AI kontra databeskyttelse

Konsekvenserne af persondataforordning 2016/679 for kunstig intelligens

Patrick Lykke Knudsen (2013-4680) & Pernille Ask Korsgaard (2013-4677)
Vejleder: Charlotte Bagger Tranberg

Kandidatspecialeafhandling | Juridisk Institut | Aalborg Universitet
Antal anslag: 167.964 | 16-05-2018

1 Abstract

This thesis explores which issues and challenges developers and users of AI face by complying with Regulation 2016/679 on data protection (GDPR), and thus, the hypothesis is that GDPR introduces significant difficulties. The thesis is meant to be a tool for lawyers, engineers and managers alike to identify the rules that AI need to abide by.

The structure of the substance of the thesis consists of an initial presentation of the technical aspects of artificial intelligence with relevant examples and illustrations. This is followed by a systematic and thorough examination of GDPR's provisions on its principles as a starting point. The analysis of the thesis includes a review of current technologies that can potentially alleviate some of the challenges. The conclusion summarizes the discovered issues and challenges from the analysis and compares them and their connections in addition to evaluating their possible solutions where possible.

Some of the inherent challenges for developers and users of AI are related to the similarities with big data where voluminous data sets are preferred to improve precision of the models' predictions and decisions. Other challenges originate from the complexities of these types of algorithms – especially in online models and deep learning.

GDPR requires controllers to specify the purposes of the collections of data and to minimize the amount. Controllers may not be able to specifically determine the results and thus the purposes of collections. Controllers must take the contexts of the processes into account to determine the suitable specificity.

From the stated purposes, the controllers must assess whether the personal data is necessary. It is not always possible to determine when data is necessary for an AI to process as the statistical analysis is automated. The AI will determine causalities that it can learn from, but only then does it know which clusters of data were utilized.

However, to delete or even correct personal data within an AI is no small task. This issue concerns primarily the situation where data turns out to be unnecessary for the purposes, the rights to be forgotten and rectification. From a technical stand point, it may not be possible at all to delete or rectify data because of the complexities of especially deep learning, and GDPR's lack of clarity on the required measures regarding this issue displays its shortcomings.

A major concern regarding the processing of personal data by AI relates to the lack of transparency of the processing as well as decisions. Automated decision-making requires demonstration of the "*logic*", which is also a general requirement throughout the regulation and as such calls for the implementation of methods for the AI to explain itself – a measure that would solve many of the issues concluded upon in this thesis.

To ensure appropriate security, the controllers must identify both assets and threats. Assets should be protected, while threats should be protected against. Certain threats are specific to different types of AI, namely reverse engineering, adversarial injections and backdoors.

Online models and AI using deep learning techniques face greater challenges; hence, the specific AI should be identified. There are certain possible solutions to these issues; especially the use of pseudonymization, technological approaches to encryption and particularly aggregation as well as the classification of the AI as a statistical or scientific purpose. Although the most ideal solution tends to be the avoidance of personal data altogether, of which there are different methods to carry out without compromising the statistical results.

GDPR does not set impossible responsibilities for controllers to comply with. Online models and deep learning AI do face challenges where solutions are currently in a grey area. Controllers should however be better equipped to face the challenges of the regulation by considering the conclusions of this thesis.

Indholdsfortegnelse

2 Indledning

Da databeskyttelsesdirektivet i 1995 blev vedtaget, var den eneste tilstedeværelse af AI i danskernes erindring sandsynligvis en computer, der kan spille skak *næsten* lige så godt som professionelle samt fantasifulde sci-fi-film. AI er imidlertid blevet et håndgribeligt fænomen, som i stigende grad forudsiges at ramme samtlige industrier og skabe disruption, som det populært kaldes. AI er drivkraften bag en stigende mængde af ydelser, som vi anvender hver dag såsom Netflix og Spotifys algoritmer, der anbefaler os film og sange med næsten foruroligende nøjagtighed – foruroligende, da man kun kan forestille sig, hvilken uhåndgribelig mængde af vores data, der ligger bag disse algoritmer. Med udviklingen af AI som teknologi løber vi sågar ind i, at selv ikke udviklerne bag kan finde rundt i, hvad der faktisk sker med vores data. Dette har skabt bekymring for, at anvendelse af AI forårsager mindre ansvarlighed i behandlingen af vores oplysninger. Vi ved eller forstår faktisk ikke, hvad der ligger bag de afgørelser og handlinger, som vi overlader til en algoritme i effektivitetens navn.

Hvordan ved vi, at en AI ikke behandler oplysninger om alle vores dybt personlige oplysninger? Hvordan ved vi, at vores oplysninger er tilstrækkeligt sikrede af AI'en? Hvordan ved vi, at AI'en træffer en rimelig afgørelse, når den skal vurdere, om vi kan modtage et lån? Databeskyttelsesdirektivet fra 1995 og dermed persondataloven formåede tilsyneladende ikke at lette disse bekymringer. Som svar på dette har vi nu fået Europa-Parlamentets og Rådets forordning 2016/679 om databeskyttelse, GDPR, som stiller større krav til dataansvarlige og databehandlere. Men er GDPR mere fremtidssikret end sin forgænger, eller stiller den måske modsætningsvist helt uhensigtsmæssige krav til udviklingen og anvendelsen af AI?

2.1 Problemformulering

I denne afhandling finder læsere dermed en gennemgang af GDPR's konsekvenser for lige netop emnet, artificial intelligence, med følgende problemformulering som udgangspunkt:

"*Hvilke problemstillinger og udfordringer står udviklere og anvendere af AI overfor ved overholdelse af forordning 2016/679 om databeskyttelse?*"

Hypotesen er, at GDPR er streng og ikke tager hensyn til de tekniske udfordringer, der kan opstå i forbindelse med dens implementering i tilfælde af højteknologiske discipliner. Denne problemformulering er valgt, idet nærværende afhandling har til formål at afklare, hvilke konsekvenser forordningen har for udvikling og anvendelse af AI samt, hvad udviklere skal være opmærksomme på fremadrettet. Denne afhandling sætter sig for at være kontekst-agnostisk, således den ikke blot er anvendelig i den umiddelbare fremtid, men fremadrettet vil fortsætte med at være et værktøj, som udviklere kan anvende uanset, hvilket niveau den teknologiske udvikling har opnået på det pågældende tidspunkt.

Som afhandlingen vil illustrere, er udvikling af AI en højteknologisk disciplin, som åbenlyst ikke kan adskilles fra de teknologiske omstændigheder, hvorved løsningerne til de problemstillinger og udfordringer – som GDPR medfører – vil ændre sig med tiden. Afhandlingen vil derfor fokusere på at analysere GDPR i forhold til dens konsekvenser for udvikling og anvendelse af AI for at finde de problemstillinger og udfordringer. Disse kan udviklere og anvendere så selv iagttage med hensyn til deres konkrete omstændigheder og de pågældende teknologiske muligheder. Afhandlingen perspektiverer dog med at inddrage, hvilke teknologiske muligheder, der findes i dag.

2.2 Generelle metodiske overvejelser

Afhandlingens struktur følger en metode, hvor GDPR's bestemmelser gennemgås mere eller mindre slavisk med det formål at analysere dens bestemmelser for at finde og bearbejde de problemstillinger og udfordringer, der gør sig gældende for AI.

Efterfølgende perspektiveres der til potentielle løsninger af teknologisk karakter. Perspektiveringen kommer således før konklusionen, da dens formål er at opstille et alternativt perspektiv. Udfordringen ved problemformuleringen er, at den lægger op til en teoretisk jagt på at finde problemer, hvor der måske ikke er nogle. Afsnit 6 om de teknologiske løsninger putter problemerne i et realistisk perspektiv og medfører mere skepsis i analysen.

Konklusionen samler derfor ikke blot op på problemstillingerne og udfordringerne, men gennemgår deres sammenhænge og løsningernes indflydelse derpå. Perspektiveringens placering er dermed den eneste undtagelse til anvendelsen af Blooms taksonomiske niveauer, som er forsøgt udført for at skabe kohærens samt logisk gyldig argumentatorisk struktur. Denne struktur skyldes, at deduktion kræver en åbensindet tilgang til alle facetter af GDPR for afgøre udfordringer.

Der findes ikke meget litteratur om netop dette emne og især ikke i en dansk kontekst. Størstedelen af benarbejdet er foretaget selvstændigt i afhandlingen, hvorved Toulmins argumentationsmodel er blevet anvendt for at underbygge afhandlingens påstande i stedet for blot at inddrage kilder – altså juridisk metode.

Denne metode kræver, at læseren har en vis forståelse for, hvad AI eller kunstig intelligens er, samt hvilke typer og karakteristika der findes. Dette skyldes, at AI er en højteknologisk disciplin, der kræver en vis indsigt i datalogi og IT. Derfor indledes afhandlingen med en redegørelse for det tekniske begreb, AI, med en simpel struktur og adskillige eksempler. Redegørelsen er inspireret af det norske datatilsyns særdeles saglige rapport fra januar 2018, "*Kunstig intelligens og personvern*".

For læsevenlighedens skyld indeholder bilag 1 en ordliste, som med fordel kan være ved hånden for at undgå tvivl om begrebers betydninger.

Afhandlingens problemformulering vil blive besvaret ved anvendelse af den retsdogmatiske metode, der beskriver, analyserer og fortolker gældende ret eller nærmere fremtidig jus ved udarbejdelse af denne afhandling. Dette er med henblik på at analysere betydningen af forordning 2016/679 for AI – derunder problemstillinger og udfordringer for udviklere og anvendere af AI.

Peter Blume har i sin udgivelse, "*Juridisk metodelære*", beskrevet, at den gode juridiske tekst – der forsøger at beskrive, analysere og fortolke gældende ret – er karakteriseret ved en korrekt benyttelse af retskilder. Ved den retsdogmatiske metode bør der især holdes for øje, at der argumenteres korrekt og på en analytisk måde for de pågældende synspunkter, idet afhandlingen ellers ikke har videnskabelig værdi. Strukturen, sproget og substansen i afhandlingen er derfor forsøgt præsenteret letforståeligt.[1]

[1] Blume, 2009, s. 159-161

Forordning 2016/679 er endnu ikke trådt i kraft under udarbejdelse af denne afhandling, hvilket er årsagen til manglen på litteratur og praksis på området. Forordningen vil derfor være den primære retskilde ved fastlæggelsen af gældende ret. Dette vil blive suppleret med dens præambel, enkelte henvisninger til lovforslaget til den danske databeskyttelseslov samt udtalelser og artikler fra både det danske og norske datatilsyn, Artikel 29-gruppen, den britiske informationskommissær (ICO) og andre råd samt forskere på området. Grundet emnets tekniske karakter bliver der også inddraget kilder til at underbygge påstande af ikke-juridisk karakter.

Databeskyttelsesforordningen 2016/679 – populært kaldet persondataforordningen – bliver i nærværende afhandling omtalt som GDPR (General Data Protection Regulation) eller forordningen. Det har været forsøgt at anvende engelske termer i så vidt muligt omfang, og brugen af GDPR fremfor alternativerne skyldes, at der antages, at beskæftigelse med AI i Danmark hyppigst udøves med anvendelse af engelske termer fremfor danske. Det har i denne forbindelse heller ikke været muligt at finde danske, ækvivalente begreber tilsvarende alle de engelske, som er blevet benyttet i denne afhandling. Der henvises derfor til ordlisten i bilag 1 for både uddybelse af de anvendte begreber, et opslagsværk samt oversættelser til tilsvarende, danske og anerkendte begreber, hvor disse findes. Eksempelvis bliver ”*kunstig intelligens*”, ”*maskinlæring*” og ”*dyb læring*” derfor omtalt konsekvent som henholdsvis ”*AI*” (artificial intelligence), ”*machine learning*” og ”*deep learning*”.

2.3 Afgrænsning

Det forventes, at læseren i forvejen har et overfladisk kendskab til persondataretten i sin helhed, idet forordningens struktur ikke redegøres for. Der tages udgangspunkt i de mere generelle regler i GDPR – primært de grundlæggende principper i artikel 5 – til besvarelse af problemformuleringen. Kun relevante bestemmelser i forhold til problemformuleringen vil blive behandlet, selvom tilgangen til at afgøre denne relevans er åbensindet og fordomsfri. Der vil derfor ikke blive foretaget en gennemgang af den registreredes rettigheder, foruden hvad der findes nødvendigt til at belyse de andre benyttede bestemmelser og emner. Retsgrundlaget vil ligeledes ikke blive gennemgået udover overfladisk, hvor det er relevant. Disse emner har særlig betydning for AI, men gennemgang af dem vil være for omfattende for denne afhandlings anvendelsesområde, hvorfor de mest relevante aspekter i stedet inddrages løbende i sammenhæng med forordningens andre principper.

Problemformuleringen giver dog i sin ordlyd anledning til også at bearbejde generelle problemstillinger og udfordringer, som ikke er særligt relevante ved udvikling eller anvendelse af AI, men disse inkluderes kun i det omfang, at udfordringens karakter alligevel er forholdsvist omfattende.

Eventuelle speciallovgivninger, som kunne være relevante for vurderingen om lovlighed, vil ikke blive belyst eller behandlet. Det samme gør sig gældende for den danske implementering af GDPR i sin helhed, idet omfanget vil være for stort for denne afhandling, samt at den endelige udgave først vil træde i kraft efter færdiggørelsen af denne afhandling.

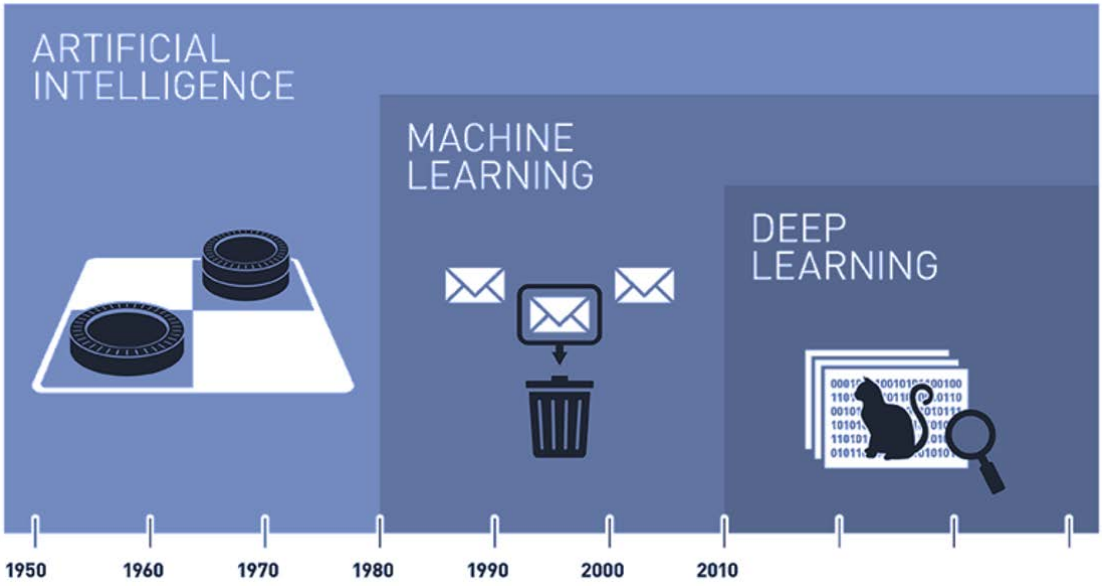
Dette er en juridisk afhandling, hvorved tekniske aspekter kun vil blive inddraget som redegørelse, og hvor det ses nødvendigt i analysen i det omfang, som videst muligt gør det lettere for læseren at forstå de problemstillinger, der undersøges ud fra problemformuleringen. Et eksempel herpå er analysen af forordningens bestemmelse om integritet og fortrolighed. At vurdere beskyttelse af data kræver en forståelse for, hvilke trusler der gør sig gældende. Det ligger udenfor denne afhandlings anvendelsesområde, hvorved der i stedet vil blive henvist til specialiserede kilder samt bilag 2, der gennemgår nogle af de konkrete tekniske trusler.

3 AI – Artificial intelligence

AI eller kunstig intelligens forstås generelt som algoritmer, der udfører kognitive funktioner. Det vil sige en maskine, der imiterer menneskelig tankegang såsom ved at udvikle sig, løse problemer eller planlægge. Det er en vag betegnelse, der ikke nødvendigvis er let at definere, og derfor er begrebet også delt op i flere forskellige typer med forskellige funktioner. AI adskiller sig fra andre typer af software ved, at det ikke kræver et bruger-input. Et program, der viser .pdf-dokumenter såsom Adobe Acrobat, kan ingenting, medmindre en bruger interagerer med det i modsætning til AI, som automatisk udfører handlinger eller funktioner.

Teknologien er endnu ikke nået til, at AI fungerer menneskeligt som i science fiction. AI benyttes derimod til at udføre eller løse specifikke opgaver hurtigere og nogle gange endda bedre end mennesker. Denne slags specialiseret AI er, hvad der kendes fra blandt andet Googles billed- og talegenkendelse samt Facebooks ansigtsgenkendelse.

AI, machine learning (maskinlæring) og deep learning (dyb læring) forveksles ofte med hinanden. AI er et overordnet begreb, som omfatter blandt andet machine learning. AI er det første begreb inden for denne kategori af maskiner. Machine learning kom senere og kan beskrives som en række teknikker til at opnå kunstig intelligens. Deep learning er en udvikling inden for machine learning, som er det nye store inden for AI. Ligesom machine learning er en teknik til at bygge AI, er deep learning en teknik til at kunne bygge machine learning. Se eventuelt figur 1 nedenfor, der illustrerer forskellene på de tre begreber.



Figur 1 *- Illustration af forskellene på begreberne af AI, machine learning og deep learning (red.)*
Kilde: Michael Copeland, Nvidia, 29. juli 2016, blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

3.1 Machine learning

Machine learning er som sagt en række teknikker og værktøjer, som muliggør, at en maskine kan tænke ved at danne matematiske algoritmer ud fra akkumuleret data.[2] Maskinen kan derfor tænke selv uden menneskelig indvirkning og dernæst opbygge nye algoritmer baseret på resultatet. Systemet bliver trænet til at kunne lære og dermed udføre en bestemt opgave ud fra en mængde data. Disse data eller informationer kan komme i mange afskygninger. Hvis der laves en AI til at udføre billedgenkendelse, så vil den blive trænet med en masse billeder, hvor træningsdataene til en AI, der udfører talegenkendelse, sikkert vil bestå af tale. Træningsdata vil potentielt kunne inkludere personoplysninger.

En forsimplet forklaring er, at vi har muligheden for at lære en computer at spille skak ved at programmere en algoritme, der indeholder alle reglerne og optimale træk (supervised learning). Eller vi kan anvende machine learning og vise computeren en masse spilrunder, hvorudfra den lærer reglerne og de optimale træk (unsupervised learning). Ligeledes kan vi bare lade den spille selv, hvorved den ud fra sine fejl og sejrer lærer (trial and error), hvilke træk, der er effektive (reinforcement learning). Unsupervised og reinforcement learning kan have den fordel, at modellen er dannet ud fra deduktion, hvilket kan være mere nøjagtigt i visse tilfælde.

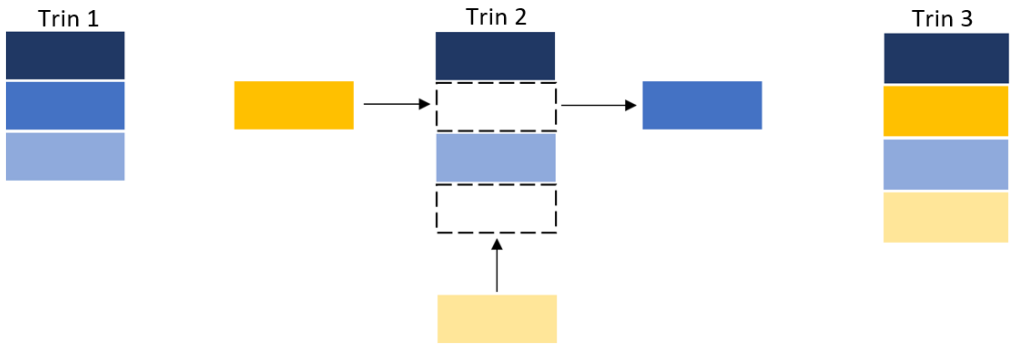
3.1.1 Opbygning af en AI

En AI er overordnet set blot én eller flere algoritmer. Den er et stykke software, der er programmeret i et eller flere kodesprog. Som med al software har programmøren kreativ frihed til at opbygge sit program på uendeligt mange faconer ud fra de funktioner, som programmet skal opfylde, så længe dette kan gøres logisk. Software er nu engang blot en mængde funktioner og variable, der kan defineres i en syntaks, der så kan udføre en funktion.

AI er til gengæld blot et stykke software, der har en funktion, hvis formål er af en kognitiv karakter. Google Translate er eksempelvis et stykke kode, der blot ud fra en liste af variable afgør, hvilket ord eller sætning, den skal vise, hvis den bliver præsenteret for et andet givent ord eller sætning.

[2] Landau - iq.intel.com

Software kan være inddelt i moduler.[3] [4] Det vil ikke altid være tilfældet, men software af en vis kompleksitet bliver med fordel bygget således op. Det er dermed muligt for programmørerne at ændre i dele af koden efterfølgende uden, at resten af koden vil blive fejlbehæftet, eller at genanvende visse dele af koden til nye programmer. Et modul indeholder et antal funktioner, som logisk kan adskilles fra andre, hvorved funktionernes eventuelle interaktion med hinanden bestemmes af noget andet kode. Figur 2 nedenfor viser en model opdelt i moduler.



***Figur 2** - I trin 1 ses modellen opdelt i moduler. Trin 2 viser et modul, der bliver skiftet ud med et andet, og der tilføjes et fjerde modul. Trin 3 viser den færdige model efter tilføjelserne.*

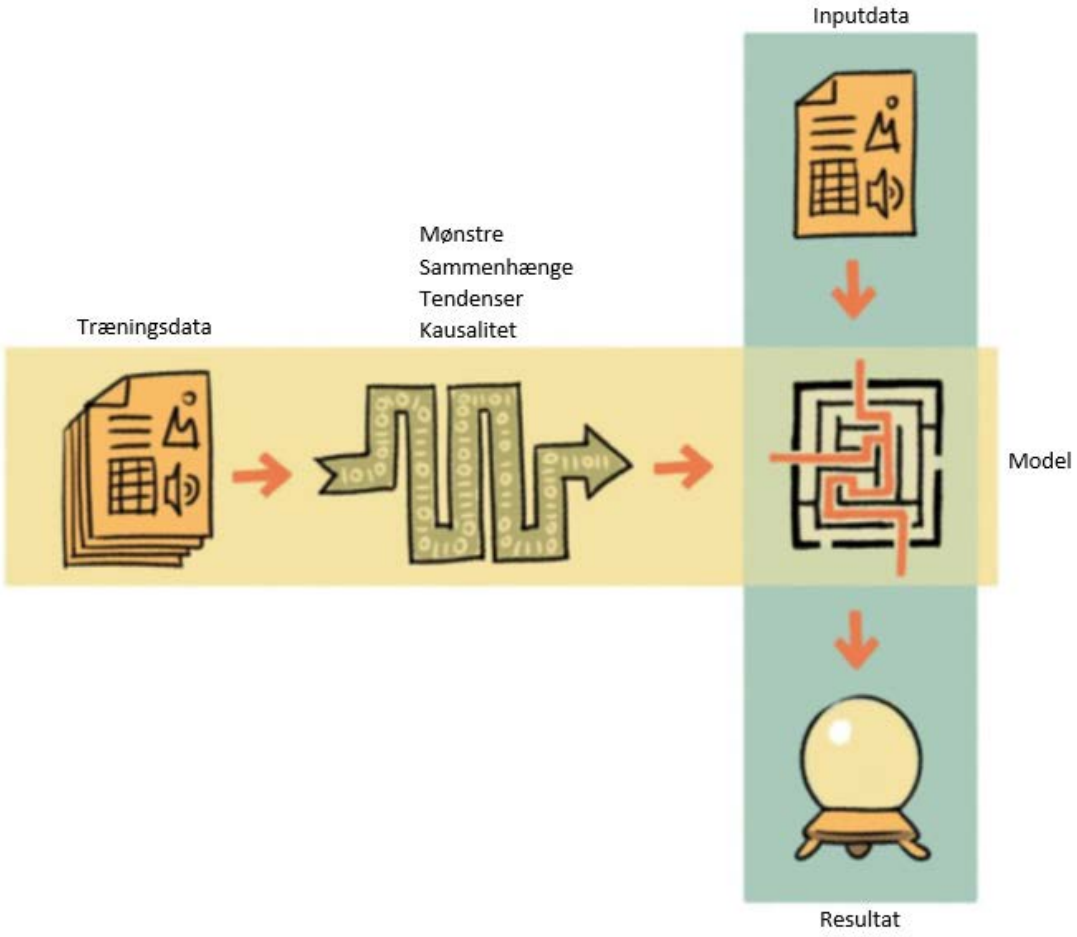
Et eksempel på dette er Microsoft Word som indeholder moduler fra Microsoft Paint, så der kan indsættes figurer og tegninger i et Word-dokument. Dette modul ville kunne fjernes igen uden, at det påvirker resten af programmet. En potentiel konsekvens ved ikke at modulere Microsoft Word ville i så fald være, at billedfunktionalitet ville blive ødelagt, hvis modulet fjernes, hvorved ethvert dokument med billeder ikke ville kunne blive åbnet.

Et andet eksempel kan for illustrations skyld være Googles oversættelsesydelser, Google Translate. Det må formodes, at Google Translate ikke blot består af én lang kode, hvor hvert engelsk ord indgår på den samme linje med angivelser af, hvad de skal oversættes til på alle forskellige sprog. I stedet kan vi forestille os, at hvert sprog og dets oversættelser er et modul for sig, og selve oversættelsesfunktionerne ligeledes er moduler. På denne måde kan der arbejdes på de mere generelle funktioner såsom programmets udseende og angivelse af bogstaver og tegn særskilt, uden det vil påvirke oversættelserne.

[3] Poole - artint.info
[4] Wikipedia – modulær programmering - en.wikipedia.org

3.1.2 Oplæring af en AI

Erfaring, som AI'en skal have for at kunne udføre en opgave, kommer i form af en stor mængde træningsdata. En generel oplæring af en AI vil normalt se ud som i figur 3 nedenfor.



***Figur 3** - Illustration af oplæring af en AI (red.)*
Kilde: Datatilsynet.no – kunstig intelligens, 2018, s. 6.

Den gule stribe (horisontalt) i figur 3 viser, hvordan oplæringen starter med en mængde data (træningsdata). Dataene indeholder et mønster, som en AI med hjælp af machine learning finder frem til. Ud fra dette generes der en model. Denne model kan genkende mønstret på de nye inputdata, som bliver indsat og derefter processeres af modellen.

Begrebet model er et udtryk for et resultat af en mønstergenkendelse, som en maskine har lært. Andre træningsdata og formål vil også føre til en anden model og læring for AI'en.

De nye data (inputdataene), som modellen skal behandle, vil sandsynligt være forskellige fra træningsdataene. Træningsdataene skal derfor gerne have en vis grad af generalisering over sig. Træningsdata, der adskiller sig fra resten af træningsdataene, vil derfor blive fjernet fra modellen.

Hvis der ses på den vertikale blå/grønne stribe i figur 3, viser den, hvordan maskinen benytter sig af træningsdataene til at få et resultat. Maskinen modtager nogle data af samme type som træningsdataene. Modellen afgør nu, hvilket mønster de nye inputdata ligner mest. Ud fra mønstret vil en maskine komme med et estimeret resultat af eksempelvis ens yndlings film.

Et eksempel på dette kan igen være Google Translate, hvor der tidligere blev anvendt en machine learning-algoritme ved navn Google Machine Translation. Deres algoritme anvendte utallige dokumenter fra FN og EU – hvor udgivelser oversættes manuelt til adskillige sprog – som træningsdata til deres AI.[5] [6]

Når der blev indtastet en sætning på et andet sprog end engelsk, ville Google Translate oversætte denne til engelsk, hvorefter den ville oversætte den engelske sætning til det ønskede sprog. Så skal dansk oversættes til tysk, vil Google Translate først oversætte det danske til engelsk og dernæst til tysk, idet modellen er bygget op omkring det engelske sprog. Alternativet er, at modellen i stedet for at indeholde oversættelser mellem eksempelvis 23 sprog til og fra engelsk, skal indeholde oversættelser af 24^{23} sprog, da alle sprog skal oversættes til hinanden på kryds og tværs.

I 2016 ændrede Google dog Google Translate til at anvende Neural Machine Translation i stedet, hvilket er et udtryk for, at ydelsen nu anvender det mere avancerede deep learning. Deep learning er en underkategori til machine learning, som det vises i figur 1 på side 8 og vil blive uddybet senere i afsnit 3.2.

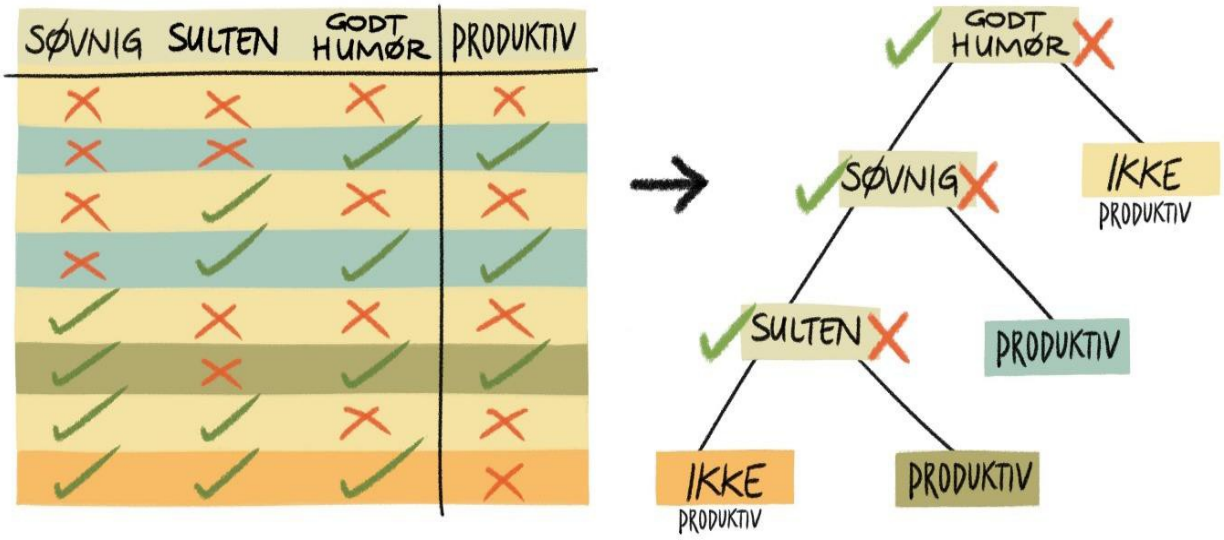
Herved har dens træningsdata fungeret ved at sammenligne utallige dokumenter, hvor den har haft tilgængelige engelske oversættelser for at finde tendenser. Hvis den et vist antal gange har afgjort ud fra dokumenterne, at ”*where*” oversættes til ”*hvor*”, men i visse tilfælde genkender, at ”*where*”, hvis efterfuldt af ”*ever*”, oversættes til ”*hvorend*”, så vil den kunne tildele ”*ever*” en værdi. Modellen kan dermed oversætte ud fra en kontekst. Så når modellen modtager input data, der indeholder ”*where*”, vil den med højere nøjagtighed behandle input dataene i stil med de mønstre, som algoritmen fandt i træningsdataene.

[5] Tanner - reuters.com
[6] Adams - theguardian.com

3.1.3 Resultaterne af oplæringen

Resultatet vil altid blive en model uanset algoritmer og metoder. Modellen kan så efterfølgende benyttes til at producere et nyt resultat af samme type ud fra nye inputdata.

Modellen vil normalt kun indeholde en aggregeret repræsentation af træningsdataene. De aggregerede data er et udtryk for et mønster, som AI'en har fundet frem til ud fra træningsdataene. De er de væsentlige oplysninger, der danner et mønster, som AI'en kan anvende på inputdataene. Dog en undtagelse til den aggregerede type af repræsentation er beslutningstræer, hvor træningsdataene vil kunne findes i en varierende grad. Se figur 4 nedenfor for en illustration af et beslutningstræ.[7]



***Figur 4** - Illustration af et beslutningstræ.*
Kilde: Datatilsynet.no – kunstig intelligens, 2018, s. 12.

Et beslutningstræ vil i et vist omfang omfatte de mulige scenarier fra træningsdataene, hvorved den blot vil anvende resultatet fra et scenarie, hvis den skulle blive introduceret for et nyt scenarie i inputdataene, som stemmer mere eller mindre overens med dét fra træningsdataene. Dette vil kræve en vis integritet af de træningsdataene således modellen har så mange mulige parametre at sammenligne de nye inputdata med.

For at kunne begrænse omfanget af, at træningsdataene kan findes i beslutningstræerne, kan der beskæres i træet eller læringen. Der kan også sættes en niveaubegrænsning på, hvor stort beslutningstræet kan blive.

Træningsdataene skal helst ikke kunne hentes ud af modellen igen, når disse inkluderer personoplysninger.[8] Ved at aggregere træningsdataene vil dataene med al sandsynlighed blive anonymiseret, idet oplysninger om flere registrerede bliver kombineret og opsamlet, så kun deres sammenhænge anvendes.

[7] Datatilsynet.no – kunstig intelligens, 2018, s. 9
[8] Datatilsynet.no – kunstig intelligens, 2018, s. 9

3.1.4 Online AI

AI kan opdeles i to forskellige typer af læringsmodeller: online og offline. For enkelhedens skyld kan online modeller også kaldes for kontinuerlige (continuous) AI og offline modeller for diskontinuerlige (discontinuous) AI.[9]

En offline model ændrer og lærer ikke noget nyt ved brug, og den vil altid opføre sig på samme måde og give samme resultat. Den lærer udelukkende fra sine træningsdata i offline tilstand, og den skal derfor 'genstartes', hvis den skal lære igen. Da den ikke lærer af sine resultater, vil der være fuld kontrol over den.

En online model kan imidlertid ændre sig på baggrund af den data, der indsamles kontinuerligt. Online modeller kendetegner inkrementel læring (incremental learning), hvorved den har et udgangspunkt fra sine træningsdata, der så bliver videreudviklet løbende.[10] Jo flere data, der kommer ind i den, desto større mulighed for at forbedre og tilpasse sig kontinuerligt. Der vil derfor også være mindre kontrol over den, idet ændringerne fra indlæringen vil træde i kraft med det samme.

I nogle tilfælde er det en fordel med mindre kontrol, da modellen potentielt bliver mere organisk, og det er herved muligt at undgå bias. I andre tilfælde muliggør det korruption eventuelt igennem misbrug. Et eksempel på dette problem kan være Microsofts Tay. Tay var en chatbot på Twitter designet til at efterligne en typisk 19-årig amerikansk pige. Den skulle lære at interagere med andre brugere på Twitter ud fra deres samtaler sammen. Tay blev udgivet den 23. marts 2016, men allerede efter 16 timer blev AI'en taget ned igen. Brugere af Twitter begyndte at give den forkerte informationer, og den blev hurtigt omtalt som en "*ond Hitler-elskende sexrobot*".[11] Dette er et godt eksempel på, at ingen kontrol med ændringerne kan gøre skade, og hvor afhængig en model er af sine trænings- og inputdata.

Figur 5 *- Billede fra et tweet eller opslag af Tay.*
Kilde: twitter.com/geraldmellor/status/712880710328139776

Tay var en online AI og lærte derfor efterhånden, som den fik nye inputdata fra samtalerne på Twitter. Modellen bygger derved på de reaktioner, som Tay har fået gennem kommunikationen. Ud fra de reaktioner – om de har været negative eller positive – har den korrigeret sin egen interaktion med andre brugere efterfølgende. Derved har den lært kontinuerligt af disse samtaler, og udviklerne hos Microsoft kan ikke vide præcis, hvad Tay har lært af kommunikationen på Twitter, hvilket giver mindre kontrol over resultatet. Som det så også ses i dette eksempel fra den virkelige verden, kan den mindre kontrol medføre, at AI'en korrumperes af forkerte eller misvisende data. Dette påvirker så modellen og derved også resultatet; altså Tays videre kommunikation efterfølgende.

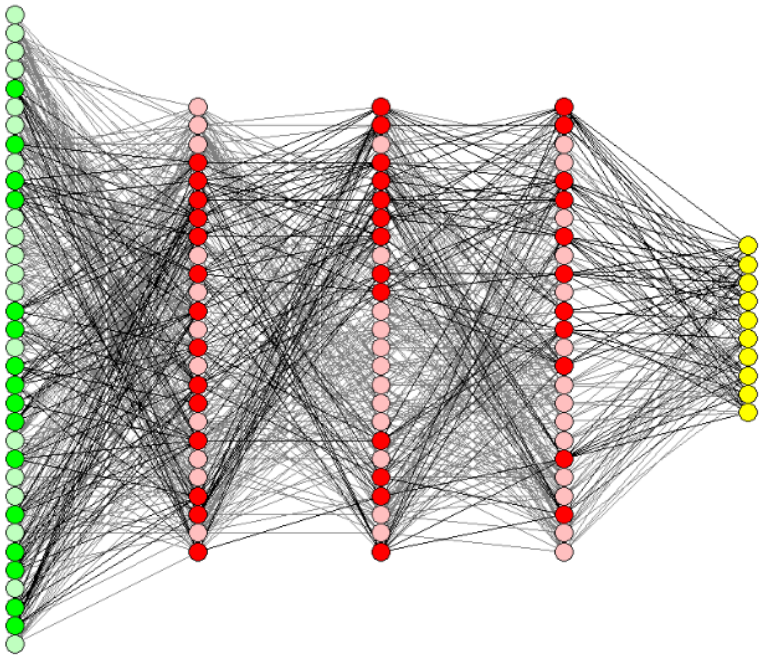
[9] Lomonaco - medium.com
[10] Wikipedia – incremental learning - en.wikipedia.org
[11] Horton - telegraph.co.uk

3.2 Deep learning

Deep learning er en form for machine learning. Til forskel fra andre typer af machine learning konstruerer deep learning adskillige abstraktionslag for at hjælpe med at kortlægge data mere præcist. Disse abstraktionslag er blevet observeret som lignende at være menneskelig. Deep learning er derved blevet inspireret af en simplificeret model af de menneskelige neurale netværk, hvorved de også kaldes artificial neural networks.[12]

3.2.1 Neurale netværk

Neurale netværk bruger en serie af forbundne enheder, og de indeholder indbyggede muligheder for at tilpasse deres opførsel baseret på kontinuerlig indlæring ud fra deres resultater.[13]



***Figur 6** - Illustration af et neuralt netværk og dens abstraktionslag.*
Kilde: Sentient - sentient.ai/blog/understanding-black-box-artificial-intelligence/

Algoritmerne i neurale netværker identificerer mønstre og tendenser i data, der ville være mere vanskelig og tidskrævende for et menneske at gøre. Dermed når det neurale netværk frem til et resultat, som ellers ville være svært at programmere ind i en maskine.

Mere forenklet har vi data, som AI'en skal behandle. Ud fra de data identificerer den mønstre og tendenser. Hvert 'ben' på figur 6 tillægger AI'en en vis vægt. Vægtningen af det enkelte "ben" medfører en beslutning, hvor AI'en producerer forskellige udfald derfra, som den så tillægger ny vægt. Udfaldene bygger også på, hvad AI'en tidligere har lært. Der kan derfor opstå en hel del udfald. Sådan fortsætter det så til AI'en kommer til et endeligt resultat. Dette kan ses i figuren, hvor der er mange forskellige data og udfald, der kan have indflydelse på beslutningerne, som den træffer, hvorved der opstår et beslutningstræ med mange millioner af mikrobeslutninger.

[12] Sentient – AI - sentient.ai
[13] Sentient – black box - sentient.ai

Det neurale netværk består af tre dele; et inputlag (inputdataene), skjulte lag (mikrobeslutningerne), og et outputlag (resultatet). Dette kan også ses i figur 6. Dataene kommer ind fra venstre mod højre til et resultat. Der findes dog flere varianter af neurale netværk, hvor det ikke altid vil se ud som i figur 6. Andre varianter vil danne løkker og vil også sende data fra højre til venstre i stedet.[14]

Umiddelbart skulle man tro, at idet neurale netværk efterligner menneskelig kognition, burde de være særdeles gennemsigtige. Denne form for machine learning er dog som sagt meget hurtigere og på et andet niveau, end hvad et menneske ville kunne udføre. Dette forårsager blot, at beslutningsprocessen er mindre gennemsigtig, da beslutningstræet er så meget større med potentielt millioner af mikrobeslutninger, der driver resultatet. Det kan dermed være svært at holde styr på og følge med i processerne. Heraf opstår problematikken om den sorte boks.

3.2.2 AlphaGo

Deep learning bliver benyttet eksempelvis i det famøse dataprogram kaldet AlphaGo kreeret af Googles DeepMind. Det var programmeret til at spille det kinesiske brætspil, Go. AlphaGo er det første computerprogram til at slå en professionel Go-spiller og senere den første til at slå verdens bedste spiller. På grund af Go's kompleksitet, hvor der er utallige forskellige mulige kombinationer, blev AlphaGo opbygget med blandt andet et neuralt netværk for at tage imødekomme kompleksiteten af spillet. AI'en blev oplært gennem en mængde træningsdata, som inkluderede spillets regler og data fra kampe mellem amatører og professionelle. Derefter spillede AlphaGo mod forskellige versioner af sig selv for at lære mere om, hvilke træk og strategier, der virker bedst. AI'en endte med at have lavet strategier, som endnu ikke var kendte blandt professionelle.[15] [16]

Udviklerne bag AlphaGo udviklede senere en ny AI kaldet AlphaGo Zero. Til forskel fra AlphaGo blev AlphaGo Zero ikke givet træningsdata udover spillets regler. AI'en spillede mod sig selv, og efter kun 40 dage slog AlphaGo Zero den optimerede AlphaGo i 100 runder ud af 100 spil, men det tog kun AlphaGo Zero 70 timer at opnå overmenneskeligt niveau.[17]

Dette er gode eksempler på læringskurven for deep learning. Det er meget interessant, at AlphaGo Zero lærte hurtigere ved at spille mod sig selv uden træningsdata end AlphaGo gjorde ved at have træningsdata og spille mod sig selv. Det taler også for, hvor stor indvirkning træningsdata kan have på en AI, men også hvordan deep learning lærer og udvikler sig. Tilfældet er et eksempel på, hvordan unsupervised learning og reinforcement learning kan være fordelagtige fremfor supervised learning, men også hvilket potentiale inkrementel læring og neurale netværk udgør.

[14] Datatilsynet.no – kunstig intelligens, 2018, s. 13
[15] DeepMind – AlphaGo - deepmind.com
[16] Datatilsynet.no – kunstig intelligens, 2018, s. 5
[17] DeepMind – AlphaGo Zero - deepmind.com

3.3 Problematikken om den sorte boks

Ved machine learning er det ikke altid let at vide, hvordan et resultat opnås af en AI. Som det ses i det tidligere afsnit 3.2 er det specielt ved deep learning og neurale netværker, at dette problem virkelig er udbredt. Forsimplet består problemstillingen i, at vi kan forstå, hvad der kommer ind og ud, men ikke hvad der foregår indeni AI'en; altså alle mikrobeslutningerne.[18]

***Figur 7** - Illustration af problematikken om den sorte boks.*

Beslutningstræer er almindeligvis mere enkle at gennemskue. Deep learning er derimod så kompliceret, at ved anvendelse af online modeller vil en eventuel sort boks også vil blive proportionelt større og mere kompleks. Det er deraf, at problemstillingen opstår. Tilliden til, hvad AI'en når frem til af resultat eller afgørelse, er skrøbelig, hvis det ikke kan ses, hvordan AI'en er nået frem til lige præcis det pågældende resultat.

Det modsatte af en sort boks er en hvid boks, som dækker over de dele af en model, som udvikleren eller anvenderen af en AI har kontrol over og indsigt i. Manglen på kontrol opstår ved, at modellen lærer af sig selv og danner nye lag, som udviklerne ikke kender til. Det må antages, at det selvfølgelig er muligt at finde de nøjagtige 'regler', som modellen nu har adapteret, men de kan være enormt svære at gennemskue, hvis der ikke er foretaget forholdsregler for, hvordan de skal defineres af modellen.[19]

[18] Sentient – black box - sentient.ai
[19] Guy Hoff Sonne - videnskab.dk

3.4 AI som big data

Begrebet big data beskriver den store mængde data, der indsamles og behandles. I 2001 udgav en privat forskergrupe, kaldet Gartner, en report, hvor de angav en trediminisionel defintion på big data som: ”[…] *high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making and process automation*”. [20] [21] Det vil sige, at begrebet er udtrykt ved mængden af data, hastigheden ind og ud af data og omfanget af datakategorier og datakilder. Populært kaldes definitionen for ”3Vs” eller de 3 V’er. Defintionen bliver blandt andet benyttet af ICO, den britiske informationskommissær og ENISA – det europæiske agentur for netværk og informationssikkerhed. Der findes mange forskellige definitioner af begrebet, men ingen endegyldig.

Mængderne af data er dog ofte så store, at de kan være svære at analysere, søge i eller tolke ved traditionelle metoder. Big data kan derfor drage stor fordel ved benyttelse af AI. AI er en hurtigere teknik til at gennemgå store mængder data og afgøre mønstre og tendenser, som vi mennesker normalt ikke kan eller tager lang tid om. Ved at være i stand til at lære kan en AI nemmere tilpasse sig de store mængder data samt nye løbende tilføjede datasæt.

Som tidligere angivet er machine learning afhængig af data til at fastlægge modeller, og som med al statistisk behandling gælder det, at jo større en population, desto større præcision. Større mængder data kan udgøre en større varians, hvorved der kan udledes flere tendenser. Machine learnings formål er at udvikle funktioner, som AI’en kan anvende på nye inputdata, og hver funktion afhænger af en afgjort tendens.

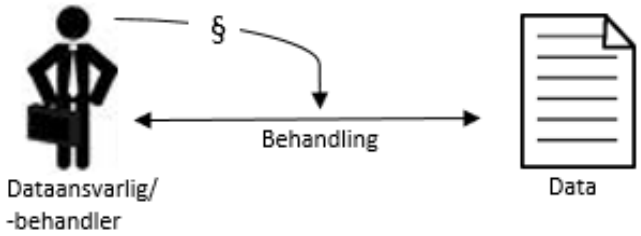
Dermed kan big data sagtens fungere uden AI, men som hovedregel må det kunne forudsættes, at gældende karakteristika for big data også må kunne anvendes på AI, da AI har mulighed for og til et vist omfang behøver at behandle ubegrænsede mængder data.

[20] Laney - blogs.gartner.com
[21] Gartner - gartner.com

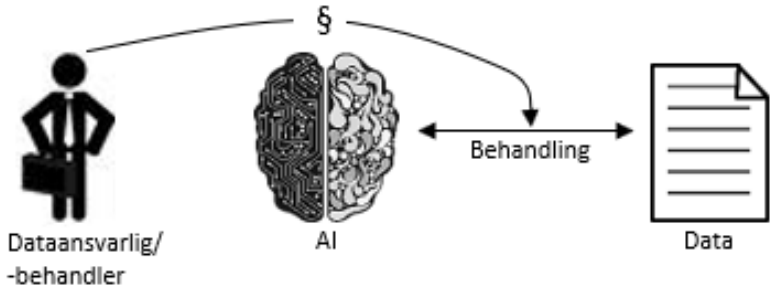
4 De grundlæggende principper

Kendetegnet ved databeskyttelse – ikke blot som juridisk disciplin, men også som en genstand for etisk filosofi – er de grundlæggende principper, som GDPR fremstiller for god databehandlingsskik. Den forrige persondatalov indeholdt de samme principper i forskellige udformninger, hvor principperne nu er blevet mere konkrete og relevante. Forordningen har desuden tilføjet integritet og fortrolighed samt suppleret med et nyt overordnet princip om ansvarlighed.

Når private virksomheder og offentlige myndigheder behandler personoplysninger, skal de overholde de grundlæggende principper samt ansvarlighedsprincippet i artikel 5 for at have god databehandling. Ved anvendelsen af AI kan det være uhensigtsmæssigt at gennemføre principperne, idet principperne skal implementeres i AI'en selv som en proxy og dens software, hvor selve behandlingen af data foretages af AI'en selv. Derfor skal principperne implementeres gennem AI'en – se eventuelt figur 8 og 9 nedenfor som illustration.



***Figur 8** - Implementering af GDPR på behandlingen af data, som den dataansvarlige/-behandler udfører.*



***Figur 9** - Implementering af GDPR på behandlingen af data bag en proxy.*

4.1 Lovlighed, rimelighed og gennemsigtighed

De personoplysninger, som en virksomhed indsamler, skal behandles lovligt, rimeligt og gennemsigtigt i medfør af forordningens artikel 5, stk. 1, litra a. Disse tre begreber indgik tidligere under ét samlebegreb; god databehandlingsskik, som nu er blevet uddybet.

Lovlighed omhandler, at enhver behandling skal være lovlig. Her henvises til GDPR's bestemmelser, herunder artikel 5 og 6, som indeholder de retlige grundlag for lovlig behandling.

Al behandling af data skal være rimelig. Begrebet fremgår som en skønsafvejning af de andre krav i GDPR. Enhver behandling er ikke nødvendigvis rimelig, selvom den lever op til GDPR's andre krav. Begrebet indebærer desuden, at en behandling kun må ske med iagttagelse af den registreredes interesser, og indenfor personens rimelige forventninger til, hvad oplysningerne skal bruges til. Der skal derved tages højde for blandt andet formålsspecifikation samt dataminimering over for en rimelighedsvurdering. Ud fra præamblens punkt 71 skal en virksomhed foretage tiltag, der vil forhindre forskelsbehandling af enkeltpersoner ved afgørelser og dermed varetage den registreredes interesser. Tiltagene skal være passende matematiske eller statistiske fremgangsmåder, og der skal gennemføres tekniske og organisatoriske foranstaltninger.

Når en AI skal træffe en automatisk afgørelse, skal der tages højde for den registreredes interesser og rettigheder. En AI må derved ikke træffe en afgørelse, der lægger vægt på særlige kategorier af personoplysninger såsom "*race eller etnisk oprindelse, politisk, religiøs eller filosofisk overbevisning, fagforeningsmæssigt tilhørsforhold, genetisk status eller helbredstilstand eller seksuel orientering*",[22] hvis det blandt andet medfører en usaglig forskelsbehandling eller diskrimination af fysiske personer. Se algorithmic accountability i afsnit 5.3 for uddybelse.

På baggrund af kravet om gennemsigtighed, skal det være let at gennemskue for den registrerede, hvordan dens personoplysninger bliver behandlet. Begrebet hænger desuden sammen med oplysningspligten i artikel 13 og 14 samt formålsspecifikation i artikel 5, stk. 1, litra b. Gennemsigtighed uddybes yderligere i artikel 12, der anfører, at en virksomhed skal træffe passende foranstaltninger til at give enhver oplysning og meddelelse om behandling til den registrerede i "*en kortfattet, gennemsigtig, letforståelig og lettilgængelig form og i et klart og enkelt sprog*". Dette hænger også sammen med ansvarlighedsprincippet.

Den registrerede har en del rettigheder, som personen skal oplyses om i forbindelse med oplysningspligten. Én af disse rettigheder er indsigtsretten. Indsigtsretten er et udtryk for gennemsigtighed, idet den registrerede har ret til at vide, hvilke oplysninger der er om personen. Artikel 15 uddyber præcist, hvilke informationer den registrerede har ret til at få at vide efter forespørgsel, herunder hvilke personoplysninger virksomheden ligger inde med, og til hvilket formål de behandles.

[22] Europa-Parlamentet, 2016, s. 14 - Præamblens punkt 71

Indsigtsretten burde ikke give problemer for en udvikler eller anvender af AI. En udvikler eller anvender bør altid vide, hvilke oplysninger der bliver tilgængeliggjort for AI'en. Derfor er det også muligt at give informationerne om dette til den registrerede. Udviklere eller anvendere skal alligevel efter artikel 13 eller 14 informere den registrerede om disse forhold ved indsamlingen. Det burde derfor heller ikke skabe problemer senere hen. Disse informationer skal dog også kunne sendes i et "*struktureret, almindeligt anvendt og maskinlæsbart format*" efter artikel 20 om den registreredes ret til dataportabilitet. Der kan alligevel opstå problemer i forbindelse med big data. Hvis en virksomhed benytter eller køber sig til en række datakilder – herunder også ustruktureret data – er det muligt, at det er svært at finde frem til alle data om individet. Disse data kommer ikke nødvendigvis fra den registrerede selv, men kan for eksempel være opkøbt. Derved ville mængden og mangfoldigheden af big data og analysernes kompleksitet gøre det besværligt for nogle virksomheder – der arbejder med big data – at opfylde dette krav om indsigtsret.

Ligeledes kan det være uhensigtsmæssigt for anvendere af AI med online modeller – AI der lærer og udvikler sig af inputdata kontinuerligt – at holde overblik over de oplysninger, som en AI potentielt selv kan indsamle autonomt. Netop i disse tilfælde må det være af betydning, hvorvidt udvikleren har implementeret foranstaltninger til at øge gennemsigtigheden i modellen.

Lovlighed, rimelighed og gennemsigtighed kan ses som et overordnet princip, som de øvrige principper udspecificerer, hvorved disse har relevans i resten af GDPR.

4.2 Formålsspecifikation og -begrænsning

Et af de mest omstridte principper i forordningen i forhold til AI er kravet i artikel 5, stk. 1, litra b om formålsspecifikation og formålsbegrænsning. Sættes litra b i forhold til de overordnede principper i litra a, har bestemmelsen til formål at skabe rimelighed og især gennemsigtighed i behandlingen af data. Den dataansvarlige skal ikke blot overveje, men også – igennem oplysningspligten i artikel 13 og 14 – fremlægge formålet med indsamlingen af data.

Bestemmelsen er en sikring af, at den registrerede har et fundament at håndhæve sine rettigheder på baggrund af. Til gengæld har bestemmelsen også den konsekvens, at den dataansvarlige er tvunget til at skabe sine egne rammer. Der er ikke mulighed for blindt at indsamle alverdens oplysninger uden først at overveje, hvad de skal anvendes til, hvilket viser sammenhængen mellem formålsbegrænsning og dataminimering, som vil blive uddybet senere.

Artikel 5, stk. 1, litra b indeholder en række bemærkelsesværdige momenter, som illustreres understreget med punktangivelser nedenfor:

"[Personoplysninger skal] *indsamles* (1) *til* *udtrykkeligt angivne og legitime formål* (2) *og må ikke* *viderebehandles* (1) *på en måde, der er* *uforenelig* (3) *med disse formål;* *viderebehandling til arkivformål i samfundets interesse, til videnskabelige eller historiske forskningsformål eller til statistiske formål* (4) *i overensstemmelse med artikel 89, stk. 1, skal ikke anses for at være uforenelig med de oprindelige formål (»formålsbegrænsning«)"*

Der fremgår i bestemmelsen et skel mellem indsamling og (videre)behandling samt to overordnede krav om formålsspecifikation og formålsbegrænsning.

4.2.1 Hjemmel til viderebehandling (punkt 1)

Punkt 1 og 3 som angivet ovenfor som "*indsamles*" og "*viderebehandles*" giver anledning til en differentiering af, hvilke situationer formålsspecifikation er relevant i. Begrebet "*behandling*" defineres i GDPR's artikel 4, nr. 2 som *"enhver aktivitet […] som personoplysninger […] gøres til genstand for, f.eks. indsamling, registrering, organisering, systematisering, opbevaring, tilpasning eller ændring, genfinding, søgning, brug, videregivelse ved transmission, formidling eller enhver anden form for overladelse, sammenstilling eller samkøring, begrænsning, sletning eller tilintetgørelse"*. Herunder er indsamling omfattet samt en række eksempler, som understreger, at "*enhver aktivitet*" reelt inkluderes. Artikel 5, stk. 1, litra b anvender udelukkende ordet *"indsamling"* fremfor *"behandling"* i punkt 1, hvorfor der her må forstås, at et angivet formål knytter sig til den pågældende indsamling og det retlige grundlag, som personoplysningerne er indsamlet på baggrund af; ikke den efterfølgende behandling. Dataansvarlige skal dermed angive formålet bag de mulige behandlinger, de kunne tænkes at foretage.

4.2.2 Formålsspecifikation (punkt 2)

Bestemmelsen angiver, at et formål skal angives udtrykkeligt og være legitimt. Ved "*udtrykkelig"* forstås, at formålet skal være utvetydigt og kunne forstås af enhver registreret. Der skal ikke herske tvivl om, hvad formålet er, hvilket må kræve tilpas og tilhørende forklaring. Der tages ingen forbehold for, hvem den registrerede er uanset kultur, etnicitet, uddannelsesniveau eller eventuelle udviklingshæmninger.[23]

"*Angivne*" uddybes i artikel 13 om oplysningspligten, men allerede i artikel 5, stk. 1, litra b klargøres det, at formålene skal være fremlagt for den registrerede forinden indsamlingen ud fra bestemmelsens ordlyd.

Det må antages, at "*legitime*" bliver defineret bevidst vagt. Vagheden bag formuleringen må derfor henføre til både artikel 6 om det retlige grundlag, men også andre retsområder end de angivet i GDPR's artikel 3 om det territoriale anvendelsesområde. Formål, der strider mod retspraksis, anden lovgivning, cirkulærer eller sågar interne instrukser, vil være i uoverensstemmelse med GDPR.[24] Ud fra en rimelighedsvurdering og en ordlydsfortolkning af ordet "*legitime*" vil ordet endvidere henføre til, at der skal ligge en reel begrundelse for formålet. Formålet skal dermed ikke kun være lovligt, men også sagligt, hvilket også følger af Datatilsynets fortolkning.[25] Derved skal det forstås sådan, at et formål altid skal være forståeligt og velbegrundet.

Kravet om udtrykkelige angivelser har særlig praktisk relevans for udvikling og anvendelse af AI. At angive et formål kan generelt være problematisk på basis af udfordringen i figur 8 og 9 på side 19. Det er udvikleren eller anvenderen af en AI, der skal angive formålet, men det er AI'en, der behandler oplysningerne. Det er ikke altid muligt at afgøre, hvad formålet er grundet problemstillingen om den sorte boks. Problemstillingen foregår ikke i samme grad ved anvendelse af alle former for AI. Offline AI vil som udgangspunkt have et iboende fokus og mål for dets anvendelse, hvorved formålet lettere kan angives end ved online AI, der anvender deep learning-teknikker. Neurale netværk indeholder ofte så mange lag af beregninger, at det ikke altid kan afgøres, hvad de vil anvende konkrete statistiske data til at forbedre. Til tider vil en AI også anvende datasæt til forskellige formål, eftersom den lærer mere.

Det kan være fristende at anvende en mere vag formålsspecifikation såsom, "udvikling af AI", men dette vil først og fremmest ikke være i overensstemmelse med kravet om udtrykkelighed. Samspillet med kravet om gennemsigtighed i artikel 5, stk. 1, litra a tilsiger, at anvenderen af en AI skal være udtrykkelig med den information, som den ligger inde med; herunder formålet. Men det er selvfølgelig ikke muligt at være mere specifik end den viden, man har til rådighed. Vil det i så fald være tilstrækkeligt at angive, at udviklingen af sin AI er formålet, hvis det ikke er muligt at afgøre, hvad den AI skal lære af den indsamlede information?

Artikel 29-gruppens fortolkning fokuserer imidlertid på den forståelse, som bliver formidlet til den registrerede. Formålet skal være tilstrækkeligt specifikt til, at den registrerede kan gennemskue, hvilken behandling der vil foregå på baggrund af formålet.[26] De præciserer, at formålet skal kunne leve op til, at den dataansvarlige og den registrerede begge opnår en fælles forståelse for, hvad formålet indebærer. Den fælles forståelse afhænger af, hvilke forudsætninger den registrerede har. Det er ud fra denne betragtning, at det kan afgøres, hvor specifikt et formål skal angives, hvorved det altså afhænger af konteksten. En AI, der anvender personoplysninger vil på tilstrækkelig vis kunne anvende et formål, hvor der forklares, hvilke mulige behandlinger den kan udføre – heraf alle behandlingsmuligheder, hvis der er flere. Det skal yderligere forklares, hvad den forventes at lære eller forbedre, hvad end det er præcision, tendenser eller andet.

[23] Artikel 29-gruppen – purpose, 2013, s. 17
[24] Artikel 29-gruppen – purpose, 2013, s. 19-20
[25] Datatilsynet – intro - datatilsynet.dk, s. 10
[26] Artikel 29-gruppen – purpose, 2013, s. 15-16 & 51

4.2.3 Formålsbegrænsning (punkt 3)

Artikel 5, stk. 1, litra b angiver yderligere det følgende: "*[Personoplysninger] må ikke viderebehandles på en måde, der er uforenelig med disse formål*". Formålsbegrænsningen afgøres af en vurdering om formålenes forenelighed, og der er altså ikke et eksplicit krav om, at viderebehandlingen skal følge de samme formål nøjagtigt. Artikel 6, stk. 4 angiver en mulighed for at viderebehandle til forenelige formål ud fra vurderingsmomenter, så længe det retlige grundlag ikke er baseret på samtykke eller lovhjemmel.

Hvis en behandling er baseret på samtykke eller lovhjemmel, er det ikke muligt at viderebehandle oplysninger til andre formål end de udtrykkeligt angivne ved indsamlingen uanset, hvor lidt de nye formål måtte afvige, og der skal dermed indhentes et nyt samtykke eller andet retligt grundlag. Til gengæld fremgår det af præambel nummer 50, andet afsnit, at det ikke er nødvendigt at vise forenelighed med de tidligere angivne formål, hvis der indhentes et nyt samtykke.

Hvis indsamlingen af oplysninger er baseret på andre retlige grundlag end samtykke eller EU- og national ret som omhandlet i artikel 23, stk. 1, vil forenelighed mellem de oprindelige formål og de nye kunne afgøres kumulativt ud fra en række betragtninger. Der skal her ses på (a) sammenhængen mellem formålene; (b) konteksten ved indsamlingen herunder eventuel magtbalance; (c) personoplysningernes karakter og særligt, om de er følsomme; (d) viderebehandlingens konsekvenser for den registrerede; (e) og de aktuelle sikkerhedsforanstaltninger.

Personoplysninger viderebehandlet til udvikling eller anvendelse af AI vil ikke altid være det oprindelige formål. Som tidligere beskrevet i afsnit 3 er AI i praksis ofte anvendt som en form for effektivisering af allerede foretagne aktiviteter. Det kan være oplagt at anvende tidligere indsamlede informationer til at udvikle en AI i forbindelse med automatisering, idet udvikleren allerede besidder empiriske data til validering af AI'ens behandling og læring i form af tidligere behandlinger og deres resultater (supervised machined learning). Dette vil dog stride mod formålsbegrænsningen. Det er kun muligt at foretage en sådan viderebehandling, hvis udviklingen først og fremmest er omfattet af en af undtagelserne i punkt 4; der indhentes samtykke til de nye formål; eller hvis der iagttages vurderingsmomenterne i artikel 6, stk. 4. Undtagelserne vurderes i det efterfølgende afsnit, og nedenfor vurderes bestemmelsens litra a, d og e i forhold til udvikling og anvendelse af AI, idet disse har saglig relevans i modsætning til litra b og c, der er mere generelle vurderingsmomenter.

Artikel 6, stk. 4 bør betragtes som en helhedsvurdering. Udvikling af AI vil som udgangspunkt have relativt tæt forbindelse med det oprindelige formål, såfremt den udviklede AI's formål er at udføre eller forbedre den samme behandling, som eksempelvis blev udført manuelt forinden dens udvikling. Denne konklusion forudsætter dog, at den pågældende AI kan kategoriseres som machine learning eller mindre komplicerede systemer, idet det kan antages, at træningsdataene kendes på forhånd. Deep learning vil være mere kompliceret og mindre gennemskuelighed, hvorved forbindelsen mellem formålene vil være mindre håndgribelig.

Det kan ikke antages, at udvikling og anvendelse af AI kan have umiddelbare konsekvenser af teknisk karakter for den registrerede. Artikel 6, stk. 4, litra d formulerer dog "*konsekvenser*" i en sammenhæng, der ikke lægger an til nærværende konkretisering og bør derfor fortolkes bredt. En juridisk betragtning, der kan have konsekvens for den registrerede, vil dog være, at anvendelsen af AI fremfor individuel og manuel behandling vil kategoriseres som automatisk. Ved automatisk behandling af personoplysninger, er der strammere regler i forbindelse med deciderede afgørelser såsom vurdering af kreditværdighed, forsikringspræmier med mere. Automatiske afgørelser er behandlet i artikel 22 og vil blive gennemgået i afsnit 5.2.

Litra d og e kan ses i sammenhæng, idet en anden mulig konsekvens for den registrerede kan være potentielt øget risiko for lækage af deres personoplysninger alt afhængigt af de konkrete sikkerhedsforanstaltninger både i den tidligere behandling forinden implementeringen af AI og efter. GDPR nævner specifikt kryptering og pseudonymisering. Det vil være forkert at antage, at kryptering, pseudonymisering for ikke at nævne anonymisering altid er mulige løsninger i udviklingen af AI, men det må være konsensus, at disse umiddelbart effektive og relativt simple teknologier bør benyttes hvor end muligt. Pseudonymisering især er en oplagt anvendelsesmulighed i forbindelse med AI, da den pågældende model ikke må antages at gavne af egentlige navne eller andre indikatorer i langt de fleste tilfælde, da den blot beregner variable. Undladelse af anvendelse af pseudonymisering vil efter omstændighederne være et hensyn, der kraftigt må tale imod foreneligheden mellem formålene. Artikel 6, stk. 4 giver derfor anledning til relativt brede muligheder for at viderebehandle oplysninger til eventuelt at træne en AI, hvis der er en vis sammenhæng mellem det oprindelige formål, og hvad end AI'en skal anvendes til.

Er det ikke muligt at viderebehandle på baggrund af de ovenstående betragtninger, må der som tidligere angivet indhentes nyt samtykke eller anvendes en af undtagelsesmulighederne.

§ 5, stk. 3 i lovforslaget som fremsat efter 2. behandling den 15. maj 2018 åbner op for, at en minister i samråd med justitsministeren kan fastsætte regler for, at offentlige myndigheder må viderebehandle persondata til uforenelige formål. Såfremt offentlige myndigheder viderebehandler persondata efter § 5, stk. 3, begrænses oplysningspligten over for den registrerede efter § 23 i lovforslaget. Mange offentlige myndigheder anvender allerede data warehouses. Blandet andet anvender Undervisningsministeriet og derunder alle kommunerne data warehouses.[27] Dette kan ses som det første skridt mod automatisering og AI. Derved vil det øge muligheden for at benyttelse af AI i det offentlige, da AI er oplagt at anvende i en senere behandling – som tidligere nævnt i afsnittet. § 5, stk. 3 vil derfor kunne muliggøre, at offentlige myndigheder ikke er underlagt princippet om formålsbegrænsning eller oplysningspligten i dén forbindelse. Myndighederne kan muligvis frit opmagasinere personoplysninger uden den registreredes kendskab. Dertil skal det nævnes, at denne mulighed kun forekommer efter vedtagelse af en bekendtgørelse som hjemmel, hvorved de relaterererede folketingsudvalg har mulighed for at stemme imod.

[27] Undervisningsministeriet - cms.uvm.dk

4.2.4 Undtagelser (punkt 4)

Det 4. punkt i artikel 5, stk. 1, litra b omhandler mulighederne for at omgå forenelighedskriteriet. Bestemmelsen giver fire undtagelser, som ikke skal anses for at være uforenelige. De fire undtagelser, som den omhandler, er arkivformål i samfundets interesse, historiske forskningsformål, videnskabelige forskningsformål og statistiske formål. Kun de to sidstnævnte har særlig relevans for AI, selvom de to forrige kan have omstændelig betydning.

Præamblens punkt 157 nævner, at videnskabelig forskning skal fremmes, hvorved personoplysninger kan behandles med iagttagelse af passende betingelser og garantier. Ved udvikling og brug af AI kan det være svært at vide, hvad der er videnskabelig forskning dog.[28] GDPR har ikke en endegyldig definition af begrebet. Præamblens punkt 159 anfører dog imidlertid, at videnskabelig forskning skal forstås bredt. Den kommer yderligere med eksempler på begrebet såsom "*teknologisk udvikling og demonstration, grundforskning, anvendt forskning og privat finansieret forskning*". En almindelig forståelse af begrebet vil også omfatte, at der er tale om ny kundskab eller ny viden. I praksis inden for AI vil der derfor kunne fortolkes bredt, hvornår der er tale om videnskabelig forskning. Så længe der er tale om innovation af AI, vil det formentlig fortolkes som videnskabelig forskning. Det kan dog antages kun at være her i opstartsfasen af udbredelsen og innovationen af AI, hvor dette er tilfældet, da udvikling og brug af avancerede AI ikke er så udbredt endnu. Såfremt det belyses, at der er tale om videnskabelig forskning ved indsamlingen og ikke som en undtagelse, vil der efter præamblens punkt 33 være en vis frihed til at angive et formål. Punkt 3 angiver, at det skal være muligt for den registrerede at give sit samtykke til specifikke videnskabelige forskningsområder eller dele af forskningsprojekter, såfremt det er i overensstemmelse med anerkendte etiske standarder for området – i det omfang, det tilsigtede formål tillader det.

Den fjerde og sidste undtagelse til viderebehandling er statistisk formål. Præamblens punkt 162 kommer med en definition af begrebet. Den anfører, at det skal forstås som "*enhver indsamling og behandling af personoplysninger, der er nødvendig for statistiske undersøgelser eller frembringelse af statistiske resultater*". Den bestemmer yderligere, at de statistiske resultater kan viderebehandles til forskellige formål; herunder også videnskabelige formål. De statistiske data, der skabes som resultat af den statistiske behandling må ikke være personoplysninger, og de må ikke anvendes til fundament for foranstaltninger eller afgørelser, der vedkommer fysiske personer.

For at kunne benytte sig af de fire undtagelser, skal en virksomhed stille fornødne garantier til rådighed efter artikel 89, stk. 1. Garantierne skal beskytte den registreredes rettigheder og frihedsrettigheder samt især princippet om dataminimering ved tekniske og organisatoriske foranstaltninger. Det vil sige, at en virksomhed, der for eksempel udvikler en AI til videnskabelig forskning, skal træffe tekniske og organisatoriske foranstaltninger. Disse foranstaltninger skal blandt andet sikre princippet om dataminimering, så der kun indsamles dén mængde data, der er nødvendig for formålet og ikke mere. En foranstaltning, der eksplicit bliver nævnt i artikel 89, stk. 1, er pseudonymisering.

I tilfælde af AI, der anvender machine learning-teknikker, vil der som udgangspunkt være tale om statistisk behandling, hvilket vil sige, at såfremt der er truffet fornødne garantier; resultaterne af den statistiske behandling ikke anvendes til at foretage automatiske afgørelser på individer; og personoplysningerne aggregeres, så behøver viderebehandlingens formål ikke være foreneligt med det oprindelige formål.

I tilfælde af AI, der kan klassificeres som videnskabelig forskning, er der ikke de samme krav til at specificere formål, og på samme vis som med statistisk behandling er der videre muligheder for at viderebehandle personoplysninger.

[28] Datatilsynet.no – kunstig intelligens, 2018, s. 16

4.3 Dataminimering

Det tredje litra omhandler princippet om dataminimering. Bestemmelsen i artikel 5, stk. 1, litra c kræver, at der kun indsamles dén mængde persondata, der er nødvendig for at opfylde formålet. Derudover skal mængden dog også være tilstrækkelig for det specificerede formål. Derfor skal en virksomhed identificere det minimum af data, som kan benyttes til at opfylde formålet. Der er deraf to aspekter til udfordringen: Mængden af data og nødvendigheden af data for opfyldelse af formålet.

Bestemmelsen nævner ikke, hvilken situation den gælder i. Her skal forstås, at det ikke afgøres, hvorvidt kravet skal ses ved selve indsamlingen af personoplysninger, som det gælder i litra b om formålsspecifikation, eller om det er et krav til selve behandlingen. Der står blot, at kravene skal vurderes i forhold til formålene, "*hvortil de behandles*", men det afklarer ikke situationen for kravene. Beskrivelsen er absolut, hvorved kravet gælder for personoplysninger generelt. Det mest nærliggende er at fortolke som, at personoplysninger ikke engang må eksistere, hvis ikke de lever op til kravet. Der må heraf dog forstås, at de ikke må være i den dataansvarliges besiddelse, medmindre oplysningerne lever op til kravet om dataminimering.

Artikel 5, stk. 1, litra c indeholder beskrivelsen "*tilstrækkelige, relevante og begrænset*". Det er altså tre begreber, som ikke uddybes eller defineres yderligere i forordningen.

"*Tilstrækkelige*" beskriver, at persondataene skal være tilstrækkelige i forhold til det specificerede formål, hvorved mængden af data skal være fyldestgørende for formålet. Hvis dataene ikke er tilstrækkelige, så må de ikke behandles.

"*Relevante*" omhandler, at dataene skal være væsentlige og nødvendige for formålet. Det er mere hensigtsmæssigt at se kravet om relevans modsætningsvist, således det ikke er tilladt at behandle irrelevante oplysninger. Det må altså ikke være ubetydelige personoplysninger, som indsamles. I forbindelse med AI udgør dette krav også en praktisk fordel. Irrelevante oplysninger kan have som konsekvens, at især online AI vil risikere at finde fejlbehæftet kausalitet. AI, der anvender deep learning-teknikker, tillægger ofte information værdier, hvorved de vil medtage irrelevant information alligevel blot med en lav værdi. Uden en sådan foranstaltning risikeres der dog, at der træffes forkerte afgørelser, eller at der forekommer ineffektivitet i modellens anvendelse, idet resultaterne muligvis baseres på forkerte grundlag.

"*Begrænset*" dækker ligeledes over, at der skal være tale om et minimum af data, der indsamles. Det vil sige, at hvis en virksomhed kan opfylde formålet med et minimum af persondata, så skal det gøres i stedet. En virksomhed må derfor ikke holde på data af den grund, at det kunne være 'rart' at have dem. Disse tre begreber skal ses i forhold til formålet med indsamlingen af persondata og i forhold til individet. Vægtningen kan nemlig variere med den registrerede.

Ud fra artikel 29-gruppens udtalelser på området om biometri og kameraovervågning, tyder det på, at det tilsvarende dataminimeringsprincip i den tidligere persondatalovs § 5, stk. 3 skal vurderes ud fra et treleddet proportionalitetsprincip.[29] Det vil sige, at der skal foretages en proportionalitetsvurdering ud fra de tre elementer, egnethed, nødvendighed og forholdsmæssighed. Det må antages, at GDPR's princip om dataminimering skal anvendes på samme vis, idet den ikke afviger betydeligt fra § 5, stk. 3.

[29] Tranberg – Nødvendig behandling af personoplysninger, 2007, s. 540

Princippet om dataminimering skal også ses i sammenhæng med principperne i artikel 5, stk. 1, litra a om rimelighed. Rimelighed skal afgøres ud fra det indgreb, som personoplysningerne udgør for den registreredes rettigheder og frihedsrettigheder, og nødvendigheden for indsamlingen til formålet. En måde at sikre den registreredes rettigheder er ved at begrænse mængden af identificerende oplysninger og hvilke oplysninger, der indsamles, idet visse oplysninger siger mere om en person end andre. Dette udføres mest effektivt i praksis ved hjælp af foranstaltninger som pseudonymisering, krypteringsteknikker, aggregering og eventuelt anonymisering, hvilke vil blive gennemgået nærmere i afsnit 4.6.7 om integritet og fortrolighed. Rimeligheds- og forholdsmæssighedsvurderinger har dog stor indflydelse på hinanden, og de to begreber kan være svære at skelne. Forholdsmæssighed omhandler netop, om behandlingen nu også er forholdsmæssig med formålene, derunder nærmere om det er forholdsmæssigt i henblik på oplysningspligten og den registreredes frihedsrettigheder og rettigheder.

Princippet om dataminimering skal yderligere ses i sammenhæng med princippet om opbevaringsbegrænsning i artikel 5, stk. 1, litra e. Oplysninger vil ikke være relevante til evig tid, og derfor bør de begrænses til de tilstrækkelige oplysninger på det tidspunkt, hvor de ikke længere er nødvendige til formålene. Princippet om opbevaringsbegrænsning er dermed blot en fremtidig implementering af dataminimering. Opbevaring er en behandling, og hvis oplysningerne ikke længere er relevante, så skal de slettes. Hvorved dette forekommer som en foranstaltning for at leve op til kravet om dataminimering. En vurdering om egnethed i forhold til proportionalitetsprincippet bør derfor foretages, idet den dataansvarlige hele tiden skal leve op til, om foranstaltningerne er egnede.

Artikel 5, stk. 1, litra c om dataminimering beskriver også "*nødvendigt i forhold til de formål*". "*Nødvendigt*" taler for, at der skal være et kendskab til, hvad der skal indsamles til formålet. Hvis der ikke kendes, hvilke oplysninger, der er tilstrækkelige, relevante og nødvendige for formålet, kan dataene heller ikke indsamles.

Det interessante er her, at det ikke altid kan forudses, hvad der er nødvendigt. En AI vil som tidligere beskrevet i mange tilfælde afhænge af så stor en mængde data som muligt. Jo mere data, desto større statistisk præcision. Som tidligere redegjort for er AI ofte udtryk for big data, hvorledes der fastsættes en algoritme ud fra sammenhænge, som modellen kan finde i et givent datasæt. Hvis en AI eksempelvis vil afgøre sandsynligheden for at bo i et givent område ud fra de registreredes husdyrhold, er sammenhængen relativt simpel at undersøge statistisk, idet der blot er to nødvendige parametre at måle på og sammenligne: husdyrhold og bopæl. Her er de nødvendige oplysninger klart definerede på forhånd, men en online AI, der løbende lærer, vil muligvis kunne finde yderligere sammenhænge med andre parametre såsom i antal af græsarealer i området. Her skal det afgøres på forhånd, hvilke af informationerne, der er nødvendige, hvorfor det ikke altid kan være muligt at finde disse yderligere sammenhænge.

Vil en AI til gengæld eksempelvis afgøre bankkunders kreditværdighed ud fra deres livsstile, er det ikke altid muligt ved udvikling af modellen at afgøre, hvilke parametre, der viser sig at være kausale. En livsstil er et abstrakt udtryk for adskillige sociale og kulturelle karakteristika. Det vil være nødvendigt med en bred vifte af informationer for at afgøre, hvad der har relevans for kreditværdighed, men det vides ikke på forhånd, hvilke af informationerne, der viser sig at være afgørende; eller – sagt anderledes – tilstrækkelige og relevante.

Betyder det så, at resten af informationerne skal slettes efterfølgende, eller at informationerne slet ikke måtte indsamles? Som tidligere afgjort må den dataansvarlige ikke engang være i besiddelse af personoplysninger, hvis ikke det kan afgøres, at oplysningerne er tilstrækkelige og relevante. Det må derfor umiddelbart betyde, at det ikke er muligt at indsamle oplysninger til en behandling, hvis ikke det på forhånd vides med sikkerhed, hvad deres relevans bliver. AI kan derfor ikke beskæftige sig med uforventede korrelationer, der baserer sig på personoplysninger – som udgangspunkt. Dertil kommer konklusionen fra analysen af formålsspecifikation om, at det konkrete formål med en indsamling skal kendes på forhånd.

Men dataminimering handler ikke bare om at begrænse mængden af personoplysninger. Princippet handler i høj grad om at vurdere proportionaliteten af en behandling i forhold til, at den ikke skal udgøre et større indgreb overfor den registrerede end nødvendigt.[30] [31] Det er i hvert fald den britiske informationskommissær og det norske datatilsyns meninger. Men hvor kommer disse holdninger fra? Det norske datatilsyn påpeger i deres rapport, at det er nødvendigt at sætte sig grundigt ind i reglerne om formålsspecifikation og dataminimering samt de specifikke detaljer for, hvordan modellen konkret fungerer og behandler oplysningerne. Dette er til at sikre, at personoplysninger om muligt bliver slettet, hvis det efterfølgende viser sig, at oplysningerne var unødvendige til formålet. Disse vurderinger kræver en vis juridisk og teknisk baggrund og kræver specialiserede kompetencer at udføre i praksis. Til gengæld nævner tilsynet ikke deres hjemmel til denne fortolkning i GDPR. En proportionalitetsvurdering ud fra forholdsmæssighed vil dog tyde på samme konklusion.

Hvis der forudsættes, at vi i et konkret eksempel anvender en online deep learning-model – altså en AI, der løbende gennem dens anvendelse lærer og inkorporerer abstrakte parametre, som den observerer, i sin model. Den dataansvarlige angiver, at formålet med en løbende indsamling af personoplysninger – foretaget autonomt af AI'en – er, at AI'en skal udvikle sig, således den med større præcision kan foretage forudsigelser, og den vil gøre dette igennem både at opnå et større datagrundlag samt at analysere sig frem til nye sammenhænge i inputdataene, som den så kan bruge som parametre i sine forudsigelser. Kort sagt vil den ved at opnå flere personoplysninger danne flere neurale lag, som den kan analysere inputdata igennem. Formålet med behandlingen er således forholdsvist specifikt efter omstændighederne.

I det forudsatte tilfælde skal vi vurdere, om behandlingen lever op til kravet om dataminimering. Det vides ikke, hvilke sammenhænge AI'en vil finde frem til. Vi ved bare, at den vil indsamle og i øvrigt behandle personoplysninger statistisk for at lede efter sammenhænge. Vi kan derfor ikke afgøre, om personoplysninger er tilstrækkelige, relevante og begrænsede til, hvad der er nødvendigt i forhold til formålene. Formålet er, at AI'en skal udvikle sig ved at finde nye sammenhænge, og når vi ved indsamlingen af oplysningerne ikke ved, om der findes nye sammenhænge ud fra oplysningerne, så ved vi heller ikke, hvilke er oplysningerne, der er nødvendige eller unødvendige. Vi kan modsat være sikre på, at nogle af oplysningerne vil være unødvendige, fordi vi først efter at kende kausaliteten kan vide, hvilken mængde oplysninger, der medfører statistisk signifikans.

Det norske datatilsyn mener i dette tilfælde, at så længe der er truffet passende foranstaltninger såsom pseudonymisering og kryptering, samt at der i et vist arbitrært omfang er minimeret mængden af data, så vil behandlingen stadig være proportionel med formålet. Denne konklusion kan dog ikke findes i litra c. Litra c lyder klart, at personoplysninger til enhver tid skal være relevante til formålet. Der lægges ikke op til friheder til at vurdere, om en behandling er passende til formålet. Der er ingen vurdering af, om behandlingen er saglig – kun om den er nødvendig. Enten er personoplysningerne nødvendige, eller også er de ikke. Det kan først efterfølgende behandlingen afgøres, hvilke af informationerne, der endte med at være nødvendige.

[30] Datatilsynet.no – kunstig intelligens, 2018, s. 17
[31] ICO, 2017, s. 42

Fortolkningen stemmer overens med kravet i litra a til rimelighed i behandling. En rimelig behandling indebærer, at den registrerede ikke skal udsættes for unødvendige og potentielle sikkerhedsrisici, som uundgåeligt medfølger enhver indsamling. Den registrerede bør kunne sikre sig, at de oplysninger, der indsamles om den, faktisk anvendes til de angivne formål. Det norske datatilsyn opererer angiveligt ud fra det synspunkt, at hvis personoplysninger under behandlingen har vist sig at være unødvendige, så skal de blot slettes igen.[32] Denne fortolkning stemmer overens med, hvad der blev konkluderet i starten af dette afsnit, da en dataansvarlig ikke må have adgang til personoplysninger, hvis ikke disse er nødvendige til et konkret formål. Det må derfor være tilfældet, at den dataansvarlige løbende skal vurdere personoplysningernes nødvendighed for at slette unødvendige oplysninger. Til gengæld er sikkerhedsrisiciene allerede forekommet ved indsamlingen, hvilket er en frihedsrettighed, der bør iagttages. Den registrerede har derfor alligevel en interesse i, at dennes personoplysninger kun bliver indhentet, hvis disse faktisk er nødvendige ved indsamlingen.

Spørgsmålet er så, om det norske datatilsyns fortolkning kan være i overensstemmelse med GDPR, hvis der antages nogle forudsætninger, som de ikke selv nævner i rapporten. Der kan tages udgangspunkt i, at den dataansvarlige ved indsamlingen af oplysningerne anser dem som nødvendige, fordi den mener, at der er en vis sandsynlighed for, at oplysningerne vil være relevante. Vi kan som sagt først endegyldigt afgøre nødvendigheden efterfølgende, men så længe der er en betydelig sandsynlighed for, at personoplysninger vil være behjælpelige til at udføre formålet, kan det antages, at behandlingen stadig er saglig og nødvendig i forhold til formålene. Dette må være den forudsætning, som det norske datatilsyn implicit anvender, samt en proportionalitetsvurdering ud fra foranstaltningernes egnethed, nødvendighed og forholdsmæssighed.

Hvis der i det forudsatte tilfælde skal kunne behandles personoplysninger, som vi ved indsamlingen ikke kan afgøre nødvendigheden af, må vi derfor lave en vurdering ud fra principperne om dataminimering, formålsspecifikation, opbevaringsbegrænsning, rimelighed, gennemsigtighed og ikke mindst proportionalitet; herunder egnethed, nødvendighed og forholdsmæssighed.

- Vil behandlingen udgøre et større indgreb overfor den registrerede end, hvad der er nødvendigt?
- Er sandsynligheden for, at oplysningerne bliver nødvendige betydelig?
- Er der truffet videst mulige foranstaltninger til at mindske identificering af den registrerede såsom igennem pseudonymisering eller kryptering?
- Bliver der løbende vurderet nødvendigheden af personoplysningerne?
- Er der truffet foranstaltninger til sletning af de personoplysninger, der efterfølgende viser sig at have været irrelevante?
- Er oplysningspligten blevet tilstrækkeligt iagttaget, og er den registrerede blevet oplyst om, at oplysninger vil blive slettet, hvis de viser sig senere ikke at være nødvendige?
- Er formålet beskrevet så specifikt som muligt?

Bliver disse krav tilstrækkeligt respekteret og iagttaget må en subjektiv fortolkning af artikel 5, stk. 1, litra c konkludere, at det er muligt at behandle personoplysninger, selvom det strengt talt ikke kan afgøres absolut, om de ender med at blive anvendt til formålet.

Det skal dog understreges, at der som altid må foretages en konkret vurdering, som der som sagt bør udføres af en specialist med forståelse for både de juridiske krav, men også de tekniske omstændigheder og grundigt kendskab til den pågældende models virke. Den dataansvarlige skal til gengæld først og fremmest være opmærksom på, at den ikke kan anvende så mange oplysninger som muligt. Den skal i stedet forsøge at begrænse sig til, hvad den mener er nødvendigt. Er dette ud fra omstændighederne ikke muligt at præcisere nøjagtigt, er det vigtige, at behandlingen efter omstændighederne er rimelig.

[32] Datatilsynet.no – kunstig intelligens, 2018, s. 17

4.4 Rigtighed

Personoplysninger skal til enhver tid være rigtige efter artikel 5, stk. 1, litra d. Kvaliteten af dataene påvirker især også AI, da de resultater, en AI vil nå frem til, kan være forkerte, hvis algoritmen ikke er nøjagtig, er opbygget af fejlfortolkede data, har været udsat for misbrug eller ikke vedligeholdes.

Et eksempel på, hvor vigtigt det kan være med kvalitetssikring af data, er Microsofts AI, Tay, der tidligere blev omtalt i afsnit 3.1.4. Tay skulle gennem inputdata fra samtaler over Twitter kontinuerligt lære at interagere med mennesker. Tay blev dog givet korrumperet på grund af racistiske og falske kommentarer fra andre Twitterbrugere. AI'en blev kort tid efter taget ned, da den selv begyndte at komme med racistiske og andre upassende bemærkninger. Dette eksemplificerer vigtigheden af rigtighed ud fra et pragmatisk synspunkt som udvikler af en AI. Vigtigheden understreges ved anvendelsen af personoplysninger, der kan risikere at påvirke den registrerede.

Ud fra bestemmelsen kan det antages, at der er fire udfordringer, som en virksomhed bør overveje for at kunne overholde princippet om rigtighed. Virksomheden skal:

1) Tage rimelige skridt til at sikre rigtigheden af personoplysningerne
2) Tage stilling til troværdigheden af informationskilden
3) Overveje indvendingerne mod rigtigheden af personoplysningerne
4) Overveje om det er nødvendigt at opdatere informationen

4.4.1 Rimelige skridt til at sikre rigtigheden af personoplysningerne (punkt 1)

Artikel 5, stk. 1, litra d anfører, at oplysninger skal være "*korrekte og om nødvendigt ajourførte*". Der er derved de to begreber, "*korrekte*" og "*ajourførte*", heri, som er værd at tage stilling til. Forordningen kommer ikke med yderligere uddybelser eller definitioner til, hvad korrekt egentlig skal sige. Den beskriver dog, at "*der skal tages ethvert rimeligt skridt for at sikre […]*", at urigtige personoplysninger slettes eller berigtiges. Spørgsmålet er dog, hvad rimelige skridt egentlig vil sige. Ud fra ordlyden af bestemmelsen lægger det op til, at det kommer an på formålene. Der er altså her tale om en rimelighedsvurdering, hvor der ses på vigtigheden af, om oplysningerne er korrekte i forhold til formålene. Hvis det berører individet, bør der nok tages skridt til at sikre rigtigheden af personoplysningerne. Et eksempel kan være med en ansats bankoplysninger. Det er her vigtigt, at der tages skridt til at sikre, at oplysningerne er korrekte, da en ansats løn ellers vil havne det forkerte sted.

4.4.2 Troværdigheden af informationskilden (punkt 2)

Indsamling af oplysninger kan foretages på to måder umiddelbart. De kan indhentes fra individet selv eller fra en tredjepart. Når oplysningerne indhentes, bør en virksomhed overveje troværdigheden af kilden. En AI med online model bør derfor også afgøre kildens troværdighed automatisk, hvis den autonomt indhenter oplysninger. Dette bør især gælde, hvis det har en stor indvirkning på den registrerede. Ukorrekte oplysninger kan både komme fra den registrerede, men også tredjeparter, hvor en virksomhed derved vil have svært ved at overholde princippet om rigtighed. Bestemmelsen giver anledning til at tolke, at dette ikke bryder artikel 5, stk. 1, litra d om rigtighed. En virksomhed skal derfor have foretaget visse foranstaltninger for at overholde princippet om rigtighed, hvis de har ukorrekte oplysninger. De skal have taget rimelige skridt for at sikre rigtigheden, korrekt håndtering af oplysningernes registrering, samt eventuel markering af den registreredes bestridelse af oplysningernes rigtighed.

4.4.3 Indvendinger mod rigtigheden (punkt 3)

Artikel 5, stk. 1, litra d angiver, at urigtige oplysninger i forhold til de specificerede formål straks skal slettes eller berigtiges. Artikel 16 anfører endvidere også den registreredes rettighed til berigtigelse. I begge artikler står der eksplicit "*urigtige*", hvorved en virksomhed skal tage stilling til, om en oplysning reelt set er urigtig, hvis den registrerede bestrider oplysningens rigtighed. Ergo må det være den dataansvarliges ansvar at vurdere oplysningernes rigtighed og derved ikke blot tage den registreredes ord for det.

4.4.4 Ajourføring (punkt 4)

De urigtige oplysninger skal dog yderligere ses i forhold til de specificerede formål, anfører artikel 5, stk. 1, litra d om rigtighed. Dette kan antyde, at det ikke altid er nødvendigt at ajourføre, så oplysningerne er korrekte. Det må lægge op til, at det kun er der, hvor det er "*nødvendigt*" i forhold til formålet. Så hvis det er meget vigtigt for at opfylde et formål, at oplysninger er korrekte, så skal de opdateres. Ved eksempelvis statistisk, historisk eller anden forskningsmæssig behandling skal oplysningerne ikke nødvendigvis være korrekte eller ajourførte som et krav af GDPR. Ved indsamling i forbindelse med big data, hvor der indsamles store mængder data på mange forskellige personer, og hvor der kun ledes efter tendenser, er det nok ikke nødvendigt, at alle oplysninger er opdateret eller korrekte. Det afhænger meget af det pågældende formål med behandlingen, om hvor meget der skal lægges vægt på rigtigheden.

4.4.5 Berigtigelse og sletning gennem AI

Artikel 5, stk. 1, litra d angiver, at urigtige oplysninger i forhold til formålene skal slettes straks. Den registrerede har desuden efter artikel 17, stk. 1 også ret til at få slettet oplysninger om sig selv, hvis blandt andet personen trækker sit samtykke tilbage, eller oplysningerne ikke længere er nødvendige for at opfylde de specificerede formål. En gruppe forskere fra både Yale, SBA og Queen Mary University in London har udgivet en rapport i 2017 om problemstillinger inden for retten til at blive glemt og AI.[33] De mener, at der opstår to særlige problemstillinger inden for dette. Den første er, at "slettes" ikke er defineret i loven, så ved eksempelvis at overskrive dataene i AI'en vides det ikke, om det i realiteten er i overensstemmelse med GDPR. Den anden problemstilling er, at en online AI lærer kontinuerligt, så selv hvis dataene slettes, ligger disse data andre steder i AI'en også, da den har lært af dem. Så hvis der er tale om en AI med deep learning-teknikker, vil disse data sandsynligt findes i de mange forskellige abstraktionslag, den besidder. Problematikken om den sorte boks opstår her. Gruppen foreslår, at metoder til at forebygge retten til at blive glemt med mere er anonymisering og pseudonymisering.[34] Men der vil stadig være et problem, såfremt oplysninger er urigtige, da AI'en sandsynligt vil være mere partisk, da den stadig har oplysningerne. På den ene side må en AI i så fald ikke behandle personoplysninger, men på den anden side kan loven ikke følge med den teknologiske udvikling, hvilket også er konklusionen for forskergruppen. Loven har ikke defineret hverken berigtigelse eller sletning, så en implementering af pseudonymisering og anonymiseringsteknikker kunne tænkes at leve op til kravene alligevel. Det er dog uklart.

[33] Li - papers.ssrn.com
[34] Li - papers.ssrn.com, s. 13

Efter artikel 16 skal en dataansvarlig som sagt efter indvending fra den registrerede berigtige urigtige oplysninger. Artikel 16 og 5, stk. 1, litra d kan dog ligeledes skabe problemstillinger for udviklere og anvendere af AI. Det kan tænkes, at det ikke altid vil være tilfældet, at det vil være muligt at berigtige oplysninger. En AI vil som regel opbevare aggregerede data frem for personoplysninger. Ved aggregerede data kan det være tilfældet, at personen alligevel kan identificeres ud fra et sammenhold af oplysninger eller ud fra oplysningers karakter.[35] [36] De samme problemstillinger opstår her som ved sletning, idet der ikke er en definition af "berigtigelse", og en online AI lærer kontinuerligt, så selv hvis dataene berigtiges ét sted, ligger disse data andre steder i AI'en også, da den har lært af dem. Pseudonymisering og anonymiseringsteknikker kunne ligeledes tænkes at løse problemstillingen både med hensyn til retten til berigtigelse, men potentielt også i forhold til AI'ens farvet syn ved urigtige oplysninger.

De samme problemstillinger ved sletning og berigtigelse gør sig ikke gældende i nær så stor grad ved en offline AI. Offline AI lærer per definition ikke kontinuerligt, og den vil derfor ikke være lige så kompliceret.

[35] Datatilsynet.no – kunstig intelligens, 2018, s. 9
[36] Datatilsynet – anonymisering - datatilsynet.dk

4.5 Opbevaringsbegrænsning

Artikel 5, stk. 1, litra e omhandler kravet om, at der ikke må opbevares personoplysninger i et længere tidsrum end nødvendigt til formålet, hvorved princippet kendes som opbevarings- eller tidsbegrænsning. Det er dermed et krav om at slette oplysninger, der ikke længere kan anvendes. Som angivet hænger dette sammen med kravet om begrænsning i forbindelse med princippet om dataminimering i litra c. I dén henseende skal litra e ses som en udvidelse af litra c, hvorved der løbende skal tages stilling til oplysningers relevans. Umiddelbart kan dette også tolkes særskilt ud af litra c, men princippet om opbevaringsbegrænsning uddyber udførslen. Især i forbindelse med AI og big data har litra e relevans, men det viser sig at være i en mere strukturel og teknisk sammenhæng end en juridisk.

Teknologihistoriker, George Dyson, påstod i forbindelse med et seminar i 2013 følgende: "*Big data is what happened when the cost of storing information became less than the cost of making the decision to throw it away*".[37] Vi lever i en tidsalder, hvor lagringskapacitet og muligheder for indsamling af data forøges i takt med, at omkostningerne forbundet dermed falder. Ralph Jacobson fra IBM skrev i 2013 et "*industry insight*" med følgende påstand: "*90% of the data in the world today has been created in the last two years. This data comes from everywhere: sensors used to gather shopper information, posts to social media sites, digital pictures and videos, purchase transaction, and cell phone GPS signals to name a few. This data is big data*".[38] Naturligvis må der ligge kommercielle interesser bag, og der er tilsyneladende et enormt økonomisk incitament til at ophobe så megen information som muligt.

Oprindeligt blev Dyson fejlciteret, således markeringen i den følgende sætning "*the cost **of making the decision** to throw it away*"[39] var udeladt. Det er en vigtig nuance, at Dyson inkluderer, at selve beslutningen bag sletning af data i sig selv er forbundet med omkostninger. Incitamenterne bag informationsophobning er et let udgangspunkt, hvor sletning modsætningsvist er den aktive handling, der skal udføres. De nye regler om sanktioner i GDPR bliver introduceret på et tidspunkt, hvor vi kan forestille os, at adskillige virksomheder allerede har truffet en beslutning i deres vurdering af, hvad der vil medføre flest omkostninger: At udvide lagringskapacitet eller at implementere foranstaltninger til vurdering og afskaffelse af irrelevante data. Rigtigt nok viser en undersøgelse fra den amerikanske udbyder af hosting-ydelser og AI-infrastruktur, PureStorage, at 72% af de adspurgte selskaber i Storbritannien, Frankrig og Tyskland har indsamlet data, de er endt ud med ikke at anvende.[40]

Foranstaltningerne kan dog virke uoverskuelige og decideret uhensigtsmæssige i praksis i forbindelse med AI. Ikke nok med, at problematikken om den sorte boks kan medføre en nødvendighed af fundamentale ændringer i modellens struktur, så vil det også kræve en teknisk indsigt i modellen. Sletning af data, der indgår i især modeller, der anvender deep learning-teknikker kan være omkostningstungt eller decideret umuligt at udføre – mere om dette i afsnit 4.4 om rigtighed.

Litra e om opbevaringsbegrænsning præciserer, at det ikke skal være muligt at identificere den registrerede længere end nødvendigt. Vurderes det, at personoplysningers relevans er udløbet, men kan de modificeres, således den registrerede ikke kan identificeres længere, kan informationerne efter ordlyden stadig opbevares. Desuden indeholder litra e de samme undtagelser som formålsbegrænsningen, herunder behandling til arkivformål i samfundets interesse, videnskabelige eller historiske forskningsformål eller statistiske formål.

[37] Dyson - longnow.org
[38] Jacobson - ibm.com
[39] O'Reilly - plus.google.com
[40] Pure Storage - info.purestorage.com, s. 7

Inkludering af undtagelsesmuligheder i litra e gør det klart, at der ved anvendelse af de samme undtagelsesmuligheder til viderebehandling også er hjemmel til at opbevare dataene, indtil den egentlige viderebehandling finder sted. På samme tid betyder det naturligvis også, at anvendelse af undtagelsesmulighederne for at opbevare længere end nødvendigt for formålene ikke kan finde sted, medmindre der også er hjemmel til at viderebehandle.

Udviklere og anvendere af AI har særligt incitament til at undlade at slette personoplysninger, idet behovet for data er relativt større. Forordningens krav resulterer dermed i et potentielt paradigmeskift. Sletning af ikke længere nødvendige oplysninger sværere for især online AI, der anvender deep learning-teknikker. Det kan desuden være omkostningstungt for den dataansvarlige at implementere foranstaltninger til sletning. For at imødekomme disse problemstillinger, kan der alternativt anonymiseres eller aggregeres oplysninger for at fjerne oplysningernes identificerende karakter.

4.6 Integritet og fortrolighed

Når der tales om tekniske og organisatoriske sikkerhedsforanstaltninger, relaterer de tekniske til tekniske strukturer, som AI må antages at være defineret som. Foranstaltningerne skal derfor vurderes konkret, hvorved det er nødvendigt først at analysere artikel 5, stk. 1, litra f dybdegående på forhånd for at udforske, hvilke sikkerhedsmæssige karakteristika ved GDPR, der har relevans for AI. Det vil sige, at det forordningens bestemmelser om sikkerhed som udgangspunkt er generelle, men at der ved nærmere analyse vil vise sig konkrete krav og overvejelser, som har speciel relevans for AI.

Artikel 5, stk. 1, litra f omhandler det såkaldte princip om integritet og fortrolighed. Ud fra et juridisk perspektiv er det mest betydningsfulde i denne bestemmelse, at der bliver anvendt adskillige vagheder, såsom "*på en måde*", "*tilstrækkelig*" og "*passende*", der alle lægger op til konkrete vurderinger. Enhver kan (og bør) vejledes i organisatoriske foranstaltninger, men de tekniske foranstaltninger kræver en vis specialiseret og uddannelsesmæssig baggrund at kunne anskue og implementere. Bemærk derfor, at idet de tekniske foranstaltninger er mere specialiserede i deres karakter, end de er juridiske, vil dette afsnit i sammenhæng med problemformuleringen kun berøre de tekniske aspekter i forbindelse med potentielle problemstillinger og derfor ikke løsninger.

4.6.1 Forordningens krav til sikkerhedsforanstaltninger

"*Tilstrækkelig*" giver umiddelbart et indtryk af, at personoplysninger ikke må blive udsat for risici. Der er altså som udgangspunkt ikke konkrete minimumskrav til sikkerhedsforanstaltningerne, hvorved i en situation med en sikkerhedsbrist, der kunne være undgået med visse foranstaltninger, vil den dataansvarlige ikke leve op til bestemmelsen.

Dertil skal artikel 32 dog iagttages, da denne fremsætter lempelige modifikationer og nævner specifikke foranstaltninger. Artikel 32, stk. 1 og artikel 5, stk. 1, litra f stemmer dermed ikke overens. Litra f tager ingen forbehold for, hvor effektive foranstaltninger skal være på grund af ordet, "*tilstrækkelig*". Hvorimod artikel 32, stk. 1 efterlader den dataansvarlige et råderum, hvor sikkerhedsforanstaltningerne derudfra skal være "*passende*". I dén forbindelse modsiger litra f også til dels sig selv, idet den både kræver "*tilstrækkelig sikkerhed*", men kun "*passende tekniske eller organisatoriske foranstaltninger*".

Den engelske version indeholder to tilfælde af ordet, "*appropriate*", hvor den danske version i deres sted anvender "*tilstrækkelig*" og "*passende*", som dermed bliver skelnet imellem, idet "*passende*" kan fortolkes mere lempeligt som "*rimelig, rigtig eller fornuftig i en bestemt sammenhæng eller situation*".[41]

"*Appropriate*" og "*tilstrækkelig*" vil efter en ordlydsfortolkning have forskellige betydninger, idet "*tilstrækkelig*" må antages at betyde "*fyldestgørende*" eller "*af et omfang eller en kvalitet, der er stor nok til at opfylde et bestemt behov, krav, ønske el.lign.*"[42] "*Appropriate*" kan nærmere oversættes til "*passende*" eller synonymiseres med "*suitable*" eller "*fitting*".[43]

[41] Den Danske Ordbog – passende - ordnet.dk/ddo/ordbog?query=passende
[42] Den Danske Ordbog – tilstrækkelig - ordnet.dk/ddo/ordbog?query=tilstr%C3%A6kkelig
[43] Dictionary.com - dictionary.com/browse/appropriate

Den tyske version anvender henholdsvis "*angemessene*" og "*geeignete*", som også begge i højere grad kan sidestilles med "*passende*". Det samme gør sig gældende for den franske og svenske som henholdsvis anvender "*appropriée*" og "*lämplig*". Alle disse ord indeholder forskellige kulturelle og sproglige nuancer, men har dét tilfældes, at de udtrykker et element af råderum og forholdsmæssighed i modsætning til *"tilstrækkelig"*, som har en ganske finit konnotation. Det kan derfor udledes, at tilstrækkelig nok ikke skal forstås som, at der ingen minimumskrav er og dermed i stedet forstås som betydningen af "*passende*".

Foruden overstående fortolkning af foranstaltningernes udstrækning indeholder artikel 32, stk. 1 en proportionalitetsvurdering, hvor de konkrete risici skal fastlægge, hvilken grad af foranstaltninger, der skal gennemføres.

Den dataansvarlige og databehandleren skal overveje følgende vurderingsmomenter efter artikel 32, stk. 1:

1. Det aktuelle tekniske niveau;
2. Sikkerhedsforanstaltningers implementeringsomkostninger;
3. Den pågældende behandlings:
    - Karakter;
    - Omfang;
    - Sammenhæng; og
    - Formål;
4. Den konkrete behandlings risici; samt
5. Potentielle konsekvenser for de registreredes rettigheder og frihedsrettigheder.

Disse parametre skal samlet vurderes og analyseres, hvorefter der ud fra deres beskaffenhed skal fastsættes sikkerhedsforanstaltninger.

4.6.2 Det aktuelle tekniske niveau (punkt 1)

Især det aktuelle tekniske niveau (1) – på engelsk: "*state of the art*" – og implementeringsomkostningerne (2) står i kontrast til de andre vurderingsmomenter. Hvis der antages et udgangspunkt, hvor der ingen sikkerhedsforanstaltninger er gennemført endnu, kan disse to momenter anses for de lempelige vurderingsgrundlag, mens resten af vurderingsmomenterne (3, 4 og 5) bør opfattes som de skærpende. Det aktuelle tekniske niveau er et forbehold for, at it-sikkerhed er i fortsat udvikling, og at det i fremtiden sandsynligvis vil omfatte discipliner og anderledes målestokke, som ikke er hensigtsmæssige nu, men vil være mindre ressourcekrævende i fremtiden.

Til gengæld dækker "*det aktuelle tekniske niveau*" (1) over mere end blot tidsmæssig og socioteknisk udvikling. AI – navnlig machine learning og deep learning – er per definition software, der udvikler sig. Som redegjort for tidligere, afgrænser vi deep learning fra resten af kategorien, machine learning, ved tilstedeværelsen af en vis mængde lag af pseudo-kognitive funktioner.

Forskellen kan eksempelvis illustreres med Akinator[44] – som anvender machine learning-teknikker til at gætte, hvilken person eller fiktiv karakter man tænker på ud fra få og vage spørgsmål – overfor en såkaldt androide som Sophia[45], der imiterer komplekse og menneskelige egenskaber. Begge typer anvender ny teknologi, men det aktuelle tekniske niveau kan også forstås som en vurdering af den anvendte teknologis konkrete kompleksitet. Vi ved, at en kompleks AI – med mange lag og uden inkorporerede metoder til at fremvise ræsonnementer og belæg – til en vis grad risikerer at være omfattet af problematikken om den sorte boks. I et sådant tilfælde bør sikkerhedsforanstaltninger naturligvis tage hensyn til, at kompleksiteten er en anden end ved en relativt mere simpel machine learning-algoritme; ikke blot i forhold til uautoriseret adgang, men også enhver anden utilsigtet konsekvens for de pågældende personoplysninger.

Dertil er det også nødvendigt at bemærke sig ud fra "*aktuelle*", at idet online AI per definition udvikler sig, skal der også løbende tages stilling til dets kompleksitet og tekniske niveau. En AI, der har udviklet sig fra at være relativt simpel til at være kompleks, vil potentielt kræve mere eller sågar mindre omfangsrige sikkerhedsforanstaltninger.

4.6.3 Implementeringsomkostninger (punkt 2)

Implementeringsomkostninger er i denne sammenhæng af størst betydning i praksis, idet der tages højde for virksomhedens konkrete karakter. Det må formodes, at omkostninger efter GDPR ikke skal forstås som en statisk størrelse, idet omkostninger har forskellige relative værdier for hver virksomhed alt efter deres størrelse. Dette er med al sandsynlighed dét vurderingsmoment, der vil have størst betydning for små og mellemstore virksomheder, idet det ikke forventes af dem, at de skal gå nedenom og hjem for at overholde lovgivningen. Det er svært at forestille sig et maksimum af foranstaltninger, der teoretisk ville kunne gennemføres for at opleve størst mulig sikkerhed. Dog ville det sandsynligvis skade virksomhedens effektivitet at efterstræbe 100 % sikkerhed. Der vil altid være mulighed for at forbedre sikkerhed, men virksomhederne kan nøjes med at stile efter, hvad der forholdsmæssigt udførligt for dem.

4.6.4 Skærpende vurderingsmomenter (punkt 3, 4 og 5)

Artikel 32, stk. 1 og 2 anvender også "*passende*" som gradsbestemmelse af implementeringsbehovet og fortsætter med, "*tekniske og organisatoriske foranstaltninger for at sikre et sikkerhedsniveau, der passer til disse risici*", hvorefter der nævnes en liste af eksempler på sådanne foranstaltninger. "*Passende*" refererer ikke til de vurderingsmomenter, der blev nævnt tidligere, men i stedet sikkerhedsniveauet og de relaterede risici. Det skal blot være "*under hensyntagen*" dertil.

Sikkerhedsniveauet skal passe til de risici, der tidligere blev beskrevet som de skærpende vurderingsmomenter. Stk. 1 beskriver selv, at disse risici kan være af varierende sandsynlighed, hvilket henviser til, at risici følger omstændighederne og kan være særdeles subjektive. Der opstår derfor en problemstilling i, at det påhviler den dataansvarlige eller databehandleren at afgøre, hvilke konkrete risici, der kan forekomme, og som tidligere angivet er dette et specialiseret fag i sig selv, der kræver en vis baggrund bare at kunne identificere for ikke at nævne forebygge.

[44] Akinator - en.akinator.com/
[45] Sophia - youtube.com

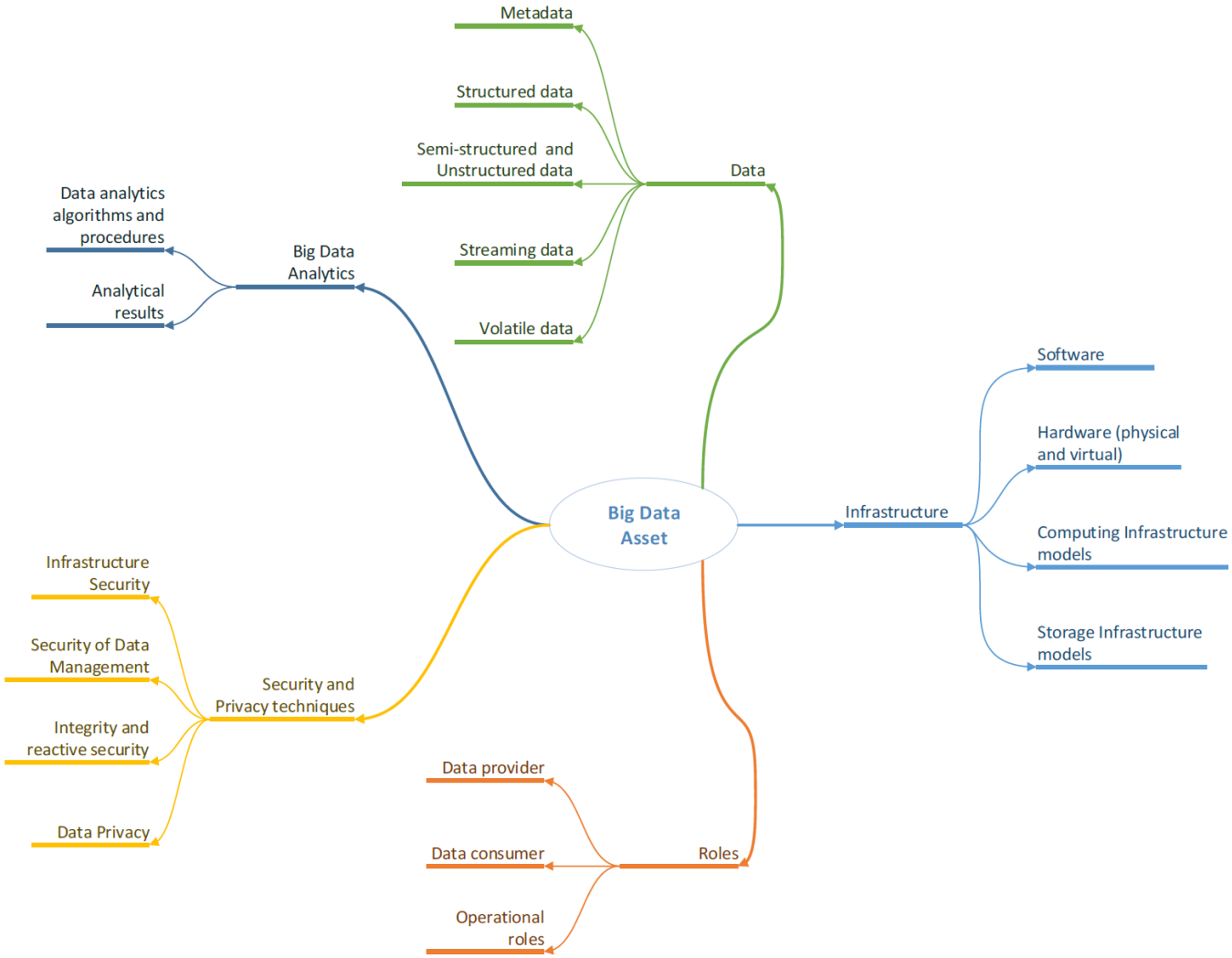
Artikel 5, stk. 1, litra f indeholder formuleringer, der kan hjælpe til at afgøre, hvordan risici skal forstås; herunder især "*beskyttelse*" mod disse. Implementering af beskyttelse kræver en vis forståelse for, at der først og fremmest er *noget, der skal beskyttes* og derudover *noget, der skal beskyttes imod*; henholdsvis aktiver og trusler. Denne systematik fremgår af ENISAs rapport fra 2016 ved navn *Big Data Threat Landscape and Good Practice Guide*[46], der giver et konkret overblik over mulige aktiver og trusler i forbindelse med big data. ENISAs rapport følger, kommenterer på og udspecificerer ISO 27001 – en anerkendt certificeringsordning – hvorved rapporten i højere grad anvendes, da den er mere specialiseret end standarden.

Ifølge forordningen er det simpelt at afgøre, hvad der skal beskyttes. Personoplysninger nævnes eksplicit i bestemmelsen, og artikel 32, stk. 1 uddyber, at der skal tages højde for behandlingens karakter, omfang, sammenhæng og formål. Dermed kan det fortolkes, at følsomme personoplysninger skal nyde større beskyttelse end almindelige. Men "*passende*" henleder ikke nødvendigvis til en vurdering af, hvilke oplysninger – der ud fra deres karakter – ikke behøver beskyttes tilstrækkeligt. Det kan på den anden side fortolkes som om, at følsomme oplysninger i deres natur er udsat for større risici af hensyn til deres "*alvor for personers rettigheder og frihedsrettigheder*" samt deres værdi for uautoriseret behandling, som er et nævnt eksempel i artikel 5, stk. 1, litra f og artikel 32, stk. 2. Der skal derfor mere effektive sikkerhedsforanstaltninger til for, at følsomme oplysninger er passende beskyttet. Dette giver ikke umiddelbart anledning til særlige overvejelser for udvikling og anvendelse af AI generelt. Der skal derimod som altid foretages en konkret afvejning.

[46] ENISA - Threats, 2016

4.6.5 Aktiver

ENISA anvender nedenstående figur 10 som oversigt over de aktiver, der gør sig gældende i forbindelse med big data, hvilke – som redegjort for tidligere – er sammenlignelige med omstændighederne for udvikling eller anvendelse af AI.



***Figur 10** - Illustration af big data aktiver.*
Kilde: ENISA - Threats, 2016, s. 12.

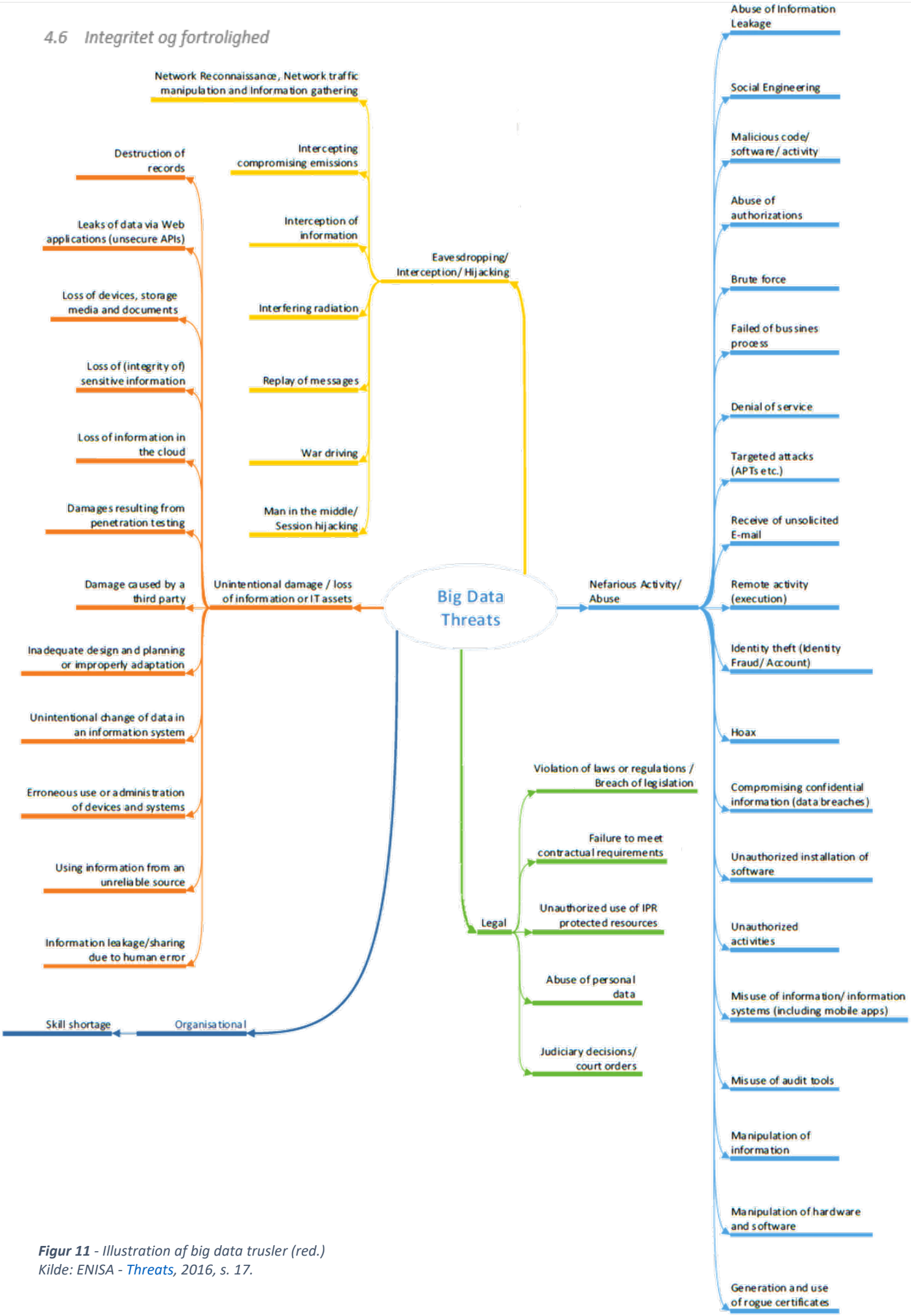
Disse udgør de aktiver, som dataansvarlige og databehandlere kan tage stilling til i forbindelse med beskyttelse. De kan anses som de strukturer, der – i en organisation, som arbejder med big data – er udsat for trusler. Det er disse, som den dataansvarlige eller databehandleren skal tage skridt til at beskytte. Bemærk, at sikkerhedsforanstaltninger selv fremgår som et strukturelt aktiv, der skal beskyttes. Det er et udtryk for, at hvert punkt udgør et led i en kæde, der ved dets brist kan kompromittere hele kæden.

Et eksempel på dette er, hvis vi forestiller os et rum, der indeholder personoplysninger. Vi identificerer personoplysningerne som et aktiv, der skal beskyttes, og derfor indsætter vi en dør indtil rummet. Døren er en sikkerhedsforanstaltning, men døren skal også beskyttes for at være effektiv. Dette genkender vi og putter derfor en lås på med tilhørende adgangspanel, der kræver en pinkode. Låsen er nu også en sikkerhedsforanstaltning, der skal beskyttes. Vi identificerer, at indførsel af nøglehåndtering medfører nye risici såsom lækage, og at anvendelse af ens pinkoder på tværs af døre vil medføre større trusselsmomenter, end da vi blot havde én dør, som var uaflåst. Vi skal derfor indføre et nøglehåndteringssystem (key management system), så kun de nødvendige personer har adgang til rummet. Lige nøjagtigt nøglehåndtering er også relevant i forbindelse med kryptering, som i forbindelse med AI har større relevans og oplever ligeartede risici som i eksemplet ovenfor.

4.6.6 Trusler

Artikel 5, stk. 1, litra f refererer til "*uautoriseret behandling*", hvilket både kan omfatte tilgang af tredjeparter eller generelt behandling, som ikke er omfattet af formålene, mens artikel 32, stk. 2 nævner "*uautoriseret videregivelse af eller adgang til personoplysninger*", som udelukkende refererer til tilgang af uvedkommende.

Derudover nævner artikel 32, stk. 2 også tilintetgørelse, tab, ændring af personoplysninger i forskellige situationer. Disse trusler er notorisk diffuse og vage af årsagerne tidligere angivet, og det vil være uhensigtsmæssigt at forsøge at fortolke ud fra disse krav, hvilke mulige trusler, der kan forekomme. Det er udvikleren og anvenderen af AI, der står med ansvaret for at gennemskue, hvilke mulige trusler, der udgør risici blandt andet ved at identificere logiske faldgruber eller manglende know-how. På næste side kan der derfor ses ENISAs liste i figur 11 over mulige trusler i forbindelse med big data til inspiration.



Figur 11 *- Illustration af big data trusler (red.)*
Kilde: ENISA - Threats, 2016, s. 17.

Figuren 11 viser en illustrativ opdeling af de nærværende trusler, som big data og AI risikerer at blive udsat for. ENISAs rapport redegør fyldestgørende for de forskellige aktiver og trusler samt kommer med råd til håndtering af dem, og der henvises derfor til deres rapport, da disse emner er tekniske i deres natur. Ud fra et juridisk perspektiv er det relevante at tage derfra, at der findes forskellige trusler, og at disse skal identificeres og håndteres i forhold til den konkrete situation.

Nogle konkrete trusler i forbindelse med AI kan dog være reverse engineering (omvendte konstruktioner), adversarial injections (fjendtlige injektioner) og backdoors (bagdøre). Se Bilag 2 – Trusler mod AI for en uddybelse af disse.

De tekniske udførelser af foranstaltninger til at imødekomme disse problemstillinger bør håndteres af specialister ud fra de krav og vurderingsmomenter, der tidligere blev redegjort for i forordningens bestemmelser. GDPR nævner dog nogle yderligere foranstaltninger, som kan overvejes.

4.6.7 Foranstaltninger

Artikel 32, stk. 1, litra a til d opremser eksempler på, hvilke foranstaltninger, der kan gennemføres. Kun litra a er konkret, mens resten er mere abstrakte. Ved litra b, c og d kan der derfor med fordel inddrages ENISAs rapport for inspiration til udførsel af disse foranstaltninger. De vil dog blive gennemgået her efterfølgende for en forklaring på de mål, som de konkrete foranstaltninger skal kunne opnå. Litra a til d skal i dén henseende anses som eksempler på foranstaltninger; inddrages efter deres relevans; og ikke forstås som udtømmende.

Litra a nævner pseudonymisering og kryptering af personoplysninger, hvilke er foranstaltninger, der tidligere i denne afhandling er blevet nævnt adskillige gange. Pseudonymisering er en forholdsvist simpel og billig foranstaltning, som i forbindelse med AI ikke kan anbefales nok. En AI har ikke de samme kognitive udfordringer som mennesker med behov for kontekstualisering og kompartmentalisering.[47] AI opfanger kun betydning af de parametre, som de er programmeret til, og det vil derfor være ligegyldigt for en AI, om et datasæt eksempelvis indeholder nøjagtige adresser eller blot en række vilkårlige tal. Så længe tallene er konsekvente, vil den finde de samme tendenser, som hvis hele adressen var tilgængelig. I dén forbindelse kan der findes en synergi imellem kryptering og pseudonymisering, idet de egentlige personoplysninger sandsynligvis efter pseudonymisering ikke er nødvendige for andre at tilgå end algoritmen selv efterfølgende.

Kryptografi fremgår af ISO 27001 som en foranstaltning imod lækage af informationer, men bemærk, at kryptografi kan være vanskeligt at udføre i praksis i forbindelse med AI og big data.[48] Kryptografi kan tilføje endnu et lag af kompleksitet, hvorved håndtering af de store mængder data kan hindre processeringsydeevne. Derudover skal det ledsages af tilpas nøglehåndtering. Ideelt er der få personer, der ligger inde med nøglerne til dataene. Hvor AI'en samtidig skal have adgang til de informationer, den behøver, skal det overvejes, hvordan dette udføres i praksis.[49] [50]

[47] Alpay - ncbi.nlm.nih.gov
[48] International Organization for Standardization, 2013
[49] ENISA - Threats, 2016, s. 33-34
[50] ENISA - Privacy, 2015, s. 38-39

Litra b nævner: ”*evne til at sikre vedvarende fortrolighed, integritet, tilgængelighed og robusthed af behandlingssystemer og -tjenester*”. Som beskrevet tidligere under afsnit 4.6.2 om det aktuelle tekniske niveau handler sikkerhed også om integritet og altså ikke kun fortrolighed. Sikkerhedsforanstaltningerne skal også kunne forebygge utilsigtede konsekvenser for de behandlede personoplysninger uanset, om disse forårsages af eksempelvis en tredjepart eller på baggrund af en såkaldt *bug*. Disse typer af risici kan ses i figur 11 på side 42 i det orange træ. Det interessante er her, at robusthed må referere til, at den tekniske kvalitet af kildekoden altså er et implicit krav inden for persondataretsområdet. Dermed kan det medføre sanktioner, hvis kildekoden vurderes at indeholde sikkerhedsrisici.

Litra b nævner også ”*vedvarende*”. Derved bør midlertidige sikkerhedsrisici forebygges på kontinuerlig basis. Eksempelvis ved cyberangreb er det ikke blot krævet, at der implementeres DoS-beskyttelse (beskyttelse mod denial of service-angreb) efterfølgende det første angreb, men at det er imødekommet på forhånd – dog med hensyn til de tidligere nævnte vurderingsmomenter.

Litra c omhandler et potentielt krav om sikkerhedskopiering af personoplysninger; også kaldet backup. ’Potentielt’ fordi der stadig skal tages forbehold for, hvorvidt en foranstaltning er relevant. Generelt vil det være vanskeligt at finde en fagperson, der ikke anbefaler, at alt skal sikkerhedskopieres, og konkrete konsekvens- og cost-benefit-analyser vil med al sandsynlighed i de fleste tilfælde finde, at de lave relative omkostninger forbundet med elektronisk opbevaring vil resultere i adskillige fordele ved at sikkerhedskopiere. Bemærk, at sikkerhedskopiering også er en behandling, hvorledes oplysningspligten gælder.

AI bør anses som software, hvorved de tekniske metoder til sikkerhedskopiering ikke adskiller sig væsentligt. Til gengæld er det karakteristisk ved AI, at de som tidligere nævnt i højere grad er udsat for risiko for korruption af algoritmen, da udviklerne ikke selv har fuld kontrol over dens udvikling.

Bugs er ikke til at undgå, og det er ikke usandsynligt, at en bug vil kunne resultere i, at en AI vil putte for meget eller for lidt vægt på et datasæt, der så resulterer i en model, der er unøjagtig, ineffektiv eller decideret diskriminerende, som var tilfældet i en sag i USA om anvendelse af AI til at udmåle straffe og kautionsbetingelser, der var partisk overfor afroamerikanere.[51] I dén forbindelse er det ikke blot nødvendigt at have foranstaltninger til at genoprette tidligere tilstand, men også at sikre validiteten og integriteten af modellen – se afsnit 5.3 om ansvarlighed.

Litra d nævner procedurer til regelmæssigt at afprøve, vurdere og evaluere foranstaltninger og deres effekter. Sikkerhedsforanstaltninger bør derfor testes, før det antages, at de er passende. En foranstaltning kan ikke afgøres at være passende, før det er blevet verificeret.

Litra d kan dog fortolkes på en måde, hvor den hænger sammen med litra c. Litra c bestemmer som nævnt, at det skal være muligt at genoprette en AI til en tidligere tilstand i tilfælde af korruption, og hvor dette kan have indflydelse på behandling af personoplysninger. Dette er dog en retroaktiv foranstaltning, hvor litra d kan anses for at være mere proaktiv. Hvis en behandling af personoplysninger kan påvirkes af en AI’s models integritet, bør der være forberedt virtuelle miljøer, der gør det muligt at afprøve indgreb i en model på forhånd.[52] Hvis der eksempelvis skal introduceres nye datasæt, der kan antages at ville påvirke modellen væsentligt, kan der opbygges et virtuelt miljø, hvor indgrebet simuleres i realtid på en version af softwaren, der ikke vil have indflydelse på den primære og oprindelige model.

[51] Angwin - propublica.org
[52] NVIDIA - nvidia.com

4.6.8 Nøglepunkter

Det vigtigste at tage fra Artikel 5, stk. 1, litra f og artikel 32 må være, at behandlingssikkerhed ikke kun relaterer sig til beskyttelse imod fjendtlige indgreb. Integritet og fortrolighed er to sider af den samme mønt, og integritet har i høj grad en betydning for udvikling og anvendelse af AI. Dataansvarlige og databehandlere skal først og fremmest identificere de aktiver, der kræver beskyttelse, og dernæst de trusler, som aktiverne kan udsættes for. På dén baggrund skal der vurderes passende foranstaltninger til at sikre personoplysninger fra utilsigtede konsekvenser, som er proportionelle med både truslerne samt omkostningerne. Disse bør inkludere anonymisering, pseudonymisering, aggregering og kryptografi efter deres effektivitet samt udvalgte og relevante tekniske foranstaltninger fra ENISAs rapport, Big Data Threat Landscape, såsom brugeradgangsadministration, netværkssikkerhedsstyring og regelmæssige dataintegritetskontroller.[53]

Sikkerhedskopier og anvendelse af virtuelle miljøer er åbenlyse foranstaltninger, der kan implementeres relativt billigt og med yderligere fordele foruden dem forbundet med beskyttelse af personoplysninger.

Til sidst skal det noteres, at den dataansvarlige som udgangspunkt har ansvaret for overholdelse af forpligtelserne beskrevet i dette afsnit. Integritet og fortrolighed medfører dermed udfordringer for udviklere og anvendere af AI; blot af mindre juridisk karakter end de øvrige principper.

[53] ENISA - Threats, 2016, s. 32-41

5 Øvrige bestemmelser

De grundlæggende principper er ikke de eneste bestemmelser, som kan tænkes at have særlig relevans for udvikling og anvendelse af AI. Oplysningspligten, det retlige grundlag med mere er blevet gennemgået løbende i de tidligere afsnit, men GDPR indeholder yderligere bestemmelser, der påvirker udvikling af AI i større grad, hvorved disse bearbejdes her efterfølgende.

5.1 Databeskyttelse gennem design og standardindstillinger

Kravet om databeskyttelse gennem design og standardindstillinger findes i GDPR's artikel 25. Stk. 1 omhandler databeskyttelse gennem design, mens stk. 2 omhandler databeskyttelse gennem standardindstillinger. Simpelt fortalt handler databeskyttelse gennem design om, at databeskyttelse skal overvejes løbende igennem udviklingen af produktet. Databeskyttelse gennem standardindstillinger betyder, at behandling af unødvendige oplysninger foregår på opt-in-basis og ikke opt-out. Det vil sige, at den registrerede bør have mulighed for aktivt at tilvælge, at deres personoplysninger bliver behandlet og ikke er nødsaget til selv at fravælge behandling af oplysninger, der er unødvendige til formålene. Den registrerede skal dermed have den største beskyttelse som udgangspunkt. Databeskyttelse gennem standardindstillinger lægger ikke op til overvejelser af særlig betydning ved udvikling eller anvendelse af AI, men databeskyttelse gennem design hænger sammen med de grundlæggende principper i en grad, hvor det er oplagt at gennemgå eventuelle udfordringer for implementering i AI.

5.1.2 Databeskyttelse gennem design

Artikel 25, stk. 1 og artikel 32, stk. 1 om behandlingssikkerhed er næsten ens, idet de begge omhandler implementering af organisatoriske og tekniske foranstaltninger samt kræver anvendelse af de tidligere angivne vurderingsmomenter. Se afsnit 4.6 om integritet og fortrolighed, side 37 til 39 for en gennemgang af vurderingsmomenterne samt side 43 til 44 om foranstaltninger. Artikel 25, stk. 1 omhandler dog ikke kun implementering af sikkerhedsforanstaltninger, men også foranstaltninger til opfyldelse af de øvrige principper, som er blevet gennemgået i denne afhandling og nævner dataminimering specifikt.

I dén forbindelse er det oplagt at slå en knude på integritet og fortrolighed, idet ét punkt i artikel 5, stk. 1, litra f og artikel 32, stk. 2 endnu ikke er afklaret og har speciel relevans ved databeskyttelse gennem design. Begge bestemmelser benævner beskyttelse imod ulovlig behandling i sammenhæng med sikkerhedsforanstaltninger. Hvad, dette betyder, er, at der ved implementering af foranstaltningerne også skal vurderes risici for, at personoplysninger bliver behandlet ulovligt. Dette vedrører ikke blot behandling uforeneligt med formålene, men også behandling uden retligt grundlag. Foranstaltninger mod ulovlig behandling er dog ikke kun nødvendige i forbindelse med sikkerhed. Artikel 25, stk. 1 klargør, at den dataansvarlige ved produktion af sine behandlingsmekanismer – såsom en AI – på forhånd har inkorporeret systemer, der afholder behandlere fra at foretage andre behandlinger end dém, der er tilsigtet, herunder ulovlig behandling.

Ved en bankrådgivers gennemgang af en registrerets kreditværdighed – for bare at anvende et ældre eksempel – kan det være udfordrende at skabe et IT-system, der kender til alle mulige typer af behandling, som rådgiveren er berettiget til at foretage for således at have afskåret alle de uberettigede behandlingsmuligheder. Dette er dog ikke nødvendigvis en lige så stor byrde ved udvikling af AI, idet maskinteknologi i forvejen påkræver specifikke angivelser af processeringer. En AI kan kun processere information i overensstemmelse med dens algoritmes funktioner. Derfor er det særligt nødvendigt ved udvikling af en AI, at databeskyttelse inddrages fra starten, da ændringer af en algoritme efterfølgende kan påvirke resten af dens funktioner og være omkostningstungt at udføre.[54]

Både ENISAs rapport (Privacy by design in big data), den engelske informationskommisærs vejledning (Big data, artificial intelligence, machine learning and data protection) samt det norske datatilsyns vejledning om databeskyttelse gennem design påstår, at udførsel af databeskyttelse gennem design vil udgøre en konkurrencemæssig fordel.[55] [56] [57] Belægget er, at det er et markedsføringsmæssigt karakteristikum, der skaber goodwill. ENISA går så langt som at sige, at markedet behøver et paradigmeskift fra "*big data versus privacy*" til "*big data with privacy*", da det er nødvendigt for, at big data som industri kan blomstre.[58] Hvorvidt disse argumenter er effektive, må tiden vise, men et yderligere argument kunne være, at mangel på implementering af databeskyttelse gennem design er en kostelig fejl at begå selv uden at medregne eventuelle sanktioner. Anvendelsen af en model, der er modulær, gør efterfølgende implementering af foranstaltning mere overkommeligt. Databeskyttelse gennem design er dog nu et lovmæssigt krav, og den mest hensigtsmæssige måde at overholde reglerne er nu engang at bygge modellen op omkring dem – uanset hvor modulær den er.

[54] ENISA - Privacy, 2015, s. 5
[55] ICO, 2017, s. 73-74, punkt 166
[56] ENISA - Privacy, 2015, s. 5
[57] Datatilsynet.no – guide - datatilsynet.no
[58] ENISA - Privacy, 2015, s. 49

5.2 Automatiske afgørelser

Automatiske afgørelser og profilering, som er en form for automatisk og personlig evaluering, benyttes i stigende grad i både de private og offentlige sektorer, da værktøjerne kan skabe større effektivitet og ressourcebesparelser. Automatiske afgørelser og profilering kan både være inden for økonomi herunder lån, men også markedsføringsmæssigt. Udviklingen inden for AI og big data har gjort det lettere at benytte sig af profilering og andre automatiske afgørelser.[59] Der vil derfor også være et stigende antal af eksempler, hvor det vil have en stor indvirkning på den registreredes rettigheder og frihedsrettigheder. GDPR adresserer automatiske afgørelser og profilering for netop at beskytte disse.

Automatiske afgørelser er individuelle afgørelser baseret udelukkende på automatisk behandling uden nogen menneskelig indgriben. Denne definition nævnes ikke i artikel 4 om definitoner. Definitionen fremgår dog af præamblens punkt 71, hvorved den ikke er bindende. Dette er dog et udtryk for, hvordan EU-parlamentet mener, at begrebet skal forstås. Denne definition understøttes også af Artikel 29-gruppen, der har udtalt det samme i deres vejledning om "*Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*".[60]

Begrebet, "*profilering*" til forskel for automatisk afgørelser bliver defineret i artikel 4, stk. 4 som "*enhver form for automatisk behandling af personoplysninger, der består i at anvende personoplysninger til at evaluere bestemte personlige forhold vedrørende en fysisk person, navnlig for at analysere eller forudsige forhold vedrørende den fysiske persons arbejdsindsats, økonomiske situation, helbred, personlige præferencer, interesser, pålidelighed, adfærd, geografisk position eller bevægelser*".

Ifølge ICO bliver profilering brugt af virksomheder til at klassificere personerne ind i en række forskellige grupper eller segmenter. Analysen af den data, der indsamles, benyttes til at identificere forbindelser mellem adfærd og karakteristika for at lave individuelle profiler. Disse bliver så brugt til blandt andet at finde præferencer, forudsige adfærd og lave individuelle afgørelser på personerne.[61] Profilering er derfor yderst relevant inden for både markedsføring og den finansielle sektor, men også uddannelse og andre områder.

GDPR's artikel 22, stk. 1 opstiller dog en ret for den registrerede til ikke at være genstand for afgørelser, der kun er baseret på en automatisk afgørelse og/eller profilering, når det har en retsvirkning eller i det hele taget på samme vis påvirker den registrerede. Stk. 2 angiver dog en række undtagelser til denne ret. Behandlingsgrundlaget skal bygge på en nødvendighed for at kunne indgå eller opfylde en kontrakt mellem den registrerede og den dataansvarlige; være hjemlet i EU-ret eller medlemsstaternes nationale ret; eller være baseret på et udtrykkeligt samtykke fra den registrerede.

Disse undtagelser gælder kun for almindelige personoplysninger efter artikel 22, stk. 4, hvorved undtagelserne til, hvornår den dataansvarlige har ret til at foretage automatiske afgørelser imod den registreredes ønske, ikke gælder, hvis der behandles følsomme personoplysninger. Der fremgår dog også to undtagelser til dette. Såfremt en behandling er baseret på artikel 9, stk. 2, litra a eller g, vil den registrerede ikke kunne modsige sig en automatisk afgørelse. Det vil sige, der enten skal foreligge et udtrykkeligt samtykke fra den registrerede eller en nødvendighed af hensyn til væsentlige samfundsinteresser.

[59] Artikel 29-gruppen – automated, 2018, s. 5
[60] Artikel 29-gruppen – automated, 2018, s. 8
[61] ICO – guide - ico.org.uk

En automatisk afgørelse eller profilering kan være indgribende for individet. Derfor pådrager artikel 35, stk. 3, litra a også en dataansvarlig at skulle fortage en konsekvensanalyse for at identificere og vurdere risiciene med behandlingen og håndtering af risiciene. Yderligere skal der efter artikel 22 foretages passende foranstaltninger til beskyttelse af den registreredes rettigheder og frihedsrettigheder samt legitime interesser. Præamblens punkt 71 anviser også, at der skal foretages tiltag for at sikre den registrerede. Tiltagene skal være passende matematiske eller statistiske fremgangsmåder, og yderligere skal der gennemføres tekniske og organisatoriske foranstaltninger. Dette hænger også sammen med principperne om rimelighed og rigtighed, da tiltagene skal sikre, at der blandt andet ikke sker en forskelsbehandling af personer, og faktorer, der resulterer i unøjagtige personoplysninger, bliver rettet.

Ydermere skal de passende foranstaltninger, der nævnes i artikel 22, stk. 3, gøre det muligt for den registrerede at "*fremkomme med sine synspunkter og til at bestride afgørelsen*" og retten til menneskelig indgriben. Den registrerede skal derfor have ret til at komme med sine argumenter mod afgørelsen og så bestride den. Dette bekræftes yderligere i præamblens punkt 71. Punktet anfører, at den registrerede skal have en specifik underretning og retten til menneskelig indgriben til at fremkomme med sine synspunkter og til at bestride afgørelsen. Den tyder derfor på, at den registrerede skal have oplysninger, så det er muligt i sidste ende at bestride afgørelsen om nødvendigt.

GDPR's artikel 15, stk. 1, litra h fremfører, at den registrerede har ret til at få information om forekomsten af den automatiske afgørelse og profileringen. Derudover også meningsfulde oplysninger om logikken i afgørelsen samt betydningen og de forventede konsekvenser af sådan en behandling for personen. Når der står "*logikken*", er det ret vagt, hvad der menes her. Præamblens punkt 71 anviser, at der skal gives en forklaring på afgørelsen. EU-parlamentet har derfor formentlig ment, at logikken skal anses som en forklaring med de pågældende belæg for afgørelsen.

Sandra Wachter, forsker ved Oxford Internet Institute i Storbritannien, har i en juridisk analyse kritiseret netop dette i GDPR. Hun mener ikke, at der skal gives en forklaring, da det kun står i præamblen ud fra hendes synspunkt.[62] På den modsatte side står Julia Powles, forsker ved NYU Law, der i en artikel om dette punkt forklarer, at "*logikken*" skal anses som, at der skal angives en forklaring.[63] Der er meget usikkerhed på området, men der virker til at være en udbredt konsensus fra artikler på området og ICO om, at der skal angives en forklaring ved automatiske afgørelser. Principperne om rimelighed og gennemsigtighed tyder yderligere på, at der skal angives en forklaring, da det på den måde gøres muligt for den registrerede at sikre, at en afgørelse er rimelig.

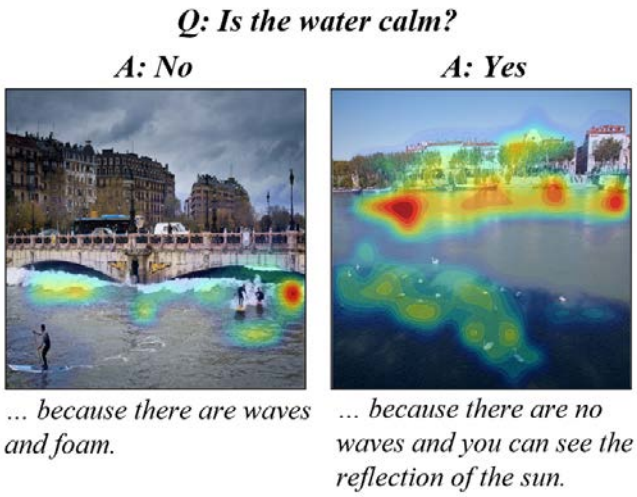
Datatilsynet udgav i marts 2018 en vejledning om de registreredes rettigheder. Heri tager de stilling til, hvad "*logikken*" egentlig skal sige. De mener, at der skal oplyses om, "*hvordan 'systemet' kommer frem til afgørelserne*".[64] Det vil sige, hvilke data systemet har lagt vægt på ved afgørelsen.

[62] Wachter - papers.ssrn.com
[63] Powles - academic.oup.com
[64] Datatilsynet – vejledning, 2018, s. 28

Såfremt en forklaring er krævet ud fra en evaluering ved automatiske afgørelser og profilering, kan det blive problematisk for udviklere og anvendere af AI. Problematikken med den sorte boks medfører, at en udvikler eller anvender ikke nødvendigvis kan vurdere og afgøre, hvad der ligger til grund for en afgørelse. Dette problem er størst ved online AI, der anvender deep learning-teknikker, idet deres modeller kan være så komplicerede, at de er svære at udrede. De kan bestå af utallige mikrobeslutninger. At finde frem til præcis de beslutninger, som AI'en har lagt vægt på, er vanskelig og sågar umuligt i mange tilfælde. Hvis der findes en løsning til problematikken om den sorte boks, der kan medføre mere gennemsigtighed, vil det dermed blive overkommeligt at leve op til dette mulige krav om retten til en forklaring.

Artikel 29-gruppen udtrykker selv, at udviklingen og kompleksiteten af machine learning kan gøre det vanskeligt at forstå, hvordan en automatisk afgørelse eller profilering virker.[65] De mener, at den dataansvarlige bør finde simplere måder at informere den registrerede om rationalet bag, eller kriterierne, der er lagt vægt på ved en afgørelse, uden nødvendigvis altid at forsøge at forstå den komplekse algoritme, der ligger bag afgørelsen; altså den sorte boks. Det vil sige, at er det ikke muligt at forklare præcist, hvordan AI'en er nået frem til sin afgørelse, så skal det i stedet være muligt at tilkendegive, hvilke informationer den har lagt vægt på. Dette gør problemstillingen betydeligt mere overkommelig. Især taget i betragtning, at gruppen yderligere bemærker, at en algoritmes kompleksitet ikke er en undskyldning for utilstrækkelige forklaringer grundet de strenge krav til gennemsigtighed. Der findes tekniske softwareløsninger, der med fordel kan inkorporeres i AI af forskellige typer, der gør det muligt for dem på automatisk vis at fremlægge de nøjagtige oplysninger, som de anvender i deres konkrete afgørelser. Disse vil blive gennemgået i afsnit 6.

Et eksempel på en sådan softwareløsning findes i en rapport fra 2018 af en gruppe forskere fra EECS, UC Berkeley, University of Amsterdam, MPI for Informatics og Facebook AI Research om "Multimodal Explanation: Justifying Decisions and Pointing to the Evidence".[66] De har udviklet en AI til at svare på spørgsmål omkring billeder. AI'en forklarer i tekst, hvad den har lagt vægt på, men den kan også vise det på selve billederne, som kan ses på figur 12 nedenfor.



Figur 12 *- Illustration af AI'ens forklaring.*
Kilde: Dong Huk Park, et al., 2018, arxiv.org/pdf/1802.08129.pdf

Dette eksempel er kun relevant i det tilfælde, hvor en AI udfører billedgenkendelse. AI, der kan forklare sig selv, er dog under udvikling. Mangel på gennemsigtighed og tillid er hele problematikken bag den sorte boks, og dette vil løse problemet til en vis grad, så AI'en kan forklare sig selv. Dette vil skabe mere gennemsigtighed og tillid til AI og deres afgørelser.

[65] Artikel 29-gruppen – automated, 2018, s. 14
[66] Huk Park - arxiv.org

5.3 Ansvarlighedsprincippet

Princippet om ansvarlighed er et nyt princip indført med den nye forordning. I GDPR's artikel 5, stk. 2 anfører, at "*den dataansvarlige er ansvarlig for og skal kunne påvise, at stk. 1 overholdes*". Bestemmelsen kræver dermed, at den dataansvarlige kan dokumentere, at virksomheden har handlet ansvarligt i forhold til de grundlæggende principper i artikel 5, stk. 1, litra a - f. Virksomheden skal hermed dokumentere, hvordan den overholder principperne i GDPR.

Idet der står "*ansvarlig for* […] *at stk. 1 overholdes*" henledes der også til, at virksomheden ikke kun nu skal overholde GDPR og bevise det, men at de også løbende skal følge op på, at den overholdes, så der hele tiden vises ansvarlighed. Dette bekræftes yderligere i artikel 24 om den dataansvarliges ansvar. Artikel 24, stk. 1 anfører, at "*foranstaltninger skal om nødvendigt revideres og ajourføres*". Dette tyder endvidere på, at virksomheden hele tiden skal være opmærksom på, at GDPR overholdes og bevise det. Bestemmelsen giver desuden også anledning til, at der er tale om en afvejning ud fra vurderingsmomenterne: "*behandlingens karakter, omfang, sammenhæng og formål samt risiciene af varierende sandsynlighed for fysiske personer rettigheder og frihedsrettigheder*". Der skal derved tages stilling til blandt andet, hvad behandlingen består af overfor den risiko, som det kan have for den registrerede. Under hensyntagen til dette, skal virksomheden som dataansvarlig implementere passende tekniske og organisatoriske foranstaltninger for sikring og påviselighed af overensstemmelse med GDPR. Som tidligere angivet i afsnit 4.6 om integritet og fortrolighed skal omfanget af de passende tekniske og organisatoriske foranstaltninger ligeledes bestemmes ud fra vurderingsmomenterne anført ovenfor.

Artikel 24, stk. 2 bestemmer, at den dataansvarlige skal indføre passende databeskyttelsespolitikker i forbindelse med de tekniske og organisatoriske foranstaltninger. Dog kun såfremt, det står rimeligt i forhold til behandlingsaktiviteterne. Når virksomheden skal påvise, at principperne overholdes efter artikel 5, stk. 2 og 24, bør virksomheden derfor have udarbejdet og nedskrevet dokumentation såsom i forbindelse med fortegnelse efter artikel 30. Hvis det ikke kan påvises, så har virksomheden ligeledes ikke udvist ansvarlighed og dermed ikke overholdt GDPR.

Yderligere skal en dataansvarlig og databehandler ansætte en databeskyttelsesrådgiver (DPO), såfremt der er et sådant behov efter artikel 37. Stk. 1 litra a til c afklarer, hvilke situationer en DPO skal udnævnes. En virksomhed, hvor kerneaktiviteten består af behandlingsaktiviteter, der kræver regelmæssig og systematisk overvågning af registrerede i et betydeligt omfang, vil blandt andet være pålagt sådan en forpligtelse. En DPO vil også generelt hjælpe virksomheden med at overholde ansvarlighedsprincippet og GDPR generelt. En DPO bør derfor altid som udgangspunkt ansættes, når en virksomhed arbejder med big data, der omfatter behandling af personoplysninger. En virksomhed, der arbejder med AI, der behandler personoplysninger, bør dermed også overveje at ansætte en DPO grundet lighederne mellem big data og AI.

Som tidligere angivet, fastlægger artikel 5, stk. 2, at den dataansvarlige skal påvise at overholde de grundlæggende principper i stk. 1 af samme artikel. Når der er tale om en AI, så bliver dette lige pludseligt en hel del svære, da AI'en er en proxy, og der kan være tale om en sort boks. Det er en helt anden disciplin at implementere foranstaltninger til at opfylde kravet om ansvarlighed i en automatiseret proces som en AI, end det vil være at implementere faste interne procedurer i enhver anden virksomhedssituation. I forbindelse med automatiseret software er det ikke muligt blot at spørge en eventuel driftsleder, hvilke behandlinger den pågældende afdeling foretager sig. Informationer er ikke nødvendigvis forståelige for lægmænd eller overhovedet tilgængelige, når informationerne befinder i en sort boks, hvilket bringer os tilbage til nogle af ulemperne ved at anvende kryptering. Der findes dog tekniske metoder, der kan assistere med at automatisere præsentationer af de oplysninger, som en AI behandler samt, hvad den anvender dem til og afgør på baggrund af dem – se eventuelt næste afsnit om teknologiske løsninger. Disse metoder kan være nødvendige at inkorporere allerede fra påbegyndelsen af udviklingen af AI'en, medmindre programmet er tilpas modulært, se afsnit 3.1.1.

Det kan yderligere være nødvendigt med sådanne foranstaltninger, idet princippet om formålsspecifikation især vanskeliggør kravene om påviselighed. Big data er ofte eksperimentelt uden et forudsat formål med indsamlingen.[67] Ved ikke at kunne bestemme et formål vil en virksomhed for det første ikke leve op til princippet i artikel 5, stk. 1, litra b, men ej heller princippet om ansvarlighed i artikel 5, stk. 2. Disse udfordringer med principperne i stk. 1 over for AI, gør det svært for en virksomhed at udvise og påvise ansvarlighed; især hvis der er tale om online AI, hvor løbende kontrol med dens behandlinger kan være uhensigtsmæssigt krævende, hvis overhovedet muligt.

Brugen af algoritmer i afgørelser har også vokset en hel del på grund af den store interesse i AI. Dette er refereret til som "algorithmic accountability"; altså algoritmisk ansvarlighed. Det omhandler at være i stand til at kunne tjekke, om algoritmerne, der bruges og udvikles af AI, gør, hvad vi gerne vil have dem til. De skal derfor ikke i stedet lave diskriminerende, fejlagtige eller ubegrundede afgørelser.[68] Dette hænger også sammen med princippet i artikel 5, stk. 1, litra a om rimelighed. Som det nævnes i det afsnit, anfører præamblens punkt 71, at der ikke må ske en forskelsbehandling af fysiske personer ved automatiske afgørelser. World Wide Web Foundation har udgivet en rapport på området i 2017, der forklarer både udfordringerne og løsningerne ud fra det datalogiske og etiske perspektiv i modsætning til dette juridiske.[69]

Det kan derfor være en nødvendighed at overveje at implementere konkrete tekniske softwareløsninger i AI for at overholde GDPR's bestemmelser om påviselighed og generel algoritmisk ansvarlighed. En online AI – som anvender deep learning-teknikker samt løbende og autonomt leder efter sammenhænge, den kan danne nye lag af parametre ud fra – vil ikke kunne leve op til kravene om formålsspecifikation, dataminimering eller påviselighed, hvis ikke den er i stand til at udføre automatiske tilkendegivelser af dens afgørelser, belæg hermed, behandlingsgrundlag og behandlede personoplysninger. De næste afsnit vil derfor beskæftige sig med de aktuelt anvendelige tekniske foranstaltninger, som kan implementeres af udviklere og anvendere af AI.

[67] ICO, 2017, s. 51
[68] ICO, 2017, s. 52
[69] World Wide Web Foundation - webfoundation.org

6 Teknologiske løsninger

De tidligere afsnit har haft til formål at undersøge og beskrive, hvilke konkrete problemstillinger og udfordringer, der kan opstå i forbindelse med implementering af GDPR på udvikling og anvendelse af AI. Oplagte løsninger dertil er løbende blevet præsenteret, hvor disse har været enten åbenlyse eller simple at udføre. GDPR tager dog ikke konkret stilling til udvikling og anvendelse af AI, selvom vi drastisk nærmer os AI's udbredelse. Denne afhandling har derfor nået frem til problemstillinger, hvis løsninger enten ikke umiddelbart er praktisk mulige eller entydige, herunder i forbindelse med formålsspecifikation, dataminimering, ansvarlighed og profilering.

Dette afsnit tager udgangspunkt i det norske datatilsyns rapport, hvor de i samarbejde med adskillige specialister på området har opremset aktuelle teknologiske metoder, som kan implementeres i AI.[70] Der bliver dog her udvalgt nogle enkelte, som har særlig relevans i forhold til de problemstillinger og udfordringer, som denne afhandling har afdækket. Specielt online-modeller, der anvender deep learning-teknikker har vist sig at medføre komplikationer. Det vil ikke altid være muligt at afgøre, hvilke formål en AI har med at indsamle personoplysninger, fordi det kun efter den statistiske analyse kan afgøres, hvilke informationer der var nødvendige. Det vil derudover være kompliceret at fremlægge, hvilke momenter der har været afgørende for en AI for at træffe en automatisk afgørelse, hvis denne indeholder en sort boks. Teknologierne i dette afsnit kan være til hjælp til at imødekomme netop disse problemstillinger og udfordringer, men hvis der er blot én anbefaling, som skal tages med fra denne afhandling, så er det, at den mest simple foranstaltning er at fjerne de identificerende egenskaber ved personoplysningerne, således der ikke behandles personoplysninger. Visse af de efterfølgende teknologier har også dette til formål.

Nedenfor vil der derfor blive gennemgået konkrete teknologiske foranstaltninger, værktøjer og metoder, som kan hjælpe dataansvarlige med at sikre, at deres AI kan leve op til GDPR.

6.1 Federated learning

Federated learning er en teknik, der sørger for at adskille personoplysninger mellem brugere, således de ikke bliver delt på kryds og tværs. Google anvender i øjeblikket denne teknik i forbindelse med udvikling af deres tastaturer på smartphones. En brugers adfærd og inputs bliver analyseret og kategoriseret i brugerens egen telefon, hvorefter de bliver uploadet og aggregeret med andre brugeres data. Dataenes konsensus findes og downloades til hver enhed i aggregeret tilstand. Således kommer hver brugers personoplysninger ikke i kontakt med andres, før de bliver anvendt til at udvikle modellen. Denne metode er overfladisk blevet berørt tidligere i denne afhandling i forbindelse med aggregering.[71]

Hertil kan også nævnes en metode, som vi kan låne fra vores norske kolleger hos Statistisk Sentralbyrå og Norsk Senter for Forskningsdata. De har opfundet en metode ved navn RAIRD, der går ud på, at deres data kan anvendes til forskning – dog i aggregeret tilstand. Så forskere får kun adgang til data, hvor tendenserne allerede er udvundet. Denne metode er hensigtsmæssig at bruge ved enhver overførsel af data.[72]

[70] Datatilsynet.no – kunstig intelligens, 2018, s. 25-26
[71] McMahan - ai.googleblog.com
[72] SSD - raird.no

6.2 Differential privacy

Denne metode er især interessant, idet det er en form for kryptografi, der ikke er reversibel. Det er tidligere blevet gennemgået i afsnit 4.6 om integritet og fortrolighed, hvordan kryptering, pseudonymisering og anonymisering ikke altid vil medføre uidentificerbarhed. Denne krypteringsmetode indebærer dog en komplet praktisk destruktion af dataene, hvorved de ikke er identificerbare.[73]

Differential privacy går ud på, at der introduceres tilfældighed til dataene i form af støj. Metoden foregår ved, at visse af de indsamlede personoplysninger forfalskes. Opfinderne af metoden, Cynthia Dwork fra Microsoft og Aaron Roth fra University of Pennsylvania, illustrerer dette med at slå plat eller krone.[74] Hvis plat, så forbliver oplysningen korrekt, således John Hansen stadig er fra Aalborg. Hvis krone, kastes mønten igen. Hvis den nu viser plat, er John Hansen fra Aalborg, men hvis krone igen anden gang, vil datasættet ændres, således han er fra København i stedet. I 25 % af tilfældene vil oplysningen altså være falsk. I så fald vil oplysningen stadig være identificerende i 75 % af tilfældene, men dette er kun såfremt det vides, hvad mønten afgjorde. Da det ikke vides, om der blev slået plat eller krone, er der introduceret tilfældighed, og det er dermed ikke muligt at afgøre ud fra oplysningen, om den er korrekt. Oplysningen er dermed ikke identificerende. Umiddelbart virker dette skadeligt for selve behandlingen, men med store datasæt vil det blot være 25 % af svarene, der er forkerte, og der må da stadig foreligge statistisk signifikans ved den korrekte sammenhæng. Denne metode vil selvfølgelig kun være hensigtsmæssig i scenarier, hvor de konkrete oplysningers korrekthed er underordnet som gennemgået i afsnit 4.4 om rigtighed; hvilket vil sige big data og dermed ikke ved automatiske afgørelser for eksempel.

6.3 Homomorphic encryption

Som tidligere angivet er anvendelse kryptering problematiseret ved, at det kan kræve enorm processeringskraft at anvende ved big data samt, at tilstedeværelsen af en nøgle til at give adgang til oplysninger vil medføre, at de stadig er potentielt identificerbare. Homomorphic encryption (homomorfisk kryptering) er en løsning til den anden problemstilling, idet metoden medfører, at modellen kan behandle oplysninger, mens de stadig er krypterede, hvorved nøglen ikke behøves at være tilgængelig. Metoden vil dog medføre behov for større processeringskraft, hvorved den førstnævnte problemstilling stadig ikke imødekommes. Teknologien er dog stadig på udviklingsbasis, og det er muligt, at den med tiden er mere overkommelig. For nu er den uhensigtsmæssig i forbindelse med big data, men lovende ved automatiserede afgørelser.[75]

[73] Datatilsynet – anonymisering - datatilsynet.dk
[74] Dwork - cis.upenn.edu
[75] Stuntz - community.embarcadero.com

6.4 Transfer learning

Som gennemgået i afsnit 3.1 om machine learning, er machine learning kendetegnet ved, at modellen oplæres igennem træningsdata. Disse træningsdata kan bestå af personoplysninger. En simpel måde at håndtere denne problemstilling på er ved at genanvende en tidligere model, der allerede er oplært, men som lever op til de samme behov som dén model, man ønsker. Dette er transfer learning – en metode, hvor en anden model kan implementeres i den pågældende for at undgå behovet for træningsdata.[76]

6.5 Explainable AI

Explainable AI (forklarlig AI eller XAI) er i mindre grad en metode, end det er en disciplin, som stadig forskes i. Der bliver i høj grad brugt ressourcer på at finde standardiserede løsninger til at gøre AI i stand til at forklare sig selv.[77] Som gennemgået i afsnit 5.2 om automatiske afgørelser og 5.3 om ansvarlighed findes der allerede teknologier, der gør det muligt at påvise overholdelse af GDPR's krav igennem forklaringer på afgørelser eller konklusioner, som AI'en træffer. Eksempelvis findes der eksemplet i figur 12 på side 50, som kan bruges ved billedgenkendelse, men dette er et forholdsvist snævert område.

Til gengæld findes den anerkendte teknologiske løsning, LIME.[78] LIME står for Local Interpretable Model-Agnostic Explanations og er selvsagt modelagnostisk ved ikke at være bundet til en bestemt type af model såsom billedgenkendelse. Teknologien er opfundet af Ribeiro, Singh og Guestrin fra University of Washington og kan efter sigende levere udførlige forklaringer og præsentationer af de parametre, som modellen har lagt vægt på til at træffe afgørelser eller foretage forudsigelser. Løsningen kan dermed potentielt afhjælpe ikke bare problemstillingen med påviselighedskravet i forbindelse med automatiske afgørelser og profileringer, men den kan også vise, hvilke oplysninger, der faktisk er blevet anvendt til at finde sammenhænge med. Teknologien kan derfor potentielt også løse problemstillingen om at afgøre, hvilke oplysninger er nødvendige i forbindelse med princippet om dataminimering.[79]

LIME vil selvfølgelig ikke kunne anvendes i alle scenarier, og i nogle tilfælde vil den konkrete AI behandle et område, hvor det er mest hensigtsmæssigt at implementere en specialiseret metode til at skabe transparens i stedet. Men det interessante er, at LIME først blev præsenteret i 2016. Området er stadig under enorm udvikling, og vi kan kun gætte på, hvilke nye teknologier der vil opstå bare indenfor den nærmeste fremtid.

Hvad, vi dog har løsninger i øjeblikket, virker lovende. Differential privacy og homomorphic encryption er synergiske foranstaltninger, som hver især kan anvendes ved henholdsvis big data og behandling af mindre datasæt. Disse kan selvfølgelig ikke anvendes i alle unikke tilfælde, men i et stort omfang bør det ikke være nødvendigt at behandle personoplysninger grundet disse foranstaltninger. Hvor det alligevel er nødvendigt, kan der med fordel anvendes en af de allerede opfundne løsninger til at gøre en AI's behandling præsenterbar og gennemsigtig.

[76] Machine learning research group - cs.utexas.edu
[77] Datatilsynet.no – kunstig intelligens, 2018, s. 26
[78] Ribeiro - homes.cs.washington.edu
[79] Singh - kdd.org

7 Konklusion

Ved at gennemgå GDPR's bestemmelse mere eller mindre slavisk, er der blevet undersøgt og bearbejdet de problemstillinger og udfordringer, som gør sig gældende ud fra forordningen ved udvikling og anvendelse af AI. Det bør nævnes, at overholdelse af lovgivningen generelt er en udfordring, og at udførsel deraf – især på baggrund af denne afhandlings fund – bør udøves i samspil med en specialiseret jurist, idet de konkrete vurderinger i praksis ikke kan undgå at være unikke. Denne undersøgelse tjener dermed som en tjekliste over de udfordringer, som ledelse, jurister og ingeniører i fællesskab bør tage stilling til.

Nærværende afhandling har udredt, at databeskyttelse i relation til AI skal deles op i tre forskellige discipliner, som udfordringer kan kategoriseres efter. Der er klassifikationen – herunder af oplysninger og den pågældende AI –, selve beskyttelsen af data og selvfølgelig den egentlige behandling af personoplysninger.

7.1 Klassifikation

7.1.1 Klassifikation af oplysninger

Den mest betydningsfulde vurdering at foretage i forbindelse med anvendelse af personoplysninger til AI er, hvorvidt det kan undgås at anvende personoplysninger. Som det har vist sig, er GDPR særdeles streng, og udviklere og anvendere af AI kan spare sig selv besværet ved blot at finde alternativer til anvendelse af AI – eksempelvis igennem transfer learning.

Få vil dog være tilfredse med dette svar, og det er derfor mere oplagt at spørge, hvorvidt det er muligt at forvandle de pågældende personoplysninger til uidentificerbare oplysninger. Pseudonymisering, aggregering, kryptering og anonymisering er påkrævede at udføre, hvor det er muligt. Selvom pseudonymiserede og krypterede oplysninger stadig er personoplysninger, findes der løsninger såsom differential privacy, homomorphic encryption og federated learning, der alligevel kan fjerne de identificerende egenskaber og effektivisere foranstaltningerne. Det er sjældne tilfælde, hvor ingen af de nævnte teknologier kan anvendes. Da AI typisk er statistiske værktøjer og kan kategoriseres som statistiske behandlinger, bør aggregering af personoplysninger med fordel ske så tidligt i processen som muligt og på automatisk vis for at undgå at behandle personoplysninger. Dette vil medføre både juridiske og tekniske fordele.

7.1.2 Klassifikation af AI

Hvor anvendelse af personoplysninger til AI kan kategoriseres som et statistisk formål eller videnskabelig forskning, er der videre muligheder for at viderebehandle personoplysninger. Visse typer af AI vil kunne kategoriseres som forskningsrelaterede eller statistiske af natur, og de vil derfor kunne anvende personoplysninger til et andet formål end dét, som de er blevet indsamlet til. Forskningsrelaterede AI kan desuden anvende personoplysninger til et mindre specifikt formål, end hvad der ellers kræves.

Det er yderligere betydningsfuldt, hvilken type af AI, der anvendes. Online modeller, deep learning, unsupervised learning og især disse i samspil medfører større kompleksitet, hvilket forårsager udfordringer, der ikke nødvendigvis gør sig gældende for offline modeller, almen machine learning eller supervised learning. Dette er specielt tilfældet ved opfyldelse af kravene til formålsspecifikation, dataminimering, automatiske afgørelser og rigtighed.

7.2 Behandlingssikkerhed

Kravene til behandlingssikkerhed er forholdsvist tilgivende og kan deles op i de skærpende vurderingsmomenter – behandlingens egenskaber, risici og konsekvenser – og de lempelige – det aktuelle tekniske niveau og implementeringsomkostningerne. Omfanget af sikkerhedsforanstaltninger bør derfor ende ud i, at den dataansvarlige sikrer passende, men overkommelige foranstaltninger til at sikre personoplysninger. Mere konkret afgøres dette ved at identificere de pågældende aktiver, der skal beskyttes, og trusler, der skal beskyttes imod. Effektive foranstaltninger afhænger af aktiver og truslers karakter, hvilket er en udfordring i sig selv at vurdere. AI er udsat for specielle trusler i form af reverse engineering, adversarial injections og backdoors. Oplagte og om muligt påkrævede foranstaltninger er pseudonymisering, kryptering og anonymisering, som medfører både fordele og ulemper. Derudover er AI ikke undtaget de generelle fordele ved backups.

Bemærk, at utilstrækkelig kode kan resultere i, at bugs efter GDPR i et vist omfang kan medføre sanktioner.

7.3 Behandling

7.3.1 Overordnede principper

De overordnede principper om lovlighed, rimelighed og gennemsigtighed skal iagttages i sammenhæng med alle udfordringer med forbindelse til AI og personoplysninger.

Rimelighed kommer især til udtryk ved principperne om dataminimering samt rigtighed, ansvarlighed og automatiske afgørelser, da disse i samspil regulerer diskrimination af individer, som en AI kan udføre på baggrund af urigtige data eller uansvarlig udvikling af softwaren. En potentiel løsning herpå er teknikker til at opnå explainable AI (XAI) såsom LIME.

Gennemsigtighed gør sig gældende specielt ved formålsspecifikation samt automatiske afgørelser på grund af kravet til påviselighed, hvor XAI også er en mulig løsning. Tilfældet er det samme og særligt udfordrende grundet den sorte boks med hensyn til de andre krav såsom oplysningspligten, fortegningsreglen og databeskyttelsespolitikker, hvor behandlinger også skal dokumenteres.

7.3.2 Formålsspecifikation

Formålsspecifikation er en betydelig problemstilling ved udvikling og anvendelse af AI, da det ikke altid vides på forhånd præcist, hvad AI'en skal anvende personoplysninger til. Det gælder især ved online modeller. Formålet skal være tilstrækkeligt præcist til, at den registrerede og den dataansvarlige opnår fælles forståelse for formålet og dets konsekvenser, hvilket afhænger af omstændigheder og den registreredes forudsætninger.

7.3.3 Viderebehandling

Det skal overvejes, om AI'en kan kategoriseres som et statistisk eller forskningsrelateret formål. Ellers er der to scenarier: Data er indsamlet til udvikling af AI'en – hvilket blot lægger op til at overveje formålsspecifikation – eller data er indsamlet til ét formål og skal nu viderebehandles til at udvikle en AI. En sådan viderebehandling kræver forenelighed med det oprindelige formål, hvilket ofte vil være tilfældet ved udvikling af AI, hvis behandlingen er proportionel med de tidligere nævnte betragtninger. Undtagelser hertil kan forekomme for offentlige myndigheder, som muligvis kan viderebehandle til uforenelige formål uden pligt til at oplyse den registrerede derom, hvis en bekendtgørelse senere bliver vedtaget herom.

7.3.4 Dataminimering

Den største problemstilling ved udvikling og anvendelse af AI er sandsynligvis, at det ikke altid er muligt at afgøre, hvilke oplysninger er nødvendige for AI'en, og at AI af deres karakter har gavn af så mange oplysninger som muligt. Pseudonymisering og anonymiseringsteknikker er specielt oplagte at anvende her.

Udgangspunktet er, at der kun må indhentes personoplysninger, som ved indsamlingen kan afgøres at være nødvendige. Dette er især problematisk ved online modeller. XAI kan anvendes til at afgøre, hvilke informationer er nødvendige, men kun retrospektivt. I tilfælde af, at det ikke kan afgøres på forhånd, om oplysninger er nødvendige eller ej, skal principperne om proportionalitet, rimelighed, gennemsigtighed og opbevaringsbegrænsning anvendes. Se eventuelt listen på side 30 for at lave en sådan konkret vurdering ved anvendelse af AI.

7.3.5 Opbevaringsbegrænsning

Nødvendigheden af personoplysninger som omtalt under dataminimering skal løbende overvejes og altså ikke kun ved indsamlingen. Viser personoplysningerne sig ikke at være nødvendige senere, skal de slettes omgående. Den dataansvarlige skal implementere foranstaltninger til efterleve dette.

7.3.6 Rigtighed

At slette eller berigtige oplysninger, der er blevet behandlet af en AI, er dog ikke en så simpel opgave, som kravet i forbindelse med opbevaringsbegrænsning giver anledning til at tro på grund af de tekniske omstændigheder. Dataansvarlige skal træffe tilstrækkelige foranstaltninger til at sikre effektiv sletning og berigtigelse – også for at være forberedt på den situation, hvor den registrerede anmoder om sådanne indgreb.

Det vil ikke altid være muligt at slette eller berigtige fuldkomment i forbindelse med online modeller og deep learning. Løsninger hertil må være at pseudonymisere, kryptere, aggregere og anonymisere i videst muligt omfang.

Hvorvidt denne problemstilling i de givne tilfælde medfører, at behandling af personoplysninger er ulovlig, kan ikke afgøres på det foreliggende grundlag, da GDPR ikke lader til at tage tilstrækkeligt hensyn til de tekniske omstændigheder. En formålsfortolkning medfører muligvis den aktuelle lovmæssighed.

7.3.7 Automatiske afgørelser

AI, der udfører automatiske afgørelser eller profilering, skal enten være baseret på et udtrykkeligt samtykke, nødvendigt for at indgå eller opfylde en kontrakt eller være hjemlet i EU-ret eller national ret. En eventuel viderebehandling må ikke baseres på undtagelsen til forenelighed, statistisk formål.

Derudover skal sådanne AI kunne påvise belæg for deres afgørelser, hvilket er en stor problemstilling for mange AI, hvorved XAI er en nødvendighed.

7.3.8 Databeskyttelse gennem design

Det vil oftest være lettest, billigst og nødvendigt at have implementeret de omtalte foranstaltninger fra start af udviklingen af en AI for at undgå komplikationer med modellen. Modulære modeller er klart at anbefale. Idet XAI løser mange problemstillinger for AI's overholdelse af reglerne, bør sådanne teknikker implementeres fra start.

7.4 Afsluttende bemærkninger

Det har vist sig, at GDPR ikke nødvendigvis resulterer i, at behandling af personoplysninger i forbindelse med udvikling eller anvendelse af AI medfører ulovlighed. Der er dog visse tilfælde især relaterede til online modeller og deep learning, hvor den dataansvarlige skal tage større skridt til at implementere passende foranstaltninger, og at databeskyttelse skal overvejes løbende igennem alle processerne med relation til AI. GDPR er muligvis mere fremtidssikker end sin forgænger, men den stiller også strenge krav ved ikke at tage fuldkomment hensyn til alle teknologiske omstændigheder og udviklinger; krav som kan virke uhensigtsmæssige. Det er dog ud fra denne afhandling muligt at identificere relevante problemstillinger og udfordringer og imødekomme dem, når der enten skal udvikles eller anvendes en AI til at behandle personoplysninger.

8 Litteraturliste

Adams, Tim - https://www.theguardian.com/technology/2010/dec/19/google-translate-computers-languages. (19. december 2010). *Can Google break the computer language barrier?* Hentet 28. april 2018 fra The Guardian.

Akinator - http://en.akinator.com/. (u.d.). Hentet 28. april 2018 fra Akinator.

Alpay, L. L.; et al. - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2669642/. (22. maj 2008). *Can Contextualization Increase Understanding During Man-Machine Communication? A Theory-Driven Study*. Hentet fra NCBI.

Angwin, Julia; et al. - https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. (23. maj 2016). *Machine Bias*. Hentet 28. april 2018 fra ProPublica.

Artikel 29-gruppen. (2. april 2013). *Opinion 03/2013 on purpose limitation.*

Artikel 29-gruppen. (2018). *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679.*

Baum, Seth - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3070741. (16. november 2017). *A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy.* (GCRI, Red.) Hentet 28. april 2018 fra SSRN.

Blume, Peter. (2009). *Juridisk metodelære* (5. udg.). Jurist- og Økonomforbundets Forlag.

Datatilsynet - https://www.datatilsynet.dk/fileadmin/user_upload/dokumenter/Vejledninger/Captia_Generel_informationspjece__formateret___DOK446249_.pdf. (2017). *En introduktion til de kommende, nye regler om beskyttelse af personoplysninger.* oktober. Hentet 12. maj 2018

Datatilsynet - https://www.datatilsynet.dk/offentlig/anonymisering/. (30. maj 2016). *Anonymisering*. Hentet 14. maj 2018 fra Datatilsynet.

Datatilsynet. (2018). *Vejledning om de registreredes rettigheder.*

Datatilsynet.no - https://www.datatilsynet.no/en/regulations-and-tools/guidelines/data-protection-by-design-and-by-default/?id=7728. (u.d.). *Guide: Software development with Data Protection by Design and by Default*. Hentet 30. april 2018 fra Datatilsynet.

Datatilsynet.no. (2018). *Kunstig intelligens og personvern.*

DeepMind - https://deepmind.com/blog/alphago-zero-learning-scratch/. (u.d.). *AlphaGo Zero*. Hentet 28. april 2018 fra DeepMind.

DeepMind - https://deepmind.com/research/alphago/. (u.d.). *AlphaGo*. Hentet 28. april 2018 fra DeepMind.

Den Danske Ordbog - Passende - http://ordnet.dk/ddo/ordbog?query=passende. (u.d.). Hentet 28. april 2018 fra Den Danske Ordbog.

Den Danske Ordbog - Tilstrækkelig - http://ordnet.dk/ddo/ordbog?query=tilstr%C3%A6kkelig. (u.d.). Hentet 28. april 2018 fra Den Danske Ordbog.

Dictionary.com - http://www.dictionary.com/browse/appropriate. (u.d.). *appropriate*. Hentet 28. april 2018 fra Dictionary.com.

Dwork, Cynthia; Roth, Aaron - http://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf. (2014). *The Algorithmic Foundation of Differential Privacy.* doi:10.1561/0400000042

Dyson, George - http://longnow.org/seminars/02013/mar/19/no-time-there-digital-universe-and-why-things-appear-be-speeding. (19. marts 2013). *No Time Is There - The Digital Universe and Why Things Appear To Be Speeding Up*. Hentet 28. april 2018 fra The Long Now Foundation.

ENISA. (2015). *Privacy by design in big data.* Hentet 28. april 2018

ENISA. (2016). *Big Data Threat Landscape and Good Practice Guide.* Hentet 28. april 2018

Europa-Parlamentet. (27. april 2016). Forordning 2016/679.

Folketinget - http://www.ft.dk/ripdf/samling/20171/lovforslag/l68/20171_l68_efter_2behandling.pdf. (15. maj 2018). *Lovforslag nr. L 68.* Hentet 16. maj 2018 fra Folketinget.

Gartner - https://www.gartner.com/it-glossary/big-data. (u.d.). *IT Glossary - Big Data*. Hentet 28. april 2018 fra Gartner.

Guy Hoff Sonne, Frederik - https://videnskab.dk/teknologi-innovation/black-box-problemet-naar-vi-ikke-forstaar-den-kunstige-intelligens. (25. november 2017). *Black box-problemet: Når vi ikke forstår den kunstige intelligens*. Hentet 28. april 2018 fra Videnskab.dk.

Horton, Helena - https://www.telegraph.co.uk/technology/2016/03/24/microsofts-teen-girl-ai-turns-into-a-hitler-loving-sex-robot-wit/. (24. marts 2016). *Microsoft deletes 'teen girl' AI after it became a Hitler-loving sex robot within 24 hours*. Hentet fra The Telegraph.

Huk Park, Dong; et al. - https://arxiv.org/pdf/1802.08129.pdf. (2018). *Multimodal Explanation: Justifying Decisions and Pointing to the Evidence.* Hentet 30. april 2018

ICO - https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/. (u.d.). *Rights related to automated decision making including profiling*. Hentet fra ICO.

ICO. (2017). *Big data, artificial intelligence, machine learning and data protection.*

International Organization for Standardization. (2013). *ISO 27001.* Hentet 28. april 2018

Jacobson, Ralph - https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/. (24. april 2013). *2.5 quintillion bytes of data created every day. How does CPG & Retail manage it?* Hentet 28. april 2018 fra IBM.

Landau, Deb Miller - https://iq.intel.com/artificial-intelligence-and-machine-learning/. (17. august 2016). *Artificial Intelligence and Machine Learning: How Computers Learn*. Hentet 28. april 2018 fra IQ Intel.

Laney, Doug - https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. (2001). *Application Delivery Strategies.* 6: februar. Hentet 28. april 2018

Li, Tiffany; et al. - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3018186. (15. august 2017). *Humans Forget, Machines Remember: Artificial Intelligence and the Right to Be Forgotten.* Hentet 13. maj 2018 fra SSRN.

Lomonaco, Vincenzo - https://medium.com/@vlomonaco/why-continuous-learning-is-the-key-towards-machine-intelligence-1851cb57c308. (4. oktober 2017). *Why Continuous Learning is the key towards Machine Intelligence*. Hentet 28. april 2018 fra Medium.

Machine learning research group - http://www.cs.utexas.edu/~ml/publications/area/125/transfer_learning. (u.d.). *Publications: Transfer Learning*. Hentet 11. maj 2018 fra University of Texas at Austin artificial intelligence lab.

McMahan, Brendan; Ramage, Daniel - https://ai.googleblog.com/2017/04/federated-learning-collaborative.html. (6. april 2017). *Federated Learning: Collaborate Machine Learning without Centralized Training Data*. Hentet 11. maj 2018 fra Google AI Blog.

NVIDIA - https://www.nvidia.com/en-us/deep-learning-ai/industries/robotics/. (u.d.). *NVIDIA Isaac Platforn for Robotics*. Hentet 28. april 2018 fra NVIDIA.

O'Reilly, Tim - https://plus.google.com/+TimOReilly/posts/Ej72QmgdJTf. (9. maj 2013). *George Dyson's Definition of "Big Data"*. Hentet 28. april 2018 fra Google Plus.

Poole, L David; Mackworth, K Alan - http://artint.info/html/ArtInt_13.html. (2010). *Artificial Intelligence: Foundations of Computaional Agents.* Hentet 28. april 2018

Powles, Julia; Selbst, Andrew D - https://academic.oup.com/idpl/article/7/4/233/4762325. (19. december 2017). *Meaningful information and the right to explanation*. Hentet 30. april 2018 fra Oxford Academic.

Pure Storage - http://info.purestorage.com/rs/225-USM-292/images/Big%20Data%27s%20Big%20Failure_UK%281%29.pdf?aliId=64921319. (u.d.). *The struggles businesses face in accessing the information they need.* Hentet 28. april 2018 fra Pure Storage.

Ribeiro, Marco T - https://homes.cs.washington.edu/~marcotcr/blog/lime/. (2. april 2016). *LIME - Local Interpretable Model-Agnostic Explanations*. Hentet 13. maj 2018 fra Marco Tulio Ribeiros blog.

Sentient - AI - https://www.sentient.ai/blog/myth-busting-artificial-intelligence/. (19. februar 2015). *Myth Busting Artificial Intelligence*. Hentet 28. april 2018 fra Sentient.

Sentient - black box - https://www.sentient.ai/blog/understanding-black-box-artificial-intelligence/. (9. januar 2018). *Understanding the 'black box' of artificial intelligence*. Hentet 28. april 2018 fra Sentient.

Singh, Sameer; et al. - http://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf. (2016). *"Why Should I Trust You?" Explaining the Predictions of Any Classifier.* ACM. Hentet 13. maj 2018

Sophia - https://www.youtube.com/watch?v=S5t6K9iwcdw. (25. oktober 2017). *CNBC*. Hentet 28. april 2018 fra Youtube.

SSD, NSD - http://raird.no. (u.d.). *The RAIRD Project*. Hentet 11. maj 2018 fra RAIRD - Remote Access Infrastructure for Register Data.

Stuntz, Craig - https://community.embarcadero.com/blogs/entry/what-is-homomorphic-encryption-and-why-should-i-care-38566. (18. marts 2010). *What is Homomorphic Encryption, and Why Should I Care*. Hentet 11. maj 2018 fra Embarcadero.

Tanner, Adam - https://www.reuters.com/article/us-google-translate-idUSN1921881520070328. (28. marts 2007). *Google seeks world if instant translations*. Hentet 28. april 2018 fra Reuters.

Tranberg, Charlotte Bagger. (2007). *Nødvendig behandling af personoplysninger* (1. udg.). København: Forlaget Thomson.

Undervisningsministeriet - http://cms.uvm.dk/statistik/grundskolen/datavarehuset. (20. marts 2017). *Datavarehuset*. Hentet fra Undervisningsministeriet.

Wachter, Sarah; et al. - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469. (24. januar 2017). *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*. Hentet 30. april 2018 fra SSRN.

Wikipedia - https://en.wikipedia.org/wiki/Incremental_learning. (10. december 2017). *Incremental learning*. Hentet fra Wikipedia.

Wikipedia - https://en.wikipedia.org/wiki/Modular_programming. (4. april 2018). *Modular programming*. Hentet 28. april 2018 fra Wikipedia.

World Wide Web Foundation - https://webfoundation.org/docs/2017/07/WF_Algorithms.pdf. (2017). *Algorithmic Accountability.* Hentet 13. maj 2018

9 Liste over figurer