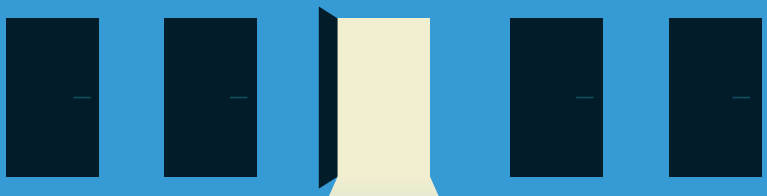
CHARLOTTE MEINDERSMA

AI MET BELEID



PRAKTISCH STUREN OP WETGEVING, ETHIEK EN DRAAGVLAK

AI MET BELEID

PRAKTISCH STUREN OP WETGEVING, ETHIEK EN DRAAGVLAK

CHARLOTTE MEINDERSMA

VANDUUREN MEDIA

ISBN: 978-94-6356-453-3
NUR: 801
Trefw.: management

Omslag: Terry Jonathans Design, Arnhem
Opmaak: Van Duuren Media, Culemborg/Barcelona
Druk: Veldhuis Media, Meppel

Eerste oplage: juni 2026

INHOUD

Inleiding ***11***

Definitie van AI *11*
Praktisch omgaan met AI *13*
Kansen, risico's, ethiek en juridische grenzen *14*

1 Wat als we AI niet meer herkennen? ***15***

AI is niet nieuw *16*
Transparantieverplichting over AI *17*
Nepnieuws *18*
Gevaar voor criminaliteit *22*
AI-bewijs *25*
AI-muziek *26*
Devaluatie van kennis en informatie *27*
Onbewust niet voldoen aan de wet *28*
Onbewust data delen *30*
Onterecht vertrouwen in de output *33*
AI-agenda *34*
Onbewuste bias *36*
Niet meer zelf afwegen of je AI wilt gebruiken *37*
Als we de AI-chatbot niet meer herkennen *38*
AI niet herkennen zorgt voor risico's *39*

2 Gemakzucht is het grootste gevaar ***41***

Verboden, maar wel makkelijk *41*
AI-antwoordmachine en expert *44*
Gevaarlijke AI-chatbots *54*
Verdwijnt de menselijk interactie? *57*
Maakt AI ons écht sneller, efficiënter of beter? *58*
Hoe AI ons werk beter maakt *61*
Deskilling *64*
Cognitive outsourcing *67*

Systeem 3: cognitieve overgave *69*
Wat doen we met de junioren? *70*
Brain fry en burn-out *72*
Handige AI-agents *74*
Ethische kant van generatieve AI *76*
Kosten van AI *77*
Gemak omarmen, gemakzucht voorkomen *80*

3 De AI-verordening als grensbewaker *81*

De kern van de AI-verordening *82*
Toezicht op naleving van de AI-verordening *101*
AI en AVG *103*
Digitale soevereiniteit *107*
AI en intellectueel eigendom *110*
Portretrecht, stemmen en deepfakes *115*
Portretrecht, stemmen en deepfakes *117*
Digital Services Act (DSA) *120*
Misleiding en reclame *122*
Wie is aansprakelijk als het fout gaat met AI? *124*

4 Een verstandig en handig AI-beleid *131*

Wie is verantwoordelijk voor het AI-beleid? *132*
Betrouwbare AI *133*
Doelstellingen bepalen *135*
Management of medewerkers? *136*
Selectie van de AI-systemen *138*
Is er een impact assessment nodig? *139*
Bepaal welke data gebruikt mag worden *142*
Wie, wat, hoe? *143*
AI-geletterdheid *144*
Naleving *145*
Klachten en suggesties *146*
Zorg voor overzicht *147*

Bronnen ***149***

Inleiding *149*
Hoofdstuk 1 *149*
Hoofdstuk 2 *150*
Hoofdstuk 3 *152*
Hoofdstuk 4 *154*

Over de auteur ***155***

Inleiding

Dat jij dit boek in je handen hebt, betekent dat je bovengemiddeld geïnteresseerd bent in AI. Dat is mooi, want AI biedt veel kansen, raakt steeds meer geïntegreerd in software en in ons leven en is precies om die redenen ook gevaarlijk. In dit boek neem ik je mee langs de filosofische en ethische dilemma's, enkele juridische grenzen en de oplossingen voor de praktijk.

Definitie van AI

Voor we met al die mooie dilemma's en oplossingen aan de slag gaan moeten jij en ik eerst op hetzelfde niveau zitten. Wanneer ik namelijk door LinkedIn scrol of een training geef, denken (te) veel mensen dat AI alleen maar generatieve AI is. Alsof er alleen tools als ChatGPT, Copilot en Gemini AI zijn. Tools waarmee we antwoorden genereren, plaatjes maken of die we code laten schrijven. Het is misschien hoe we AI het makkelijkste herkennen: er was eerst niets, we zetten deze tool in en magisch is er opeens iets. We begrijpen helemaal niets van hoé dat dan precies werkt, maar dat het werkt, dat kunnen we zíen.

Online en in de literatuur tref je veel verschillende definities van AI aan. Bijvoorbeeld dat het menselijke intelligentie na zou bootsen. Het hangt simpelweg af van iemands vak, uitgangspunt of AI-gebruik, welke defini-

tie iemand gebruikt. In dit boek probeer ik zo dicht mogelijk bij de juridische definitie van AI te blijven.

Volgens de Europese AI-verordening (*AI Act*) zijn er zeven componenten die er samen voor zorgen dat er sprake is van een AI-systeem:

- Het is een op een machine gebaseerd systeem. Dat kan hardware en/of software zijn. Het gaat hier immers niet om wat een mens doet, maar wat een computer doet.
- Dit systeem heeft een zekere mate van autonomie. Het systeem bepaalt geheel of gedeeltelijk zelf wat het doet en hoe het werkt.
- Het kan aanpassingsvermogen vertonen. Dit is geen vereiste om over AI te mogen spreken, maar het is wel een optie. Een systeem dat zelflerende capaciteiten heeft, valt dus eerder onder de noemer AI-systeem.
- Een AI-systeem kan een algemeen of (zeer) specifiek doel hebben waarvoor het is geprogrammeerd en waarvoor het wordt ingezet. Hoe dan ook moet het een doel hebben.
- AI heeft inferentievermogen. Het kan van input afleiden wat de output moet worden. Dit maakt een AI-systeem anders dan een algoritme. Het gaat hier ook om de manier waarop AI leert: soms meer door mensen aangestuurd tot volledig zelfstandige *deep learning*.
- Er zijn vier verschillende soorten output mogelijk: voorspellingen, inhoud, aanbevelingen of beslissingen.
- We spreken van AI-output wanneer deze van invloed kan zijn op fysieke of virtuele omgevingen. Het kan dus zijn dat een persoon of een computer beïnvloed wordt of reageert op AI-output.

AI is daarmee ook niet zomaar kunstmatige intelligentie die menselijke intelligentie probeert na te bootsen. Juist omdat AI altijd een computergestuurd systeem is, kan het eigenlijk alleen maar rekenen. Het kan alleen verbanden leggen die mathematisch kloppen. Denk maar aan generatieve AI, waaraan je zo goed merkt dat het niet exact de opdracht uitvoert, maar een kansberekening maakt. Daardoor krijg je een uitkomst waarvan de kans het grootste is dat die klopt, op basis van de data die het AI-systeem voor handen had. Maar dat klopt dus niet altijd. De menselijke laag die soms nodig is, zit er niet overheen.

Als je zou googelen, of het een AI-chat zou vragen, dan komen er verschillende definities van wat AI is naar boven. Het is goed om te weten dat in dit boek bovenstaande definitie uit de AI-verordening[i] gebruikt wordt, die overeenkomt met de definitie van de OECD (*Organisation for Economic Co-operation and Development*).

Praktisch omgaan met AI

AI is er en zal er blijven. Banen zullen veranderen en sommige zullen verdwijnen, zoals dat altijd al gegaan is. Volgens het World Economic Forum[ii] gaan er door AI 92 miljoen banen verdwijnen. Wees gerust, door AI gaan er juist ook 170 miljoen banen bij komen. Ook al verdwijnen er banen, werkeloos zullen we niet worden door AI. Het gaat ons (werkende) leven echter wel veranderen en daar moeten we ons op aanpassen. Maar ik kan je geruststellen: dat gaat veel minder dramatisch zijn dan velen je doen geloven. Die veranderingen gaan langzamer dan je misschien zou verwachten. Het zijn immers nog steeds de mensen die aan de knoppen zitten en beslissingen nemen. AI gaat zeker het een en ander veranderen en heeft dat ook al gedaan, maar dat gaat veel geleidelijker dan waar we nu soms bang voor worden gemaakt.

AI wordt meer en meer geïntegreerd in bestaande systemen en het wordt steeds makkelijker om AI te 'bedienen'. Waar je eerst nog vacatures zag voor prompt engineers, is kunnen prompten op een gegeven moment enerzijds veel minder belangrijk, omdat de tools daar zelf al in zullen begeleiden, en wordt het anderzijds een vaardigheid die we van iedereen verlangen. Net zoals we ook verwachten dat iedereen kan e-mailen en met Word kan omgaan. AI zullen we hierdoor ook steeds slechter herkennen.

Die integratie zorgt zowel voor kansen als gevaren. Wanneer mensen vaardiger met AI kunnen omgaan, zorgt het voor betere resultaten en verkleint het risico's. Maar zodra AI alomtegenwoordig is en/of we niet eens meer doorhebben dat we het aan het gebruiken zijn, worden de risico's groter.

Om risico's te verkleinen moet je vooral niet het gebruik van AI verbieden. Mensen zoeken namelijk altijd een omweg. Mensen zijn als water: het zoekt de makkelijkste weg met de minste weerstand. Als de weg daardoor langer wordt, is dat geen probleem. Het maakt de risico's echter wel groter.

Daarom is het van belang om praktisch en pragmatisch met AI om te gaan. Het liefst niet enkel van bovenaf opgelegd en al helemaal niet met als argument dat het 'van de wet moet', maar juist op een manier die past binnen de organisatie. Dat betekent dat we van medewerkers goed moeten weten hoe ze AI nu al gebruiken, wat ze ervan verwachten en wat hun wensen zijn. Beleid en communicatie moeten daar op aansluiten. Dat is de enige manier om ervoor te zorgen dat medewerkers überhaupt bereid zijn om AI op een correcte manier te gebruiken.

Kansen, risico's, ethiek en juridische grenzen

Om op de juiste manier met AI om te kunnen gaan, moet je eerst een aantal andere stappen nemen.

Het begint bij de kansen en mogelijkheden die AI biedt, de wensen vanuit klanten of medewerkers en de visie van de organisatie. Juist om het potentieel van AI zo goed mogelijk in te zetten, wil je eerst grenzeloos kunnen denken en brainstormen. Allereerst wil je juist alle opties open houden.

Daarna kun je voorzichtig afbakenen. Wat is onze eigen filosofie over AI? Hoe willen we omgaan met de impact op mens en milieu? Welke risico's zijn er eigenlijk en hoe willen we die mitigeren? Als allerlaatste kijk je vervolgens of er nog aanpassingen nodig zijn om ook binnen de juridische kaders te blijven.

Dit is precies de manier waarop ik je in dit boek wil meenemen. Ongemerkt maak je vervolgens een AI-beleid voor je organisatie, waarmee je AI op verantwoorde wijze gebruikt én je meteen voor een deel zorgt voor de AI-geletterdheid van je organisatie.

1

Wat als we AI niet meer herkennen?

Mensen die zeggen dat ze geen AI gebruiken moet je niet geloven. Misschien dat ze het niet actief en bewust zelf inzetten, misschien dat ze het zelfs liever niet zouden wíllen gebruiken, maar ze doen het wel.

Het eerste contact met de klantenservice van een gemiddeld groot bedrijf, spellingcorrectie en woorden aanvullen in WhatsApp, de suggesties in Spotify, YouTube of Netflix… het zijn allemaal interacties met AI. Dan hebben we het nog niet gehad over de integratie van Copilot in Microsoft 365 of AI-integraties in e-mailprogramma's. Zelfs op LinkedIn wordt gevraagd of je post geoptimaliseerd moet worden met AI en Marktplaats vraagt ook of ze met AI een tekst zullen suggereren of dat je het liever op de oude manier doet. Soms zie je dus niet eens meer dat je gebruikmaakt van AI, omdat het er niet bij vermeld wordt. Bij andere software staat het er misschien nog bij, zodat je het kunt herkennen. En dan hebben we natuurlijk nog de specifieke AI-tools of software die expliciet verkocht wordt vanwege AI-functionaliteit. AI is dan nog onderdeel van de marketing, maar dat gaat langzaam verdwijnen. AI gaat, net zoals bij WhatsApp en Spotify, volledig geïntegreerd raken.

AI is niet nieuw

Is het wel nodig dat we nog leren prompten en dat we om leren gaan met tools? AI voelt nieuw, omdat het met de komst van ChatGPT in 2022 opeens voor iedereen met een computer en een internetverbinding beschikbaar werd. Het idee van artificiële intelligentie werd echter al in 1955 door Alan Turing geïntroduceerd in *Computing Machinery and Intelligence*. Toen was het idee nog dat mensen data inzetten, analyseren en op basis daarvan beslissingen nemen, en eigenlijk is dat nog steeds de basis van wat we nu artificiële intelligentie noemen. In de jaren zeventig ontstonden de eerste neurale netwerken, die al meer lijken op hoe hersenen werken. Simpel gezegd zijn dat meerdere algoritmen die aan elkaar verbonden zijn en reageren op de feedback die ze krijgen, en zodoende ook zichzelf weer kunnen verbeteren.

AI wordt al vele jaren in veel software en systemen gebruikt, zonder dat het genoemd werd en dus zonder dat we het doorhadden. Denk aan vele functies in Photoshop, spraakherkenning en transcribeertools, spamfilters in e-mailprogramma's en allerlei chat- of berichtentools die onze zinnen vanzelf al aanvullen. Maar het zit bijvoorbeeld ook in software die data voor ons verwerkt, analyseert en beslissingen neemt of voorstellen doet. Weten we of dat AI is? Eigenlijk alleen maar wanneer dat in de marketing voor die software zo is uitgelegd of verkocht. Nu is dat nog nuttig, omdat mensen graag willen dat iets met AI werkt. Als het er niet meer bij vermeld wordt, zullen we meestal niet meer beseffen dat de oplossing die met de software wordt geboden, door AI is gerealiseerd. Het wordt allemaal normaal. De oplossing is immers belangrijker dan of het met behulp van AI tot stand komt of niet. Van de AI-verordening mag dat niet altijd. Mensen op wie AI wordt toegepast moeten daarvan altijd op de hoogte worden gebracht. Om aan de AI-verordening te voldoen moeten we in elk geval weten of in de software die we gebruiken AI verwerkt zit. En als we dat niet meer beseffen, dan kunnen we de risico's van het gebruik ook slechter inschatten én zullen we dus ook minder goed informeren over het gebruik van AI, wanneer dat invloed heeft op mensen.

AI niet meer herkennen heeft dus op veel partijen invloed: de organisatie, de medewerkers en de mensen die geraakt worden door AI-content of waarover door middel van AI beslissingen worden genomen.

Transparantieverplichting over AI

De AI-verordening schrijft voor dat wanneer we interactie hebben met AI-systemen, dat AI-systeem zichzelf kenbaar moet maken als AI. Dat moet natuurlijk bij AI-chatbots, maar bijvoorbeeld ook wanneer je digitaal solliciteert en bijvoorbeeld met een AI-tool een video moet opnemen of een assessment moet doen. Of misschien wil je wel eens kijken of je door te *vibecoden* een eigen website of app kunt bouwen. Ook dat AI-systeem moet wel even melden dat het een AI-systeem is. Is er een website die aanbiedt om documenten voor je te genereren? Als dat met AI gedaan wordt, moet dat erbij vermeld worden.

Voor AI-content zijn de eisen iets minder streng. Het AI-systeem waarmee de content wordt gemaakt of bewerkt moet wel machine-leesbaar, dus in de metadata, vermelden dat het met AI is gegenereerd of bewerkt. Het moet dus vooral voor andere systemen herkenbaar zijn, niet zozeer voor mensen. Gaat het echter om deepfakes, dus om audio- of visueel materiaal dat lijkt op iets dat in werkelijkheid bestaat of zou kunnen bestaan, dan moet het er wél bij gemeld worden. Als het gaat om bijvoorbeeld kunst of satire mag het echter het genot niet te veel belemmeren. Een kleine vermelding, zolang het nog maar leesbaar is, is dan voldoende.

In het geval van tekst hoeft er weer geen vermelding bij te staan, wanneer het gecontroleerd of geredigeerd is door een mens.

Er zijn dus wel wat verplichtingen om ervoor te zorgen dat we AI kunnen herkennen, maar dat zorgt er nog niet voor dat we altijd alle AI moeten kunnen herkennen. Daarnaast is het nu eenmaal zo dat echt niet iedereen zich aan deze regels houdt, bewust of onbewust, per ongeluk of expres.

Wat we bovendien weten over andere transparantieverplichtingen, zoals de cookiemelding, de privacyverklaring en de regels rondom transparantie over reclame, is dat we die mededelingen bewust overslaan of wegklikken of er zelfs blind voor worden. Op sociale media vallen de labels over samenwerkingen en de hashtags zoals AD, collab of sponsored ons niet meer op. Een goede juridische, morele verplichting. Meer kun je ook zeker niet doen. Maar dat betekent nog niet dat het net zo effectief is als we zouden willen.

Dat betekent dus dat we AI-systemen en AI-content op een gegeven moment niet of minder goed zullen herkennen en dat heeft maatschappelijke en commerciële gevolgen.

Nepnieuws

Nepnieuws, of eigenlijk desinformatie en misinformatie, is een steeds groter probleem. Deels door partijen, zoals politiek en sociale media die daar baat bij hebben (desinformatie), maar ook per ongeluk door mensen die bepaalde informatie zijn gaan geloven en dat verspreiden (misinformatie), zoals bepaalde complottheorieën.

Achteraf aantonen dat de informatie onjuist is, helpt maar een klein beetje. Mensen zijn al emotioneel betrokken bij de eerste informatie die ze ontvingen. Van andere feiten moeten ze dus vooral overtuigd worden. Dat is lastig wanneer er geen emotie bij gepaard gaat. Dat merk je ook wanneer mensen roddelkanalen volgen. Dat doen ze vooraf met het idee dat ze heus wel weten dat lang niet alles klopt, dat het een hoop speculatie is, maar toch blijft er wat van hangen en geloven we het in elk geval ten dele: waar rook is, is vuur.

De alomtegenwoordigheid van AI is van grote invloed op zowel het maken als verspreiden van nepnieuws. Sociale media gaan goed op berichten waar veel op wordt gereageerd: berichten die controversieel zijn of waar mensen boos om worden, waar lekker discussie over kan ontstaan. Mensen blijven dan maar reageren. Het nepnieuws tegengaan doe

je dan ook niet door een bericht te *factchecken* en met de feiten onder dat bericht te reageren. Daarmee verspreid je eigenlijk het nepnieuws alleen maar meer. Ook als je op een andere manier over dat onderwerp publiceert om het nepnieuws tegen te spreken, geef je juist aandacht aan dat nepnieuws. We zorgen dan zelf voor een soort Streisand-effect. Barbra Streisand trok met een rechtszaak aandacht naar een foto van haar huis, om te voorkomen dat mensen zouden weten welk huis van haar was. Vóór die rechtszaak hadden maar vier mensen die foto gezien én was het nog niet gekoppeld aan haar naam. Door die rechtszaak zagen vele duizenden mensen die foto en wisten dus ook meteen dat het om haar huis ging. Zo werkt dat ook met het proberen tegen te gaan van nepnieuws door er feiten tegenover te zetten. Door het delen van die feiten vraag je namelijk ook meteen aandacht voor de informatie die niet juist is. Liever verspreid je die informatie helemaal niet.

Daarnaast is het gemakkelijk om met AI nepnieuws te fabriceren. Bijvoorbeeld over de Gaza-oorlog of de oorlog tussen Iran en Israël. Juist wanneer het voor reguliere media lastig is om in een bepaald gebied verslag te doen, kijken we meer naar de informatie via andere bronnen, waarvan we de betrouwbaarheid niet altijd direct kunnen verifiëren. Soms lijken beelden bijvoorbeeld van ooggetuigen te komen. Ook media hebben moeite om echt van nep te onderscheiden. Op basis hiervan nemen mensen echter wel standpunten in.

Een rechtbanktekening van Aloys Oosterwijk (volgende pagina, bovenaan) werd door *Hart van Nederland* aangepast. Vermoedelijk zodat de tekening beter op de achtergrond, achter de presentatoren, zou passen. Aan de tekening werd een persoon toegevoegd, die er in werkelijkheid niet was. In dit geval niet alleen een inbreuk op auteursrecht, maar het zorgt ook voor een verkeerde voorstelling van zaken. Hier maakt dat misschien niet zo heel veel uit, maar het is wel degelijk nepnieuws. Als bijvoorbeeld relevant zou zijn hoeveel mensen er in de zaal aanwezig waren, dan doet het er wel toe dat er nog een extra persoon in beeld is.

De PVV Noord-Brabant ging nog iets verder. In eerste instantie zie je misschien vooral dat de achtergrondkleur is aangepast. Maar vooral de gezichtsuitdrukkingen zijn sterk veranderd. Dat geeft een totaal ander beeld over deze twee personen. Dat is wel problematisch nepnieuws. Lang heeft het niet online gestaan. Nadat de NOS[i] de video van de PVV met deze gemanipuleerde rechtbanktekening was opgevallen en erover had opgebeld, werd het binnen vijftien minuten offline gehaald.

Nu lijkt dit misschien allemaal evident en ben jij van mening dat je het wel herkend zou hebben. Of de gevolgen zijn niet zo groot, waardoor je het ook niet zo problematisch vindt. Maar we halen onze informatie steeds meer uit AI. "Even aan Chat vragen", hoor ik sommige mensen wel eens zeggen. ChatGPT of het AI-antwoord van Google vertrouwen we alsof het een normale, betrouwbare bron is. Die bron ga je niet meer controleren. Je hebt eerder bepaald of de specifieke bron in het algemeen betrouwbaar is. Onderzoeken die in wetenschappelijke tijdschriften zijn gepubliceerd, die vertrouwen we. Publicaties in vakbladen worden gecheckt door een redactie met kennis. Het lijkt erop alsof we AI-systemen ook op die manier aan het beoordelen zijn. Perplexity en Google vertrouwen we als betrouwbare bron, ChatGPT niet altijd. Maar zou je een AI-tool wel op die manier als bron moeten zien? Weet je nog, hoe we vroeger te horen kregen dat we niet zomaar op Wikipedia mochten vertrouwen? Juist omdat iedereen die pagina's kan maken en aanpassen. We vertrouwen wel literatuur, omdat de auteur al een autoriteit is en er kennelijk ook een uitgever is die erachter staat. Maar een AI-antwoord heeft niet dezelfde waarde. Die bouwt immers op basis van statistiek wat het antwoord moet zijn en begrijpt de inhoud niet daadwerkelijk. Je krijgt van AI ook niet elke keer hetzelfde antwoord, ondanks dat je wel elke keer exact dezelfde vraag stelt, met dezelfde formulering. Meerdere antwoorden op exact dezelfde vraag komen voor ongeveer 75 procent met elkaar overeen. Je moet een door AI gegenereerd antwoord dus goed controleren. Klik alsjeblieft op dat linkje bij het resultaat.

Volgens het AI-overzicht van Google, waar linkjes bij staan naar de bronnen, is het prima om baby's jonger dan zes maanden water te geven. Het is echter gevaarlijk om kinderen onder de zes maanden zomaar extra water te geven, omdat je dan het risico loopt dat ze te weinig voeding binnen krijgen, omdat hun maag al vol zit van het water.

Als je doorklinkt op de linkjes, blijkt daar ook iets anders te staan. De linkjes zelf geven wel een gevoel van betrouwbaarheid. Het suggereert dat de informatie van die specifieke pagina vandaan komt. Als die bron wel een betrouwbare is, dan klik je niet door, maar ga je ervan uit dat de tekst die je in het AI-zoekresultaat leest, inderdaad klopt.

Ja, het is over het algemeen prima om baby's bij warm weer extra water te geven, maar het is niet altijd nodig, vooral niet bij borstvoeding.

Voor baby's die borstvoeding krijgen:

- Extra water is meestal niet nodig, omdat moedermelk voldoende vocht bevat, zelfs bij warm weer.
- Je kunt je baby wel vaker aanleggen om te zorgen dat ze voldoende vocht binnenkrijgen.
- Zorg ervoor dat je als moeder zelf voldoende drinkt, aangezien de voeding van de baby hierdoor beïnvloed wordt.

Voor baby's die flesvoeding krijgen:

- Je kunt je baby extra water geven, maar doe dit met mate (paar lepeltjes per keer).
- Te veel extra water kan de eetlust van je baby verminderen, waardoor ze minder voedingsstoffen binnenkrijgen.
- Geef bij voorkeur water in plaats van andere dranken zoals vruchtensap, dat kan suiker bevatten.

Algemene tips voor warm weer:

- Zorg voor verkoeling, bijvoorbeeld door een lauw badje te geven of een nat washandje over het gezichtje en nekje te halen.
- Kleed je baby luchtig aan en vermijd direct zonlicht.
- Geef je baby geen koude dranken, maar lauwwarm water.
- Let goed op signalen van uitdroging, zoals een droge mond of verminderde plasluiers.
- Raadpleeg bij twijfel altijd een arts of consultatiebureau.

Nepnieuws kan gemakkelijk je organisatie binnensluipen. Het is laagdrempelig om AI om een antwoord te vragen. Zeker als er snel een antwoord nodig is, iemand het werk saai vindt of de druk hoog is. Als er tools voor worden ingezet die zichzelf presenteren als AI, dan is er nog een bepaald bewustzijn dat de antwoorden misschien niet kloppen. Bij Google is dat al lastiger. Dat kennen we al lang, daar hebben we al een bepaald gevoel bij, niet iedereen herkent even goed dat het antwoord bovenaan door AI is gegenereerd, juist omdat er wel een bron bij vermeld staat. Daardoor komt er onbewust onjuiste informatie de organisatie binnen. Het is belangrijk om medewerkers hiervan bewust te maken.

Gevaar voor criminaliteit

Ook criminelen gebruiken AI graag. Het kost weinig en kan veel. Zoals het bouwen van malware of het nabootsen van stemmen. Dit betekent dat medewerkers extra goed getraind moeten worden om grote problemen te voorkomen.

De hack bij Odido kwam door een phishingmail. De hackers deden zich voor als de ICT-afdeling. Dit is gelukt omdat er mensen waren die kennelijk niet verwacht hadden dat hackers zich als collega's van ICT zouden voordoen. En een e-mailadres of telefoonnummer kan natuurlijk *gespooft* worden. Nu zou je kunnen zeggen dat het misschien ook wel stom was om hierin te trappen. Anderzijds zijn de systemen kennelijk dermate slecht beveiligd, dat één mailtje genoeg is om aan de juiste informatie te komen.

Veel gevaarlijker gaat het zijn als dit soort zaken niet meer op deze manier per mail komen, maar het mis gaat omdat mensen denken in te loggen in bepaalde software, bijvoorbeeld. Of omdat ze denken in een videocall te zitten met een collega, terwijl dat eigenlijk een AI-deepfake blijkt te zijn.

Dit zijn risico's die echt niet alleen onze ouders of grootouders treffen, die wat minder digitaal vaardig zijn. Juist jongere generaties die met internet zijn opgegroeid en heel vertrouwd zijn met computers en smartphones, schatten risico's laag in. Bovendien werd veel software steeds gemakkelijker, waardoor ze niet allemaal goed begrijpen hoe het echt werkt of zelf problemen kunnen oplossen als de software ze daar niet in begeleidt. Het is een illusie te denken dat zij gevaren beter herkennen.

Met AI wordt ook CEO-fraude veel makkelijker. De crimineel doet zich voor als CEO of een manager; iemand van wie je absoluut opdrachten aanneemt en uitvoert, *no questions asked.* Vaak is het een e-mail waarbij gevraagd wordt om met spoed iets uit te voeren en het vooral discreet te doen omdat het bijvoorbeeld om bedrijfsgevoelige informatie gaat. Meestal moet er daarom geld worden overgemaakt naar een bankrekeningnummer dat niet bekend is bij de financiële administratie.

In 2018 verloor bioscoopketen Pathé op deze manier 19 miljoen euro. De algemeen directeur en de CFO werden per mail benaderd, zogenaamd door de CEO van het Franse moederbedrijf met het verzoek om geld over te maken ten behoeve van de financiering van een bedrijf in Dubai. Als alles in Dubai rond zou zijn, zou het geld weer teruggestort

worden. De algemeen directeur en de CFO voerden samen overleg, vonden het 'wonderlijk' maar voldeden toch aan het verzoek. Ze kwamen er pas achter dat ze opgelicht waren toen ze vanwege een ander onderwerp contact hadden met de Franse tak van het bedrijf. De algemeen directeur en de CFO zijn ontslagen. De CFO heeft dat nog aangevochten, maar verloren.[ii] Ontslag op staande voet had niet gemogen, maar ontbinding van de arbeidsovereenkomst wegens verwijtbaar handelen wel.

Met Mythos, een AI-model van Anthropic, is het mogelijk om gemakkelijk en snel beveiligingslekken op te sporen. Het model kan niet alleen losse kwetsbaarheden vinden, maar ze in samenhang benutten om complete aanvalsketens te construeren. Op het moment van publicatie van dit boek heeft nog maar een beperkt aantal bedrijven toegang tot het model. Op het moment dat het echter in verkeerde handen komt, zijn de gevolgen groot. Bedrijven moeten hierop voorbereid zijn. De reactiesnelheid bij het ontdekken van een lek moet omhoog. AI wacht immers ook niet.

Het Franse AI-bedrijf Mistral is ook bezig met het ontwikkelen van AI om beveiligingslekken op te sporen. Zij werken nu al samen met enkele Europese banken om hen te helpen op die manier lekken sneller te vinden en te dichten.

Het is dus belangrijk om technische waarborgen te treffen en wat drempels op te werpen. Niet iedereen hoeft bij elk systeem of bij elk document te kunnen. Denk praktisch na over wat verstandig is om af te schermen.

Maak onderling ook goede afspraken en zorg voor voldoende kennis, training en instructie. Niet als moetje, zodat er iets van een lijstje afgestreept kan worden, maar juist om te voorkomen dat mensen iets doen omdat ze niet beter weten. Maak duidelijke afspraken over wat je wel of niet per e-mail of in een videocall deelt. Voor sommige zaken moet misschien even de moeite genomen worden om bij elkaar langs te lopen. Of als iedereen remote werkt, hanteer dan een vierogenprincipe en maak afspraken over hoe daarmee omgegaan wordt. De manier waarop dat wordt uitgevoerd maakt immers ook nog uit voor hoe makkelijk een derde dat kan manipuleren.

AI-bewijs

Photo or it didn't happen. Maar hoe doe je dat, als je die foto's of video's met AI kunt manipuleren?

De verzekering oplichten met AI-beeld is zo gebeurd. Zij willen tenslotte meestal foto's en/of een video als bewijs. Het specifieke product of de specifieke situatie hoeft niet live door een persoon bekeken te worden. Digitaal bewijsmateriaal is meestal voldoende. Met behulp van AI kun je dat echter allemaal nabootsen. Een live videocall dan? In beginsel is dat ook geen oplossing, omdat je ook dan met AI het product of de situatie kunt manipuleren. De oplossing zou dan moeten zijn bijvoorbeeld dat zo'n verzekering eigen software heeft die per se gebruikt moet worden, waardoor er bij een live verbinding misschien toch geen gemanipuleerd beeld door kan komen of dat het anders in elk geval beter gedetecteerd zou kunnen worden.

Ook webshops vragen soms liever om digitaal bewijsmateriaal in de zin van foto's en video's dan dat ze een product terug moeten laten sturen. Als het product inderdaad stuk is, komen immers de kosten van de retourzending voor rekening van de webshop. Ze laten dus liever niet iets onnodig opsturen. Is een product er zo slecht aan toe dat het bij ontvangst toch linea recta de vuilnisbak in moet, dan kiezen webshops er geregeld voor het product niet meer terug te laten sturen. Er zullen natuurlijk van die gehaaide mensen zijn die weten welke webshop welk beleid heeft en daar gebruik van maken. Voorheen wellicht met foto's die dan op internet gevonden zijn, nu door foto's en video's met een enkele druk op de knop te genereren.

Kon je eerst aan dat soort beeldmateriaal nog wel zien dat het AI was, omdat er vaak toch wat foutjes in zaten, dat is nu verleden tijd. Het beeldmateriaal wordt steeds beter. Als er voldoende moeite wordt gedaan om goed beeldmateriaal te produceren, dan kun je het met het blote oog niet meer zien. En als we nu zo ver nog niet zouden zijn, dan wel over een paar maanden.