardoor de enigste detectieregel die effectief is.

Een tweede **menselijke factor** betreft de experts, die continu systemen en netwerken in de gaten houden. Zij plukken de vruchten van securitytools die voorzien zijn van AI-capaciteit. Zoals we gezien hebben in deze paper, wordt het makkelijker het aantal false positives te reduceren, kunnen SOC-medewerkers zich richten op meldingen met een hoge prioriteit en krijgen analisten weer plezier in hun werk. En dat komt weer ten goede aan de digitale weerbaarheid van een organisatie.

Dat wordt mogelijk gemaakt door **technologie**. Leveranciers in ons werkveld doen er alles aan om oplossingen geavanceerder te maken met AI.
AI-based detectie is daar een goed voorbeeld van. Tesorion brengt allereerst met je organisatie in kaart op welke punten van de infrastructuur nieuwe AI-based securitytechnologie iets kan toevoegen. Kijkend naar de primaire processen, de key assets en de kwetsbaarheden. Doordat wij werken met verschillende vendoren, kunnen wij altijd de juiste oplossing aandragen en implementeren.

En over die **processen** gesproken, laat ons die dan ook meteen doorlichten. In hoeverre sluiten processen aan op wet- en regelgeving? Hoe is de awareness onder medewerkers en hoe kunnen we blijvend verbeteren? Zijn securityplannen en noodscenario's nog up to date en wordt hierin rekening gehouden met de kracht van AI? Benut je de kansen van AI, maar ook: ben je voldoende bewust van de gevaren van AI?

De impact van AI op de samenleving is nog lang niet op een hoogtepunt en die parallel kunnen we ook trekken voor de securityomgeving. Het is wellicht wat overweldigend allemaal en de dreigingen die op ons af komen zijn realistisch en soms ook angstaanjagend. Zorg ervoor dat dit niet verlammend gaat werken. Mobiliseer al je technologiepartners en benut de kracht van AI om digitaal nog sterker voor de dag te komen.



Voorbeeld van een AI gegenereerd beeld waarbij de vraag aan de beeldgenerator was:

"a photo of Connected ai as a defender of digital devices"

Wil je meer weten over wat Tesorion voor jouw bedrijf kan betekenen?

Wij maken graag vrijblijvend een afspraak met je om te kijken of we je kunnen helpen met een nog betere beveiliging of het vergroten van de awareness.

www.tesorion.nl

Informatiebeveiliging
ISO/IEC 27001
NEN 7510
www.dekra-uacl.com
DEKRA
Gecertificeerd

TESORION

Ravenswade 2A, 3439 LD Nieuwegein, Nederland | 088-274 78 80 | info@tesorion.nl | www.tesorion.nl