Henk-Jan van der Molen is onder andere docent bij de Security Academy voor CISSP-, CISM- en CISA-modules en is werkzaam bij het SECO Institute voor het ontwikkelen van examens. Henk-Jan is bereikbaar via: henk-jan.van.der.molen@securityacademy.org.

Hoe betrouwbaar zijn uw big data en AI-systemen?

Informatiebeveiliging voor big data en AI-systemen is belangrijk om de informatiekwaliteit te garanderen, maar ook om big data en AI-systemen te beschermen tegen cybercriminelen. Dit artikel beschrijft de beveiligingsmaatregelen die relevant zijn in deze context. Deze maatregelen zijn niet compleet of voor elke situatie geschikt. Momenteel worden zowel voor big data als AI-systemen nog normen ontwikkeld. Voor een adequate beveiliging is daarom voorlopig maatwerk nodig op basis van een risicoanalyse. Het blijft een menselijke afweging waar en onder welke voorwaarden AI-systemen kunnen worden ingezet.

If you know the enemy and know yourself,
you need not fear the result of a hundred battles.
If you know yourself but not the enemy,
for every victory gained you will also suffer a defeat.
If you know neither the enemy nor yourself,
you will succumb in every battle

Sun Tzu, The Art of War

IT is onmisbaar geworden in onze samenleving. Ook al nadert de zogenaamde Wet van Moore (1) zijn houdbaarheidsgrens, op korte termijn groeit de rekenkracht van computers nog. Voor bedrijven is succesvolle innovatie nodig om te blijven bestaan. Klanten verwachten namelijk steeds meer 'non stop service' en een persoonlijke benadering van organisaties. Betere service vergt meer digitalisering en meer informatie-uitwisseling tussen processen. Ook door de toename van IoT-apparatuur groeit de beschikbare data. De meeste data die nu beschikbaar is, is in de laatste jaren gegenereerd (2). Surfend op die vloedgolf van data worden big data en AI-systemen ontwikkeld die steeds meer sturende taken overnemen.

Met de digitalisering van processen neemt de druk op de beveiliging toe en intussen heeft de AVG de privacyregels aangescherpt. Zo zijn privacy by design en privacy by default verplicht en moeten profileringsalgoritmes transparant zijn. Naast opzet en bestaan vereist de AVG ook dat de effectieve werking van beveiligingsmaatregelen aantoonbaar is. Omdat er voor beveiliging geen onbegrensd budget is, moeten organisaties de prioriteit leggen bij de beveiliging van de belangrijkste bedrijfsprocessen.

Beveiliging focust vaak op geheimhouding, maar stuurinformatie moet vooral betrouwbaar zijn. Het artikel start daarom met vier architectuurprincipes die worden uitgewerkt in generieke maatregelen om betrouwbare informatie te produceren. Daarna worden de ontwikkelingen van het big data en AI-vakgebied samengevat. Op basis van verschillen met traditionele IT-systemen sluit het artikel af met extra beveiligingsmaatregelen om big data en AI-systemen operationeel te houden en verder te kunnen ontwikkelen.

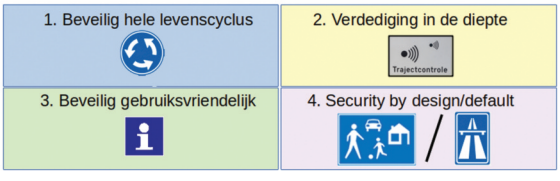
Architectuur en informatiebeveiliging

Informatiebeveiliging (IB) bestaat uit drie componenten. Allereerst is dat 'vertrouwelijkheid': alleen de doelgroep heeft toegang tot informatie. Dit wordt vaak ingevuld met encryptie, waarbij alleen de doelgroep de cryptomiddelen heeft om vercijferde informatie te gebruiken. Het tweede component is 'beschikbaarheid': informatie is beschikbaar als de organisatie die nodig heeft. Dit vergt bijvoorbeeld back-upvoorzieningen. Tenslotte geeft 'integriteit' aan dat de data of informatie authentiek, juist, volledig en actueel is. In dit artikel is 'data' de input van informatiesystemen en 'informatie' de output. Kwaliteit en betrouwbaarheid worden als synoniem gebruikt voor integriteit van informatie.

Maatregelen voor integriteit liggen minder voor de hand. De Gebruikers Acceptatie Test (GAT) toetst bijvoorbeeld of een nieuw informatiesysteem de gewenste informatie produceert, maar standaard negeert een GAT vragen over informatiekwaliteit. Denk hierbij aan vragen als:

- Kunnen alleen de juiste personen data invoeren en muteren?
- Zijn de informatievoorzieningsprocessen voldoende transparant?
- Voldoet de beveiliging van het informatiesysteem?
- Welke kwaliteit heeft de brondata en de daarmee geproduceerde informatie?

Een architectuur bevat coherente en consistente regels, richtlijnen en standaards die samenhangende componenten als geheel beschrijven (3). Een IT-architectuur beschrijft sturingsprincipes voor de ontwikkeling en het gebruik van de informatievoorziening. Vier architectuurprincipes worden hierna uitgewerkt in generieke maatregelen om betrouwbare informatie te produceren (zie figuur 1).

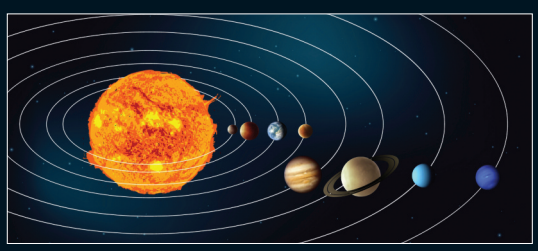


Figuur 1 - Vier architectuurprincipes voor informatiekwaliteit.

Principe 1: bescherm informatie van creatie tot vernietiging

Het beleid moet relevante eisen in wet- en regelgeving signaleren en bij projecten voorschrijven dat beveiligingseisen zo vroeg mogelijk worden vastgesteld. Na het vaststellen van de gewenste betrouwbaarheidsniveaus ligt de focus op governance: wie kan namens de eigenaar data invoeren, lezen, muteren of vernietigen op basis van need to know/least privilege? De procedures voor HR, systeembeheer, configuratiebeheer, wijzigingsbeheer en beveiligingseisen van transacties bepalen welke rechten aan verantwoordelijkheden worden gekoppeld. Voor de veiligheid bij aanschaf en inzet van IT-middelen is beleid nodig, gericht op leveranciersmanagement, telewerken en gebruik van bedrijfsmiddelen.

Om data te kunnen vinden en hergebruik te stimuleren is metadata nodig die de inhoud en context van data beschrijft op basis van de FAIR-principes: 'Findability', 'Accessibility', 'Interoperability' en 'Reusability'. Master Data Management en de WBP hebben deze principes eerder geïntroduceerd. Datamanagement start met een unieke identificatie van data-elementen, datagroepen en relaties daartussen, zodat metadata eenduidig kan worden geregistreerd in een data dictionary. Dit is ook relevant voor ketens, omdat eenduidige afspraken nodig zijn om de impact van foute informatie te beperken. Een ketenorganisatie heeft namelijk maar beperkt zicht op de informatiebehoefte en de risico's voor de processen van andere ketenpartijen. Meer eisen voor metadata staan bijvoorbeeld in de ISO 11179. Het beheer van metadata is een verantwoordelijkheid van de chief data officer.



Casus: wetenschap heeft betrouwbare data nodig

De wetenschap onderkende al vroeg het belang van betrouwbare gegevens. Aristarchos van Samos beweerde al in 265 v.Chr. dat de planeten om de zon draaien, maar de geocentrische visie van Aristoteles hield stand. Copernicus herintroduceerde het heliocentrische model in 1543, maar zijn wiskundige theorie was strijdig met de toenmalige uitgangspunten van de wetenschap.

Pas in de 17e eeuw begon men de Bijbel minder letterlijk te interpreteren en ontwikkelde de wetenschap zich tijdens de Verlichting. Johannes Kepler formuleerde in 1609 zijn bewegingswetten voor hemellichamen. Na de uitvinding van zijn telescoop in 1610 kon Galileo Galilei veel nauwkeurige waarnemingen doen die het heliocentrische model bevestigden. De Inquisitie dwong Galileo echter het heliocentrische model af te zweren, omdat dit zou conflicteren met de Bijbel.

Als contrast: in februari 1897 legde het Huis van Afgevaardigden van de Amerikaanse staat Indiana in een wetsontwerp de rekenwaarde van π met bijna 2% afwijking vast als 3,2. Deze 'Pi-wet' kwam op het laatste moment niet door de Senaat en werd dus niet van kracht.

Encryptie kan het onbevoegd lezen en muteren van informatie voorkomen en met PKI-certificaten kunnen systemen of personen elkaar authentiseren. De datakwaliteit kan ook worden bewaakt met controletellingen, hashing en elektronische handtekeningen. Effectieve controles op informatiekwaliteit moeten zo dicht mogelijk bij de productie worden ingericht en het effect van afwijkingen nauwkeurig kunnen inschatten. Gedurende de levensduur van informatie kunnen beveiligingsincidenten de kwaliteit ervan verminderen. De organisatie moet kwetsbaarheden en incidenten daarom tijdig signaleren, afhandelen en evalueren. Dit vergt allereerst een goede preventie, zoals firewalls, regelmatig testen van back-up- en restoreprocedures en het vastleggen van forensische informatie. Omdat reguliere antivirus steeds ineffectiever wordt, hebben grote organisaties een security operations center (SOC) met een systeem voor incident en event management (SIEM) om real time abnormaal netwerkverkeer te kunnen signaleren. Vervolgens is een procedure nodig om gecontroleerd van incident naar crisismanagement te kunnen gaan, zijn er periodieke oefeningen nodig om medewerkers risicobewust te maken en is het belangrijk om te trainen als crisisteam.

Om risico's goed te managen moet de PDCA-cyclus functioneren, zodat de organisatie consequent zoekt naar conformi-

teit met geldende normen en leert van geconstateerde afwijkingen. Omdat een organisatie niet alle incidenten kan detecteren, moeten controles regelmatig worden ingepland en alle significante risico's afdekken. Organisaties moeten ook bewaken dat de juiste controles op het juiste moment worden uitgevoerd.

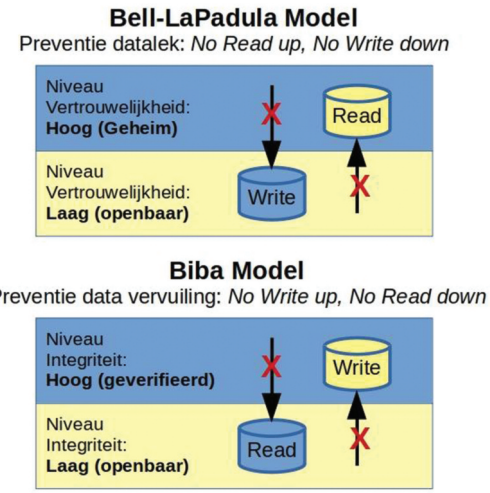
Organisaties die incidenten, klachten en auditbevindingen als startpunt willen gebruiken voor continue verbetering, kunnen daarvoor de 4xO-methode toepassen:

O1: wat is de omvang van het issue of waar is correctie nodig?
O2: wat is de oorzaak (Engels: root cause analysis) of wat bepaalt de kans op herhaling?
O3: welke oplossing met correctieve en preventieve acties is proportioneel met het risico?
O4: hoe en wanneer testen we of deze oplossing operationeel voldoet?

Principe 2: verdediging in de diepte

Voor staatsgeheimen bestaan meerdere vertrouwelijkheidsniveaus. Beveiliging wordt daarom vaak in lagen gevisualiseerd: elke laag bevat specifieke maatregelen. Voor informatiekwaliteit zijn meerdere niveaus ongebruikelijk, maar soms is wel bekend dat data vervuild is en moet worden opgeschoond.

Voor hogere niveaus van informatiekwaliteit zijn meer en zwaardere maatregelen nodig. Relevante maatregelen zijn onder andere fysieke en logische toegangscontroles, screening van medewerkers, wachtwoordbeleid, controles op informatiekwaliteit, data governance, controletechnische functiescheiding, logging van mutaties, audits op de naleving van normen en audits op de effectiviteit van maatregelen. Naast de herkomst, eigenaar en betekenis, kan metadata het betrouwbaarheidsniveau voor datakwaliteit aangeven, inclusief de houdbaarheid in plaats en tijd.



Figuur 2 - Bell-LaPadula-model versus Biba-model.

Beveiligingsmodellen voor logische controles zijn bijvoorbeeld Bell-LaPadula (voor vertrouwelijkheid) en Biba (4) (voor integriteit van informatie), zie figuur 2. Beide modellen gaan uit van centrale controle en zijn ouder dan het internet, maar zijn nog steeds waardevol. Het beveiligingsbeleid bijvoorbeeld geeft alleen aan wat een rol mag doen, terwijl een beveiligingsmodel verklaart waarom bepaalde acties wel of niet zijn toegestaan.

Het Bell-LaPadula-model voorkomt datalekken met het credo: 'no read up, no write down'. Dat wil bijvoorbeeld zeggen dat niemand informatie kan lezen die hoger gelabeld is dan zijn screening. Daarnaast kan iemand met de juiste screening geen geheime informatie kopiëren naar een locatie met een lager vertrouwelijkheidsniveau.

Het Biba-model voorkomt gegevensvervuiling met het credo: 'no write up, no read down'. Iemand met toegang tot laagwaardige informatie kan die niet kopiëren naar locaties voor hoogwaardige informatie. Tegelijkertijd mag een uitvraag voor hoogwaardige informatie geen laagwaardige data bevatten.

Principe 3: gebruiksvriendelijke beveiliging

Bekijk het systeem als gebruiker. In de ideale situatie is onveilig handelen moeilijker dan veilig handelen. Gebruikers moeten voldoende opleiding hebben gehad, beschikken over begrijpelijke documentatie en een helpdesk kunnen benaderen met vragen en klachten. De metadata van geproduceerde informatie moet worden vastgesteld op basis van de doelen en belangen van de doelgroep. De processen van de informatievoorziening moeten transparant zijn, zodat gebruikers kwetsbaarheden in de logica beter kunnen signaleren en kunnen nagaan hoe betrouwbaar hun informatie is. Om de complexiteit te verminderen, moeten informatiesystemen modulair zijn. Informatiesystemen kunnen ook functionaliteit bieden waarmee gebruikers zelf eenvoudig afwijkingen in brondata kunnen signaleren.

Anderzijds vormen de belangen van medewerkers voor (on)betrouwbare informatie een uitgangspunt voor controles op de brongegevens en geproduceerde informatie. Op

basis van risicoscenario's en pre mortem-analyses kunnen audits worden gepland en uitgevoerd.

Principe 4: security by design en security by default

Beveiligingsmaatregelen moeten het risico van incidenten zoveel mogelijk beperken. Dat begint bij dataminimalisatie: wat je aan gevoelige data niet (meer) nodig hebt, hoef je ook niet te beveiligen. Dit lijkt in tegenspraak met big data, maar relevante maatregelen hiervoor zijn vastgestelde bewaartermijnen, logische toegangsrechten op basis van 'need to know' en encryptie van data (5).

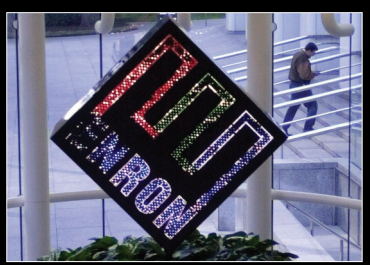
Om vanuit de privacy risico's voor betrokkenen de risicobeperkende maatregelen te bepalen is een Privacy Impact Analyse (PIA) nodig (6). Voorbeelden van privacymaatregelen zijn pseudonimiseren van data, het borgen van k-anonimiteit (voorkomt heridentificatie) en het recht op inzage, correctie en regie op het gebruik van eigen data. Een PIA beoordeelt ook de rechtmatigheid van gegevensverwerkingen.

Behalve de juiste brondata in het afgesproken formaat is ook een goed proces nodig om daar betrouwbare informatie van te maken. Als dit principe niet goed is ingericht, kunnen fouten in data of algoritmes incidenten veroorzaken. Bijvoorbeeld de Mars Climate Orbiter van $ 135 miljoen crashte in 1999, omdat één stuurcomputer communiceerde in 'imperial units' (pound) en andere in SI-eenheden (kilo) (7). Daarom gebruikt elke Java Virtual Machine dezelfde rekenfuncties en gegevensformaten, zodat een Java-programma op elk platform dezelfde uitkomsten produceert.

Om veilige software te ontwikkelen bestaan methodes zoals Secure Software Development (8). Met open source is de kwaliteit van software, inclusief de controles op informatiekwaliteit, zichtbaar en zijn die controles eventueel ook aan te passen. Daarnaast is sjoemelen met software een stuk lastiger met open source.

Van big data naar AI-systemen

De ontwikkeling van het AI-vakgebied kende verschillende stadia. Eind jaren tachtig werd business intelligence ingezet om de aandelenkoersen te voorspellen. Rond 1995 werd data mining toegepast voor direct marketing, fraudedetectie en kredietprofilering. Na de opkomst van het internet werden e-commerce-toepassingen ontwikkeld om automatisch gepersonaliseerde diensten te leveren. In de biochemie en astronomie werden intussen met succes steeds grotere modellen ontwikkeld en in 2011 werd big data in bijna elk vakgebied toegepast.



Casus: manipulatie van beursgevoelige informatie

Enron was een Amerikaans energiebedrijf dat in 1980 vooral gas leverde. Begin jaren negentig begon Enron echter miljarden te investeren in andere producten en diensten zoals water, telecommunicatie, metaal, chemie, internet en verzekeringen. Enron werd hét voorbeeld voor iedereen die in de nieuwe economie snel veel geld wilde verdienen.

Maar Enron gaf meer uit dan er binnenkwam en hield verliezen met schimmige constructies uit de boeken. De Amerikaanse financiële controlecommissie (SEC) stelde vast dat Enron zijn winsten jarenlang fors had overdreven en 20 miljard dollar rood stond. Ook bleek dat managers miljoenen hadden gestolen en honderden dochterondernemingen hadden opgezet om belasting te ontduiken. In december 2001 was de energiereus failliet en stonden 21.000 werknemers op straat. In januari 2002 startte een strafrechtelijk onderzoek naar Enron, nam de directie ontslag en pleegde een ex-topman zelfmoord. Dit schandaal genereerde wereldwijd veel aandacht voor corporate governance en zorgde indirect voor verscherpte wetgeving op dit vlak zoals Sarbanes-Oxley, Basel II en de Nederlandse Code Tabaksblat.

Arthur Andersen was een van de oudste en meest gerenommeerde accountantskantoren van de Verenigde Staten, maar keurde als accountant van Enron de jaarrekening goed, terwijl dat overduidelijk onterecht was. Toen de Amerikaanse financiële toezichthouder de administratie opvroeg, bleek dat Arthur Andersen veel documenten had vernietigd. Door het Enron-schandaal bestaat Arthur Andersen niet meer.

Met name in het speldomein verliepen de AI-ontwikkelingen stormachtig. In maart 1997 won IBM's supercomputer Deep Blue van de regerend wereldkampioen schaken, Gary Kasparov. In januari 2016 versloeg het AlphaGo Zero-computersysteem de wereldkampioen Go (Lee Sedol), terwijl Go vele malen complexer is dan schaken (9). Volgens de Wet van Moore groeide de verwerkingscapaciteit van computers over de periode 1997 tot 2016 ongeveer met een factor 6000. AlphaGo Zero werkt dan ook fundamenteel anders en leert door tegen zichzelf te spelen (10). Menselijke Go-spelers bereikten het huidige spelniveau in ongeveer 3000 jaar, AlphaGo Zero had maar 3 dagen nodig om daar boven uit te stijgen. In juli 2019 won het Pluribus AI-systeem voor het eerst van de beste menselijke pokerspelers (11). Voor AI is de uitdaging van poker de combinatie van toeval, tactiek en onvolledige informatie. Ook Pluribus leerde poker door 8 dagen tegen zichzelf te spelen. Leren van data uit spelletjes is effectief, omdat het beste algoritme simpel te bepalen is (hint: die wint het meeste).

Machine learning is succesvol in het speldomein, maar daarbuiten is het trainen van AI-systemen veel moeilijker. Behalve het verzamelen van grote hoeveelheden betrouwbare data, bepaalt het domein namelijk mede het succes van de AI-toepassing. Voor bijvoorbeeld weersverwachtingen en aardbevingen gelden de fundamentele onzekerheden van de Chaos-theorie, waardoor heel nauwkeurige voorspellingen onmogelijk zijn. Omdat mensen soms niet-rationele besluiten nemen, geldt dat ook voor economische voorspellingen. Momenteel kunnen de vijf bestaande AI-vakgebieden alleen hun eigen, specifieke problemen oplossen (12).

- Symbolisten gebruiken beslissingsbomen en inductieve logica om te bepalen welke regels het beste de waargenomen data beschrijven en welke nieuwe informatie voorspellingen kan verbeteren. Het algoritme is helder, maar is gevoelig voor overfitting (uitzonderingen in trainingsdata), lokale optima en (nog onbekende) uitzonderingen in de praktijk.
- Connectionisten maken een netwerk van elektronische neuronen om een kunstmatig brein te bouwen. Zo'n netwerk kan leren kankercellen te herkennen op foto's, maar waarom bepaalde kankercellen wel of niet worden herkend (het algoritme) is onduidelijk.
- Genetische algoritmes simuleren evolutie, waarbij oplossingen worden gecodeerd als genen. In de volgende generatie worden de beste genen met willekeurige mutaties gecombineerd, om hopelijk betere oplossingen te krijgen.
- Het Theorema van Bayes geeft aan hoe nieuwe waarnemingen (data) de waarschijnlijkheid beïnvloeden van een oorzaak (hypothese). De uitslagen van (objectieve) medische testen veranderen bijvoorbeeld de geschatte (subjectieve) kans dat een bepaalde ziekte de oorzaak is van de waargenomen symptomen.
- Analogisten bouwen 'support-vector machines' die voor objecten efficiënt analogieën kunnen afleiden uit gewogen informatie van vergelijkbare objecten. Netflix bijvoorbeeld attendeert Alice op een film als Bob die film goed vond en van dezelfde films houdt als Alice. Het probleem is: welke subset van 100.000+ eigenschappen is relevant?

Casus: neuraal netwerk produceert nepnieuws

Menselijke invoer: 'Russia has declared war on the United States after Donald Trump accidentally ...'
AI: '... fired a missile in the air. Russia said it had 'identified the missile's trajectory and will take necessary measures to ensure the security of the Russian population and the country's strategic nuclear forces.' The White House said it was 'extremely concerned by the Russian violation' of a treaty banning intermediate-range ballistic missiles. The US and Russia have had an uneasy relationship since 2014, when Moscow annexed Ukraine's Crimea region and backed separatists in eastern Ukraine.'

De OpenAI Research Group California ontwikkelde voor neurale netwerken het zelflerend algoritme GPT-2 dat het volgende woord voorspelt in teksten (13). Ook na training met 40 GB aan internetteksten produceert GPT-2 nog veel semantische onzin, maar begin 2019 kwam er ook een hoogwaardig nepnieuwtje (zie casus). Nepnieuws is politiek belangrijk zoals de Brexit, Trump en de MH17-ramp laten zien. Door meerdere 'alternatieve

waarheden' te publiceren en onderzoeksjournalistiek te bestrijden wordt de aandacht voor de objectieve waarheid verdund. Met de huidige AI-systemen is het bestrijden van nepnieuws moeilijk (14). GPT-2 komt uit de stal van Elon Musk en is oorspronkelijk tegengehouden vanwege dit nepnieuws. In 02.2019 alsnog vrijgegeven.

Buiten het speldomein zullen AI-toepassingen zich waarschijnlijk geleidelijk ontwikkelen door het proces van probleemoplossing steeds meer te automatiseren. Vroeger was het ontwikkelen van wetenschappelijke theorieën mensenwerk en kostte bijvoorbeeld het handmatig uitrekenen van ingewikkelde formules veel tijd. Tegenwoordig ondersteunt software het bewijzen van hypothesen steeds beter. Wetenschappers richten zich daarom steeds meer op het zodanig formuleren van problemen dat software maximaal kan ondersteunen bij de oplossing ervan. De heilige graal is één AI-systeem te ontwikkelen dat alle problemen acceptabel kan oplossen, maar dat kost waarschijnlijk nog jaren. Grovers kwantumalgoritme om efficiënt te zoeken in databases kan big data en AI-ontwikkelingen versterken. Maar ondanks Nevens wet (de Wet van Moore op anabolen) vergt de ontwikkeling van voldoende krachtige kwantumcomputers meer tijd (15).

Beveiliging van big data en AI

Naast de eerder genoemde generieke maatregelen voor informatiekwaliteit hebben big data en AI-systemen specifieke beveiliging nodig. Zoals bekend zijn big data en AI-systemen erg data-intensief. Wie de beste en meeste trainingsdata heeft, kan de beste big data en AI-systemen ontwikkelen. Omdat AI-systemen zowel data gebruiken voor de training en de werking, zijn ze extra gevoelig voor laagwaardige data (16).

Om de kwaliteit en 'scheefheid' van gebruikte datasets objectief te meten, is naast normen zoals de ISO8000 een bewezen objectieve dataset als 'waterpas' referentie nodig. Voor commerciële bedrijven die innovaties ontwikkelen voor verschillende doelgroepen is dat laatste onpraktisch, vooral als concurrenten meekijken. De datakwaliteit bepaalt namelijk hoe vertrouwelijk big data moet blijven voor concurrenten. Toch moeten bedrijven de informatiekwaliteit beheersen, omdat blindvaren op scheve datasets leidt tot discutabele resultaten (17). De bijsluiter van geproduceerde informatie moet consistent de datakwaliteit vermelden.

AI-systemen worden ook voor cyberaanvallen ingezet. In 2008 ontwikkelde de Carnegie Mellon universiteit een proof of concept-systeem dat automatisch malware kon genereren uit een software-update die buffer overflows repareert (18). In 2009 werd circa 60 procent van de gemelde kwetsbaarheden in software actief misbruikt (19). In 2011 verslechterde deze situatie: ongeveer 97 procent van alle kwetsbaarheden werd binnen twee weken na publicatie misbruikt (20). In januari 2019 meldde het Australische Cyber Security Center dat uit patches voor besturingssystemen binnen 48 uur malware werd ontwikkeld en verspreid (21).
De Carnegie Mellon universiteit won in 2016 met het AI-systeem Mayhem de DARPA Cyber Grand Challenge. Omdat Mayhem volledig automatisch werkte, vond het de meeste kwetsbaarheden in de aangeboden software – ondanks een vroegtijdige crash (22). Deze successen voorspellen dat AI-systemen steeds vaker cyberaanvallen zullen uitvoeren die bovendien effectiever en moeilijker te detecteren zijn.

Casus: 'stealth' malware produceert nepinformatie

Stuxnet was het eerste Cyber Weapon of Mass Destruction dat in juni 2010 werd ontdekt door een fabrikant van antivirussoftware (23). De ontwikkeling van deze geavanceerde worm kostte naar schatting 10 manjaren. Stuxnet infiltreert specifieke Siemens-apparatuur die onder andere ultracentrifuges besturen en zou ontwikkeld zijn om het Iraanse nucleaire programma te saboteren. Stuxnet misbruikt vijf beveiligingsgaten (waarvan destijds vier onbekende Zero Days) om zichzelf te verspreiden via USB-sticks en via LAN's. Stuxnet bevatte een gestolen digitaal certificaat voor installatie en actualiseert zichzelf via een eigen P2P-netwerk. Na installatie verbergt de worm zichzelf via een 'rootkit' en verwijdert zichzelf van USB-sticks na drie infecties. Stuxnet manipuleert een Siemens-stuurprogramma zodanig dat het toerental van de ultracentrifuges op de achtergrond steeds meer varieert, terwijl het dashboard een constant toerental blijft aangeven. Door Stuxnet-besmettingen is in Iran een onbekend aantal ultracentrifuges kapotgeslagen. De broncode van Stuxnet is al in volgende malwaregeneraties toegepast. De gedachte om fake stuurinformatie te produceren is later toegepast in de sjoemelsoftware van Diesel Gate.

Zowel de brondata als kennis over AI-systemen zijn aantrekkelijk voor cybercriminelen (24). Meer nog dan bij big data-systemen

zijn voor AI-systemen daarom zware beveiligingsmaatregelen nodig, zoals een SOC/SIEM met Intrusion Prevention System, systeem hardening, periodieke pentesten en Red Teaming. Daarnaast zijn beproefde normen nodig om de beveiliging adequaat in te richten en te testen.

IT-systemen toetsen kan bijvoorbeeld al met de Common Criteria (ISO15408), maar die zijn niet specifiek bedoeld voor big data en AI-systemen. De JTC 1/SC 42 heeft in juli 2019 drie ISO-normen gepubliceerd voor big data-systemen en ontwikkelt samen met de NEN nog twaalf normen, waarvan tien voor AI-systemen (25).

AI-systemen moeten goed blijven functioneren als ze zelf worden aangevallen. Een AI kan namelijk tijdens een aanval gewoon doorgaan (waardoor ongelukken kunnen ontstaan) of zichzelf uitschakelen, maar met beide uitkomsten is een DOS-aanval succesvol. AI-systemen moeten dus hun doelen bijstellen als er een situatie ontstaat waarin de AI slecht gaat functioneren. Daarvoor moeten AI-systemen complexiteit kunnen reduceren en niet-relevante data kunnen uitfilteren. Met andere woorden: een AI-systeem heeft 'situational awareness' nodig om in een onvoorspelbare wereld effectief te kunnen functioneren. Met de huidige AI-systemen staat dit nog in de kinderschoenen. Zo heeft BMW onlangs data van autosensoren gepubliceerd om gevaarlijke situaties beter te onderkennen (26). Totdat duidelijk is hoe AI-systemen effectief kunnen worden beveiligd, moet hun autonomie mogelijk worden ingeperkt om het restrisico acceptabel te houden.

Conclusie

Informatiebeveiliging voor big data en AI-systemen is belangrijk om informatiekwaliteit te garanderen, maar ook om commerciële innovatie mogelijk te maken. Dit artikel beschrijft welke beveiligingsmaatregelen in deze context relevant zijn. Deze maatregelen zijn niet compleet of in elke situatie geschikt. Zowel voor big data als AI-systemen worden nog normen ontwikkeld. Een adequate beveiliging is voorlopig maatwerk op basis van een risicoanalyse, ook omdat big data en AI-kennis zeer aantrekkelijk zijn voor cybercriminelen. Voor organisaties met big data en AI-systemen is beveiligingsexpertise dus een must-have.

Zoals bekend is een beveiliging nooit perfect. Elke beveiliging kan daarom worden verbeterd na beveiligingsincidenten, gemelde klachten, nieuwe dreigingen voor de organisatie en door systeemwijzigingen te toetsen op hun beveiligingsimpact. Het management moet de beveiliging ook actualiseren als de bedrijfseisen, contractuele eisen, normen of wet- en regelgeving veranderen.

AI kan bestaande aanvalstechnieken verbeteren en snel nieuwe cyberaanvallen lanceren, ook tegen AI-systemen zelf. AI-ontwikkelaars moeten onderkennen dat hun uitvindingen zowel voor goede, als slechte doelen kunnen worden ingezet. Het zijn uiteindelijk mensen die moeten bepalen in welke situaties en onder welke voorwaarden AI-systemen kunnen worden ingezet. In het huidige IT-tijdperk wordt informatiebeveiliging vaak uitgelegd als: hoe beschermen we systemen tegen falen? Maar het ethische principe daarachter is: hoe beschermen we mensen tegen falende systemen?

Referenties

1) De wet van Moore: elke 18 maanden verdubbelt de rekenkracht van computers
2) 'Big Data – for better or worse', SINTEF, 2013
3) Gebaseerd op definities van NORA online en Capgemini
4) Official CISSP CBK ISBN 0-8493-8231-9, blz 324 – 326
5) 'Privacy by design in big data', dec 2015, ENISA
6) Zie de PIA voor Big Data, www.rijksoverheid.nl
7) 'Lost in Translation', aug 2009, https://sma.nasa.gov/
8) www.forumstandaardisatie.nl
9) 'How does the complexity of Go compare with Chess?', 27 jan 2018, www.quora.com
10)' Demystifying AlphaGo Zero', nov 2017, https://arxiv.org
11) AI Beats Professionals in Six-Player Poker', www.cmu.edu
12) The Master Algorithm – Pedro Domingo ISBN 978-0-141-97924-3 blz 21
13) 'Better Language Models and Their Implications', feb 2019, https://openai.com/
14) 'Facebook bestrijdt nepnieuws', NRC 16-04-2019
15) Qutech Annual Report 2018 bevat voornamelijk proof-of-principle vooruitgang; 'Quantum Supremacy Is Coming', 18 juli 2019, Quanta Magazine
16) Zie de morele ontsporing in 2016 van Tay, de chatterbot van Microsoft
17) AI recruiting tool that showed bias against women', Okt 2018, www.reuters.com
18) Automatic patch-based exploit generation', 2008, www.cs.cmu.edu
19) Fortinet Threatscape Report, juni 2009, www.fortiguard.com
20) IBM X-Force 2010 trend and risk report (2011)
21) Assessing Security Vulnerabilities and Applying Patches', Jan 2019, www.cyber.gov.au
22) Mayhem the Machine That Finds Software Vulnerabilities', jan 2019, https://spectrum.ieee.org
23) Dissecting a Cyberwarfare Weapon', mei 2011, https://ieeexplore.ieee.org
24) It is difficult to think of a major industry that AI will not transform' – Andrew Ng
25) https://www.iso.org/committee/6794475.html en www.nen.nl/ai
26) BMW openbaart autodata', NRC 4 juni 2019