BCG BOSTON CONSULTING GROUP

Analyse af kunstig intelligens i et sikkerhedsperspektiv

Analyserapport

JANUAR 2020

DIGITALISERINGSSTYRELSEN

ERHVERVSSTYRELSEN

FE CENTER FOR CYBERSIKKERHED

Analysens hovedkonklusioner

AI anvendes allerede i dag i flere danske organisationer, men brugen er i mange tilfælde relativt basal

- Mange løsninger er i højere grad at betragte som maskinlæring snarer end reel 'intelligens', og ses ofte ifm. automatisering og beslutningsunderstøttelse
- Flere virksomheder ser AI som *meget kritisk* for deres drift, hvor det i mange tilfælde er en integreret del af forretningen; mindre kritisk i myndigheder

Kapitel 1: Brug af AI i driften giver organisationer flere nye angrebsflader, men traditionelle sikkerhedstiltag er stadigvæk i fokus

- Kun få identificerede sikkerhedstiltag er unikke for brugen af AI, mens en række traditionelle tiltag ændrer karakter i den forbindelse; de traditionelle tiltag bør i mange henseender fortsat være førsteprioritet for virksomheder og myndigheder
- Der er særligt mange nye, AI-specifikke tiltag ifm. forebyggelse og beskyttelse, hvoraf størstedelen af disse er tekniske og skal tænkes ind i AI-løsningen allerede i udviklingsstadiet (dvs. *security-by-design*)
- Brugen af AI udsætter organisationer for fire nye, specifikke angrebstyper, som dog alle primært er teoretiske trusler illustreret i akademisk forskning, men som endnu ikke er set i praksis; disse er *data poisoning*, *adversarial attacks*, *backdoor attacks* og *data extraction*

Kapitel 2: Brug af AI som forsvar er fortsat relativt umodent, men forventes at dominere markedet for softwareløsninger om 5-10 år

- Flere danske virksomheder udbyder i dag AI-understøttede sikkerhedsløsninger
- Brugen af AI er oftest en videreudvikling af eksisterende software, da størstedelen af de identificerede løsninger eksisterer både med og uden AI
- Flere eksperter forventer at AI-baserede sikkerhedsløsninger vil være standarden indenfor 5-10 år – både i Danmark og internationalt
- Inklusionen af AI i sikkerhedssoftware har potentiale til at øge effektiviteten af it-sikkerheden i organisationer ved bl.a. at automatisere tidskrævende processer ift. opdagelse og respons samt øge evnen til at detektere hidtil ukendte angreb

Kapitel 3: Brug af AI som angrebsmiddel har potentiale til at øge effektiviteten (volumen, hastighed, personalisering) af cyberangreb

- Der eksisterer på nuværende tidspunkt kun få veldokumenterede eksempler på brugen af AI i cyberangreb, men det er bl.a. set benyttet ifm. stemmeefterligning ved CEO fraud
- AI forventes indenfor de næste par år at øge mængden og kompleksiteten af traditionelle angreb (fx ved politisk manipulation og spear-phishing) samt at muliggøre angreb med direkte fysiske konsekvenser i stor skala (fx ifm. hacking af selvkørende biler og droner)
- Brugen af AI i cyberangreb har indflydelse på en lang række af de sikkerhedstiltag, virksomheder og myndigheder i øjeblikket benytter, fx stiller det højere krav til træningen af brugere og detaljeringsgraden ifm. detektionen af anomalier

Formål, baggrund og metode for analysen

Formål: At fremme danske virksomheder og myndigheders informations-sikkerhed i forbindelse med brugen af kunstig intelligens

Denne analyse fokuserer på samspillet mellem informationssikkerhed og kunstig intelligens (AI), og belyser dette fra tre forskellige vinkler:

1. AI som angrebsmål
2. AI som forsvar
3. AI som angrebsmiddel

Analysen beskæftiger sig både med danske virksomheder og myndigheder samt det danske marked for it-sikkerhedsløsninger, men den drager samtidig relevante paralleller og eksempler ind fra andre lande. Der er fokus på emnet anno 2019/20 med fremtidsperspektiver på visse områder, fx i forhold til brugen af AI som angrebsmiddel og dennes indflydelse på risikobilledet samt udbredelsen af AI-baserede sikkerhedsløsninger.

Analysen er udarbejdet på baggrund af:

- Interviews med internationale og nationale eksperter og forskere
- Interviews med internationale og nationale it-sikkerhedsudbydere
- Interviews med danske virksomheder og myndigheder, som anvender kunstig intelligens
- Danmarks Statistiks undersøgelse af virksomheders IT-anvendelse (VITA)
- Eksisterende litteratur indenfor både kunstig intelligens og informationssikkerhed

Ifm. projektet er der, udover denne analyse, også udarbejdet en vejledning, som har til formål at *sikre at danske virksomheder og myndigheder tager de fornødne forholdsregler, når de anvender kunstig intelligens*, som kan findes her 'Vejledning: Tiltag til at sikre brugen af kunstig intelligens'[1].

1. Vejledningen kan findes på sikkerdigital.dk Kilde: BCG analyse

Baggrund: Øget anvendelse af og politisk fokus på AI og informationssikkerhed i Danmark

Analysen er udarbejdet i et samarbejde mellem Digitaliseringsstyrelsen, Erhvervs-styrelsen, Center for Cybersikkerhed, Globeteam og Boston Consulting Group, og står i forlængelse af *National strategi for kunstig intelligens* samt *National strategi for cyber- og informationssikkerhed*, der har til formål at styrke hhv. brugen af kunstig intelligens og informationssikkerheden i Danmark.

Analysen og den tilhørende vejledning adresserer overlappet mellem de to strategier gennem et øget fokus på informations-sikkerhed for virksomheder og myndigheder, som anvender AI. For yderligere information om informationssikkerhed se fx

- Sikkerhedstjekket (ERST)
- Cyberforsvar der virker (CFCS & DIGST)
- Sikkerdigital.dk (ERST & DIGST)
- Kravkataloget (DIGST)
- Informationssikkerhed i leverandørforhold (CFCS & DIGST)

DIGITALISERINGSSTYRELSEN ERHVERVSSTYRELSEN GLOBETEAM BCG BOSTON CONSULTING GROUP CENTER FOR CYBERSIKKERHED

Analysen anvender den nationale strategis definition af kunstig intelligens (AI); som i stigende grad anvendes i både virksomheder og myndigheder

Definitionerne af AI er mange; i denne analyse anvendes definitionen anvendt i National strategi for kunstig intelligens

Bred forståelse: "AI findes og er udbredt"

- "Typiske AI-kapabiliteter i dag inkluderer tale-, billed- og videogenkendelse mv." — Stanford University
- "Kunstig intelligens anvendes på en række områder, fx i søgemaskiner, til tale- og billedgenkendelse eller til at understøtte droner og selvkørende biler" — OECD, Europa-Kommissionen
- "Egentlig kunstig intelligens findes endnu ikke og har nok lange udsigter" — FE

Smal forståelse "AI findes endnu ikke"

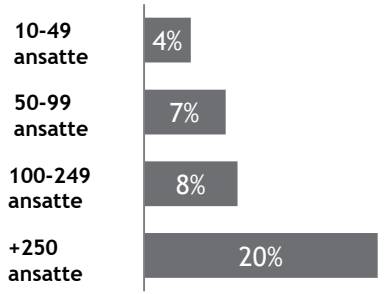
Definition på kunstig intelligens som anvendes i denne analyse

"Kunstig intelligens systemer baseret på algoritmer—dvs. matematiske formler—der ved at analysere og finde mønstre i data kan identificere den mest hensigtsmæssige løsning. Langt de fleste systemer varetager specifikke opgaver på afgrænsede områder til fx kontrol, forudsigelse og vejledning. Teknologien kan udformes til at tilpasse sin adfærd ved at observere, hvordan omgivelserne påvirkes af tidligere handlinger"

National strategi for kunstig intelligens, EU-kommissionen og OECD

Brug af AI er allerede udbredt blandt danske virksomheder og myndigheder i dag…

Danske **virksomheder** som anvender maskinlæring eller AI:



Eksempler på danske **myndigheder** der bruger eller ser potentiale i at bruge AI:

Miljø- og Fødevareministeriet, Landbrugsstyrelsen · Erhvervsstyrelsen · Region H · Sund & Bælt · Frederiksberg Kommune · Nævnenes Hus · Helsingør Kommune · Lægemiddelstyrelsen, Danish Medicines Agency

…og **indenfor de næste 3-5 år**, forventer både virksomheder og myndigheder **øget brug af AI**

- **88% af adspurgte ledere i det private** forventer, at AI i høj grad vil spille en betydelig rolle i deres virksomhed indenfor 5 år
- **~50% af danske virksomheder** forventer, at AI vil have en stor indvirkning på endnu ukendte dele af deres forretning i fremtiden
- **55% af alle danske kommuner** forventer at have implementeret AI indenfor de næste tre år
- **48% af adspurgte ledende embedsmænd** forventer, at AI i høj grad vil ændre offentlige arbejdsopgaver indenfor 5 år

Kilde: National strategi for kunstig intelligens (2019); Moderniseringsstyrelsen (2018): Kortlægning af analytics i staten; Microsoft (2019): Kunstig intelligens i den offentlige sektor; Microsoft (2018): Artificial Intelligence in Europe – Denmark

AI anvendes overordnet set til fem forskellige formål – med flere eksempler på hver i både danske virksomheder og myndigheder

Eksempler

Anvendelsesområder for AI

Forudsige udfald
Forudse begivenheder og udfald før de sker

Automatisere processer
Udføre opgaver uden menneskelig involvering

Generere indsigter
Identificere og forstå nye mønstre og trends i data

Personalisere indhold
Skræddersy brugeroplevelse til individuelle kunder

Foreslå løsninger
Foreslå løsninger til allerede definerede problemstillinger

Private virksomheder

Virksomheders brug af AI i dag[1]

Forudsige udfald

Vestas
Anvender data fra sensorer i deres møller til at foretage forebyggende vedligeholdelse

corti
Analyserer medicinske samtaler real-time og forudsiger tilfælde af eksempelvis hjertestop

Danfoss
Anvender visuel AI til en app, der kan tage et billede og analysere, om en reservedel skal skiftes

92%

Automatisere processer

Ørsted
Anvender 'Robot process automatisering' til at automatisere tidligere manuelle opgaver

SEB
Automatiserer tidligere manuelle opgaver og frigiver dermed tid til kundekontakt

whaii
Automatiserer virksomheders screeningproces ifm. rekruttering og matcher kandidater med jobopslag

88%

Generere indsigter

Lundbeck
Finder tendenser på tværs af datasæt, der normalt ville være meget tidskrævende eller umulige at finde

Carlsberg
Anvender sensorer til at måle smagsaromaer, som resulterer i nye øl

CHR. HANSEN
Bruger kunstig intelligens til at udvikle nye ingredienser til fødevarer eller landbrugsprodukter

80%

Personalisere indhold

SAXO BANK
Personaliserer klientoplevelsen baseret på den enkelte brugers adfærd

nøie
Leverer personaliserede hudplejeprodukter ved at identificere den optimale kombination af ingredienser

Interflora
Identificerer intention, anledning og relation ifm. gave-givning mhp. mersalg

64%

Foreslå løsninger

MAERSK
Anvender sensorer og AI-software til at spare penge på brændstof

PFA PENSION
Bruger sprogteknologi til at gruppere forespørgsler fra kunder og foreslå svar

Aguardio
Samler data om vandforbrug for private kunder og hoteller mhp. at foreslå vandbesparende adfærd

32%

Offentlige myndigheder

Erfaring/potentiale hos myndigheder[2]

Forudsige udfald

Politiet
Undersøger mulighederne i at anvende machine learning til risikobaseret udvælgelse af, hvor tilsyn skal foretages

Sund & Bælt
Bruger droner til at tage billeder af bro-betonen, som bliver analyseret med algoritmer mhp. at finde skader, som har brug for reparation

Automatisere processer

Nævnenes Hus
Tester koncept hvor AI bruges til at aflæse indkomne klagesager og automatisk fremsøge lignende sager og afgørelser

norddjurs kommune
Fordeler og journaliserer indgående post automatisk ved at bruge maskinlæring til at klassificere posten

Generere indsigter

Miljø- og Fødevareministeriet, Miljøstyrelsen
Bruger maskinlæring til at analysere satellitfotos af markarealer for at monitorere landbrugsaktivitet

Region H
Tester løsning, hvor data om postoperative patienter analyseres mhp. at varsle evt. komplikationer

Personalisere indhold

Helsingør Kommune
Chatbotten 'HelsingørBot' hjælper borgere med at starte byggesager rigtigt op ved at guide borgeren i den rigtige retning.

Københavns Kommune
Ifm. borgeres afbetalingsordninger vil en AI-model skulle sørge for, at borgeren ender ved den rette sagsbehandler uden behov for viderestilling

Foreslå løsninger

Erhvervsstyrelsen
Tester koncept hvor kunders bevægelser på hjemmesiden bruges til at forudsige og foreslå løsninger, hvis kunden ringer til support

Københavns Kommune
Arbejder med machine learning algoritme som skal lede bilister hen, hvor de har størst sandsynlighed for at finde en ledig parkeringsplads

xx% *Andel af vsh. der bruger AI*

Potentiale *Erfaringer*

1. Rundspørge blandt 25 danske virksomheder. 2. Baseret på sammenlignelig opdeling i anvendelsesområder hos Moderniseringsstyrelsen (2018).
Kilder: BCG analyse; Microsoft (2018): Artificial Intelligence in Europe – Denmark; National strategi for kunstig intelligens (2019); Microsoft (2018): Kunstig intelligens i Danmark - Potentialer og barrierer; Moderniseringsstyrelsen (2018): Kortlægning af analytics i staten

Indholdsfortegnelse

Dette dokument

Analyserapport
Tre kapitler med konklusioner vedr. AI's indflydelse på informations-sikkerhed – fra et dansk perspektiv

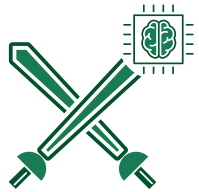
Vejledning
Råd til hvordan informations-sikkerheden sikres ved brug af AI
Se 'Vejledning: Tiltag til at sikre brugen af kunstig intelligens'

1 **AI som angrebsmål**
Hvordan brugen af AI giver anledning til nye trusler, og hvordan disse kan adresseres

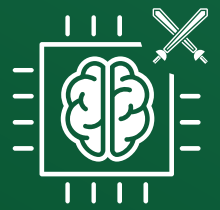
s. 6

2 **AI som forsvar**
Hvordan AI kan anvendes til at effektivisere informations-sikkerheden

s. 25

3 **AI som angrebsmiddel**
Hvordan hackeres brug af AI kan ændre trusselsbilledet, og hvordan dette kan adresseres

s. 38

Kapitel 1



AI som angrebsmål

Hvordan brugen af AI giver anledning til nye trusler, og hvordan disse kan adresseres

Grundlæggende begreber defineres og de væsentligste trusler ifm. brugen af AI kortlægges og prioriteres - for senere at kunne identificere relevante tiltag til at adressere disse

Trusler og risici forbundet med brug af AI

- Definition af kernebegreber i risikobaseret tilgang til sikkerhed
- Kortlægning af trusler forbundet med brug af AI-løsninger, samt afgrænsning og prioritering af risikoscenarier analysen vil fokusere på

Tiltag der kan øge informationssikkerheden ifm. brug af AI

- Overblik over generelle sårbarheder i danske virksomheder og myndigheder
- Identificering af tiltag der øger sikkerheden ved at adressere trusler forbundet med brug af AI

Danske organisationers kendskab til AI-sikkerhedsudfordringer

- Vurdering af i hvilket omfang danske virksomheder og myndigheder forholder sig til sikkerheds-udfordringer ifm. brug af AI
- Kvalificering af virksomheders og myndigheders selvevaluering med ekspertinterviews

En risikobaseret tilgang til informationssikkerhed hviler på en række kernebegreber, som i analysen defineres med udgangspunkt i ISO 27000

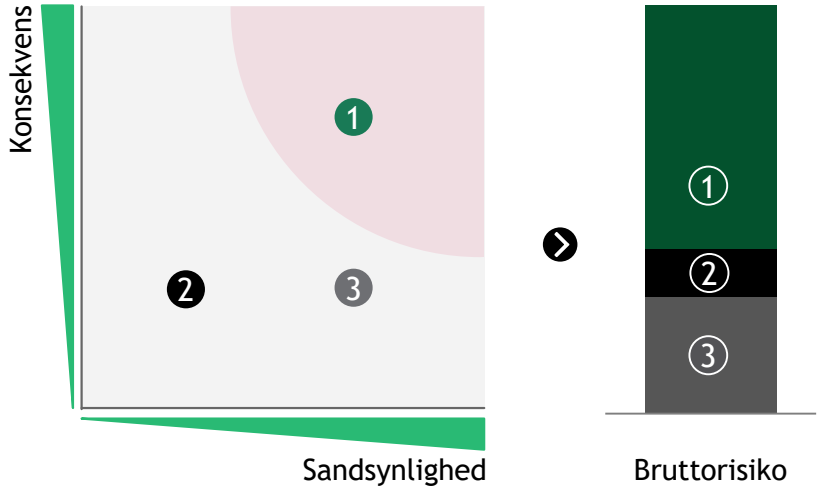
Eksempel

1 En organisation er udsat for tre trusler

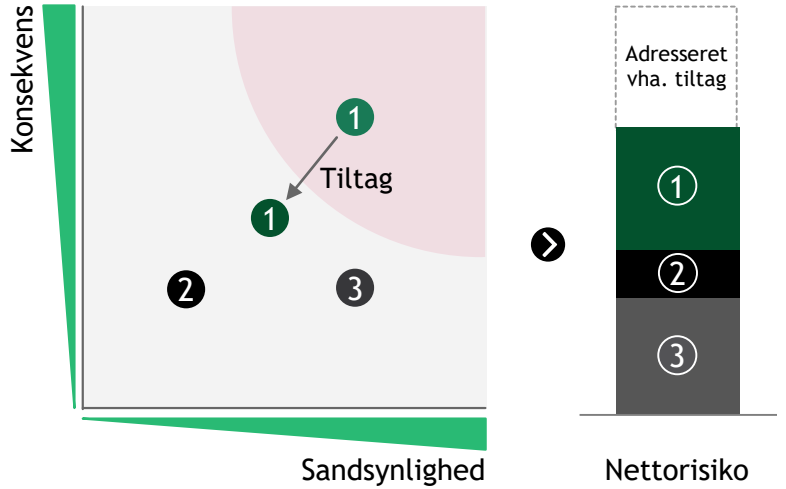
- A Indbrud
- B Brand
- C Oversvømmelse

Eksempel

2 Risici udgøres dels af sandsynligheden for, at trusler udnytter sårbarheder og dels af konsekvensen heraf



3 Organisationen kan indføre tiltag for at mindske risici



Trussel er en potentiel årsag til et uønsket scenarium, der kan skade den givne organisation.

Sårbarhed er en fejl eller svaghed, som trusler kan udnytte.

Sandsynlighed er en vurdering af, hvorvidt det kan forventes, at truslen materialiserer sig for den givne organisation – ved at udnytte sårbarheder

Konsekvens er den effekt, en given trussel vil have for den givne organisation, hvis den materialiseres – uafhængigt af sandsynligheden herfor. Kan fx være tabt omsætning eller tab af borgernes tillid.

Bruttorisiko er den kombinerede sandsynlighed og konsekvens forbundet med en trussels udnyttelse af sårbarheder – før mitigerende tiltag er indført.

Tiltag modificerer risici ved at mindske sandsynligheden for og/eller konsekvenserne af dem. Tiltag kan fx være en forsikring, teknologiske foranstaltninger eller kontrolprocesser.

Nettorisiko er den tilbageværende risiko forbundet med en trussel, som en organisation er udsat for, efter tiltag mod truslen er indført, fx fordi et tiltag ikke 100% dækker alle sårbarheder

Note: Baseret på definitioner hos ISO (27000)

Trusselskatalog

Metode

Standarder for risikostyring ved brug af AI er endnu ikke udarbejdet. Flere nationale og internationale standardiseringsorganer har dog igangsat dette arbejde, bl.a.

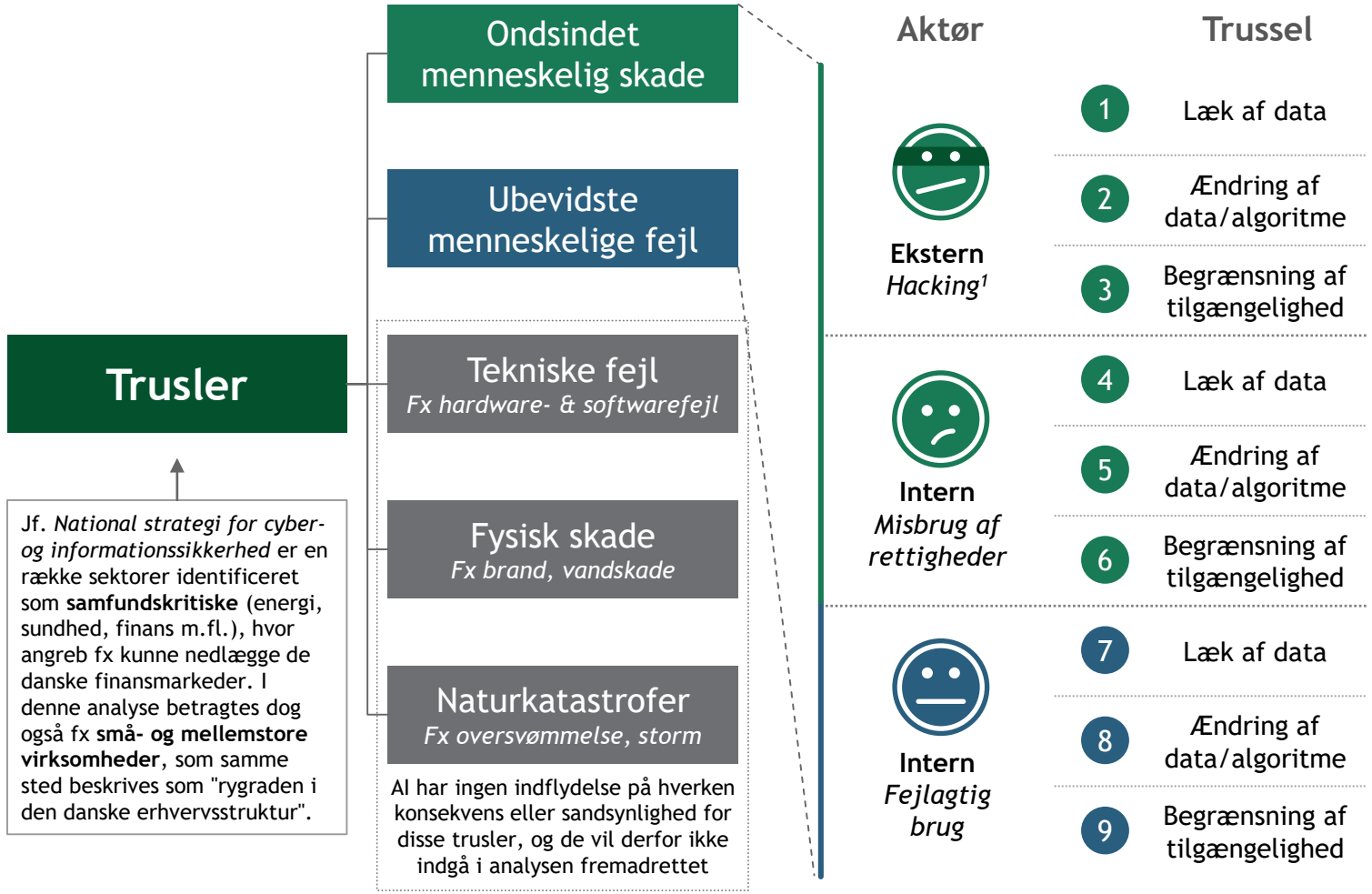
- ISO
- NIST (USA)
- ENISA (EU)

I denne analyse er ISO 27005's trusselsliste for 'information security' brugt som udgangspunkt for at sikre en udtømmende oversigt, mens Center for Cybersikkerheds opdeling af interne (både bevidste og ubevidste) og eksterne aktører er brugt til at definere de prioriterede trusselsområder.

Dette er suppleret med trusler specifikt associeret med AI vha.

- Ekspert interviews
- Nyhedsartikler
- Industrirapporter

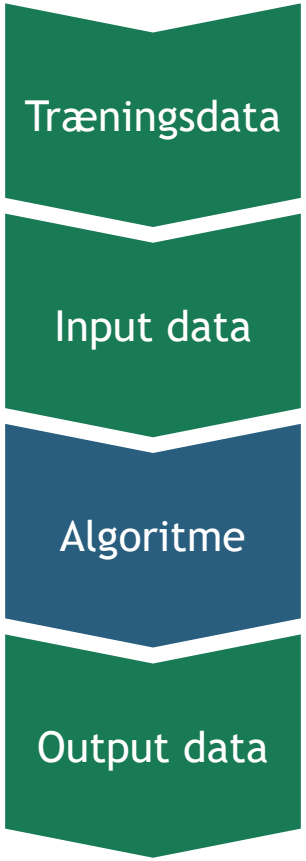
Analysen fokuserer på tre slags trusler forårsaget af tre forskellige aktører – dvs. i alt ni trusler



1. Hacking dækker både over cyberspionage, cyberkriminalitet, cyberaktivisme og cyberterror jf. Center for Cybersikkerheds definitioner. Kilde: ISO 27005 – Annex C; CFCS (2019): Cybertruslen mod Danmark; CFCS (2019): Cybertruslen fra bevidste og ubevidste insidere; NIST; ENISA; ekspert interviews

Brugen af kunstig intelligens har introduceret fire nye angrebsflader – angreb målrettet disse er dog krævende og enten ikke opdaget eller udført i stor skala

AI-proces



Angrebstype

Data poisoning (forgiftning af data) sker, når en aktør ændrer i træningsdata med formålet at manipulere output. Ændring i træningsdata kan både ske ved at ændre direkte i data (af interne eller eksterne aktører) eller, hvis algoritmen er selvlærende, ved at sende den en stor mængde inputdata af en bestemt art. Et kendt eksempel på sidstnævnte er Microsofts chatbot Tay, hvor Twitter-brugere på under 24 timer fik 'lært' Tay at skrive vulgært om en række ømtålelige emner.

Adversarial attacks (manipulering af data) sker, når hackere, typisk uden adgang til algoritmen, algoritmens parametre eller træningsdata, formår at 'snyde' modellen til at opnå en ønsket klassifikation ved at sende inputdata med en marginal ændring (fx ændre én pixel i et billede eller ændre bølgelængden på tale). Ofte kan ændringen ikke umiddelbart opdages af mennesker. Formålet kan bl.a. være *evasion*, hvor man ønsker at undgå en bestemt klassifikation, fx i en løsning hvor AI bruges til ansigtsgenkendelse.

Backdoor attacks (bagdørsangreb) sker ifm. AI typisk ved, at hackere eller interne får uretmæssig adgang til produktionskoden til algoritmen og ændrer koden (dvs. implementerer en *backdoor*) for at muliggøre et ønsket udfald ved en atypisk eller teoretisk umulig kombination af input variable. Det kan fx være input hvor *antal års joberfaring* > *alder*, som resulterer i en unormalt lav forsikringspræmie.

Data extraction (dataekstraktion) sker ved, at en aktør benytter information fra output til ét af to formål; 1) at bygge en surrogat/kopi model eller 2) at forstå bias i modellen. Ved førstnævnte kan den nye model fx sælges (som IP), bruges til at underbyde offeret på markedet, eller til at rekonstruere træningsdata. Ved sidstnævnte kan identifikationen af bias gøre det muligt at identificere kombinationer af inputdata, som resulterer i et ønsket udfald (også kaldet *impersonation*), hvilket bl.a. kan bruges i adversarial attacks.

Udbredelse

At udnytte de nye angrebsflader er **mere ressourcekrævende** for hackere end traditionelle angreb

Angrebene er **endnu ikke set i stor skala i praksis**, hvilket kan skyldes at de enten ikke er udført, opdaget eller offentliggjort

Basale sikkerhedstiltag mod traditionelle angreb bør som udgangspunkt være på plads før tiltag mod disse nye angrebsflader prioriteres

Kilde: BCG analyse; Patrick Hall: Proposals for model vulnerability and security; ekspertinterviews; nyhedsartikler

Seks trusler prioriteres pba. kvalitativ vurdering af risiko på tværs af myndigheder og virksomheder

Trusselsvurderinger er specifikke for hver organisation og afhænger af en lang række faktorer

Både konsekvens ved og sandsynlighed for en given trussel varierer fra organisation til organisation og fra løsning til løsning, ligesom konsekvens kan vurderes på flere forskellige parametre afhængig af organisationens primære formål, fx. tillidsmæssige vs. økonomiske.

Liste af faktorer der påvirker risikoen (ikke-udtømmende):

- **AI løsning:** Des vigtigere AI er for virksomheden eller myndighedens funktion, des større risiko vil organisationen være udsat for (fx hvis løsningen bliver utilgængelig)
- **Type af data:** Eksempelvis øges risikoen ved sensitivt data (fx sundhedsdata), IP data (fx forskningsresultater), data om drift som er en forretningshemmelighed (fx ruteplanlægning) samt data af økonomisk anvendelighed (fx kreditkortdata)
- **Grad af tillid:** Des vigtigere tillid blandt brugere/borgere er for organisationen, des større et problem er brud på denne (fx ifm. læk af data)
- **Populationsstørrelse:** Et sikkerhedsbrud i en stor organisation vil typisk både få mere opmærksomhed og have større negativ effekt på den samlede populations tillid, da flere vil blive ramt; direkte eller indirekte
- **Ansattes kompetencer:** Kompetente ansatte og generelt kendskab til AI vil i særdeleshed reducere sandsynligheden for fejl, men vil også øge evnen til at respondere på angreb

De ni trusler prioriteres således pba. en kvalitativ evaluering af risiko

Prioriterede risikoscenarier

1. Hacking: Læk af data
2. Hacking: Ændring af data/algoritme
3. Hacking: Reduktion af tilgængelighed
4. Misbrug: Læk af data
5. Misbrug: Ændring af data/algoritme
8. Fejl: Ændring af data/algoritme

Ikke-prioriterede risikoscenarier

6. Misbrug: Reduktion af tilgængelighed
7. Fejl: Læk af data
9. Fejl: Reduktion af tilgængelighed

Kilde: Ekspertinterviews; nyhedsartikler; BCG analyse

Kilder til evaluering

Ekspert-interviews

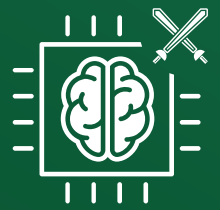
Interviews med udbydere

Faglitteratur

Bevæggrunde for evaluering

1. Grad af risiko for danske virksomheder og myndigheder generelt, vurderet pba. faktorer på venstre side
2. Trusler specifikke for brug af kunstig intelligens

Kapitel 1



AI som angrebsmål

Hvordan brugen af AI giver anledning til nye trusler, og hvordan disse kan adresseres

Pba. de identificerede trusler beskrives tre kategorier af sårbarheder, som typisk giver anledning til udførslen af disse trusler; 33 tiltag beskrives til at adressere sårbarhederne

Trusler og risici forbundet med brug af AI

- Definition af kernebegreber i risikobaseret tilgang til sikkerhed
- Kortlægning af trusler forbundet med brug af AI-løsninger, samt afgrænsning og prioritering af risikoscenarier analysen vil fokusere på

Tiltag der kan øge informationssikkerheden ifm. brug af AI

- Overblik over generelle sårbarheder i danske virksomheder og myndigheder
- Identificering af tiltag der øger sikkerheden ved at adressere trusler forbundet med brug af AI

Danske organisationers kendskab til AI-sikkerhedsudfordringer

- Vurdering af i hvilket omfang danske virksomheder og myndigheder forholder sig til sikkerheds-udfordringer ifm. brug af AI
- Kvalificering af virksomheders og myndigheders selvevaluering med ekspertinterviews

Tre arketyper af sårbarheder er forbundet med de identificerede trusler

Teknologisk infrastruktur

Forældede systemer, manglende monitorering og kontrol, manglende opdateringer, umoden teknologi, fejl i implementering eller opsætning samt mangel på fx firewall er alle eksempler på sårbarheder i den teknologiske infrastruktur

Adgange og rettigheder

Utilstrækkelig governance og kontrol samt mangel på løbende opdatering af brugeradgange og -privilegier kan blive udnyttet bevidst af outsiders og insiders, og det er samtidig en kilde til ubevidste fejl begået af insiders

Adfærd og retningslinjer

Mangelfuld uddannelse og træning af brugere, administratorer og ledere i retningslinjer, handlingsplaner og sikker adfærd generelt er en udbredt kilde til sårbarheder lige såvel som, at mangel på tekniske kompetencer kan øge risikoen for ubevidste fejl

Relation til risikoscenarier

1	Hacking: Læk af data	●	◕	◕
2	Hacking : Ændre data/algoritme	●	◕	◕
3	Hacking : Tilgængelighed	◕	◑	◔
4	Misbrug: Læk af data	◔	●	◑
5	Misbrug : Ændre data/algoritme	◔	●	◑
8	Fejl: Ændre data/algoritme	◔	◑	●

Kilde: BCG analyse; ekspertinterviews

● *Sårbarhed tæt forbundet med risikoscenarie* ○ *Sårbarhed ikke relevant for risikoscenarie*

Et godt it-forsvar skal adressere alle tre grupper af sårbarheder

Læs mere om de specifikke tiltag på de følgende sider og i 'Vejledning: Tiltag til at sikre brugen af kunstig intelligens'[1]

	Forebyggelse & beskyttelse *Før angrebet sker*	Opdagelse *Når angrebet sker*	Respons *Efter angrebet er sket*
Organisatorisk	• **Risikostyringens ramme** defineres, herunder bl.a. afklaring af acceptabel risiko, allokering af ressourcer, fastlæggelse af metode til at måle risici mv. • **Risikovurdering** foretages for hvert (AI) projekt for at kortlægge og tydeliggøre de risici, som organisationen er udsat for • **Retningslinjer og politikker** udformes for organisationen for grundig og løbende monitorering og styring af risici • **Roller og ansvar** fordeles, så dedikerede kompetencer afsættes, og ansvar tydeligt kan henføres til enkeltpersoner på ledelses-/bestyrelsesniveau **A**		
	• Træning af medarbejdere • Awareness-kampagne • Test af brugerkompetencer • Positivliste af software og devices • Risikovurdering af leverandører • Styring af brugere og adgange • Styring af opdateringer og vedligehold • Styring af modeludvikling og -træning AI **B**	• Træning af medarbejdere • Awareness-kampagne **C**	• Træning af medarbejdere • Handlings- og kommunikationsplan • *Business continuity*-plan • Gendannelsesplan • Beredskabsøvelse • Feedback-loop af *lessons learned* **D**
Teknisk	• Penetrationstest af systemer • Malwarebeskyttelse • Datakryptering • Etablering af netværksgrænser og barrierer • Sikring af modellens robusthed AI • Undgå læk af modelparametre AI • Modificering af inputdata AI • Træn på manipuleret (*adversarial*) data AI • Begrænsning af inputdata AI • Øge validiteten af modeller m/hårde labels AI **E**	• Penetrationstest af systemer • Monitorering af aktivitet • Overvågning af input og output AI • Benchmarking af model AI **F**	• Nedlukning og isolation af inficerede systemer • Undersøgelse af angreb via logs • Backup-løsning **G**

1. Vejledningen kan findes på sikkerdigital.dk Kilde: BCG analyse; ekspertinterviews; CFCS; National Cyber Security Centre (NCSC); NIST; Center for Internet Security (CIS)

AI *Tiltag kun relevante ifm. brugen af AI*

Oversigt over tiltag, adresserede sårbarheder og udbredelse (1/5)

Type	Tiltag	Beskrivelse	Adresserede sårbarheder: Teknologisk infrastruktur	Adresserede sårbarheder: Adgange og rettigheder	Adresserede sårbarheder: Adfærd og retningslinjer	Udbredelse[1] AI
A	Risikostyringens ramme	Rammen for organisationens risikostyring skal defineres, og processen skal drives af ledelsen. Centrale aspekter i rammen inkluderer bl.a., at organisationen fastlægger sin acceptable risiko, allokerer ressourcer til risikostyring (økonomi, medarbejdere, lederes tid), prioriterer forskellige typer af konsekvenser (økonomiske, menneskelige, miljømæssige, omdømmemæssige) samt fastlægger begreber og metode til at måle risici (hvordan vurderer vi konsekvenserne og sandsynligheden for en given hændelse?)	✓	✓	✓	■■□□□
	Risikovurdering	Organisationens digitale risici skal undersøges med samme grundighed som finansielle, juridiske og andre risici undersøges, fx vha. STRIDE modellen (særligt anvendelig for AI) til at identificere alle relevante trusler	✓	✓	✓	■■□□□
	Retningslinjer og politikker	Retningslinjer kan ses som ledelsens kommunikation af risikostyring til organisationen; kan ifm. informationssikkerhed eksempelvis indeholde regler vedr. software- og hardware-køb, krav om sikkerhedstjek til visse jobs, og regler for brug af sociale medier	✓	✓	✓	■■■□□
	Roller og ansvar	Gør det klart for alle i organisationen, hvem der har ansvaret for hvad ift. brugen af it og ved sikkerhedshændelser; sørg for at ansvaret også er forankret på ledelsesniveau	✓	✓	✓	■■■□□
B C D	Træning af medarbejdere	Træning gennem e-learning moduler, on-site træning, introforløb etc. Tilpasset den enkeltes rolle og ansvar (fx rapportering af hændelser for brugere, *security by design* for udviklere og viden om relevante trusler for beslutningstagere)			✓	■■□□□
B C	Awareness-kampagne	Ansatte påmindes løbende om sikkerhedspolitikker og -retningslinjer, fx via mails og videoer fra ledelsen, information om cyberangreb og i sikkerhedsmøder			✓	■■■□□
B	Test af bruger-kompetencer	Som en eksamen ifm. træninger eller ved *code of conduct*-sessioner udsættes brugere for reelle sikkerhedsdilemmaer og tests af deres kendskab til retningslinjer og politikker			✓	■□□□□
	Positivliste af software og devices	Liste over devices som er godkendt til arbejdsrelateret brug (fx kun krypterede USB'er), og software som er tilladt at anvende på arbejdsdevices	✓		✓	■■■□□

1. Udbredelse defineret som grad af anvendelse hos DK virksomheder/myndigheder i dag pba. interviews med brugere og eksperter; 0 = Eksempler på brug ikke fundet eller kun fundet "på papiret"; 5 = Implementeret som effektfuldt tiltag hos alle. Kilde: BCG analyse; NCSC; NIST; CFCS; ekspertinterviews

AI *AI-specifikt tiltag*

Oversigt over tiltag, adresserede sårbarheder og udbredelse (2/5)

Type	Tiltag	Beskrivelse	Adresserede sårbarheder: Teknologisk infrastruktur	Adresserede sårbarheder: Adgange og rettigheder	Adresserede sårbarheder: Adfærd og retningslinjer	Udbredelse[1]	AI
B	Risikovurdering af leverandører	Hvis services (fx cloud lagring, AI-udvikling, konsulentbistand) outsources, er det vigtigt at sikre, at leverandøren også lever op til organisationens sikkerhedskrav, hvilket bør indgå i evalueringen af udbydere sammen med fx kvalitet og pris. Ifm. udvikling/hosting af AI er det særligt vigtigt, da en AI-model ikke altid let kan 'flyttes' til en ny udbyder	✓	✓	✓	■■□□□	
	Styring af brugere og adgange	Ved at begrænse brugeres adgang til kun at dække nødvendige programmer og data isoleres omfanget af en hændelse, ligesom risikoen for misbrug eller fejl af insidere begrænses. Ligeledes gælder, at tidligere brugere skal slettes og admin-rettigheder begrænses. For at bekræfte brugeres identitet benyttes fx to-faktor-godkendelse, komplicerede passwords og adgangskort e.l. til fysisk adgang		✓		■■■□□	
	Styring af opdateringer og vedligehold	Også kendt som *patch management*. Så snart opdateringer til software og operativsystemer er tilgængelige, skal disse installeres hos alle brugere for at begrænse eksponering til kendte sårbarheder. Det kan overvejes at lade sikkerhedsteamet teste opdateringer, førend disse installeres	✓			■■■■□	
	Styring af model-udvikling og -træning	For at undgå brugerfejl og sikre tilstrækkelig dokumentation for en AI-model benyttes versionsstyring og logning ifm. udvikling, fx i form af snapshots af modellen og dens parametre, træningsdata, performance mm.			✓	■■■□□	AI
D	Handlings- og kommunikationsplan	I tilfælde af hændelser skal det stå klart for alle, hvem der skal underrettes, hvordan dette eskaleres om nødvendigt, og hvordan det kommunikeres bredt til organisationen			✓	■■□□□	
	Business continuity-plan	I tilfælde af en hændelse skal en *business continuity*-plan beskrive, hvordan det sikres, at interne processer kan opretholdes, og hvad der evt. skal prioriteres. Fx kan det ifm. en AI-model til *automatisering af processer* beskrives, hvordan disse processer håndteres manuelt ved nedbrud, samt hvem der er ansvarlige for hvad			✓	■■□□□	
	Gendannelsesplan	Nært forbundet med *business continuity*-planlægning. En gendannelsesplan fokuserer på at genetablere kritiske systemer (fx gennem backups), således at processer returnerer til normaltilstand hurtigst muligt			✓	■■□□□	

1. Udbredelse defineret som grad af anvendelse hos DK virksomheder/myndigheder i dag pba. interviews med brugere og eksperter; 0 = Eksempler på brug ikke fundet eller kun fundet "på papiret"; 5 = Implementeret som effektfuldt tiltag hos alle. Kilde: BCG analyse; NCSC; NIST; CFCS; ekspertinterviews

AI *AI-specifikt tiltag*

Oversigt over tiltag, adresserede sårbarheder og udbredelse (3/5)

Type	Tiltag	Beskrivelse	Adresserede sårbarheder: Teknologisk infrastruktur	Adresserede sårbarheder: Adgange og rettigheder	Adresserede sårbarheder: Adfærd og retningslinjer	Udbredelse[1] AI
D	Beredskabsøvelse	I stil med penetrationstest af systemer men fokuseret på organisationens handlinger efter en hændelse. Detekteres bruddet, og i så fald hvor hurtigt? Bliver det afskærmet? Følger de ansatte de etablerede retningslinjer og handlingsplaner?	✓		✓	■■□□□
	Feedback-loop af *lessons learned*	Handlinger i løbet af en hændelse (eller test) journalføres, så indsatsen kan analyseres bagefter, og evt. ændringer kan implementeres i handlings- og kommunikationsplan. Evt. fokusområder kan adresseres med trænings- eller awareness-kampagner	✓	✓	✓	■■■□□
E F	Penetrationstest af systemer	Hyre venligtsindede hackere til at forsøge at trænge ind i nuværende sikkerhedssetup for at identificere sårbarheder. Typisk fokuseret på den teknologiske infrastruktur, men kan også adressere fx adfærd ved at efterlade inficerede USB'er med organisationens logo på parkeringspladsen. Et beslægtet tiltag herunder er *bug bounty programmes*, hvor hackere, fx via en ekstern leverandør, tilbydes en dusør for at finde fejl og sårbarheder	✓	✓	✓	■■□□□
E	Malware beskyttelse	Beskytter mod ondsindet software (fx virus) ifm. mails, downloads osv.	✓		✓	■■■■□
	Datakryptering	Kryptering af data minimerer konsekvenserne ved eksempelvis læk af data, da fortrolig eller værdifuld information i det lækkede data kan holdes hemmeligt, så længe data ikke dekrypteres. Værdifuld data bør krypteres både ved opbevaring og transit (fx fra hjemmenetværk til arbejdsnetværk). Ifm. selvlærende AI-modeller kan kryptering af data forsinke og fordyre læringsprocessen, hvorfor data ofte dekrypteres fx én gang om måneden mhp. træning, hvorefter data krypteres igen	✓	✓	✓	■■■□□
	Etablering af netværksgrænser og -barrierer	Klar afgrænsning af interne og eksterne netværk begrænser eksponeringen til angreb fra internettet. Derudover bør kritiske systemer isoleres og beskyttes, adgang til interne netværk beskyttes med fx VPN, og netværk generelt monitoreres for atypisk aktivitet	✓			■■■□□

1. Udbredelse defineret som grad af anvendelse hos DK virksomheder/myndigheder i dag pba. interviews med brugere og eksperter; 0 = Eksempler på brug ikke fundet eller kun fundet "på papiret"; 5 = Implementeret som effektfuldt tiltag hos alle. Kilde: BCG analyse; NCSC; NIST; CFCS; ekspertinterviews

AI *AI-specifikt tiltag*

Oversigt over tiltag, adresserede sårbarheder og udbredelse (4/5)

Type	Tiltag	Beskrivelse	Adresserede sårbarheder: Teknologisk infrastruktur	Adresserede sårbarheder: Adgange og rettigheder	Adresserede sårbarheder: Adfærd og retningslinjer	Udbredelse[1]	AI
E	Sikring af modellens robusthed	Modeltræning, krydsvalidering, udvælgelse af træningsdata osv. skal alt sammen ske med formålet at gøre modellen så robust som muligt. Dette reducerer modellens sårbarhed ift. kategorisering ved små ændringer i datainput, hvilket gør modellen sværere at manipulere for hackere. Vigtigt element af *Security-by-design*	✓			■■■□□	AI
	Undgå læk af modelparametre og -beregninger	Mængden af information, som en hacker får fra output af AI-modellen, skal begrænses. Fx kan en model oplyse sandsynligheden for, at et billede forestiller en panda vs. en bil vs. en bog; alternativt kan modellen blot oplyse den mest sandsynlige kategori. Ved førstnævnte kan hackeren nemmere ved at lave små ændringer i sit input finde frem til, hvilke der forårsager størst ændringer i output kategoriseringen	✓			■□□□□	AI
	Modificering af inputdata	Ved at ændre modellens inputdata (fx ændre filtype, komprimere fil, ændre størrelse) tilføjes tilfældig støj, som gør det sværere for hackere at benytte *adversarial attacks*	✓			■□□□□	AI
	Træn på manipuleret (*adversarial*) data	For at forhindre *adversarial attacks* trænes modellen på en stor mængde af manipuleret data med korrekte kategoriseringer; fx data med bevidst støj såsom ændrede pixels	✓			■□□□□	AI
	Begrænsning af inputdata	Ved af definere hvilket inputdata AI-modellen kan modtage, begrænses muligheden for at aktivere skjulte mekanismer i modellen via *backdoors* i data i forsøget på at opnå et bestemt output. Samtidig kan en AI-model potentielt opfange en lang række inputs, som vi mennesker ikke kan opfange, fx kan Siri og Alexa opfange frekvenser, det menneskelige øre ikke kan høre (kendt som *dolphin attacks*), hvilket også kan filtreres	✓			■■□□□	AI
	Øge validiteten af modeller med hårde labels	AI-modeller med 'hårde labels' (fx kategoriseringer i ja/nej, 0/1, kat/hund/hest) er sårbare overfor *adversarial attacks*, hvis hackere kan identificere de specifikke ændringer, der skal til for at ændre kategoriseringen. *Label smoothing*, *emsemble modeller* og *defensiv destillering* gør alle modellen mere moderat i sine konklusioner og derved mere robust overfor udsving i input og/eller evt. fejl og mangler i træningsdata	✓			■□□□□	AI

Fælles for de nederste fem tiltag ovenfor er, at de alle har vist væsentlige begrænsninger ift. at stoppe misbrug – og derfor til dato primært er set i akademiske tests snarere end i praksis

1. Udbredelse defineret som grad af anvendelse hos DK virksomheder/myndigheder i dag pba. interviews med brugere og eksperter; 0 = Eksempler på brug ikke fundet eller kun fundet "på papiret"; 5 = Implementeret som effektfuldt tiltag hos alle. Kilde: BCG analyse; NCSC; NIST; CFCS; ekspertinterviews

AI *AI-specifikt tiltag*

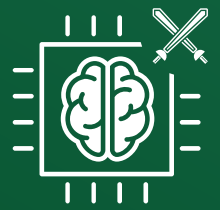
Oversigt over tiltag, adresserede sårbarheder og udbredelse (5/5)

Type	Tiltag	Beskrivelse	Adresserede sårbarheder: Teknologisk infrastruktur	Adresserede sårbarheder: Adgange og rettigheder	Adresserede sårbarheder: Adfærd og retningslinjer	Udbredelse[1]	AI
F	Monitorering af aktivitet	Monitorering af bruger- og netværksaktivitet for at identificere unormale mønstre eller kendte signaturer fra hackere, fx ifm. *data loss prevention* og *intrusion detection*. Dette er det primære område, hvor AI ses benyttet som en del af it-forsvaret. Når ondsindet eller unormal aktivitet opdages, kan det fx blive rapporteret ind i et overordnet SIEM-system (Security Information and Event Management)	✓			■■■■□	
	Overvågning af input og output	I forlængelse af ovenstående monitorering kan både inputdata til og output fra AI-modellen også tjekkes for anomalier, fx i form af ændringer i frekvensen af specifikke udfald eller i antallet af forespørgsler. Det kan fx benyttes til at identificere *data extraction* angreb	✓			■□□□□	AI
	Benchmarking af model	Andre velkendte modeller, fx tidligere anvendte modeller eller simple lineære modeller, kan benyttes til at verificere output af en kompleks eller ny AI-model. Hvis forskellen mellem de to modeller er stor, kan årsagen hertil undersøges	✓			■□□□□	AI
G	Nedlukning og isolation af inficerede systemer	Automatisk system som ved opdagelse af uretmæssig adgang eller uønsket adfærd indenfor netværket skiller systemet i forseglede containere og slukker inficerede dele. Således minimeres risikoen for at angreb spredes samtidig med, at skaden på inficerede systemer forsøges begrænset	✓			■■□□□	
	Undersøgelse af angreb via logs	Logs fra systemer og udstyr benyttes reaktivt til at analysere et brud og identificere den bagvedliggende årsag. Eksempler på relevante logs inkluderer DNS, firewall og VPN. I visse tilfælde kan det også være relevant at undersøge logs over ændringer i logs (for at detektere evt. manipulation)	✓			■■■□□	
	Backup-løsning	Backups kan eksistere enten i form af en kopi af data (til evt. gendannelse) eller som et alternativt system ved begrænsning af tilgængelighed (fx via standby-aftaler). Backups af AI-systemer besværliggøres af modellernes natur – og i særdeleshed for modeller der løbende lærer på input, hvor træningen vil skulle gentages på en evt. backup	✓			■■■■□	

1. Udbredelse defineret som grad af anvendelse hos DK virksomheder/myndigheder i dag pba. interviews med brugere og eksperter; 0 = Eksempler på brug ikke fundet eller kun fundet "på papiret"; 5 = Implementeret som effektfuldt tiltag hos alle. Kilde: BCG analyse; NCSC; NIST; CFCS; ekspertinterviews

AI *AI-specifikt tiltag*

Kapitel 1



AI som angrebsmål

Hvordan brugen af AI giver anledning til nye trusler og hvordan disse kan adresseres

Danske virksomheder og myndigheder evalueres pba. deres kendskab og tilgang til at adressere de identificerede trusler – både pba. eksperters udsagn og selvevalueringer

Trusler og risici forbundet med brug af AI

- Definition af kernebegreber i risikobaseret tilgang til sikkerhed
- Kortlægning af trusler forbundet med brug af AI-løsninger, samt afgrænsning og prioritering af risikoscenarier analysen vil fokusere på

Tiltag der kan øge informationssikkerheden ifm. brug af AI

- Overblik over generelle sårbarheder i danske virksomheder og myndigheder
- Identificering af tiltag der øger sikkerheden ved at adressere trusler forbundet med brug af AI

Danske organisationers kendskab til AI-sikkerhedsudfordringer

- Vurdering af i hvilket omfang danske virksomheder og myndigheder forholder sig til sikkerheds-udfordringer ifm. brug af AI
- Kvalificering af virksomheders og myndigheders selvevaluering med ekspertinterviews

Metode

Analysen af hvordan virksomheder og myndigheder forholder sig til sikkerhedsudfordringer i forbindelse med brugen af kunstig intelligens er baseret på:

- 16 interviews med virksomheder
- 13 interviews med myndigheder
- 18 interviews med danske eksperter, forskere og leverandører af it-sikkerhedsløsninger
- Survey-data fra Danmarks Statistik vedr. ~4.000 danske virksomheders it anvendelse (VITA)

Virksomhedernes og myndighedernes egen vurdering af deres informationssikkerhed er testet og kvalificeret vha. interviews med eksperter, forskere og leverandører.

Resultaterne er perspektiveret til danske virksomheder generelt via repræsentativt survey-data fra Danmarks Statistik, der indeholder spørgsmål om bl.a. virksomheders brug af maskinlæring, investering i it-sikkerhed mm.

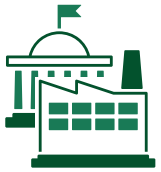
Bias i interviewees

Virksomheder og myndigheder er udvalgt pba. kendte eksempler på brug af AI, og har selv ønsket at svare på spørgsmål om it-sikkerhed. Udsagn fra interviews kan derfor ikke nødvendigvis forventes at være repræsentative for virksomheder og myndigheder generelt.

Konklusion: Digitaliseringen udvikler sig hurtigere end informationssikkerheden i Danmark

Eksperter ser mulighed for at forbedre sikkerheden i DK

- Alle adspurgte forskere, it-sikkerhedsleverandører og andre eksperter vurderer, at information-sikkerheden hos danske virksomheder og myndigheder kan forbedres
- Danmark er på niveau med nabolande hvad angår informationssikkerhed, men Danmark er længere fremme ift. digitalisering, hvilket har skabt et gab mellem de to
- Gabet afspejles også i BCG's *Digital Acceleration Index*, hvor danske virksomheder generelt scorer 9%-point lavere end internationale virksomheder ift. it-sikkerhed, men 3%-point højere hvad angår udbredelsen af AI-løsninger
- Barriererne for højere informationssikkerhed er først og fremmest kompetencer og manglende forståelse for egen organisations risici



Iflg. virksomheder/myndigheder selv er sikkerheden høj

- 26 af 29 adspurgte virksomheder og myndigheder vurderer, at modenheden af deres informations-sikkerhedsstrategi er høj eller meget høj
- 11 af 16 virksomheder ser deres AI-løsninger som noget eller meget kritiske for deres drift, mens andelen er 3 af 12 af myndigheder - konsekvensen af utilgængelighed størst hos virksomheder
- 9 af 16 virksomheder ser deres data som meget følsomt, mens det samme gælder for 9 af 12 myndigheder
- 8 af 16 virksomheder rapporterer, at de har forholdt sig anderledes til informationssikkerhed end de plejer som følge af brugen af AI, mens dette ikke er tilfældet for nogen myndigheder
- Større virksomheder har generelt implementeret flere, mere omfattende sikkerhedstiltag – for flere startups/SMV'er er sikkerheden ved brug af AI dog i højere grad tænkt ind i løsningen fra start

Virksomheder der bruger AI rapporterer oftere sikkerhedsbrud

- Blandt danske virksomheder rapporterede 6% sikkerhedsbrud i 2018; blandt de virksomheder der bruger maskinlæring er andelen 11% - hos begge forventes dog store mørketal
- Årsagen til forskellen i andelene der rapporterer sikkerhedsbrud kan skyldes en række faktorer, fx at brugere af AI opdager flere brud, er mere åbne om brud eller er mere digitaliserede og sårbare
- Blandt alle danske virksomheder havde 35% i stigende grad investeret i it-sikkerhed i løbet af 2018, mens andelen er 24%-point højere hos virksomheder, der bruger maskinlæring
- De mest gængse sikkerhedstiltag er alle +20%-point mere udbredte hos virksomheder, der bruger maskinlæring, end blandt danske virksomheder generelt

Kilde: BCG analyse

Eksperter og leverandører deler betragtning om, at der er mulighed for forbedring af sikkerhedsniveauet i Danmark; brugen af AI øger behovet herfor

Sikkerhedsniveau:
Plads til forbedring – især ift. Danmarks grad af digitalisering

"Danske virksomheder har slet ikke forstået risikoen, det gælder især små virksomheder men også de større"

"Når jeg kommer ud til virksomheder i dag, så siger jeg det samme til dem, som jeg sagde for 25 år siden"

"Når virksomheder er med helt fremme i den teknologiske udvikling, så forstår jeg simpelthen ikke, at man ikke er længere fremme med sikkerheden"

"IoT og digitalisering er overalt; alting er forbundet til internettet, så alle skal vide noget om it-sikkerhed i dag"

"It-sikkerhed er på vej op på agendaen hos ledelserne, men det vil tage nogle år"

"Vi er langt fremme med cybersikkerheden i Danmark, men vi er endnu længere fremme med digitalisering, hvilket skaber et gab imellem de to"

"Jeg tror, vi ville score lavere på informationssikkerhed end både Sverige og Norge"

BCG's Digital Acceleration Index[1]
- Danske virksomheder scorer 9%-point lavere end det globale gennemsnit på it-sikkerhed*
- Til gengæld scorer danske virksomheder 3%-point højere på brug af kunstig intelligens*
- Globalt scorer offentlige myndigheder 6%-point lavere end virksomheder på it-sikkerhed og 11%-point lavere på AI

**Bemærk lavt antal observationer i Danmark (21)*

Barrierer for forbedring:
Kompetencer, måling af risiko og økonomi er de største barrierer

"Det handler jo grundlæggende set om økonomi; ledelsen fokuserer på omsætning, og det er en svær business case at vise, at man skal bruge 2 mio. kr. på sikkerhed"

"Små virksomheder tror det kun sker for Maersk og Demant, men alle der har penge er i hackernes søgelys"

"Vi mangler kompetencer i Danmark; både bredt ift. brug af teknologi – og især kunstig intelligens – men også specialister indenfor it-sikkerhed"

"Ledelsen kan ikke se ned gennem ledelseslagene og ned i maskinen - de kender ikke den faktiske risiko. De monitorerer omsætningen tæt men ikke sikkerheden. Det svarer til at tjekke sikkerheden i et fly uden at tjekke, om motoren er i orden"

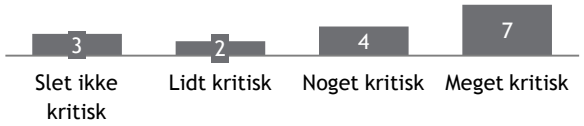
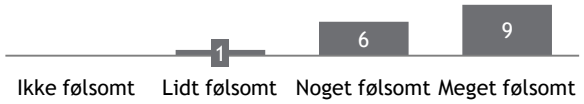
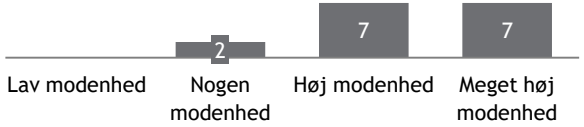
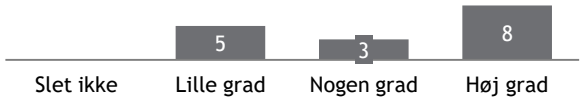
"Jeg savner, at organisationer tør dele information om deres sikkerhedsbrud"

"Små virksomheder tænker kun på deres produkt"

"Barrieren for større sikkerhed er penge"

1. Standardiseret selvevaluering af virksomheder og myndigheder på 35 parametre; n(Danmark) = 21; n(global – privat) = 2.779; n(global – offentlig) = 232
Kilde: BCG Digital Acceleration Index; BCG analyse; ekspertinterviews

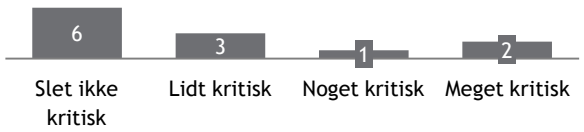
Adspurgte virksomheders egen vurdering indikerer høj grad af kendskab til sikkerhedsudfordringer ifm. brug af AI og høj modenhed generelt

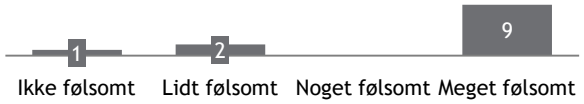
	Konklusioner	*Data*	*Citater*
Formål med AI	• AI bruges typisk til flere formål samtidig • Blandt flere virksomheder anvendes en AI-løsning i direkte interaktion med eksterne parter, fx kunder	**Formål med brug af kunstig intelligens[1]** 	*"Brugen af maskinlæring bliver en central del af vores forretning, og vi har flere digitale lighthouses"* *"Vi har en del projekter med selvlærende elementer på bedding"*
Vigtighed af AI	• De fleste virksomheder betegner AI-løsningen som relativt kritisk, men kan stadig operere uden AI-løsningen • AI mest kritisk hos virksomheder, som decideret sælger en AI-løsning	**Vigtighed af AI løsning for drift** 	*"Det vil være yderst kritisk hvis løsningen er utilgængelig"* *"Hvis løsningen går ned, kan vi falde tilbage på gamle procedurer. Det vil koste penge, og ventetiden i kundeservice vil blive længere, men værre er det ikke"*
Følsomhed af data	• Næsten alle adspurgte virksomheder ser deres data som "noget" eller "meget" følsomt	**Følsomhed af data[2]** 	*"Læk af data vil altid være et problem"* *"Vores data er absolut forretningskritisk for vores kunder"*
Modenhed af sikkerhed	• Næsten alle adspurgte virksomheder ser modenheden af deres informationssikkerhedsstrategi som "høj" eller "meget høj"	**Modenhed af informationssikkerhedsstrategi generelt** 	*"Én dårlig sag er én for meget"* *"Vi er ikke top of class, men vi er rimelig godt med"* *"Der er et vist element af tillid i vores tilgang"*
Sikkerhed ift. AI	• Halvdelen af de adspurgte virksomheder har i "høj grad" forholdt sig til sikkerhedsudfordringer ifm. brug af AI	**Forholdt sig til sikkerhedsudfordringer ifm. brug af AI** 	*"Vi har forholdt os til sikkerhed i utrolig høj grad - måske i for høj grad"* *"Vi har implementeret tiltag til alle rimelige grænser - men man kan jo altid gøre mere"*

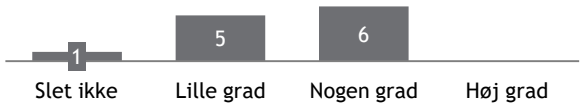
1. Visse virksomheder bruger kunstig intelligens til flere forskellige formål. 2. Hvis virksomheder både behandler mere og mindre følsomt data, angiver vi her følsomheden af det mest følsomme data. Kilde: Interviews med 16 danske virksomheder

Adspurgte myndigheder behandler meget følsomt data i ikke-kritiske AI-løsninger; myndighederne vurderer selv høj modenhed af sikkerhedsløsninger

	Konklusioner	*Data*	*Citater*
Formål med AI	• "Automatisering af processer" mest udbredt • "Personalisere indhold" er ikke et udbredt formål, og ud over chatbots er AI-løsningen sjældent i direkte interaktion med brugere	**Formål med brug af kunstig intelligens[1]**	*"Vi har en chatbot sat i produktion, så har vi testet nogle løsninger og har nogle andre, som vi har en ambition om at sætte i drift"* *"Vi har forskellige proof-of-concepts"*
Vigtighed af AI	• AI-løsningerne er sjældent kritiske for driften, men bliver formentlig sværere at undvære på sigt i takt med at de varetager vigtigere funktioner	**Vigtighed af AI løsning for drift**	*"Pengene kommer til at mangle i kassen, hvis løsningen er ude af drift"* *"Det ville ikke være særlig kritisk lige nu men måske mere på sigt"*
Følsomhed af data	• For næsten alle myndigheder behandles data af meget følsom karakter	**Følsomhed af data[2]**	*"Det er potentielt meget følsomt"* *"Noget af vores data er meget følsomt, andet er ligegyldigt"*
Modenhed af sikkerhed	• Alle adspurgte myndigheder vurderer modenheden af deres informationssikkerhed til at være "høj" eller "meget høj"	**Modenhed af informationssikkerhedsstrategi generelt**	*"Vores modenhed er enormt høj, vi har et virkelig højt niveau"* *"Vi har en fornemmelse af, at vores modenhed er høj, vi har ikke set samme modenhed hos andre"*
Sikkerhed ift. AI	• Sikkerhed ifm. brug af AI falder som regel ind under generel informations-sikkerhed • I visse tilfælde implementeres få yderligere tiltag	**Forholdt sig til sikkerhedsudfordringer ifm. brug af AI**	*"Vi har haft sikkerheden bygget ind fra start"* *"Vi har ikke implementeret nogen særlige tiltag - men det har givet anledning til at genbesøge vores it-sikkerhed helt generelt"*







1. Visse myndigheder bruger kunstig intelligens til flere forskellige formål. 2. Hvis myndigheder både behandler mere og mindre følsomt data, angiver vi her følsomheden af det mest følsomme data. Kilde: Interviews med 13 danske myndigheder (én myndighed udeladt ovenfor pga. manglende information)

VITA-data indikerer, at danske virksomheder, som bruger AI, har rapporteret markant flere brud end andre – men også tager sig flere sikkerhedsforbehold

	Virksomheder der bruger AI	Alle virksomheder	
Har rapporteret sikkerhedsbrud (%) *Muligvis underestimeret grundet selvevaluering (lavt ift. andre kilder)*	11	6	+5
Stigende investering i sikkerhed seneste år (%)	59	35	+24
Brug af sikkerhedsforanstaltning... (%)			
• Informationssikkerhedspolitik	74	51	+24
• Retningslinjer for medarbejdere	83	62	+21
• Uddannelse og træning	57	31	+26
• Risikoanalyse	66	42	+24
• Avancerede tekniske tiltag	59	33	+26
• Krav til leverandører	73	51	+22
Identificeret som barriere for informationssikkerhed... (%)			
• Har oplevet begrænsninger generelt	13	6	+7
• Manglende kompetencer[1]	29	48	-19
• Usikkerhed om gevinst ved investering[1]	26	45	-19
• Manglende økonomiske resourcer[1]	29	30	-1
• Mangel på specifikke løsninger på markedet[1]	23	30	+2
• Andre forhindringer[1]	67	49	+18
	n = 342	*n = 3.954*	

Øvrige konklusioner

- Blandt virksomheder, der har rapporteret sikkerhedsbrud i det forgangne år, har 17% oplevet begrænsninger ift. at anvende sikkerhedsløsninger - blandt virksomheder generelt er andelen 6%
- Andelen af virksomheder, som har investeret mere i informationssikkerhed, er steget fra 2017 til 2018 både for virksomheder generelt (28% ⟶ 35%) og virksomheder der bruger AI (52% ⟶ 59%)
- 52% af virksomheder, som har rapporteret brud i det forgangne år, har samtidig øget deres investering i sikkerhed - ca. samme andel som året før
- Andelen af virksomheder, der bruger AI, er højere for større virksomheder end mindre virksomheder
- Virksomheder, der bruger AI, bruger det i større omfang i deres drift (85%) end i deres produkt eller service (43%)

1. Baseret på lavt antal besvarelser; 49 for virksomheder med AI og 328 for alle virksomheder. Note: Besvarelser vægtet ift. total population; *n* = 3.954 er det faktiske antal besvarelser mens *n* = 342 er det estimerede *n* for den vægtede sub-population af virksomheder, der bruger AI. Kilde: VITA-data fra Danmarks Statistik

Kapitel 2

AI som forsvar

Hvordan AI kan anvendes til at effektivisere informations-sikkerheden

Sikkerhedsløsninger med brug af AI identificeres og rangeres pba. deres modenhed, og udviklingen på området beskrives mhp. at perspektivere til den danske it-sikkerhedsbranche

Tilgængelige AI-baserede sikkerheds-løsninger

- Kortlægning og kategorisering af nuværende løsninger ift. formål og modenhed
- Scenarier for den fremtidige udvikling på området

Status for den danske IT sikkerheds-branche

- Kortlægning af den danske IT sikkerhedsbranche i dag samt brugen af AI løsninger
- Vurdering af fremtidsperspektiverne for branchen

Gevinster ved og barrierer for disse løsninger

- Identifikation af gevinster og barrierer pba. interviews – for både brugere og leverandører
- Prioritering af barrierer for udvikling og anvendelse af AI-baserede sikkerhedsløsninger

Identificerede AI-baserede sikkerhedsløsninger kategoriseres i tre grupper pba. deres formål

Ved at scanne store mængder data og derigennem kortlægge virksomhedens infrastruktur kan AI systemer komme med forudsigelser om, hvor et angreb mest sandsynligt vil indtræffe, således at forebyggende tiltag kan implementeres.

Identificerer angreb vha. fleksible regler ift. hvad der konstituerer 'normal' brug af fx netværk. AI-systemer kan således identificere unormal adfærd uden en fast definition for hvad der er 'unormalt' - og opdaterer løbende sin egen definition.

Hjælper i tilfælde af angreb ved fx at isolere ramte områder af et netværk, forsøge at omdirigere angrebet væk fra vigtige områder, eller skabe en virtuel kopi af netværket i et forsøg på at opholde angrebet mens angriberen identificeres.

Kilde: BCG (2018): Artificial intelligence is a threat to cybersecurity. It's also a solution; BCG analyse; ekspert interviews; NIST

Metode

Kortlægningen af sikkerhedsløsningerne er baseret på omfattende litteratursøgning (akademiske artikler, intl. sikkerhedsorganisationer, udbyderhjemmesider mv.), ekspert-interviews samt nyhedsartikler.

Opdeling af sikkerhedsløsningerne i hhv. *forebyggelse & beskyttelse*, *opdagelse* og *respons* følger NIST's 'Cybersecurity Framework'.

Løsningers modenhed er vurderet af eksperter på området baseret på nedenstående kriterier. Modenhed af kategorierne er verificeret ved hjælp af eksisterende litteratur.

Løsningers **modenhed** vurderes på en skala fra 1-5 pba. følgende kriterier:

- Hvor længe har løsningen eksisteret?
- Er den udbredt nationalt?
- Findes der standarder for konfigurering/anvendelse?
- Findes der undervisnings-/kursusmateriale i anvendelse?
- Betragtes den som *best practice* blandt eksperter?
- Er leverandør-/produktporteføljen på markedet stor?

Forebyggelse og beskyttelse

Løsningers **modenhed** vurderes på en skala fra 1-5 pba. følgende kriterier:

- Hvor længe har løsningen eksisteret?
- Er den udbredt nationalt?
- Findes der standarder for konfigurering/anvendelse?
- Findes der undervisnings-/kursusmateriale i anvendelse?
- Betragtes den som *best practice* blandt eksperter?
- Er leverandør-/produktporteføljen på markedet stor?

AI styrker sikkerheden allerede inden angreb gennem adfærdsanalyse og skærpet adgang

Modenhed	Løsning	Beskrivelse
3	Identity and access management	• Sikrer at kun de rette vedkommende har adgang til information, fx vha. passwords og biometrisk scanning
	Malware prevention	• Beskytter mod software designet til at kompromittere computere eller netværk, fx virusser og ransomware, hvor AI øger sandsynligheden for opdagelse før infiltrering
	Intrusion prevention	• Beskytter bagvedliggende infrastruktur. Er ofte indbygget i firewalls eller implementeret som en teknologi i sig selv
2	Vulnerability management / virtual patching	• Identificerer sårbarheder pba. analyse af fx telemetri. AI kan øge analysevolumen samt hjælpe med prioritering
1	Application protection	• Beskytter applikationer mod angreb, og kan vha. AI automatisk opdatere beskyttelsen til at reflektere ændringer i applikationen
	Endpoint protection	• Som *intrusion prevention*, men beskytter enkeltstående enheder (fx mobiler, bærbare computere, tablets)
	Adaptive authentication	• Bekræfter at en bruger ikke er en robot eller hacker fx vha. biometrisk data eller koder, men med individuelt tilpassede krav baseret på AI-baseret risikoanalyse af fx lokation
	Data loss prevention	• Finder mønstre i data og blokerer dem, før de forlader netværket (fx CPR numre), hvor AI er med til at øge både hastigheden og effektiviteten

Kilde: Ekspert interviews; BCG analyse; nyhedsartikler

Opdagelse

Løsningers **modenhed** vurderes på en skala fra 1-5 pba. følgende kriterier:

- Hvor længe har løsningen eksisteret?
- Er den udbredt nationalt?
- Findes der standarder for konfigurering/anvendelse?
- Findes der undervisnings-/kursusmateriale i anvendelse?
- Betragtes den som *best practice* blandt eksperter?
- Er leverandør-/produktporteføljen på markedet stor?

AI øger evne til at identificere angreb pba. forskellige former for mønstergenkendelse

Modenhed	Løsning	Beskrivelse
4	Governance, risk and compliance analysis	• Vurdering af risiko- og compliance-niveau ud fra indsamlede metrikker og læring fra *best case* scenarier
3	Malware detection	• Identificering af ondsindet software baseret på opførsel i *sandbox* (dvs. isolerede) miljøer eller ved analyse af softwarens kode førend softwaren eksekveres (fx vha. signaturer)
	Intrusion detection	• Analyserer netværk primært baseret på identifikation af signaturer i koden, og kan intelligent opdateres vha. maskinlæring
2	User behaviour analysis	• Analyserer anvendelsesmønstre for aktive brugere for at identificere uregelmæssig adfærd, hvilket kan gøres i langt højere omfang og med større præcision med AI
	Abnormality detection	• Identificerer om aktivitet adskiller sig fra normale brugsmønstre
	Industry anomaly detection[1]	• Beskytter primært industrielle netværk med SCADA systemer (fx el- og vandforsyning, lufthavne)
	Fraud detection[2]	• Opdager svindel ved at analysere mønstre der indikerer uønskede transaktioner
1	Data loss detection	• Scanner automatisk efter læk af følsom data baseret på evaluering af de faktiske data der har forladt netværket – med øget præcision og hastighed vha. AI

1. Meget industrispecifikt – og kun brugt i store organisatoner. 2. Lovpligtigt i finanssektoren - anvendes ikke andre steder. Kilde: Ekspert interviews; BCG analyse; nyhedsartikler

AI hjælper med at prioritere trusselshåndtering efter angreb og øger responshastighed

Respons

Modenhed	Løsning	Beskrivelse
2	Security orchestration, automation and response	• Identificerer andre angreb vha. andre it-løsninger, men iværksætter så selv en automatiseret protokol for at adressere angrebet, og kan med IA trænes til at duplikere en analytikers handlinger ved angreb
	Security incident analysis	• Analyserer tidligere angreb og alerts (interne og eksterne) for at identificere koordinerede angreb
1	Self-defending networks	• Automatiserer netværkshandlinger som omkonfigurering, segmentering og nedlukning ved angreb, for at begrænse skaden og opretholde netværksfunktionalitet
	Automated incident response	• Igangsætter automatisk handlinger til at adressere angreb, men prioriterer også angreb, så analytikere kan fokusere på de største trusler

Visse tiltag i *Forebyggelse & beskyttelse* også relevante her da aktiv beskyttelse = respons

Løsningers **modenhed** vurderes på en skala fra 1-5 pba. følgende kriterier:

- Hvor længe har løsningen eksisteret?
- Er den udbredt nationalt?
- Findes der standarder for konfigurering/anvendelse?
- Findes der undervisnings-/kursusmateriale i anvendelse?
- Betragtes den som *best practice* blandt eksperter?
- Er leverandør-/produktporteføljen på markedet stor?

Kilde: Ekspert interviews; BCG analyse; nyhedsartikler

Løsninger og modenhed identificeret i analysen modsvarer den eksisterende litteratur – *Opdagelse* mest udbredt og hvor AI er mest anvendeligt

	Denne analyse			*Eksternt benchmark*
	Løsninger	Modenhed		*Udnyttelsesgrad blandt organisationer i dag*[1]
	(antal)	*(sum)*	*(gns.)*	
Forebyggelse & beskyttelse	8	15	1,9	34% Forebyggelse & beskyttelse
Opdagelse	8	19	2,4	51% Opdagelse
Respons	4	6	1,5	18% Respons

AI anvendelse

> *Det er især **opdagelse**, hvor der er plads til AI, og hvor det allerede bliver brugt meget*

> *AI er god til at tage store mængder data og finde strukturer og mønstre i det, og derfor særligt anvendelig til **opdagelse***

> *AI fungerer ikke som et menneske. Det er vigtigt at indse. Folk er nødt til at forstå hvad AI kan og ikke kan bruges til*

> *AI kan give en fordel, men hacking er meget konkret - det skal løses konkret*

1. Baseret på svar fra 850 ledere fra 10 lande (inkl. USA, Sverige) og flere forskellige industrier, på spørgsmålet "Vurdér din organisations brug af AI i cybersecurity for the følgende områder [høj, medium, lav]" – tallene repræsenterer andelen af respondenter der angav *Høj*.
Kilde: Ekspert interviews; BCG analyse; nyhedsartikler; Capgemini (2019): Reinventing Cybersecurity with Artificial Intelligence

Anvendelsen af AI-understøttede sikkerhedsløsninger er stigende og forventes at blive standarden i fremtiden

Begrænset anvendelse

- Alle adspurgte udbydere tilbød ingen eller få AI-baserede løsninger for 5-10 år siden
- AI-baserede ydelser begrænset til brug af maskinlæring
- Første AI-understøttede ydelser kørte lokalt på computeren - modsat cloud nu
- *"Vi begyndte at tilbyde fuldt AI-understøttede løsninger for 2 år siden. To af vores konkurrenter kunne godt se værdien og fulgte hurtigt trop"*

Anvendt – men lav modenhed

- De fleste internationale og en række nationale udbydere på det danske marked tilbyder AI-baserede løsninger
- Løsningerne er dog fortsat relativt dyre, og er således primært set i større organisationer
- *"Løsninger i dag er ikke baseret på ægte kunstig intelligens men i stedet på maskinlæring"*

Standard på markedet

- De fleste udbydere ser AI som en essentiel del af it-sikkerhed i fremtiden – dog i symbiose med menneskelig intelligens
- It- og fysisk sikkerhed vil muligvis i højere grad smelte sammen i takt med stigende brug af droner og selvkørende biler – forsvaret må følge med
- *"Ingen ikke-AI baserede løsninger om 5-10 år"*
- *"Det [AI] vil være en del af næsten alle produkter"*

Kilde: Interviews med eksperter og it-sikkerhedsudbydere; BCG analyse

Kapitel 2

AI som forsvar

Hvordan AI kan anvendes til at effektivisere informations-sikkerheden

Udbredelsen af de nævnte løsninger blandt danske it-sikkerhedsudbydere kortlægges, og der sættes fokus på fremtidsperspektiver for branchen og de forventede succesfaktorer fremadrettet

Tilgængelige AI-baserede sikkerheds-løsninger

- Kortlægning og kategorisering af nuværende løsninger ift. formål og modenhed
- Scenarier for den fremtidige udvikling på området

Status for den danske IT sikkerheds-branche

- Kortlægning af den danske IT sikkerhedsbranche i dag samt brugen af AI løsninger
- Vurdering af fremtidsperspektiverne for branchen

Gevinster ved og barrierer for disse løsninger

- Identifikation af gevinster og barrierer pba. interviews – for både brugere og leverandører
- Prioritering af barrierer for udvikling og anvendelse af AI-baserede sikkerhedsløsninger

Flere udbydere på det danske marked tilbyder allerede løsninger der anvender AI

AI anvendelse

	Ingen	Delvis	Hovedsagelig
Software	**20%** af udbydere	16% af udbydere	**14%** af udbydere
Rådgivning[1]	**34%** af udbydere	**13%** af udbydere	**3%** af udbydere

70 udbydere i alt

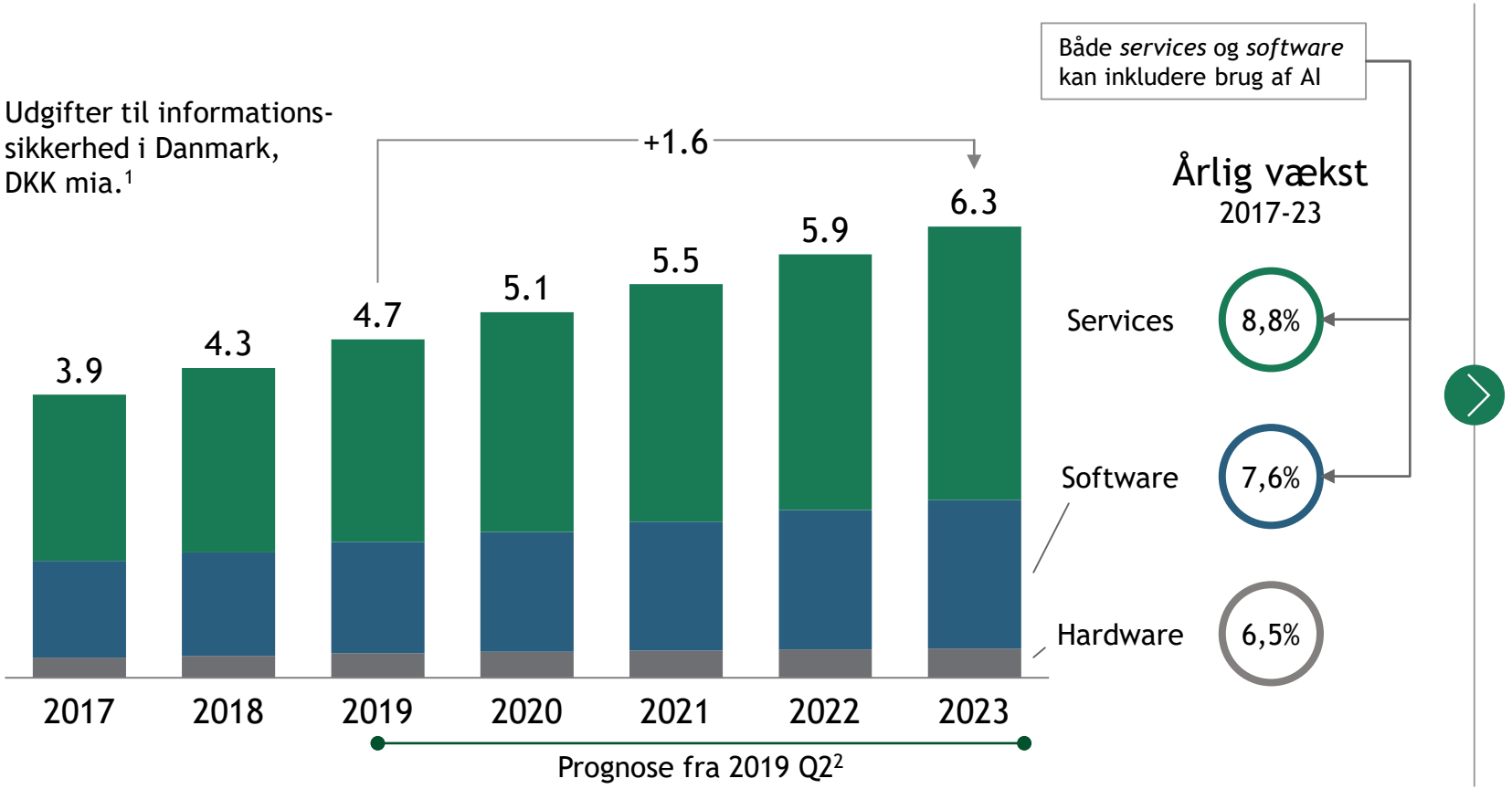
Metode

Analysen er afgrænset til virksomheder med it-sikkerhed som deres primære forretnings-område

Udgangspunktet er Erhvervsstyrelsens kortlægning af den danske it-sikkerhedsbranche fra 2019, som efterfølgende er beriget med informationer omkring produkttype og AI-anvendelsesgrad pba. interviews, information fra udbydernes egne hjemmesider og gennemgang af markedsanalyser

1. 'Delvis' og 'Hovedsagelig' referer for rådgivning til, i hvor høj grad AI-understøttede løsninger italesættes som del af det overordnede løsningstiltag på udbyderens hjemmeside. Kilde: Interviews; udbyderes egne hjemmesider; Erhvervsstyrelsens kortlægning af den danske it-sikkerhedsbranche (2019)

Udgifter til informationssikkerhed i Danmark forventes at stige med ~1,6 mia. DKK over de næste 4 år, hvoraf AI vil udgøre en stadig stigende andel



Fremtidens succesfaktorer

- *Lige nu er penetrationen af AI på it-sikkerhedsmarkedet ca. 20-30%. Jeg vil forvente, at det er 80-85% om 5 år*
- *Nye løsninger bygget på AI står stærkere end legacy produkter, der forsøger at indarbejde AI*
- 48% af de adspurgte estimerer at budgettet til AI-baseret it-sikkerhed vil stige - i gns. med 29% i FY2020[3]
- *Der har manglet en prisvenlig model til danske SME'er, fordi business casen har været for dårlig for mange udbydere*
- I dag automatiserer 1/3 af organisationer 40% eller mere af deres sikkerheds-opgaver – det forventes at stige til 2/3 indenfor de næste 3 år
- *Der er kommet mere fokus på Danmark og København som iværksætterhub*

1. Omregnet fra USD til DKK vha. Gartner Exchange Rate database. 2. Prognose estimeret af Gartner pba. primær og sekundær research, forespørgselsanalyser og interviews med industrieksperter. 3. Baseret på svar fra 850 ledere fra 10 lande (inkl. USA, Sverige) og flere forskellige industrier. Kilde: Ekspert interviews; BCG analyse; Gartner (data); Capgemini (2019): Reinventing Cybersecurity with Artificial Intelligence; ServiceNow (2017): Global CISO Study

Kapitel 2

AI som forsvar

Hvordan AI kan anvendes til at effektivisere informations-sikkerheden

Med udgangspunkt i markedsudbredelsen i Danmark til dato, identificeres de primære gevinster ved og barrierer for øget brug af AI-baserede sikkerhedsløsninger

Tilgængelige AI-baserede sikkerheds-løsninger

- Kortlægning og kategorisering af nuværende løsninger ift. formål og modenhed
- Scenarier for den fremtidige udvikling på området

Status for den danske IT sikkerheds-branche

- Kortlægning af den danske IT sikkerhedsbranche i dag samt brugen af AI løsninger
- Vurdering af fremtidsperspektiverne for branchen

Gevinster ved og barrierer for disse løsninger

- Identifikation af gevinster og barrierer pba. interviews – for både brugere og leverandører
- Prioritering af barrierer for udvikling og anvendelse af AI-baserede sikkerhedsløsninger

Gevinsterne ved AI-baserede sikkerhedsløsninger kan være øget effektivitet, øget præcision og lavere omkostninger

Færre brud

- Genkender mønstre fra tidligere angreb
- Adapterer til nye trusler
- Proaktiv tilgang til trusselshåndtering

"Det vigtige for mig er, at den [AI-løsningen] lærer. At den er bedre i dag, end den var i går, så den fx fanger, at det du er ved at gøre er underligt ift. dine kollegaers ageren"

Flere identificeres

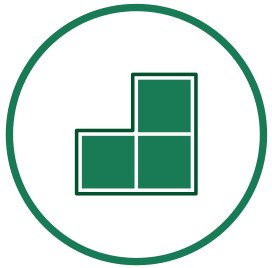
- Større andel af brud identificeres
- Muliggør håndtering af større mængder data
- *Potentielt flere falske positiver absolut*

"Den traditionelle løsning opdagede 60% af truslerne, hvorimod den AI-baserede fangede 95-99%"

Opdages hurtigere

- Automatiseret identifikation og prioritering af mulige trusler
- Lærer fra *best case* eksempler og optimerer processer herefter

AI reducerer tiden det tager at opdage sikkerhedsbrud og trusler med 12% i gns.[1]



Tilpasses brugeren

- Lærer den enkelte brugers adfærd og tilpasser sig derefter
- Personliggør sikkerhedstiltag

"Analyse af brugernes adfærd er centralt i produktet. Det ligger i selve agenten"

Lavere omkostninger

- Prioriterer indsats fra sikkerhedseksperter
- Mindsker tabet ifm. angreb pga. hurtigere respons
- *Høje etablerings-omkostninger*

"Jeg skulle ansætte mange flere, hvis vi skulle tjekke alle hændelser igennem manuelt"

"Det [AI] gør jobbet meget sjovere for medarbejdere, som ikke længere skal reagere på 10.000 alarmer, men kan fokuseres deres tid og energi"

1. Baseret på svar fra 850 execs fra 10 lande (inkl. USA, Sverige) og flere forskellige industrier. Kilde: Ekspert interviews; BCG analyse; nyhedsartikler; Forrester; Breakout Vendors: Security Automation And Orchestration (2017); Capgemini (2019): Reinventing Cybersecurity with Artificial Intelligence

Mangel på kompetencer og problemer med at kvantificere business case ses som største barrierer

Alle adspurgte udbydere og eksperter angav **efterspørgsel** som begrænsende faktor – snarere end udbud

Barriere	Uddybning	Nævnt i interviews
Manglende kompetencer	• Manglende AI-specifik viden indenfor virksomheden • It-sikkerhedsbranchen tiltrækker generelt for få talenter	**100%**
Usikker business case	• Svært at kvantificere business case førend virksomheden har oplevet et angreb • Mangel på eksempler til sammenligning	**80%**
Begrænset ledelsesfokus	• Reaktiv fremfor proaktiv tilgang • Ofte en "*implement and forget*" tilgang til IT-løsninger	**73%**
Lav transparens	• Sikkerhedsniveau uigennemsigtigt for topledelsen • Uvidenhed om hvorvidt nuværende løsninger er forældede • Svært at sammenligne kvaliteten af konkurrerende løsninger	**53%**
Manglende tillid	• Lav tillid til modenheden af AI-baserede løsninger	**27%**
Legacy systemer	• Inkompatibilitet af AI-baseret løsning med virksomhedens eksisterende legacy systemer og fysiske infrastruktur, fx logning, gør implementering langsom og omkostningsfuld	**20%**

Kilde: Interviews med eksperter og udbydere; BCG analyse

Kapitel 3

AI som angrebsmiddel

Hvordan hackeres brug af AI kan ændre trusselsbilledet, og hvordan dette kan adresseres

Udbredelsen af AI-baserede cyberangreb beskrives med fokus på ét konkret eksempel for at fremhæve relevansen for danske virksomheder og myndigheder

Nye typer af AI-baserede cyberangreb

- Udbredelsen af AI-baserede cyberangreb i dag og beskrivelse af problematikken ifm. detektion
- Detaljeret eksempel på AI-angreb

Indflydelse på det samlede it-risikobillede og effekt på sikkerhedstiltag

- Revideret risikobillede med effekten af AI-angreb på volumen, hastighed mv.
- Fokus på udviklingen indenfor AI og cyberangreb
- Effekten af AI-baserede angreb på sikkerheds-løsninger og tiltag samt deres relevans

Der er kun få detekterede eksempler på brug af AI i cyberangreb

Udbredelse

Der er til dato kun offentliggjort få angreb med AI, men det betyder ikke nødvendigvis, at de ikke finder sted:

- Da AI bl.a. kan bruges til at maskere cyberangreb, detekteres potentielt færre
- Ofre for cyberangreb kender ikke altid til angrebenes tekniske karakteristika, fx brugen af AI
- Cyberangreb offentliggøres ikke nødvendigvis, og tekniske elementer beskrives sjældent

Eksempler

Identificerede cyberangreb hvor AI er inkorporeret:

- Tysk virksomhed frarøvet 2 mio. DKK ved CEO fraud angreb (læs mere på næste side)
- MedTech-selskab angrebet ved brug af AI, hvor AI'en skjulte sin aktivitet over en lang periode
- Advokatselskab angrebet ved brug af AI, som på få minutter infiltrerede over 20 ansattes interne kommunikation

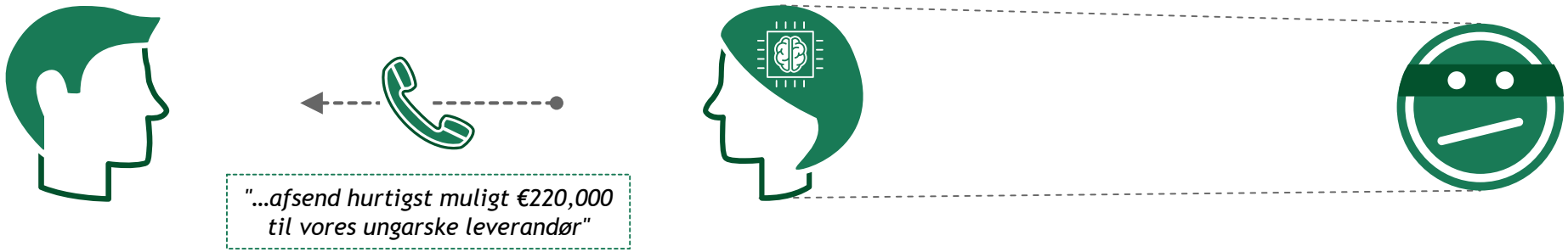
Kilde: Darktrace (2018): AI-Driven Cyber-Attacks; BCG analyse

Eksempel: AI-genereret stemmeefterligning benyttet til at frarøve tysk firma for 2 mio. DKK

Offer
Direktør for datterselskab

Våben
"Direktør for moderselskab"

Gerningsperson
Profitmotiveret cyberkriminel



1) Cyberangrebet begynder hos gerningspersonen, som identificerer et offer og kortlægger offerets relationer
2) Der udvælges en relation til at agere våben, og et neuralt netværk fodres med læringsdata (telefonopkald eller offentlige optagelser), hvorved en replikation af direktørens stemme genereres (dvs. personalisering)
3) Det afgørende opkald foretages, og kapitaloverførslen initieres

> *Datterselskabets britiske direktør genkendte sin chefs svage tyske accent samt det karakteristiske toneleje, da han hørte stemmen i telefonen*

FORSVARETS EFTERRETNINGSTJENESTE

Udbredelsen af denne type cyberangreb bekræftes af Forsvarets Efterretningstjeneste, som har øget opmærksomheden på anvendelsen af neurale netværk i cyberangreb, da disse er den drivende kraft bag automatiseret stemme-, video- og tekstgenerering. Der fokuseres bl.a. på brugen af *deep fakes* i forbindelse med manipulation af demokratiske processer.

Kilde: WSJ (2019): Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case; interview med Forsvarets Efterretningstjeneste

Kapitel 3

AI som angrebsmiddel

Hvordan hackeres brug af AI kan ændre trusselsbilledet, og hvordan dette kan adresseres

Med udgangspunkt i et eksperiment illustreres effekten på risikobilledet ved introduktionen af AI i angreb, hvorefter det illustreres, hvordan organisationer bør tage højde herfor

Nye typer af AI-baserede cyberangreb

- Udbredelsen af AI-baserede cyberangreb i dag og beskrivelse af problematikken ifm. detektion
- Detaljeret eksempel på AI-angreb

Indflydelse på det samlede it-risikobillede og effekt på sikkerhedstiltag

- Revideret risikobillede med effekten af AI-angreb på volumen, hastighed mv.
- Fokus på udviklingen indenfor AI og cyberangreb
- Effekten af AI-baserede angreb på sikkerheds-løsninger og tiltag samt deres relevans

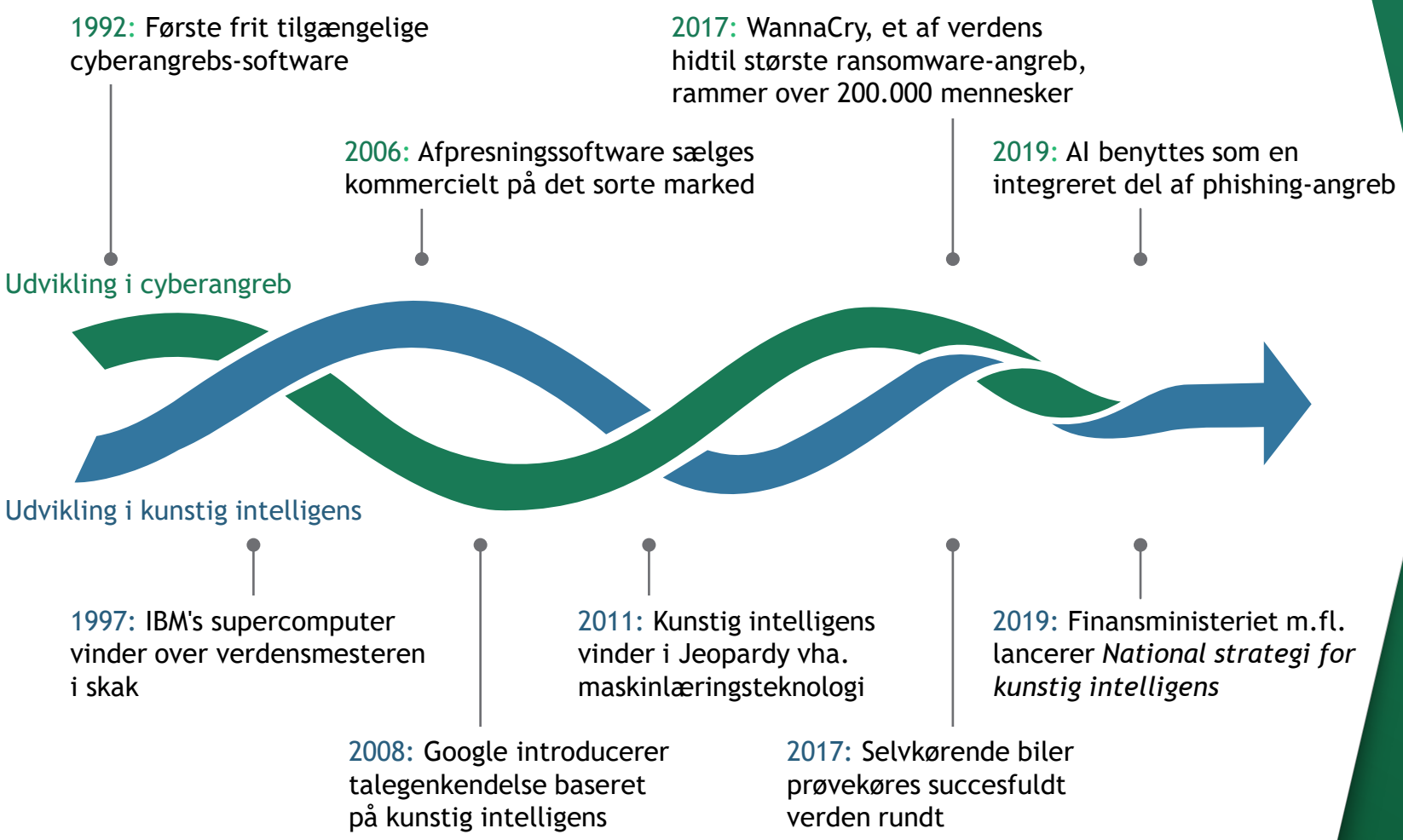
Kunstig intelligens øger truslen fra typiske cyberangreb – såsom phishing – ved både at øge hastigheden, kvaliteten samt variationen af angreb

Eksempel på phishing-angreb

	Udvælgelse	Research	Personalisering
Eksisterende teknologi	Kunstig intelligens muliggør **hurtig** identifikation samt udvælgelse af optimale phishing-mål, fx ved genkendelse af særlige karakteristika vha. maskinlæring	Kunstig intelligens kan foretage **bredere** og **hurtigere** analyser af tilgængelig data om offeret, hvorved både **kvaliteten** samt **hastigheden** af angrebet forbedres	Kunstig intelligens foretager automatisk skriftlig efterligning af nære venner og/eller kollegaer ud fra tilgængelig data om deres sprogbrug, hvorved både **hastigheden** samt **kvaliteten** højnes
Eksperiment: AI vs hacker[1]	*Tilgængelig AI kan gøre hastigheder af identifikation **4x hurtigere***	*Succesraten for tilgængelig kunstig intelligens vurderes at være over **5x højere***	*Tilgængelig AI genererer personaliseret indhold mindst **4x hurtigere** end et menneske*
Nye variationer	Maskinlæring medfører, at udvælgelses-kriterierne løbende bliver mere sofistikerede som følge af, at AI'en lærer af sine tidligere succesfulde angreb	Deep learning kan potentielt sammenkoble data fra fx sociale og professionelle netværk mhp. at kortlægge offerets personlighed i en hidtil uset grad	Kunstig intelligens muliggør både stemme- og videoefterligning; efterligning af stemmeføring er allerede automatiseret, fx ved Google Duplex

1. Thomas Brewster (2016): Who's Better At Phishing, Me Or Artifical Intelligence. Kilde: Miles Brundage et al (2018): The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation; Darktrace (2018): AI-Driven Cyber-Attacks; World Economic Forum (2019): 3 ways AI will change the nature of cyber attacks

Kunstig intelligens bliver en integreret del af cyberangreb i takt med teknologiens modning



Kilde: BCG analyse; Miles Brundage et al (2018): The Malicious Use of Artificial Intelligence; Protiviti (2017): The Evolution of Autonomous Vehicles; Danmarks Radio (2018): Det har robotterne lært, mens du har levet

Mulige fremtidsscenarier

Øget frekvens & kvalitet
Automatiseret identifikation af sårbarheder samt sofistikerede angrebsteknikker eliminerer det eksisterende tradeoff mellem frekvens og kvalitet, fx ved større udbredelse af spear-phishing

Politisk manipulation
AI-genereret indhold kan målrettes, personaliseres og distribueres via sociale medier, fx via 'deep fakes' såsom AI-genererede videoer med falske udtalelser

Fysiske konsekvenser
Kombinationen af selvstyrende fartøjer og udbredelsen af 'Internet of Things' øger risikoen for cyberangreb med fysiske konsekvenser, fx manipulation af læringsdata for selvkørende biler

"AI medfører ikke bare hurtigere og klogere angreb. Vi er formentlig slet ikke i stand til at begribe, hvordan AI vil transformere cyberangreb."

Justin Fier, Director, Darktrace

Brug af AI i cyberangreb giver ikke nødvendigvis anledning til nye tiltag, men øger vigtigheden af visse tiltag

	Forebyggelse & beskyttelse *Før angrebet sker*	Opdagelse *Når angrebet sker*	Respons *Efter angrebet er sket*
	AI-baserede cyberangreb kan forventes at være så avancerede, at virksomheder ikke kan vide sig sikre på deres informationssikkerhedsløsninger, hvorfor fokus i stedet vil være på at implementere barrierer, som øger omkostningerne for potentielle trusselsaktører og derved mindsker attraktiviteten som offer • **Risikostyringens ramme** opdateres mhp. fastholdelse af organisationens risikoappetit i forhold til det potentielt større ressourceforbrug som følge af AI • **Risikovurdering** af nye typer risici forbundet med AI-cyberangreb for at kortlægge og tydeliggøre organisations sårbarheder som følge af denne type angreb		
Organisatorisk	Kunstig intelligens fordrer mere omfattende forebyggelse af cyberangreb, da nye usete variationer af cyberangreb kan forekomme • Informationsniveauet om AI-cyberangreb kan øges ved **awareness-kampagner**, fx ifm. øget risiko for personaliseret spear-phishing • Da sårbare netværksadgange kan udnyttes af AI-cyberangreb, kan **styring af brugere og adgange** fx opgraderes med biometrisk udstyr	AI øger kompleksiteten ved opdagelse af cyberangreb, da neurale netværk kan maskere sig som almindelig netværksaktivitet • Evnen til at kende sande positive angreb fra falske positive kan forbedres ved **træning af medarbejdere**	Kunstig intelligens medfører umiddelbart ikke nye prioriteringer indenfor responsmæssige tiltag, men det stiller højere krav til fx planer for kommunikation og systemgendannelse, da et eventuelt angreb kan være mere omfangsrigt end normalt
Teknisk	Kunstig intelligens medfører højere risiko for fjendtlig identifikation af egne sårbarheder, da enorme datamængder analyseres hastigt • Omfanget af sårbarheder, og dermed konsekvensen ved angreb, kan minimeres ved at **etablere netværksgrænser og barrierer** • Da effektiviteten af cyberangreb generelt øges ved brug af AI, vil det være givtigt at styrke **datakryptering**, så skaden ved et eventuelt angreb begrænses	AI medfører behov for nøje granskning af egen dataaktivitet, da AI'en kan genkende og efterligne ofrets normale aktivitet • Der stilles højere krav til **logning af aktivitet**, da der skal indsamles mere data til mere sofistikeret monitorering • AI kan inkorporeres i **monitoreringen af aktivitet,** hvilket øger evnen til at opdage såkaldte *low & slow* angreb, hvor hackerens aktivitet efterligner offerets	AI medfører, at cyberangreb inficerer it-systemer hurtigere, da søgningen efter værdifuld data bliver langt hurtigere • Da få sekunder kan have store konsekvenser, når AI er inkorporeret i et angreb, vil **automatisk nedlukning og isolation af inficerede systemer** være særligt vigtigt

Kilde: BCG analyse

Disclaimer

The services and materials provided by Boston Consulting Group (BCG) are subject to BCG's Standard Terms (a copy of which is available upon request) or such other agreement as may have been previously executed by BCG. BCG does not provide legal, accounting, or tax advice. The Client is responsible for obtaining independent advice concerning these matters. This advice may affect the guidance given by BCG. Further, BCG has made no undertaking to update these materials after the date hereof, notwithstanding that such information may become outdated or inaccurate.

The materials contained in this presentation are designed for the sole use by the board of directors or senior management of the Client and solely for the limited purposes described in the presentation. The materials shall not be copied or given to any person or entity other than the Client ("Third Party") without the prior written consent of BCG. These materials serve only as the focus for discussion; they are incomplete without the accompanying oral commentary and may not be relied on as a stand-alone document. Further, Third Parties may not, and it is unreasonable for any Third Party to, rely on these materials for any purpose whatsoever. To the fullest extent permitted by law (and except to the extent otherwise agreed in a signed writing by BCG), BCG shall have no liability whatsoever to any Third Party, and any Third Party hereby waives any rights and claims it may have at any time against BCG with regard to the services, this presentation, or other materials, including the accuracy or completeness thereof. Receipt and review of this document shall be deemed agreement with and consideration for the foregoing.

BCG does not provide fairness opinions or valuations of market transactions, and these materials should not be relied on or construed as such. Further, the financial evaluations, projected market and financial information, and conclusions contained in these materials are based upon standard valuation methodologies, are not definitive forecasts, and are not guaranteed by BCG. BCG has used public and/or confidential data and assumptions provided to BCG by the Client. BCG has not independently verified the data and assumptions used in these analyses. Changes in the underlying data or operating assumptions will clearly impact the analyses and conclusions.


BCG
BOSTON
CONSULTING
GROUP
bcg.com