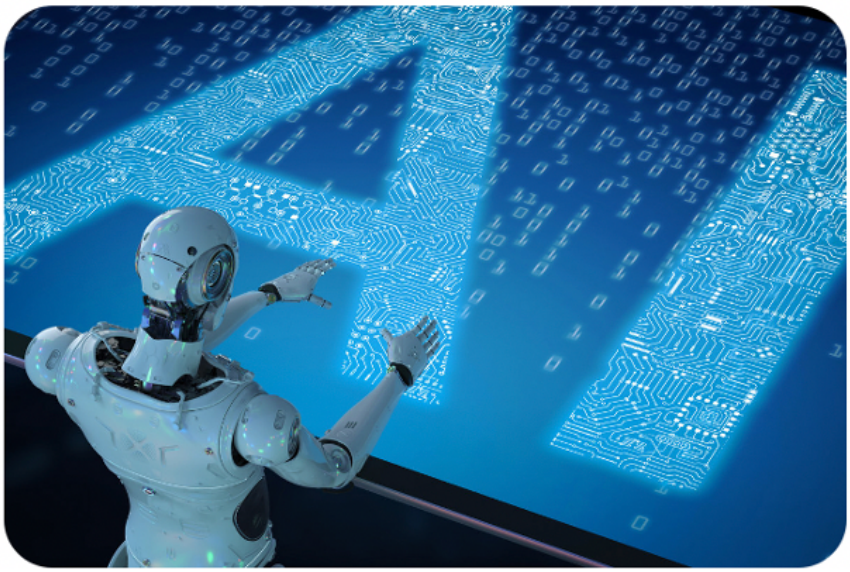
Overholdelse af GDPR ved udvikling og anvendelse af AI set ud fra et compliance-perspektiv

- med fokus på udfordringer samt løsninger og dokumentation

af SHERVIN SHIRAZI



Kandidatafhandling

Cand.merc.jur

Juridisk vejleder
Karsten Revsbeck

Økonomisk vejleder
Kirstine Emilie Gaard Brix

Forfatter
Shervin Shirazi

Aarhus Universitet
Business and Social Sciences
[1. juni 2021]

Antal: **119.991**

Indholdsfortegnelse

Abstract

This thesis discusses how controllers comply with the General Data Protection Regulation (GDPR) when developing and deploying artificial intelligence (AI). A new model called ”the extended simplified AI-lifecycle” will be introduced, that facilitates the use of AI in organizations and public authorities without violating rules across the different phases of the AI-lifecycle. Especial attention will be paid to potential data protection challenges that may arise during the first three phases of the AI-lifecycle and their respective possible solutions.

Firstly, the paper states a legal and technical definition of the term AI. Google Translate is used as an example to illustrate the workings of machine learning and some of its weaknesses. The findings indicate that the functionality of an algorithm is highly dependent on its complexity. For example, neural networks are in most cases incomprehensible because their inner workings are unknown. Additionally, discriminative machine learning will be explained. The conclusions reached in this paper are that discrimination is caused by imbalanced training data sets and training data that reflect past discrimination.

Secondly, the paper also explores inherent challenges and requirements concerning the development and deployment of machine learning in relation to GDPR.

a) Machine learning often will fall within the material scope of the GDPR cf. art. 2(1), however, it may be difficult to discern whether data is personal data cf. art. 4(1).

b) The need for controllers to determine the purpose of the AI solution with adequate precision as as early as possible.

c) Furthermore, machine learning and art. 5(1b) are incompatible. Additional processing of personal data to train, test and build the models in order to profit from data hampers GDPR compliance.

d) The controllers must follow the data minimisation principle in art. 5(1c), when collecting personal data. However, again this principle is difficult to comply with because the precision is dependent on the legitimacy of art. 5(1b).

e) The data subject is protected in connection with automated decision-making, provided that the processing fulfills the requirements in art. 22. Additionally, profiling in art. 4(4) most likely interferes with principles in art. 5 such as art. 5(1a) and 5(1d).

The last findings derive from the trade-offs. The right to data protection is established as a fundamental right in the EU Charter for Fundamental Rights art. 8 and TFEU art. 16. Machine learning must

comply with privacy. Machine learning also has to be fair cf. art. 5(1a), accurate cf. art. 5(1d) and explainable cf. art. 13 (2f)/art. 14(2g)/art.15(1h) and not unfairly, inaccurate or unexplainable. Overall, the controllers need to prevent high-risk factors such as serious discrimination, especially as a result of imbalanced training data sets and/or training data that reflect past discrimination and profiling. However, the analyses proves that this is not always the case because the IT-developers are able to tweak outputs by removing, adding or changing variables in an algorithm.

Finally, the paper contains technical measures based on privacy by design in art. 25(1) that can be implemented to ensure that the AI-specific areas are mitigated. The most adequate measures to mitigate data minimisation, fairness, accuracy and explainability are (1) data minimisation techniques, (2) mitigation measures, (3) accuracy measures and (4) explainable AI measures. *Firstly,* data minimisation techniques, such as privacy-preserving methods, make the distinction between personal and non-personal data less blurry. *Secondly,* mitigation measures can be used to add or remove data regarding under- or overrepresented groups. *Thirdly*, accuracy measures improve recruitment overall. *Fourthly*, explainable AI measures help to reduce the complexity of AI-systems.

This thesis goes further beyond addressing the above-mentioned challenges, requirements and technical measures. It also addresses how to ensure that trade-offs are identified and solved. This means that the controllers both need to strike the right balance between privacy, fairness, accuracy and explainability and at the same time implement adequate technical measures to ensure that they are compliant with the legislation. The controllers also need to be able to demonstrate how to strike the right balance to be compliant with the accountability principle in art. 5(2). In order to achieve all of this, another new model called ”the trade-off-model” will be introduced and discussed.

1 Introduktion

Artificial intelligence (AI) har et kæmpe potentiale og redder liv. Teknologien kan forudsige hvem der højst sandsynligt vil dø af COVID-19[1], bekæmpe kræft[2] eller terrorangreb[3], fremme performance hos OL-atleter[4] og skabe programmet AlphaGo til at besejre den ypperste Go-mester i historien[5] mv.

Der er således mange gavnlige effekter ved at implementere AI og de økonomiske effektivitetsgevinster er heller ikke til at tage fejl af i Europa-Parlamentets storrapport.[6] Imidlertid har vi en problemstilling, der hedder: efterlevelse af reglerne i databeskyttelsesforordningen[7] (herefter *compliance*). Compliance er altså en udfordring for såvel private virksomheder som offentlige myndigheder. I takt med at man efterlever reglerne, så kan AI tillige føre til diskrimination af det enkelte individ. Dette fordrer, at man ofte bliver nødt til at holde visse regler op mod privatlivsbeskyttelsen. I Storbritannien tør Information Comissioner's Office (ICO) også at være åbne om, at der foreligger nogle fundamentale *trade-offs* mellem krav til databeskyttelse på den ene side og de grundlæggende drivkræfter bag AI-baserede løsninger på den anden.[8] Et trade-off handler derfor om at foretage en afvejning. ICO har bl.a. lanceret et nyt "*Auditing Framework for AI*", hvorefter der både fremhæves nogle "*AI-specifikke*" udfordringer samt vigtigheden af "*Governance og accountability*", da dette generelle krav er nødvendigt med komplekse og risikable teknologier som AI.[9]

I Danmark anerkender man mulighederne og har derfor publiceret "*National strategi for kunstig intelligens*"[10]. I Norge anerkender man til gengæld udfordringerne og har som følge heraf publiceret "*Nasjonal strategi for kunstig intelligens*"[11]. Disse potentielle lovgivningsmæssige og investeringsmæssige tiltag fra Danmark, Norge og øvrige medlemsstater vil i øvrigt blive støttet af EU-Kommissionen, i det omfang formålet er at fremme anvendelsen af AI og håndtering af risiciene.[12]

Der findes derfor gode argumenter for at komme i gang med sin compliance set ud fra virksomheders og offentlige myndigheders synsvinkel. AI indebærer nogle af de højeste risici for det enkelte individ

[1] Datalogisk Institut: *Computer fortæller om man dør af COVID-19* (2021)
[2] Nielsen: *Kunstig intelligens styrker kræftbehandling* (2018)
[3] https://www.digitaltrends.com/cool-tech/machine-learning-v-for-victory-terrorist-identification/ (01.03.2021)
[4] Soper: *How Olympic athletes use machine learning and data analysis to reach peak performance levels* (2016)
[5] DeepMind: *AlphaGo – The Movie | Full Documentary* (01.03.2021)
[6] Figur 3 i EU-Parliament: *The impact of the GDPR on AI* (2020), s. 7
[7] Europa-Parlamentets og Rådets forordning (EU) 2016/679 af 27. april 2016 om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger og om ophævelse af direktiv 95/46/EF (herefter GDPR)
[8] ICO: *Trade-offs* (2019)
[9] ICO: *An overview of the Auditing Framework for Artificial Intelligence and its core components* (2019)
[10] https://www.regeringen.dk/nyheder/2019/national-strategi-for-kunstig-intelligens/ (01.03.2021)
[11] Kommunal- og moderniseringsdepartementet: *Nasjonal strategi for kunstig intelligens* (2020)
[12] EU-Kommissionen: *Hvidbog om kunstig intelligens – en europæisk tilgang til ekspertise og tillid* (2020), s. 1

og kan potentielt føre til nogle af de største bøder både nu og i fremtiden. Årsagen hertil er, at udvikling og anvendelse af AI, herunder særligt machine learning, hurtigt kan komme i konflikt med nogle af de helt grundlæggende behandlingsprincipper i GDPR.[13] Flere af datarettighederne kan tilsvarende blive stærkt udfordret.[14] Der er således et stort behov for at skabe en klarere retsstilling på AI-området. Undervejs i afhandlingen kan man om nødvendigt slå relevante tekniske begreber op.[15]

1.1 Problemformulering

På baggrund af ovenstående introduktion vil der i afhandlingen blive belyst nogle af de væsentligste udfordringer, der kan opstå i forbindelse med at blive mest muligt compliant med AI. Dernæst belyses de løsninger og dokumentation, der er nødvendige for at sikre denne compliance. Med afsæt heri har afhandlingen til formål at undersøge:

- *Hvordan man kan sikre overholdelse af GDPR ved udvikling og anvendelse af AI, herunder særligt machine learning set ud fra et compliance-perspektiv. Undersøgelsen tager udgangspunkt i udvalgte bestemmelser, som hovedsageligt fokuserer på udfordringer samt løsninger og dokumentation.*

1.2 Afgrænsning

Formålet med afhandlingen er kun at gå i dybden med overholdelsen af GDPR ved udvikling og anvendelse af machine learning set ud fra et compliance-perspektiv. Det er derfor kun udvalgte bestemmelser, som afhandlingen vil gå i dybden med. I opsummeringen (afsnit 5.1.2), findes forklaringer på, hvilke bestemmelser der vælges, og hvorfor det er dem, der tillægges stor juridisk betydning i arbejdet med compliance.

Selvom afhandlingen går i dybden med en række udvalgte bestemmelser, bliver de ikke gennemgået udtømmende eller udførligt. De vil hovedsageligt blive analyseret og diskuteret ud fra et afgrænset fokus på udfordringer i indsamlingsfasen, udviklings- og træningsfasen og anvendelsesfasen af AI. Løsninger og dokumentation heraf vil ske i forankringsfasen. Det betyder i praksis, at essensen eller uddybningen af de udvalgte bestemmelser alene vil blive analyseret i et omfang, der skaber de bedste forudsætninger for at dykke nærmere ned i det afgrænset fokusområde. Det bemærkes i øvrigt, at der

[13] Norsk datatilsyn: *Kunstig intelligens og personvern* (2018), s. 14
[14] Bødeniveauet for at overtræde bl.a. behandlingsprincipperne og datarettighederne følger af art. 83, stk. 5, litra a og b.
[15] Bilag 3

potentielt altid kan ske en udvikling og anvendelse af machine learning i hver af de tre første faser i *AI-livscyklussen* (afsnit 3.2.3 herunder 3.2.2). Dette forhold skal derfor i afhandlingen ikke begrænses til kun udviklings- og træningsfasen samt anvendelsesfasen.

Afgrænsningen er præget af, at der er en sidetalsbegrænsning på afhandlingen. Derfor foreligger der øvrige emner, som naturligt skulle have været indgået, men som helt eller delvist alligevel ikke vil blive behandlet. Det drejer sig bl.a. om følgende emner: *de øvrige behandlingsprincipper* (art. 5, stk. 1, litra e[16] og f[17]), *lovlig behandling af almindelige og følsomme personoplysninger* (art. 6, stk. 1 og 9, stk. 2)[18], *forskning og statistik*[19], *risikovurderingen*[20]*, konsekvensanalysen* (art. 35)[21], *adfærdskodekser og certificering* (art. 40-43)[22] og *øvrige datarettigheder*[23].

Afhandlingen centrerer sig primært om forholdet mellem de registrerede (herefter datasubjekterne) og de(n) dataansvarlige. Disse er begreber som GDPR opstiller. "*Datasubjektet*" er en identificeret eller identificerbar fysisk person, jf. art. 4, nr. 1, og det er denne, der skal beskyttes. Den "*dataansvarlige*" er i denne afhandling enten en privat virksomhed eller en offentlig myndighed. Det er disse, der fastsætter formålene og hjælpemidlerne med behandlingen af personoplysninger. Dette følger af art. 4, nr. 7.[24] Når der i afhandlingen skrives "de(n)" dataansvarlige, henvises der til, at det ikke kan udelukkes, at to eller flere dataansvarlige i fælleskab fastlægger formålene med og hjælpemidlerne til behandlingen af personoplysninger. I så fald er man ifølge art. 26, jf. art. 4, nr. 7 fælles om dataansvaret ved udvikling og anvendelse af machine learning.

En privat virksomhed inddrages i afhandlingen. Virksomheden fungerer som dataansvarlig ved udviklingen og anvendelsen af deres AI-løsning, som anført i afsnit 2.2. Casestudiet har en understøttende funktion i afhandlingen ved at uddybe og nuancere de analyser og diskussioner, der bliver foretaget løbende ved besvarelsen af problemformuleringen. Da det har en understøttende funktion, har det ikke været meningen med afhandlingen at begrænse det til kun én dataansvarlig og heller ikke til

[16] EU-Parliament: *The impact of the GDPR on AI* (2020), s. 48*f*
[17] ICO: *Known security risks exacerbated by AI* (2019)
[18] Blume & Motzfeldt: *Databeskyttelse og udvikling af kunstig intelligens* (2020), s. 3
[19] Ibid. s. 3*f*
[20] Risikovurderingens faser *kan* se ud som dem, der fremgår af Olsen: *Håndbog i dataansvarlighed* (2020), s. 410*ff*
[21] ICO: *Data Protection Impact Assessments and AI* (2019)
[22] EU-Parliament: *The impact of the GDPR on AI* (2020), s. 69
[23] ICO: *Enabling access, erasure, and rectification rights in AI* (2019)
[24] Korfits & Lotterup: *Databeskyttelsesforordningen og databeskyttelsesloven med kommentarer* (2020), s. 271-280

private virksomheder. Offentlige myndigheder[25] kan som dataansvarlige således også få gavn af resultaterne. Meningen med casestudiet har været at bruge den praktiske virkelighed til at fortælle noget generelt om retstilstanden. Dette er med til at *udvikle* vores forståelse af de begreber, vi har inden for persondataretten.

Databeskyttelsesloven[26] (herefter DBL) eller lex specialis vil ikke blive behandlet, da problemformuleringen centrerer sig om balancegangen mellem brugen af AI på den ene side og retten til privatlivsbeskyttelse i henhold til GDPR på den anden. De(n) dataansvarlige skal dog altid være opmærksom(me) på, at GDPR efter betragtning 8[27] åbner for et nationalt råderum i dansk ret, hvor visse afvigelser og præciseringer ikke kan udelukkes.

2 Metoder og retskilder mv.

2.1 Retsdogmatisk formål

Den retsdogmatiske metode har til formål at beskrive, analysere og systematisere gældende ret[28] *(de lege lata).* Denne metode er benyttet til at indplacere AI i eksisterende lovgivning. Der tages afsæt i GDPR med henblik på at fastsætte, hvad gældende ret er. GDPR er en forordning, der almengyldig og bindende i hver medlemsstat i EU[29] og har forrang frem for national ret[30]. Der er tale om en retsakt, der har en høj retskildeværdi. Som nævnt i forrige afsnit er der dog givet et frirum til national udfyldning, hvorefter der i Danmark er udarbejdet en DBL.[31]

Præambelbetragtninger vil blive inddraget i høj grad som fortolkningsbidrag. Præamblen skal – ved tvivlsspørgsmål i denne afhandling – anses for at være af stor betydning. Dette illustreres af, at der er sket en nøje gennemgang af de enkelte betragtninger ved udformningen af GDPR.[32] Justitsministeriets betænkning nr. 1565/2017 over GDPR (herefter bet. 1565/2017) vil også bruges som fortolkningsbidrag til uddybning af reglerne i GDPR. Begge lovforarbejder har en høj retskildeværdi, da de dels

[25] Revsbech m.fl.: *Forvaltningsret – almindelige emner* (2016), s. 17*f*, hvorefter "*offentlige myndigheder*" skal forstås som organer, der kan henføres under den offentlige forvaltning. En meget stor del af den samlede offentlige forvaltning foregår inden for rammerne af statslige ministerier, kommuner og regioner.
[26] Lov nr. 502 af 23. maj 2018
[27] Om præambelbetragtninger (afsnit 2.1)
[28] Madsen: *Retsdogmatisk forskning* (2021), s. 2
[29] Sørensen m.fl.: *EU-retten* (2014), s. 96
[30] Ibid. s. 174
[31] Blume: *Den nye persondataret* (2018), s. 29*f*
[32] Ibid. s. 53

angiver hjemmelsgrundlaget og dels nærmere angiver lovgivers motiver.[33] [34] Herved opnås en større forståelse for, hvad meningen med lovgivningen er.

Det danske datatilsyn (Datatilsynet) er den centrale uafhængige myndighed, der fører tilsyn med, at de databeskyttelsesretlige regler bliver overholdt.[35] Uden datatilsyn ingen databeskyttelse. De skal dog ikke kun sikre, at reglerne fungerer og træffe afgørelser. Det betones endvidere, at de også har pligt til – via vejlednings- og rådgivningsfunktionen – at finde praktiske og pragmatiske løsninger over for de(n) dataansvarlige.[36] Det forudsætter it-teknologisk kompetence[37] om AI, hvilket i øjeblikket må anses for at være en mangelvare i medarbejdersammensætningen hos tilsynet[38]. Dette kan virke bekymrende i forhold til de udfordringer, AI kan medføre for datasubjekternes rettigheder og frihedsrettigheder. Af disse anførte grunde – og fordi de ikke har udgivet en vejledning om AI endnu – vil fortolkningsbidrag fra det danske datatilsyn blive inddraget i mindre grad.

Afhandlingen har et internationalt fokus. Derfor vil der i tilfælde af modstridende interesser mellem det danske datatilsyn og Art. 29-gruppen blive lagt vægt på sidstnævnte. Dette kan godt lade sig gøre, idet Art. 29-gruppens udtalelser i vidt omfang stadigvæk kan bruges.

Art. 29-gruppen var en uafhængig europæisk arbejdsgruppe, der var knyttet til det tidligere databeskyttelsesdirektiv (herefter DBD).[39] [40] Art. 29-gruppen er i dag blevet til Det Europæiske Databeskyttelsesråd (herefter EDPB), men der er flere bestemmelser, der er videreført fra direktivet til forordningen.[41] I denne afhandling er der kun benyttet retningslinjer fra Art. 29-gruppen, som er tiltrådt af EDPB, og hvor bestemmelserne indholdsmæssigt er identisk.[42]

Europa-Parlamentets storrapport, det norske datatilsyns vejledning og det spanske datatilsyns vejledning inddrages i højere grad ved fortolkning af de udvalgte bestemmelser i GDPR. Disse bidrag beskriver, hvordan persondata bliver udfordret af AI. Herudover tillægges det vægt, at dokumenterne

[33] Hansen & Werlauff: *Den juridiske metode* (2016), s. 246
[34] Andersen: *Ret & Metode* (2002), s. 144
[35] https://www.datatilsynet.dk/om-datatilsynet (10.03.2021)
[36] Blume: *Den nye persondataret* (2018), s. 188
[37] Ibid.
[38] Datatilsynet: *Årsberetning 2020*, s. 14*f*
[39] Direktiv 95/46/EF af 24. oktober 1995
[40] https://edpb.europa.eu/our-work-tools/article-29-working-party_da (20.03.21)
[41] Bet. 1565/2017, s. 31
[42] https://edpb.europa.eu/sites/edpb/files/files/news/endorsement_of_wp29_documents.pdf (22.03.2021)

tager afsæt i GDPR og Art. 29-Gruppens retningslinjer.[43] [44] De pågældende valg af udenlandske retskilder skyldes, at det danske datatilsyn endnu ikke har forholdt sig til emnet.

Det britiske datatilsyn har de senere år publiceret en bred vifte af materiale om AI, der på hver sin måde er med til at beskrive de udfordringer, som AI indebærer. Afhandlingen tillægger disse AI-specifikke risici betydelig vægt, da der generelt er fokus på en risikobaseret tilgang til databeskyttelse.

De fem sidstnævnte fortolkningsbidrag giver alle velbegrundede bud på, hvilke retlige udfordringer AI kan medføre, hvorfor de har en vis retskildemæssig betydning. Det bemærkelsesværdige er imidlertid, at ICO er den eneste aktør, der for alvor tør beskrive løsningerne – og ikke kun udfordringerne.[45] Løsningsafsnittet vil derfor naturligt lægge vægt på ICO's fortolkningsbidrag.

Retspraksis eller administrativ praksis bliver ikke inddraget som retskilde til at fastlægge gældende ret, da der ikke er fundet domme eller afgørelser på AI-området.

Retspolitiske overvejelser *(de lege ferenda)* forekommer kun i et vist omfang. Hovedvægten vil klart være på gældende ret. Overordnet set bevæges der inden for et grænseområde, hvor der i det væsentligste er tale om fortolkning[46], men hvor det læner sig tæt op ad noget, der bør laves om.

2.2 Kvalitativt casestudie

Der inddrages et kvalitativt casestudie med Farmbrella[47], som er en Software as a Service (SaaS) virksomhed med fokus på agrosektoren. Farmbrella er en digital platform, hvor brugere og virksomheder kan matches ved brug af AI.[48]

Casestudiet har givet indsigt i, hvilke processuelle og praktiske udfordringer machine learning har givet anledning til for Farmbrella. Virksomheden er nøje udvalgt, da de befinder sig på et stadie, hvor der skal tænkes compliance i udvikling og anvendelse af deres AI-løsning. De har nemlig brug for at

[43] Norsk datatilsyn: *Kunstig intelligens og personvern* (2018), s. 3
[44] EU-Parliament: *The impact of the GDPR on AI* (2020), s. 38*ff*
[45] F.eks. gives der en række anbefalinger fra det norske datatilsyn, der i højere grad forklarer, *hvad* der skal gøres frem for, *hvordan* det skal løses set ud fra et teknologisk perspektiv. I øvrigt er det skitseret ganske kortfattet. Norsk datatilsyn: *Kunstig intelligens og personvern* (2018), s. 27*f*
[46] Revsbech m.fl.: *Forvaltningsret – almindelige emner* (2016), s. 155
[47] https://farmbrella.dk (24.04.2021)
[48] Casestudiet om Farmbrella uddybes endvidere i bilag 2

vide, hvilke udfordringer det indebærer og hvilke løsninger og dokumentation, der kræves af dem. "Casestudie designet" er derfor yderst relevant for problemformuleringen.[49] Endvidere er det med til at svare på spørgsmål om hvordan og hvorfor, der sker en udvikling og en anvendelse af machine learning.[50]

Fordelen ved et kvalitativt casestudie er, at den kan tilbyde detaljerede informationer om et komplekst område.[51] Omvendt skal man være varsom med at generalisere ud fra de udfordringer Farmbrella har oplevet. Det skyldes, at der ikke foreligger fuld repræsentativitet[52] – selvom studiet til en vis grad kan være repræsentativt nok. Yderligere kan AI antage mange former samt have mange forskellige funktioner og anvendelsesformål, hvorfor konteksten også kan være anderledes.[53]

2.3 Litteratur

Ydermere inddrages relevant økonomisk og juridisk litteratur fra lærebøger, videnskabelige artikler, rapporter mv. Disse bidrag er med til at forklare, hvordan AI-teknologien fungerer og hvordan lovgivningen er sammensat på AI-området. Litteratur kan skabe en væsentlig forståelsesramme, men er imidlertid i overvejende grad udtryk for subjektive holdninger. Endvidere kan retslitteratur *(jurisprudence)* ikke betragtes som en retskilde.[54]

2.4 Struktur

Målet for de(n) dataansvarlige er som bekendt at blive mest muligt compliant med AI.

For det første foretages der en juridisk og teknologisk afgrænsning af AI (**afsnit 3**). Her vil de(n) dataansvarlige få et fornødent indblik i teknologien. Derudover introduceres den *udvidede forenklede AI-livscyklus,* der fungerer som grundstammen i forhold til overholdelsen af GDPR ved udvikling og anvendelse af AI.

---

[49] Baxter & Jack: *Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers* (2008), s. 556
[50] Ibid. s. 545 og Palmer & Balderston: *A Brief Introduction to Qualitative Research* (2006), s. 17
[51] Almeida: *Strengths and Limitations of Qualitative and Quantitative Research Methods* (2017), s. 379
[52] Casestudiet indeholder nemlig kun én dataansvarlig. Ibid. s. 377
[53] *Præsentation af juridisk værktøjskasse for ansvarlig AI* (2020), s. 7 i KL og Kammeradvokatens samlede værktøjskasse for ansvarlig AI: https://videncenter.kl.dk/viden-og-vaerktoejer/informationssikkerhed-og-gdpr/juridisk-ai-vaerktoejskasse/ (25.02.2021)
[54] Hansen & Werlauff: *Den juridiske metode* (2016), s. 166

Dernæst undersøges hvilke udfordringer til udvikling og anvendelse af machine learning, der er særlig relevante i forhold til GDPR (**afsnit 4**). I underafsnittende bliver de(n) dataansvarlige præsenteret for en række spørgsmål, som de med fordel kan stille sig selv. Spørgsmålene tager udgangspunkt i en systematisk gennemgang af nogle af de væsentligste regler, som GDPR opstiller i relation til compliance.

Endelig undersøges hvilke løsninger, der bør implementeres i forankringsfasen (**afsnit 5**). I underafsnittende bliver de(n) dataansvarlige også her præsenteret for en række spørgsmål – men denne gang i relation til *accountability* og *den risikobaserede tilgang til databeskyttelse.* Heri indgår også en ny *trade-off-model*. Dette AI-værktøj bygger på de forudgående analyser og bidrager til at dokumentere ens compliance på AI-området, den dag de nationale tilsynsmyndigheder fører tilsyn med, om reglerne er overholdt.

3 Artificial intelligence

3.1 Juridisk afgrænsning

Det, der var problemet med det tidligere DBD fra 1995 og persondataloven[55] fra 2000 var, at de ikke tog højde for internettet, sociale medier, hjemmesider mv. Disse teknologier er i dag en integreret del af vores hverdag som følge af den teknologiske udvikling. Da man i 2018 begyndte at lancere GDPR, samlede man op på alle de nyskabelser, der var sket siden man fik DBD. GDPR indeholder imidlertid ikke begrebet "*artificial intelligence*" [56]. GDPR er med andre ord teknologineutral[57], hvilket ifølge betragtning 15 skal sikre, at man forebygger risiko for omgåelse.

EU-kommissionen har for nylig anno 2021 fremsat nyt forslag om regulering af AI med tilhørende bilag.[58] I forslagets art. 3, nr. 1 kommer EU-Kommissionen med det tætteste bud på en juridisk definition, og den er baseret på flere års forarbejder. Den lyder som følger: "'*artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as*

[55] Lov nr. 429 af 31. maj 2000
[56] EU-Parliament: *The impact of the GDPR on AI* (2020), s. 35
[57] ICO: *Project explain: Interim report* (2019), s. 7
[58] https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence (14.05.2021)

content, predictions, recommendations, or decisions influencing the environments they interact with;". Annex I oplister heraf en række teknikker (modelvalg) omfattet af definitionen, hvori supervised og unsupervised learning nævnes som de første.

Udviklingen fra 1995 til i dag skal ses som et udtryk for, at jura halter bagud i forhold til teknologien. Opfindelser skabes først og reguleres efterfølgende. Imidlertid er mange bestemmelser i GDPR meget relevante i forhold til AI[59], hvilket bl.a. skyldes automatiseret behandling af personoplysninger i stor skala og brugen af profilering i automatisk beslutningstagning[60].

3.2 Teknologisk afgrænsning

Rent teknisk kan "*artifical intelligence*" defineres på mange forskellige måder.[61]
Det bedste bud på en international og toneangivende teknisk definition er ISO/IEC: "*An interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning*".[62] Fælles for de fleste AI-baserede teknologier er, at de tager afsæt i datasystemer, som kan lære af egne erfaringer og løse opgaver i forskelligartede situationer – egenskaber man tidligere har tiltænkt som unikke for mennesker.[63] Definitionen indikerer en række ligheder mellem måden "robotten" løser opgaver på, og måden mennesker løser opgaver på. Der er dog alligevel en væsentlig forskel. AI kan ofte løse opgaven med en hastighed og præcision, der langt overstiger menneskets evne, hvorfor teknologien i vidt omfang medfører tidsbesparelser og frigørelse af ressourcer.[64]

AI er en teknologi, der ofte er drevet af machine learning – selvom det ikke altid involverer machine learning. Begrebet AI vil i afhandlingen blive brugt som samlebetegnelse for de forskellige metoder, der hører ind under AI. Machine learning vil være den metode, som der tages udgangspunkt i. Selvom fokus er på machine learning, skal de(n) dataansvarlige være opmærksom(me) på, at andre former for AI også kan give anledning til databeskyttelsesretlige udfordringer i relation til compliance.[65] *Supervised* og *unsupervised* learning vil til gengæld være de primære modelvalg, som vil blive

[59] EU-Parliament: *The impact of the GDPR on AI* (2020), s. 35
[60] ICO: *Explaining decisions made with AI – Part 1* (2020), s. 11
[61] Ibid. s. 6
[62] ISO/IEC 2382-28:1995(en), Information technology — Vocabulary — Part 28: Artificial intelligence — Basic concepts and expert systems (14.05.2021)
[63] Norsk datatilsyn: *Kunstig intelligens og personvern* (2018), s. 4
[64] Afsnit 1
[65] ICO: *Guidance on AI and data protection* (2020), s. 7

gennemgået, hvoraf *decision trees, neural networks* og *clustering* også vil blive inkluderet. Andre modelvalg eller øvrige komponenter heri[66] vil kun blive berørt i mindre omfang eller slet ikke blive berørt. Disse tekniske analyser ved udvikling og anvendelse af modelvalgene inden for machine learning vil have det primære formål at styrke juraen i relation til de i afsnit 4 og 5 nævnte spørgsmål. På den måde står de juridiske vurderinger ikke alene, men begrundes også ud fra teknologien.

3.2.1 Hvordan fungerer machine learning i praksis og hvad er nogle af faldgruberne?

Machine learning[67] er en form for AI, som gør et system i stand til at lære og forbedre sig på baggrund af erfaringer, uden at systemet skal pogrammeres manuelt. Machine learning som metode indeholder to centrale underkategorier, hvilket er *supervised learning* og *unsupervsied learning.* Supervised learning er en model, som er trænet på et datasæt, som indeholder labelled data til forskel for unsupervised learning, der er trænet på et datasæt uden labelled data eller instruktioner.[68]

Google Translate (GT) er også machine learning[69] og er et godt eksempel på at illustrere, hvordan supervised learning fungerer i praksis, og hvad faldgruberne er.[70] Hvis et datasubjekt (A) eksempelvis skal oversætte det danske ord: ”hvad” til det engelske ord: ”what”, så er A's user input[71] ”hvad”. Training input[72] er derimod noget helt andet. Hvis det antages, at Google ejer 30 mio. computere på verdensplan, vil de 30 mio. computere (ca. 15 exabyte data) repræsentere de træningsdata, som Google anvender som training input i AI-modellen. Disse træningsdata føres herefter ind i en læringsalgoritme[73]. Læringsalgoritmen er en meget vigtig del af processen, fordi den styrker AI-modellen til at lære og ikke mindst til at blive intelligente. Når A skriver sit user input ind, kan modellen herefter hente informationer ned fra læringsalgoritmen (som altså er baseret på den viden, der kommer fra træningsdatene). AI-modellen er derfor nu i stand til at komme op med en plausibel forudsigelse, der gerne skulle give det output[74], som A ønsker. Altså outputtet ”what”. I dette forsimplede eksempel, er der slet ikke brug for så mange computere til at kunne frembringe det givne output. Mindre data kunne sagtens have givet et lige så præcist eller accurate[75] output. Hvis kompleksiteten imidlertid var

[66] Bilag 1
[67] Bilag 3
[68] ICO: *Explaining decisions made with AI – Part 1* (2020), s. 7
[69] Neil Nie: *Understanding Artificial Intelligence and Its Future*, TEDxDeerfield (2017)
[70] Figur 6 i EU-Parliament: *The impact of the GDPR on AI* (2020), s. 7
[71] Bilag 3
[72] Ibid.
[73] Ibid.
[74] Ibid.
[75] Ibid.

betydeligt højere, – som følge af meget lange og avancerede sætningskonstruktioner fra A – vil behovet for træningsdata tilsvarende stige. Det er også derfor, at Google sætter et maksimum for, hvor mange ord, A kan få lov til at oversætte ad gangen i GT. Det er for at undgå et alt for upræcist output, der kan skabe en lav explainability[76], og dermed risikere at afskære A fra at ville gøre brug af AI-løsningen.

GT er endvidere baseret på neural networks[77] i læringsalgoritmen, hvilket ofte ikke er særlig explainable.[78] Godt nok giver modellen et output, men det er svært at vide, hvorfor den rent faktisk har kommet frem til det givne output. For at illustrere måden neural networks fungerer på i praksis, kan der igen tages udgangspunkt i førnævnte eksempel med GT. Variablene x^1, x^2, x^3, x^4 er de variable, som Google indsætter i modellen. Det er med andre ord de data, man kender (dvs. Googles 15 exabyte træningsdata), der anvendes som training input. Med de data giver modellen herefter noget vægtning w^1, w^2, w^3, w^4 på de pågældende training input. Vægtningen er det, man i den sidste ende får en funktion f ud af. Slutteligt kan man få ens outcome $\hat{y}$. Explainability er lav, fordi vægtene som nævnt ikke kan forklare hvordan eller hvorfor, den gør, som den gør (også fordi vægtningen er afhængig af de andre variables vægte). De andre variables vægte er i dette tilfælde alt det andet, som Google indsætter som training input. Beregningerne er altså så komplekse, at de opleves som skjulte. Det er derfor at systemet er kaldt for en "black-box"[79]. Udfordringen er at "åbne" den sorte boks og gøre det synligt, hvad der de facto sker i den. Udfordringen er derfor at skabe en øget transparens.[80]

Forsimplet sagt er det den måde, GT fungerer på. Systemet vil blive givet et labelled datasæt bestående af input data og output data. Derfor fodrer man reelt set output fra algoritmen ind i systemet. Det betyder i praksis, at systemet allerede kender til outputtet i algoritmen, inden det begynder at arbejde på det.[81] Det er også derfor, at man som datasubjekt på få sekunder kan få sit output: "what" at vide. I unsupervised learning derimod, er systemet kun givet input data. Derfor er der ikke på samme måde et target[82]. Systemet skal selv prøve at forstå tingene ud fra de pågældende input data, som man indsætter. Måden, man gør dette på, er ved at finde frem til skjulte mønstre i datasættet. Det

[76] Bilag 3 og afsnit 4.5.4
[77] Bilag 3
[78] Figur 11 i FPF: *The Privacy Expert's Guide To AI and Machine Learning* (2018), s. 19
[79] Bilag 3
[80] https://www.kmd.dk/indsigter/kunstig-intelligens-skal-kunne-forklare-sine-anbefalinger (26.03.2021)
[81] FPF: *The Privacy Expert's Guide To AI and Machine Learning* (2018), s. 10
[82] Bilag 3

er herved, at clustering[83] kommer til udtryk, hvor man ikke har nogen information om target variablene. Man overlader det derfor til algoritmen at definere outputtet. Derfor ved man ikke på forhånd, hvad man leder efter eller hvor mange grupper, der er. Man går mere undersøgende til værks med det formål at opdage nogle skjulte mønstre i dataene. Styrken ved unsupervised learning er derfor, at den kan give et værdifuldt indblik i den underliggende struktur af et datasæt. Dette er ikke muligt med supervised learning. F.eks. bruger Netflix clustering til at lave nye anbefalinger til film ved at identificere hvilke nye film, der er relateret til film, som brugeren tidligere har set.[84] På den anden side har clustering den svaghed, at den er sensitiv over for visualiseringer, hvor accuracy ikke altid er helt så skarp sammenlignet med supervised learning.

3.2.1.1 Hvordan kan machine learning føre til diskrimination?

En af de største udfordringer ved udvikling og anvendelse af machine learning er diskrimination. Manglende overholdelse af art. 5, stk. 1, litra a's rimelighedsprincip har stærk tilknytning hertil, men forbud mod diskrimination er ikke kun forbeholdt denne bestemmelse. Diskrimination kan også optræde i forbindelse med øvrige behandlingsprincipper eller i andre sammenhænge[85]. Fokus er imidlertid på GDPR. Hvad begrebet *fairness*[86] indebærer uddybes nærmere i afsnit 4.5.2.

Med henblik på at illustrere hvordan machine learning rent teknisk kan føre til diskrimination, kan man forestille sig, at en bank har udviklet et AI-system. Formålet er at regne kreditrisikoen ud på potentielle kunder. Banken vil bruge systemet til at godkende eller afvise låneansøgninger. Med henblik på at træne systemet op, har banken indsamlet store mængder af data. Disse data indeholder en lang række oplysninger om tidligere skyldnere herunder variable som erhverv, indkomst, alder og om de har tilbagebetalt deres lån eller ej. Under testen ønsker banken at kontrollere eventuelle biases om køn og finder ud af, at systemet giver kvinder lavere kreditscore, hvilket fører til færre godkendte lån. Måden, hvorpå AI-systemet førte til diskrimination, skyldtes formentlig to ting. Den ene er *ubalancerede træningsdata*, mens den anden er *træningsdata som afspejler tidligere diskrimination*.[87]

---

[83] Ibid.
[84] Castañón: *Machine Learning Methods that Every Data Scientist Should Know* (2019)
[85] F.eks. ikke-diskriminationslovgivningen i Danmark i form af *lov om etnisk ligebehandling*, *ligebehandlingsloven*, *lov om forbud mod handicapdiskrimination mv*. Ikke-diskriminationslovene er ikke lex-specialis, da de ikke regulerer brug af data. De går således ikke forud for GDPR. Det er dog lovgivning, der skal overholdes, jf. art. 5 om *lovlighed*.
[86] Bilag 3
[87] ICO: *Human bias and discrimination in AI systems* (2019)

I de ubalancerede træningsdata indeholdte træningsdataene en større andel af skyldnere, som er mænd, fordi tidligere var det sådan, at færre kvinder ansøgte om at få optaget et lån. Derfor har banken ikke tilstrækkelige data om kvinder. Andelen af mænd versus kvinder i træningsdataene er derfor ikke afbalanceret. Eftersom populationen for mænd var overrepræsenteret i træningsdataene, vil modellen i højere grad være opmærksom på de statistiske forhold, som forudsiger tilbagebetalingsevnen for mænd – og mindre opmærksom på tilbagebetalingsevnen for kvinder. Modellen performer således bedre, når det er mænd og ikke kvinder. Dette må anses for at være diskrimination over for den kvindelige population, der er underrepræsenteret i træningsdataene. Ubalancerede træningsdata kan også give anledning til udfordringer i forhold til en *ansigtsgenkendelsesmodel* trænet i et uforholdsmæssigt stort antal ansigter, der hører til bestemte etniciteter og køn (f.eks. hvide mænd). Her vil modellen performe bedre, når man genkender individer i denne målgruppe og performe dårligere, når man genkender individer i en anden målgruppe (f.eks. sorte kvinder), fordi de netop er underrepræsenteret i træningsdataene.[88] I detailhandlen begynder man allerede nu at tage ansigtsgenkendelser til nye højder ved brug af AI.[89] Ikke noget med at ansigtsanalyser i sig selv kan føre til diskrimination, men meget tyder på, at der tillige er tale om biometriske data, jf. art. 4, nr. 14. Biometriske data er følsomme oplysninger, der som udgangspunkt er forbudte i medfør af art. 9, stk. 1. Tiden vil vise, hvilke konsekvenser disse arrangementer vil få i fremtiden.[90] Ikke desto mindre er der fra lovgivers side allerede tænkt over, at sådanne behandlingsaktiviteter hurtigt kan sprede sig både "*regionalt, nationalt eller overnationalt*" i henhold til betragtning 91, 1. pkt.

Årsagen til diskrimination kunne også have været forårsaget af den anden diskriminationsform: træningsdata som afspejler tidligere diskrimination. Hvis kvinders låneansøgninger historisk set tidligere blev afvist oftere end mænds låneansøgninger (på baggrund af f.eks. race eller køn), kan det være, at en model baseret på disse slags træningsdata vil reproducere de samme diskriminationsmønstre. Disse problemer kan opstå selvom træningsdataene nu og her ikke indeholder beskyttede karaktertræk som race og køn. Flere forskellige features i træningsdata (såsom beskæftigelse) er ofte korreleret med disse beskyttede karaktertræk. De kaldes også *proxy variable*[91] og gør det muligt for modellen at reproducere samme mønstre for diskrimination forbundet med race og køn – selvom hensigten var anderledes. Det er derfor, at machine learning er mere slagkraftig end traditionelle statistiske

[88] Ibid.
[89] Kjær: *Ugens Startup: Med avanceret ansigtsanalyse vil Justface Retail give fysiske butikker data-superkræfter* (2021)
[90] Rækkevidden af de juridiske konsekvenser ved ansigtsgenkendelse er kimen til stor debat. Se hertil: https://www.zetland.dk/historie/sekdrB5l-moV7PnW6-fbf2b (26.05.2021)
[91] Bilag 3

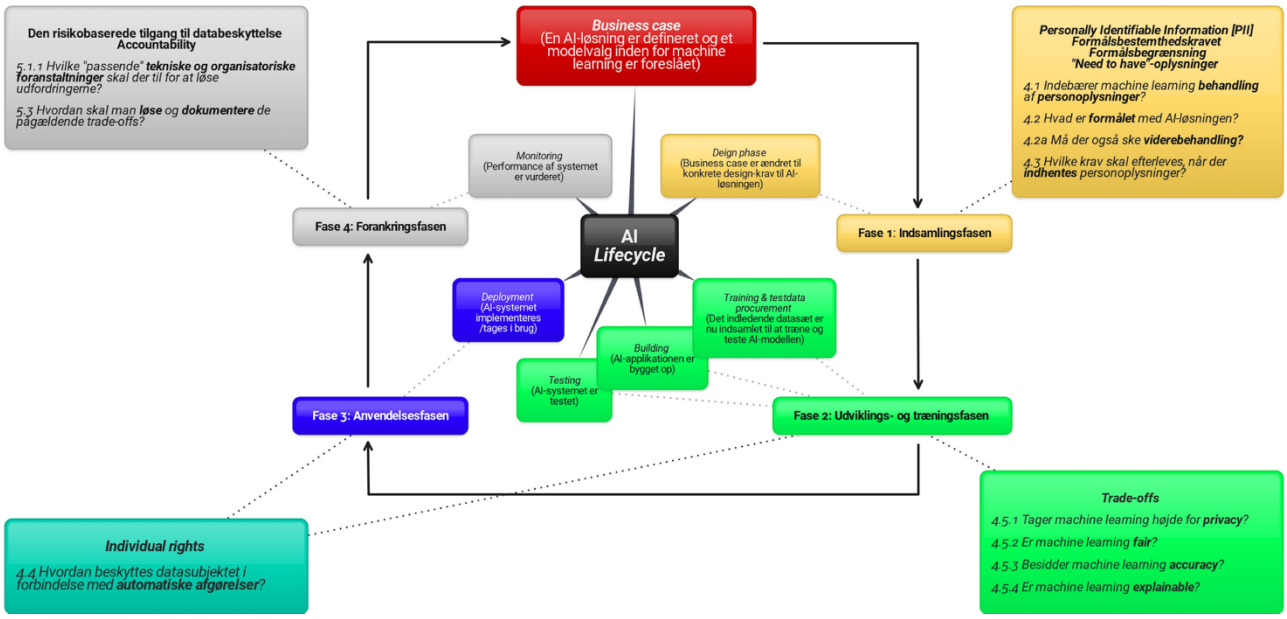
modeller, fordi den er bedre til at afdække skjulte mønstre i data, hvilket også inkluderer mønstre som afspejler diskrimination.

3.2.2 Hvilken model giver mening at bruge i den specifikke kontekst og hvilken betydning har modelvalget inden for GDPR?

Ovennævnte tekniske analyser viser, at de forskellige modeller kan antage mange forskellige funktioner, anvendelsesformål, faldgruber og i øvrigt forskellige former for diskrimination. Når det er sagt, er et modelvalg ikke altid bedre end et andet modelvalg. Om det skal være supervised eller unsupervised learning beror på en konkret vurdering alt afhængig af, hvilken opgave AI-løsningen skal løse. Dog skal man vide, at det pågældende modelvalg influerer på, hvordan de videre juridiske analyser ser ud. Så selvom faktum er, at GDPR's krav giver forskellige udfordringer i forskellige faser, er de juridiske analyser ikke nødvendigvis de samme i de forskellige faser – det afhænger af konteksten.

Til illustration ovenfor kan man forsimplet sige, at fordi man – i eksemplet med Google Translate (afsnit 3.2.1) – går ind og laver en plausibel forudsigelse, har man som udgangspunkt besluttet sig for, at det ikke skal være unsupervised learning, idet supervised learning er karakteriseret ved, at man har et target. I unsupervised learning har man ikke et target. Derfor ved man ofte tidligt i processen hvilken model, der giver mening at bruge i den specifikke kontekst. Svaret er dog mere nuanceret end dette, idet der er i modelleringen løbende foretages en vurdering af, hvilken model der passer bedst til den givne business case (afsnit 3.2.3). Det er her, at man i udviklings- og træningsfasen går ind og tester modellen af med henblik på at finde ud af, hvad der rent faktisk fungerer i praksis. På den måde tester man flere metoder af og finder eksempelvis ud af, at en model passer bedre med dataene og dermed har en højere præcision end en anden model. Så selvom modelvalget ofte er givet tidligt i processen, har det den betydning, at det kan nå at ændre sig meget undervejs i de forskellige faser. Det kan på sin vis også skiftes ud i de senere faser, hvoraf man eventuelt helt eller delvist bliver nødt til at starte forfra i f.eks. indsamlingsfasen eller udviklings- og træningsfasen. Mange ting kan ske – både bevidst eller ubevidst. Alt er afhængig af, hvilke data modellen passer bedst på. Modelvalget har derfor i den grad betydning inden for GDPR og kan ændre på konteksten i forskelige faser.

3.2.3 *Hvordan kan man sikre overholdelse af GDPR ved udvikling og anvendelse af machine learning?*



*Tabel 1: Den **"udvidede forenklede AI-livscyklus"** (med inspiration fra ICO's **"forenklede AI-livscyklus"**) har til formål at knytte de forskellige faser bag udvikling og anvendelse af modelvalget inden for machine learning op på de særlig udvalgte bestemmelser fra GDPR. Alle disse bestemmelser hidrører fra enten spørgsmålene i afsnit 4 eller spørgsmålene i afsnit 5.*

Det anbefales, at de(n) dataansvarlige med fordel følger den udvidede forenklede AI-livscyklus i arbejdet med at sikre overholdelse af GDPR ved udvikling og anvendelse af machine learning set ud fra et compliance-perspektiv (se tabel 1). Tanken bag ICO's egen forenklede AI-livscyklus er alene at fremhæve de faser, hvor de AI-specifikke risici højst sandsynligt manifesterer sig eller hvor de foreslåede kontrolmekanismer sandsynligvis er mest effektive.[92] Den nye udvidede forenklede AI-livscyklus, som er genstand for nærværende afhandling, tager dog skridtet videre. Udover at videre-bringe tankerne fra ICO's forenklede AI-livscyklus, kobler modellen også de databeskyttelsesretlige spørgsmål i afsnit 4 (udfordringerne) og afsnit 5 (løsningerne/dokumentation) op på:

- ***Fase 1: Indsamlingsfasen,***
- ***Fase 2: Udviklings- og træningsfasen,***
- ***Fase 3: Anvendelsesfasen og***
- ***Fase 4: Forankringsfasen.***

De fire faser vil i det følgende være udgangspunktet for analysen af de juridiske problemstillinger. Modellen kan virke kompliceret nu, men den indeholder centrale delspørgsmål i afsnit 4 og 5, som

[92] ICO: *An overview of the Auditing Framework for Artificial Intelligence and its core components* (2019)

bør fremhæves. Der er tale om en indledende model, hvorfra der vil ske en nærmere uddybning heraf i resten af afhandlingen. Bemærk i øvrigt at tekstboksen "individual rights" som vist i modellen (i modsætning til de tre andre tekstbokse) er forbundet til både udviklings- og træningsfasen samt anvendelsesfasen[93]. Det skyldes, at individual rights[94] (f.eks. explainability i afsnit 4.5.4) primært er relevant både ved iagttagelse af f.eks. oplysningspligten[95] i udviklings- og træningsfasen[96] og f.eks. indsigtsretten i anvendelsesfasen, såfremt datasubjektet anmoder om indsigt.

Alt i alt har de(n) dataansvarlige et fornødent overblik over, hvilke udvalgte bestemmelser fra GDPR, der stammer fra hvilke spørgsmål. Fordi faserne bliver sat i kontekst med de regler, GDPR opstiller i relation til compliance, er den udvidede forenklede AI-livscyklus et stærkt værktøj til at nå målet om at blive compliant med AI (afsnit 2.4). Når det er sagt, er den udvidede forenklede AI-livscyklus – som navnet antyder – stadigvæk "forenklet", hvorfor man skal være opmærksom på, at de forskellige spørgsmål også kan risikere at optræde under nogle af de andre faser. F.eks. er dataminimering (afsnit 4.3) primært relevant i indsamlingsfasen, men kan også være det både i udviklings- og træningsfasen samt anvendelsesfasen. De(n) dataansvarlige skal herudover huske, at selvom afhandlingens problemformulering fordrer et fokus på nogle af de vigtige bestemmelser inden for GDPR i relation til compliance, er der ikke noget til hinder for, at man som dataansvarlig(e) indføjer yderligere bestemmelser til AI-livsscyklusen, hvis det har relevans for arbejdet.

4 Hvilke udfordringer og krav til udvikling og anvendelse af machine learning er særligt relevante i forhold til GDPR?

4.1 Indebærer machine learning behandling af personoplysninger?

Det materielle anvendelsesområde i art. 2, stk. 1 foreskriver, at GDPR finder anvendelse på behandling af personoplysninger, der helt eller delvis foretages ved hjælp af automatisk databehandling, og på anden ikkeautomatisk behandling af personoplysninger, der er eller vil blive indeholdt i et register.

[93] I øvrigt illustreret ved to stiplede linjer i modellen.
[94] Bemærk at en individual right også kan være art. 22-rettighed (afsnit 4.4)
[95] Oplysningspligten indgår som en komponent i begrebet explainability (afsnit 4.5.4)
[96] Man kan imidlertid foranlediges til at tro, at oplysningspligten altid skal opfyldes ved *indsamlingen* af personoplysninger (indsamlingsfasen). I praksis kan det diskuteres (afsnit 4.5.4)

Det forudsættes, at de fire undtagelser i art. 2, stk. 1, jf. art. 2, stk. 2, litra a-d, ikke finder anvendelse, og at behandlingen falder inden for det territorielle anvendelsesområde i henhold til art. 3.

Behandlingsbegrebet i art. 4, nr. 2 er defineret som enhver aktivitet, som personoplysninger gøres til genstand for og art. 4, stk. 1 vedrører de identificerede eller identificerbare fysiske personer [Personally Identifable Information eller PII]. Brug af pseudonymisering, jf. art. 4, nr. 5 og kryptering, jf. bl.a. art. 32, stk. 1, litra a, er stadig personoplysninger, da pseudonymiseringen er reversibel og krypteringen kan dekrypteres igen inden for organisationen eller af andre aktører.[97]

Machine learning bliver ofte omfattet af art. 2, stk. 1's anvendelsesområde, da elektronisk behandling anses som enhver behandling af personoplysninger gennem automatiserede midler, eksempelvis en computer, en mobiltelefon eller en router.[98] Undtagelsen heraf er, hvis alle data i modellen ikke er PII. Machine learning er, som nævnt i afsnit 3.2.1 underopdelt i modeller, som løbende bruger data til at lære af erfaring – både ved indsamlingsfasen, træningsfasen og ved anvendelsesfasen. Der udføres dermed en behandling i art. 4, nr. 2's forstand ved f.eks. at indsamle, registrere, organisere, systematisere, søge og bruge data.

Anvendelsen af art. 4, stk. 1 kan diskuteres ved udvikling og anvendelse af machine learning. Det skyldes, at det ikke altid er givet, at der er tale om en personoplysning, når algoritmen bygges op. I Farmbrellas arbejde[99] med at blive compliant var det omdiskuteret, hvorvidt udviklerens mappings/database arkitektur og datapunkter med unikke ID'er af brugerne (herunder det kodesprog indeholdt i de forskellige datatabeller i algoritmen), var en art. 4-oplysning. På den ene side kunne man ved første øjekast foranlediges til at tro, at oplysningerne (isoleret set i hver datatabel) var anonymiserede. På den anden side pegede en nærmere juridisk vurdering i retning af, at der var tale om pseudonymisering. Det skyldtes det forhold, at oplysningerne (1), som kan trækkes ud af én data-tabel (2), indirekte kan henføre det til en anden datatabel (3), indirekte kan henføres det til en tredje datatabel (4), for til sidst at henføre det til den endelige user og dermed finde ID'et frem inden for virksomheden.

[97] Olsen: *Håndbog i dataansvarlighed* (2020), s. 143
[98] Council of Europe & European Union Agency for Fundamentals Rights: *Handbook on European data protection law* (2018), s. 99
[99] Casestudie om Farmbrella i afsnit 2.2 og bilag 2

Ovennævnte diskussion tyder på, at denne problemstilling i forhold til AI-baserede algoritmer skaber en vis forvirring, da der i virkeligheden kan være tale om "mixed datasets", hvor ikke alle personoplysninger er pseudonymiseret men delvist anonymiserede og dermed til dels ikke omfattet af GDPR.[100] De(n) dataansvarlige skal under alle omstændigheder huske, at hvis machine learning gør det muligt at omdanne anonymiserede data til personoplysninger, anses denne data som PII, jf. FFD regulation betragtning 9, 3. pkt. og dermed omfattet af GDPR. Om ikke andet er sondringen mellem "personal versus non-personal data"[101] en særlig problemstilling, som Farmbrella og andre dataansvarlige bør være opmærksomme på, sådan så de rigtige oplysninger indgår i udarbejdelsen af den generelle (*person*)risikovurdering[102] og *eventuelle* konsekvensanalyse(r)[103]. Herved bemærkes, at en risikoanalyse skal laves *inden* AI-foranstaltninger implementeres (afsnit 5.1.1).

4.2 Hvad er formålet med den machine learning, der skal bygges op?

Art. 5, stk. 1, litra b, 1. led foreskriver, at "*personoplysninger skal indsamles til udtrykkeligt angivne og legitime formål*" jf. betragtning 39, 6. pkt.[104] Det betyder konkret, at det skal være klart, hvorfor personoplysninger bliver indsamlet, og formålet skal være sagligt.[105] Det er altså de(n) dataansvarliges opgave at definere formålet med en vis præcision.[106] Her fastslår Art. 29-gruppen i øvrigt, at formålet skal være så åbenbart og udtrykkeligt, at det er sikret, at alle involverede har den samme utvetydige forståelse af formålene.[107] Eksempelvis opfylder formålsbeskrivelsen "administration" inden for den offentlige forvaltning ikke kravet om udtrykkelighed, da den er for generel eller vag.[108]

Det kan diskuteres, om machine learning og formålsbestemthedskravet fungerer sammen. Først og fremmest gik det hurtigt op for udvikleren fra Farmbrella[109], at denne ved udviklingen af machine learning fik seriøse problemer med at definere formålet i henhold til art. 5, stk. 1, litra b præcist nok, da det var uvist, hvorvidt personoplysninger i modellen senere kunne blive indsamlet til andre formål end først antaget. Yderligere var det svært at vide, hvilken model, der skulle bruges til fastlæggelse

[100] Europa-Parlamentets og Rådet forordning (EU) 2018/1807 af 14. november 2018 om en ramme for fri udveksling af andre data end personoplysninger i Den Europæiske Union, Den Europæiske Unions Tidende (L 303/59)
[101] https://www.carve.dk/2019/09/09/no-personal-data-no-gdpr-right-yes-but/ (27.02.2021)
[102] Olsen: *Håndbog i dataansvarlighed* (2020), s. 405-430
[103] Ibid. s. 431-450
[104] Bet. 1565/2017, s. 93
[105] Blume: *Den nye persondataret* (2018), s. 86
[106] Bet. 1565/2017, s. 83
[107] WP203 (2013), s. 39
[108] Bet. 1565/2017, s. 83
[109] Casestudie om Farmbrella i afsnit 2.2 og bilag 2

af formålet. Dette besværliggøres ovenikøbet yderligere, såfremt der var tale om unsupervised learning herunder særligt clustering[110], hvor algoritmen på egen hånd gradvist finder frem til det optimale mønster i datasættet.

De omtalte argumenter ovenfor taler for, at formålsbestemthedskravet brister på dette område. Her vil indsamlingen af oplysninger været ukendt for udvikleren, hvorfor denne har svært ved på et tidligt stadie at kunne efterleve bestemmelsen ved indsamlingen. Det er først for alvor ved AI-modellens egentlige behandling af personoplysninger, at formålet kan fastsættes. Derfor giver det ikke altid mening for udvikleren at fastsætte formålet helt fra starten – det er op til "robotten"/AI'en.

Det bemærkes, at ordlyden "*udtrykkeligt angivne […] formål*" i art. 5, stk. 1, litra b uddybes i art. 13 og 14. Efter art. 13, stk. 1, litra c skal de(n) dataansvarlige "*på det tidspunkt, hvor personoplysningerne indsamles*" give datasubjekterne oplysninger om de(t) pågældende formål (dvs. med det samme, når indsamlingen påbegynder). Denne tidshorisont kan dog ved machine learning på samme måde føre til udfordringer med den konsekvens, at tiden strækkes unødigt langt ud i strid med bestemmelsen, jf. førnævnte begrundelser. Dette er endnu en understregning af, at AI og formålsbestemthedskravet kommer i konflikt med hinanden og har svært ved at passe sammen. Diskussionen fortsætter i afsnit 4.5.4 for så vidt angår de *yderligere oplysninger* i stk. 2.

Endvidere følger det af art. 5, stk. 1, litra b, at "*personoplysninger må ikke viderebehandles på en måde, der er uforenelig med disse formål*". De må altså gerne anvendes til et andet formål – blot det ikke er uforeneligt, hvilket tillige fremgår af betragtning 50, 1. pkt. Meningen med dette er konkret, at der skal skabes gennemsigtighed i behandlingen, hvilket også følger mere generelt af art. 5, stk. 1, litra a. Man kan spørge, hvis interesser tilgodeses? Det gør datasubjektets. Hvem har interessen i selve behandlingen? Det har de(n) dataansvarlige. Derfor er det centralt, at datasubjektet kan forudse, hvad de indsamlede oplysninger bliver brugt til *(forudberegnelighed)*[111]. AI-modellen er imidlertid skabt til at "tænke selv" på baggrund af store mængder af træningsdata, som er indeholdt i læringsalgoritmen, hvilket kan være i strid med bestemmelsen, hvis den af egen drift behandler personoplysninger til formål, som er uforenelige med de oprindelige formål.[112] Funktionaliteten af en algoritme udvander derfor forudberegneligheden for datasubjektet, hvilket må betegnes som problematisk. Art. 5, stk.

[110] Om clustering (afsnit 3.2.1)
[111] Olsen: *Håndbog i dataansvarlighed* (2020), s. 193
[112] Kunckel m.fl.: *Kunstig intelligens, GDPR og andre juridiske udfordringer* (2018)

1, litra b suppleres af art. 6, stk. 4 til vurderingen af, hvornår der er tale om et nyt formål, der lovligt kan anvendes til viderebehandling. En viderebehandling er lovlig, såfremt den enten er baseret på (1) et *samtykke* til viderebehandlingen, (2) *EU-retten* eller medlemsstaternes *nationale ret* i overensstemmelse med art. 23, stk. 1 eller hvis (3) det nye formål kan anses for ikke at være uforeneligt med det oprindelige formål på baggrund af *ikke-uforenelighedstesten* i art. 6, stk. 4, litra a-e.[113]
Sidst men ikke mindst skal de(n) dataansvarlige være opmærksom(me) på, at art. 5. stk. 1, litra b, 2. led, jf. art. 89, stk. 1 ikke skal anses for at være uforenelig med de oprindelige formål.

Tekniske fagfolk må ofte opfatte art. 5, stk. 1, litra b som en hæmsko for udvikling og anvendelse af AI, da det er med til at indskrænke muligheden for at træne, teste og bygge modellerne op til at finde frem til nye sammenhænge i data. Bekymringerne forstærkes ved udvikling og anvendelse af unsupervised learning, hvor det ikke er til at vide, hvad modellerne vil give svar på (afsnit 3.2.1). Modelvalget kan tilmed nå at ændre sig undervejs i de forskellige faser (afsnit 3.2.2). Alle disse ting indskrænker mulighederne betydeligt. Der kan tillige opstå udfordringer i forbindelse med videresalg af almindelige personoplysninger i algoritmen, idet det kræver et princip om *granularitet*[114]. Når man bruger personoplysningerne til flere forskellige formål, skal man indhente samtykke til hvert enkelt formål. Dette er dog ikke kun forbeholdt samtykke. Det gælder også for de andre behandlingsgrundlag i art. 6, stk. 1, litra a)-f).

4.3 Hvilke krav skal efterleves, når der indhentes personoplysninger til algoritmen?

Faktum er, at AI-systemer ofte kræver store mængder af data. Dataansvarlige skal dog overholde dataminimineringsprincippet, hvis nogle af disse data involverer personoplysninger. Det betyder i praksis, at personoplysningerne i medfør af art. 5, stk. 1, litra c skal være "*tilstrækkelige, relevante og begrænset til, hvad der er nødvendigt i forhold til de formål, hvortil de behandles*". Det er derfor ulovligt at indsamle personoplysninger i algoritmen uden noget reelt formål og/eller behandle irrelevante personoplysninger. Dataminimeringsprincippet indebærer desuden et krav om nødvendighed og proportionalitet i beslutningen om designet i designfasen/indsamlingsfasen.[115] Det betyder, at der kun må behandles personoplysninger, hvis det er: 1) nødvendigt til efterlevelse af de specifikke

[113] Bet. 1565/2017, s. 94
[114] Datatilsynet: *Vejledning - Samtykke* (2021), s. 8*f*
[115] Olsen: *Håndbog i dataansvarlighed* (2020), s. 457

formål, de er indsamlet til *(nødvendighed)* og 2) hvis mængden af personoplysninger står i et rimeligt forhold til formålet, som personoplysningerne er blevet indsamlet til *(proportionalitet)*.[116]

Som nævnt i afsnit 3.2.1, er brændstoffet i AI-modellen den træningsdata, den bliver fodret med. Disse træningsdata er med til at skabe ny viden og optimere modellens ydeevne ud fra store mængder af data. Allerede her kommer man i konflikt med efterlevelsen af art. 5, stk. 1, litra c. Udfordringen ved at efterleve dataminimeringsprincippet har i øvrigt også tæt sammenhæng med de føromtalte udfordringer med at efterleve formålsbestemthedskravet på et tidligt stadie ved indsamlingen. Det bliver uhyre vanskeligt at vurdere hvilke personoplysninger, der findes tilstrækkelige, relevante og begrænset til, hvad der er nødvendigt i forhold til de dataindsamlinger, der finder sted. Årsagen hertil er, at formålet ganske enkelt *ikke er fastslået* endnu (afsnit 4.2). Det kan også være, at *formålet ændrer sig*, i takt med at læringsalgoritmen gradvist bliver "klogere" på baggrund af de erfaringer, den kan trække ud af træningsdataene. Brug af læringskurver til at lave trinvis tests af prædiktionsevnen kan være en løsning til at vurdere, hvornår man har nok data til formålet – og bør også efterprøves[117]. Efter art. 25, stk. 1 bidrager dette til effektiv implementering af dataminimineringsprincippet[118].

4.4 Hvordan beskyttes datasubjektet i forbindelse med automatiske afgørelser?

Beskyttelsen af datasubjektet i forbindelse med automatiske afgørelser skal især ses i lyset af art. 22. Det fremgår af bestemmelsen, at datasubjektet har en rettighed *(individual right)*: "*til ikke at være genstand for en afgørelse, der alene er baseret på automatisk behandling, herunder profilering, som har retsvirkning eller på tilsvarende vis betydeligt påvirker den pågældende*". Der foreligger en række betingelser, der skal være opfyldt, før art. 22, stk. 1 som udgangspunkt træder i kraft.

Den første betingelse er, at afgørelsen er *fuldautomatisk*. Kendte eksempler på fuldautomatiske afgørelser findes i den offentlige (digitale) forvaltning i forhold til årsopgørelserne, dele af SU-området og indsigt i CPR-registret. Hvis der derimod er tale om en kombination af en automatisk afgørelse og et element af menneskelig indgriben, falder man uden for art. 22's anvendelsesområde.[119]
Den anden betingelse er, at afgørelsen får "*retsvirkning*" *eller* "*på tilsvarende vis betydeligt påvirker datasubjektet*". GDPR definerer imidlertid ikke disse to begreber, men ordlyden gør det klart, at

[116] Ibid. s. 211-213
[117] Blume & Motzfeldt: *Databeskyttelse og udvikling af kunstig intelligens* (2020), s. 5
[118] Ibid. og afsnit 5.1.1
[119] Motzfeldt: *Retssikkerheden bør følge med den automatiserede forvaltning* (2018), s. 230

virkningen skal være alvorlig.[120] Dette kan være udelukkelse fra et lån, forsikring eller forhindring i at søge et job.[121] Betragtning 71, 1. pkt. nævner i den sammenhæng et automatisk afslag på en "*onlineansøgning om kredit*" eller "*rekrutteringsprocedurer uden nogen menneskelig indgriben*".
Det kan diskuteres, hvorvidt der foreligger en tredje betingelse. Det antages i den danske bet. 1565/2017, at der gælder en yderligere betingelse om, at der skal foreligge en evaluering af datasubjektets personlige forhold.[122] Denne betingelse eksisterer imidlertid ikke i de retningslinjer Art. 29-gruppen kommer med. Det der ifølge bet. 1565/2017 taler for, at der foreligger en tredje betingelse er, at art. 22 har overskriften: "*Automatiske individuelle afgørelser*", hvilket skal sidestilles med, at afgørelsen skal vedrøre personlige forhold.[123] Det, der til gengæld taler imod en tredje betingelse, er den begrundelse, der følger af betragtning 71, stk. 1's ordlyd "*[…] kan omfatte en foranstaltning, som evaluerer personlige forhold […]*." Her må udtrykket "*kan*" fortolkes i sammenhæng med art. 22's ordlyd "*herunder profilering*"[124], der fastslår, at automatiske afgørelser enten kan foretages med eller uden profilering[125]. Evaluering af personlige forhold "skal" derfor ikke nødvendigvis inddrages som en tredje betingelse i henhold til præambelbetragtningen.

Diskussionen ovenfor viser, at der er betragtninger, der taler for og imod en tredje betingelse. Imidlertid må det lægges til grund, at det er Art. 29-gruppens opfattelse, som de(n) dataansvarlige skal følge, da pligterne skal ses i et mere internationalt perspektiv. Det bemærkes, at der er sket en videreførelse af ordlyden i art. 22, stk. 1, hvorfor bestemmelsen fortsat er brugbar.[126]

Efter Art. 29-Gruppens retningslinjer, kræver art. 22-rettigheden ikke, at datasubjektet gør indsigelse, da udtrykket "*ret*" skal fortolkes som et generelt forbud[127]. Argumentet er, at denne fortolkning styrker datasubjektets *kontrol* over sine personoplysninger, hvorfor datasubjektet beskyttes i højere grad[128]. Af disse grunde anerkendes et generelt forbud også andre steder i den internationale litteratur[129]. Der foreligger imidlertid tre undtagelser i art. 22, stk. 2, litra a-c til det generelle forbud i art.

[120] WP251 (2018), s. 21
[121] Olsen m.fl.: *Eksponeret* (2018), s. 132
[122] Bet. 1565/2017, s. 377
[123] Ibid.
[124] Profilering er defineret i art. 4, nr. 4 og fremgår også af betragtning 71, 2. pkt.
[125] WP251 (2018), s. 8
[126] Bet. 1565/2017, s. 378 og afsnit 2.1
[127] WP251 (2018), s. 19
[128] Ibid. s. 20
[129] Brkan: *Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond* (2019), s. 99

22, stk. 1. Undtagelserne vil ikke blive vurderet særskilt, men fokus vil være på de "*passende foranstaltninger*", der skal gennemføres af de(n) dataansvarlige i medfør af art. 22, stk. 3 til at opfylde art. 22, stk. 2, litra a og c (afsnit 5.1.1.4). Disse foranstaltninger kan tilsvarende benyttes til at fravige art. 9-forbuddet i art. 22, stk. 4 forudsat, at art. 9, stk. 2, litra a eller g bl.a. også finder anvendelse.

Profilering kan som nævnt finde sted uden automatiske afgørelser, men profilering og automatiske afgørelser er ikke nødvendigvis særskilte aktiviteter. En automatisk afgørelse kan i starten vise sig ikke at være baseret på profilering – f.eks. ved at blive pålagt hastighedsbøder fra hastighedskameraer – men senere risikere at være profilering – f.eks. ved at vurdere, om billisten har begået tidligere overtrædelser.[130] Profilering kan i praksis bruges til *at* beslutte hvilke kandidater, der passer bedst til en given stilling. Se f.eks. afsnit 5.1.1.3 om CV-filtreringssystemet til profilering af kvalificerede kandidater. Profilering kan benyttes til *at* beslutte, om lånansøgningen skal accepteres. Se f.eks. afsnit 3.2.1.1 om én banks vurdering af kreditrisiko over for potentielle kunder eller 4.5.3 om en anden banks profilering. Ydermere kan profilering anvendes til *at* beslutte, om der kan tegnes en forsikring eller *at* træffe automatiske kørselsbeslutninger mv.[131] Selvom disse tiltag kan øge effektivitet og mindske omkostninger, kan de samtidig også medføre betydelige risici for datasubjekternes rettigheder og frihedsrettigheder – f.eks. risikoen for, at profileringen fører til diskrimination.[132] Dette kom bl.a. til udtryk ved de(n) mulige diskriminationsform(er) i bankens godkendelse/afvisning af låneansøgninger (afsnit 3.2.1.1). AI-specifikke risici som følge af profilering behøver således ikke kun at være begrænset til art. 22 i dette afsnit eller explainability i afsnit 4.5.4, idet begrebet profilering tillige kan risikere at optræde i øvrige trade-offs såsom fairness (mere herom i afsnit 4.5.2) og accuracy (mere herom i afsnit 4.5.3). Dette afspejles også i betragtning 72, 1. pkt., der bl.a. nævner "*databeskyttelsesprincipper*" i forbindelse med profilering.

4.5 ***Trade-offs***

4.5.1 Tager machine learning højde for ***privacy****?*

Det følger udtrykkeligt af art. 1, stk. 2, at GDPR har til formål at "*beskytte fysiske personers grundlæggende rettigheder og frihedsrettigheder, navnlig deres ret til beskyttelse af personoplysninger*". Beskyttelse af personoplysninger er en grundfæstet rettighed ifølge Chartrets art. 8[133] og art. 16 i

[130] WP251 (2018), s. 8
[131] Olsen m.fl.: *Eksponeret* (2018), s. 131
[132] Ibid.
[133] Den Europæiske Unions Charter om grundlæggende rettigheder, Den Europæiske Unions Tidende (2010/C 83/02)

TEUF[134]. Retten til respekt for privatliv og familieliv fremgår af art. 8 i EMRK[135]. Selvom menneskerettighederne ikke er genstand for en selvstændig analyse, skal de(n) dataansvarlige altid sikre, at deres algoritmer ikke risikerer at være i strid med bestemmelserne, navnlig retten til *privacy*[136].

Alle menneskerettigheder skal altid overholdes. Der er ikke nogen dele af GDPR, der er undtaget fra dette, jf. betragtning 4. Det fremgår endvidere af betragtning 75, 1. pkt. at risiciene for fysiske personers rettigheder og frihedsrettigheder kan opstå som følge af behandling af personoplysninger (herunder ofte machine learning (afsnit 4.1)), som kan føre til *fysisk, materiel eller immateriel skade*, *navnlig forskelsbehandling/diskrimination* mv.[137] Behandlingsprincipperne er ofte dem, der kan føre til ulovlig forskelsbehandling. Når ICO f.eks. nævner "accuracy vs privacy" (afsnit 5.2.1.1) som et trade-off, er det altid en konkret afvejning, der afgør, hvordan man finder den bedste balance mellem henholdsvis rettighederne og de grundlæggende behandlingsprincipper.[138]

*4.5.2 Er machine learning **fair** i forhold til datasubjektet?*

Det fremgår af art. 5, stk. 1, litra a, at "*personoplysninger skal behandles lovligt, rimeligt og på en gennemsigtig måde i forhold til datasubjektet*". Bestemmelsen fastsætter en standard om *god databehandlingsskik,* som skal udfyldes af tilsynsmyndighederne.[139] Når det skal vurderes, hvorvidt machine learning er fair i forhold til datasubjektet, er princippet om rimelighed *(fairness)* relevant. GDPR sondrer mellem to typer af fairness: *informational fairness* og *substantive fairness*.[140]

Informational fairness er forbundet med idéen om transparens, hvoraf betragtning 60, 2. pkt. anfører, at "*de(n) dataansvarlige bør give den registrerede eventuelle yderligere oplysninger, der er nødvendige for at sikre en fair og gennemsigtig behandling under hensyntagen til de specifikke omstændigheder og forhold, som personoplysninger behandles under*". Her må det overvejes, hvorvidt det er nødvendigt at skabe transparens ved at få adgang til et AI-systems træningssæt. Formålet herved vil være at identificere mulige årsager til unfairness som følge af ubalancerede træningsdata eller biased data. Denne nødvendighed kan i øvrigt skærpes, hvis læringsalgoritmen viser sig uigennemsigtig.

---

[134] Traktaten om Den Europæiske Unions Funktionsmåde, Den Europæiske Unions Tidende (2012/C 326/01)
[135] Den Europæiske Menneskerettighedskonvention, vedtaget i 1950
[136] Bilag 3
[137] Derfor tyder det på, at datasubjektet efter art. 82, stk. 1 som udgangspunkt har ret til erstatning for den forvoldte skade, hvis vedkommende bliver genstand for *diskrimination af en algoritme* som følge af en overtrædelse af GDPR.
[138] ICO: *Trade-offs* (2019)
[139] Bet. 1565/2017, s. 92
[140] EU-Parliament: *The impact of the GDPR on AI* (2020), s. 44*f*

Substantive fairness udspringer til gengæld af betragtning 71, 6. pkt., der vedrører tre tilfælde hvor *profilering*, kan være en udfordring i relation til princippet om fairness. Første tilfælde er faktorer, der resulterer i "*unøjagtige personoplysninger*" (mere herom i afsnit 4.5.3 nævnte "Husky versus Wolf-eksperiment" samt "accuracy kontra statistical accuracy"). Det andet tilfælde er *"at risikoen for fejl minimeres*". Her har der været en lignende episode i Storbritannien, hvor en kvindelig læge blev låst ude af damernes omklædningsrum. Et sikkerhedssystem havde fejlagtigt profileret kvinden som en mand, fordi AI-modellen associerede titlen "Dr." med mænd.[141] Tredje tilfælde er profilering, der er nævnt som en risikofaktor for datasubjekternes interesser og rettigheder. Dette kan være "*forskelsbehandling af fysiske personer på baggrund af race eller etnisk oprindelse, politisk, religiøs eller filosofisk overbevisning, fagforeningsmæssigt tilhørsforhold, genetisk status eller helbredstilstand eller seksuel orientering*". Her var en udfordring for Farmbrella bl.a. algoritmens evne til at tænke diversitet ind i løsningen, herunder særligt i forhold til ikke at forskelsbehandle eller diskriminere på baggrund af især køn eller alder, da der som nævnt er tale om en høj aldersgruppe samtidig med, at de fleste direktører inden for landbrugsbranchen er mænd[142]. Man kan imidlertid rejse spørgsmålet, om Farmbrella har lov til at ignorere betragtning 71 fuldt ud, idet variable som *køn* og *alder* ikke indgår heri? Det følger af EU charterets art. 21, stk. 1, at enhver forskelsbehandling på baggrund af køn og alder er forbudt. Der er derfor risiko for, at disse variable kan føre til diskrimination, selvom det ikke udtrykkeligt følger af betragtning 71. På baggrund heraf er det derfor anbefalingsværdigt for Farmbrella at forebygge dette aspekt i AI-løsningen, således at unfair forskelsbehandling undgås. Måden, udvikleren skal gøre dette på, er ikke ligetil, da det vil kræve et trade-off. Denne diskussion uddybes nærmere i afsnit 5.2.1.2.

*4.5.3 Besidder machine learning **accuracy**?*

Art. 5, stk. 1, litra d fordrer, at "*personoplysninger skal være korrekte og om nødvendigt ajourførte*" og kræver, at de(n) dataansvarlige tager "*ethvert rimeligt skridt for at sikre, at personoplysninger, der er urigtige i forhold til de formål, hvortil de behandles, straks slettes eller berigtiges.*" Bestemmelsen indebærer et princip om *rigtighed (accuracy)*. Det betyder, at AI-systemer skal være af den

[141] ICO: *Big data, artificial intelligence, machine learning and data protection* (2017), s. 20
[142] Casestudie om Farmbrella i afsnit 2.2 og bilag 2

fornødne datakvalitet, up-to-date og hverken være ukorrekte/vildledende[143]/urigtige/unøjagtige/upræcise eller indeholde biases, jf. eksemplet med en bank[144] og Husky versus Wolf-eksperimentet[145] (herefter Husky vs Wolf) nedenfor.

En bank ønsker at bruge machine learning til at profilere en række kunder, der søger om banklån. Når banken skal bestemme, hvordan AI-systemet skal trænes i udviklings- og træningsfasen, kræves det, at de(n) dataansvarlige har *korrekte* data til at modtage præcise outputs fra modellen. Her skal de indsamlede kundedata være mest muligt *nøjagtige* og disse personoplysninger skal kun indsamles fra datakilder[146] med *korrekte og up-to-date* oplysninger. Sideløbende er det helt afgørende, at banken husker at teste, hvorvidt AI-modellen samtidig sikrer den fornødne privacy og ikke giver diskriminerende outcomes.[147] Mere herom i afsnit 5.2.1.2.1. Når AI-systemet herefter er trænet fuldt op og operativ i anvendelsesfasen, kan banken bruge resultaterne som en del af den lånevurdering, der bliver foretaget af datasubjekterne. Banken skal også løbende efterse de outputs, som AI-modellen kommer med og om nødvendigt være ajourførte/up-to-date i henhold til art. 5, stk. 1, litra d, 1. led.

Et væsentligt problem med både princippet om accuracy og fairness er, at træningsdata som indeholder *biases* ofte bliver ved med at være i modellen. Disse biases kan typisk opstå i den situation, hvor supervised learning indhenter store mængder af træningsdata, som bliver labelled forkert. Visse dele af træningsdatasættet kan også fejlkategoriseres. Hvis disse ting sker, vil kvaliteten af træningsdatasættet aftage. Når datakvaliteten af træningsdata forringes, forringes datakvaliteten af det endelige output også, når modellen rammer virkeligeden i anvendelsesfasen. Når outputtet forringes, vil det også anses for *unfair* i strid med art. 5, stk. 1, litra a. Dette kan eksemplificeres med Husky vs Wolf, hvor deep learning[148] (som er en del af neural networks) blev anvendt til at træne et system til at kende forskel på billeder af sibiriske huskier og ulve. Hver gang deep learning blev brugt til at identificere billeder af ulve, opdagede forskerne, at systemet henlagde opmærksomheden på tilstedeværelsen af sne (eller mangel af samme) på billedet. Systemet "lærte", at hvis der skulle være tale om en ulv, var det afhængig af, om der var sne på billedet. Derfor støttede det på en falsk korrelation. Systemets evne til at kende forskel på dyrearterne var mislykket. Eksperimentet illustrerer, hvordan

[143] Blume: *Den nye persondataret* (2018), s. 92*f*
[144] EDPB: *Guidelines 4/2019 on art. 25 DPbDD* (2019), s. 21*f*
[145] FPF: *The Privacy Expert's Guide To AI and Machine Learning* (2018), s. 14
[146] Datakilder kan f.eks. være fra datasubjektet selv, en myndighed, en anden virksomhed, sociale medier eller fra egne IT-systemer. Olsen: *Håndbog i dataansvarlighed* (2020), s. 149
[147] EDPB: *Guidelines 4/2019 on art. 25 DPbDD* (2019), s. 22
[148] Bilag 3

et biased datasæt kan føre til unøjagtige oplysninger, som i den sidste kan skade datasubjekterne. Ganske vist er eksperimentet møntet på dyr, men det kunne i princippet have været et stort antal datasubjekter. Derfor bør man – i øvrigt tidligst muligt efter art. 25, stk. 1 – sikre, at træningsdata ikke er labelled forkert med henblik på imødegå ringe datakvalitetsoutput eller inaccuracy.[149]

Datasubjekterne har i øvrigt altid mulighed for at gøre brug af *retten til berigtigelse*, hvis inaccuracy opstår alligevel. Et eksempel herpå er, hvis en AI-løsning inden for markedsføring forudsagde, at en person var forælder (selvom personen rent faktisk ikke havde børn).[150] Her ville outputtet være inaccurate i forhold til de formål, hvortil de behandles. I dette tilfælde vil personen i medfør af art. 16 have ret til at bede de(n) dataansvarlige om at rette op på disse AI-outputs.

Det kan imidlertid diskuteres, om art. 16 har en stor nok gennemslagskraft til at beskytte datasubjektet. Det problematiske er nemlig, at AI-outputs i mange tilfælde vil generere personoplysninger, hvor der slet ikke er en faktuel oplysning at rette op på. F.eks. kunne et AI-system forudsige, at personen har højere sandsynlighed for at blive forælder i løbet af de næste 3 år.[151] Denne forudsigelse kan hverken være accurate eller inaccurate i forhold til en faktuel oplysning. Noget tyder altså på, at art. 16 har problemer med at beskytte datasubjektet i dette tilfælde. Det betyder samtidig også, at AI-systemers output i mange tilfælde ikke har til hensigt til at blive anset som faktuelle oplysninger om datasubjektet. I stedet skal de repræsentere et statistisk informeret gæt om noget, der måske kan siges at være korrekt om datasubjektet – enten nu eller i fremtiden. Dette fænomen betegnes undertiden som *statistical accuracy*[152]. Her skal udvikleren af en AI-løsning sikre, at man kategoriserer personoplysninger som et statistisk informeret gæt, og *ikke misfortolker* dem som faktuelle oplysninger.

Det bemærkes, at et AI-system ikke nødvendigvis behøver at være 100% statistically accurate for at være i overensstemmelse med art. 5, stk. 1, litra d. Dog skal det siges, at en forbedret stastistical accuracy af et AI-systems outputs vil have positiv indvirkning på fairness i art. 5, stk. 1, litra a. Årsagen hertil er, – som nævnt i afsnit 4.5.2 og 4.2 – at der skal skabes den fornødne transparens og forudberegnelighed. Det skal være noget som datasubjektet skal kunne se ind i – og med rimelighed skal kunne forvente. En for lav statistical accuracy er ikke forventeligt.

[149] Blume & Motzfeldt: *Databeskyttelse og udvikling af kunstig intelligens* (2020), s. 5*f*
[150] ICO: *Accuracy of AI system outputs and performance measures* (2019)
[151] Ibid.
[152] ICO: *Guidance on the AI auditing framework* (2020), s. 46

*4.5.4 Er machine learning **explainable** i forhold til datasubjektet?*

Indledningsvist kan der rejses spørgsmål om, hvorvidt datasubjektet overhovedet har en ret til at få en forklaring på en automatisk afgørelse. Udover art. 13, stk. 2, litra f *bør* de(n) dataansvarlige også efter art. 14, stk. 2, litra g give datasubjektet meningsfulde oplysninger om forekomsten af en automatisk afgørelse efter art. 22, (da vurderingen *uomtvisteligt* påhviler de(n) dataansvarlige). Dette indebærer evnen til at forklare *logikken*, *betydningen* og de *forventede konsekvenser* af den automatiske afgørelse.[153] En ret til at få en forklaring på en automatisk afgørelse fremgår imidlertid ikke udtrykkeligt af art. 22, art. 13 eller art. 14. Der kan dog være brugbar viden at hente i betragtning 71, 4. pkt., hvoraf kan udledes, at datasubjektet har "*retten til […] at få en forklaring på den afgørelse, der er truffet*". På baggrund heraf forudsættes det, at denne ret er gældende.[154] Explainability er således en målestok for de(n) dataansvarliges evne til at forklare brugen af AI til datasubjekterne.[155]

Endvidere kan det diskuteres, hvornår de meningsfulde informationer skal tilvejebringes for datasubjektet. Blume og Motzfeldt fastslår uden nærmere begrundelse, at det med hjemmel i art. 13 og 14 skal ske ved indsamlingen af personoplysninger i indsamlingsfasen.[156] Dette argument lader til at være bygget på ordlyden i art. 13, stk. 2, som kræver, at informationen er tilvejebragt "*på det tidspunkt, hvor personoplysningerne indsamles*". Imidlertid kan en lignende vending ikke findes i art. 14, stk. 2, hvilket gør forfatternes argument mangelfuld i tilfælde af, at data ikke er indsamlet hos datasubjektet selv. Yderligere refererer art. 13, stk. 2, litra f og art. 14, stk. 2, litra g til "*forekomsten af en automatisk afgørelse*", hvilket tyder på, at afgørelsen *allerede* finder sted. Yderligere specificerer art. 15, stk. 1 heller ikke tidspunktet for, hvornår datasubjektet kan få indsigt i de tilsvarende oplysninger, der følger af litra f. På baggrund heraf konkluderes, at indsamlingen af informationerne ikke kræves tilvejebragt i indsamlingsfasen. I praksis understøttes konklusionen tillige på, at oplysningspligten er yderst svær at efterleve i indsamlingsfasen, da formålet sjældent er fastsat endnu (afsnit 4.2). Det samme må siges om tilstrækkeligheden, relevansen og nødvendigheden (afsnit 4.3). Der kan således udvises kritik af forfatternes udsagn på baggrund af GDPR og praksis.

---

[153] Rettighederne skal i øvrigt ses i sammenhæng med kravet om *transparens* i art. 5, stk. 1, litra a. Transparens er forudsætningen for, at datasubjekterne kan udøvede deres rettigheder. Blume & Motzfeldt: *Databeskyttelse og udvikling af kunstig intelligens* (2020), s. 6

[154] Dette er til trods for, at præambelbetragtninger *ikke* er juridisk bindende. Om ikke andet benyttes de hyppigt af EU-Domstolen. Brkan: *Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond* (2019), s. 115

[155] ICO: *Guidance on the AI auditing framework* (2020), s. 7

[156] Blume & Motzfeldt: *Databeskyttelse og udvikling af kunstig intelligens* (2020), s. 6

Denne konklusion er sket på baggrund af en almindelig ordlydsfortolkning[157] og derfor inden for rammerne af GDPR. Der er derfor ikke tale om retspolitik. Når det er sagt, anbefales det imidlertid, at førstnævnte ordlyd i art. 13, stk. 2 *fjernes*, da den på baggrund af ovennævnte begrundelser ikke findes hensigtsmæssig i en AI-kontekst. Her vil der til gengæld være tale om retspolitik, da lovgivningen bør laves om.

Når der skal træffes afgørelser ved udvikling og anvendelse af machine learning, er overholdelse af ovenstående bestemmelser ikke uden udfordringer. Det skyldes, at beregningerne kan være så komplekse, at de opleves som skjulte (altså den føromtalte "black-box"-problematik). Denne problematik har varierende betydning, alt efter hvor høj en kompleksitet AI-modellen har. *Decision trees*[158] [159] er ofte enkle at forklare (og dermed har en relativ høj explainability) til forskel fra *neural networks*[160], som ofte er vanskelige at forklare (og dermed har en forholdsvis lav explainability). Der er således forskel på kompleksiteten eller explainability, om man vil. Dette gælder også, selvom man bruger de samme træningsdata.[161] Der kan dermed tegnes et billede af, at den bagvedliggende logik er svær at forklare for de(n) dataansvarlige, hvilket har en stærk sammenhæng med kompleksiteten af AI og de ovenfor skitserede black-box-udfordringer. De praktiske problemer heraf anerkendes også af Edwards og Veale inden for den internationale litteratur. Kompleksiteten skal dog ikke ifølge dem være et argument for at undlade at forklare forekomsten af den automatiske afgørelse.[162] [XAI]-foranstaltninger (afsnit 5.1.1.4) kan derfor være brugbare løsninger. Modsat vil betydningen og de forventede konsekvenser af en afgørelse ofte afhænge af, hvor stor et kendskab udvikleren har til den kontekst, AI-systemet anvendes i. Disse vil formentlig være mindre komplekse, da udvikleren typisk vil have et indgående kendskab herfor.

[157] Afsnit 2.1
[158] Bilag 3
[159] Decisions trees er ofte et simpelt modelvalg. Såfremt datamængden er overkommelig, vil det være muligt at bevæge sig gradvist opad træet for at se på hvilke kriterier, der ligger til grund for outputtet. En stor udfordring er imidlertid, at en øget indsamling af data i indsamlingsfasen alligevel kan gøre det vanskeligt for et menneske at se ind i. Norsk datatilsyn: *Kunstig intelligens og personvern* (2018), s. 12. Dette viser, at decision trees alligevel kan risikere at være en black-box og være uigennemsigtige. Selv et simpelt modelvalg afhænger af kompleksiteten.
[160] Neural networks forklares i afsnit 3.2.1 og de udfordringer Google Translate i øvrigt oplevede.
[161] Norsk datatilsyn: *Kunstig intelligens og personvern* (2018), s. 12*f*
[162] Brkan: *Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond* (2019), s. 110

5 Hvilke slags løsninger og dokumentation bør implementeres ved udvikling og anvendelse af machine learning?

5.1 Princippet om ansvarlighed (accountability) og den risikobaserede tilgang til databeskyttelse

Det fremgår udtrykkeligt af art. 5, stk. 2, at "*den dataansvarlige er ansvarlig for og skal kunne påvise, at stk. 1 overholdes*." Principperne i art. 5, stk. 1 kan betragtes som art "grundlov" og det skal være synligt, at de(n) dataansvarlige efterlever princippet om ansvarlighed *(accountability)*.[163] Denne "synlighed" er – ved gennemgangen af de i afsnit 4 nævnte behandlingsprincipper – imidlertid blevet sløret ganske betragteligt ved udvikling og anvendelse af machine learning. De tekniske udfordringer ved AI-systemer gør det svært at efterleve art, 5, stk. 2's ordlyd. Her er et afgørende element efter art. 24, stk. 1/art. 5, stk. 2 ikke kun at *sikre* men også at *påvise,* hvordan man løser udfordringerne.[164] Her er det påkrævet, at de(n) dataansvarlige implementerer tekniske og organisatoriske foranstaltninger (**afsnit 5.1.1**) til at løse eller mitigere de risici, der måtte forekomme.[165] Yderligere bliver man nødt til – afhængigt af modelvalg (afsnit 3.2.2), design, træning, opbygning, testning og ibrugtagning[166] – at identificere, løse og finde en *passende balance* mellem trade-offs (**afsnit 5.2 og 5.3**).[167]

5.1.1 Har de(n) dataansvarlige "passende" tekniske og organisatoriske foranstaltninger til at løse udfordringerne?

Databeskyttelse gennem *design* i medfør af art. 25, stk. 1[168] indebærer, at de(n) dataansvarlige har pligt til at gennemføre passende tekniske og organisatoriske foranstaltninger, der er designet med henblik på bl.a. "*effektiv implementering*" af databeskyttelsesprincipperne i art. 5, stk. 1. Derfor vil der i det følgende blive foretaget en gennemgang af de tekniske AI-foranstaltninger (**afsnit 5.1.1.1 – 5.1.1.3)**, som findes relevante til at reducere de AI-specifikke risici, som dataminimering, fairness, accuracy har givet anledning til for datasubjekterne. Det er imidlertid ikke et krav, at de(n) dataansvarlige skal efterkomme kravene alene ved implementering af rent tekniske foranstaltninger – også

[163] Blume: *Den nye persondataret* (2018), s. 85
[164] ICO: *Guidance on the AI auditing framework* (2020), s. 13
[165] AEPD: *Audit Requirements for Personal Data Processing Activities involving AI* (2021), s. 10
[166] Dvs. i indsamlingsfasen, udviklings- og træningsfasen samt anvendelsesfasen i AI-livscyklussen (afsnit 3.2.3)
[167] ICO: *Guidance on the AI auditing framework* (2020), s. 12
[168] Databeskyttelse gennem *standardindstillinger* fremgår af art. 25, stk. 2, men vil ikke blive behandlet i afhandlingen.

organisatoriske foranstaltninger kan være relevante at inddrage.[169] Det primære fokus i afhandlingen vil dog være på de tekniske foranstaltninger. Det følger endvidere art. 25, stk. 1, in fine, at der skal ske "*integrering af de fornødne garantier i behandlingen for at opfylde kravene i denne forordning og beskytte de registreredes rettigheder*". Her vil foranstaltninger om explainable AI (herefter [XAI]-foranstaltninger (**afsnit 5.1.1.4**)) blive inddraget, sådan så de(n) dataansvarlige også har værktøjer til at opfylde ordlyden heri.

De passende foranstaltninger, der er nævnt ovenfor, skal ses som et udtryk for en *risikobaseret tilgang* til databeskyttelse. Det betyder, at man baserer sine relevante tiltag på de risici, der kan identificeres.[170] Efter det danske datatilsyns vejledning, skal den brede formulering af "*databeskyttelse gennem design*" fortolkes som en *helhedsorienteret tilgang*[171] til databeskyttelse. Et af argumenterne er, at databeskyttelse bliver en integreret del af organisationen. Herved sikres en *proaktiv tilgang* til databeskyttelse, hvorved identificering og løsning af udfordringerne kan ske på et tidligt tidspunkt.[172] Der er derfor *bred enighed* om, at databeskyttelse gennem design udgør et væsentligt redskab til at reducere de databeskyttelsesrisici, der navnlig kan krænke datasubjekternes rettigheder og frihedsrettigheder ved udvikling og anvendelse af machine learning.[173] [174] [175]

Art. 25 angiver i øvrigt kun overordnede retningslinjer for hvilke tiltag, der skal være tale om, hvorfor der overlades et stort råderum til de(n) dataansvarlige til selv at vælge.[176] Endvidere må det bemærkes, at foranstaltningerne skal gennemføres både på tidspunktet for fastlæggelse af midlerne til behandling og fra første dag behandlingen påbegynder.[177]

*5.1.1.1 Hvilke **dataminimeringsteknikker** skal til for at sikre **dataminimering**?*

Der findes forskellige teknikker til at minimere personoplysninger i både indsamlingsfasen, i udviklings- og træningsfasen samt i *inferensfasen*[178] (dvs. anvendelsesfasen), hvor AI-systemet tages i brug. Anbefalingen fra ICO er, at dataminimeringen efter bedste evne bliver gennemført tidligst muligt i

[169] Datatilsynet: *Behandlingssikkerhed* (2018), s. 26
[170] Bet. 1565/2017, s. 417
[171] Datatilsynet: *Behandlingssikkerhed* (2018), s. 24
[172] Ibid. s. 27
[173] EDPB: *Guidelines 4/2019 on art. 25 DPbDD* (2019), s. 6 og 25*f*
[174] Norsk datatilsyn: *Kunstig intelligens og personvern* (2018), s. 24
[175] EU-Parliament: *The impact of the GDPR on AI* (2020), s. 66*f*
[176] Bet. 1565/2017, s. 417
[177] Ibid. s. 416
[178] Bilag 3

indsamlingsfasen for at undgå skabelsen af store datasæt indeholdende udelukkende PII. De dataminimeringsteknikker, de(n) dataansvarlige anbefales at anvende i udviklings- og træningsfasen (men hvis muligt i indsamlingsfasen), er *future selection methods* og *privacy-preserving method*s.[179]

Selve anvendelsen af de informationer, læringsalgoritmen har fra træningsdataene, indebærer et datasæt, der indeholder et sæt features for hvert datasubjekt. Disse bruges til at generere forudsigelsen eller klassificeringen, men de inkluderede features i et datasæt er dog ikke nødvendigvis relevante for opgaven. Eksempelvis vil ikke alle økonomiske og demografiske features være nyttige til at forudsige en kreditrisiko. I den forbindelse kan teknikere såsom data scientists med fordel anvende nogle forskellige standard features selection methods[180] til at vælge de features, som lige netop vil være nyttige at inkludere i en AI-model. Feature selection methods vil derfor være med til at sikre overholdelse af dataminimeringsprincippet (afsnit 4.3).

Privacy-preserving methods kan tillige anvendes til at opfylde art. 5, stk. 1, litra c. Det kan i praksis gøres ved at *ændre* på træningsdataene med det formål at reducere det omfang, hvori de kan spores tilbage til bestemte individer (samtidig med at de bevarer deres anvendelighed med henblik på at træne velfungerende modeller). Herved ændres også værdierne af de datapunkter, der hidrører fra datasubjekter på en random måde. Dette har nogle store fordele. *For det første* er det med til at beskytte/bevare nogle af de statistiske egenskaber ved disse features. *Dernæst* er færre random data med til at gøre identificerbarheden mere gennemsigtig – også for Farmbrella. Udvikleren kan nu bedre skille de pseudonymiserede og anonymiserede oplysninger ad i det blandede datasæt i vurderingen af, om der reelt er tale om en art. 4-oplysning (afsnit 4.1). Sondringen mellem personal versus non-personal data bliver altså mere tydelig og transparent at se ind i. Privacy-preserving methods er dermed et stærkt dataminimeringsværktøj til både at sikre dataminering og reducere genidentifikation.

*5.1.1.2 Hvilke **mitigerende foranstaltninger** skal til for at sikre **fairness**?*

Det fremgår direkte af betragtning 71, andensidste pkt., at de(n) dataansvarlige bør "*gennemføre tekniske* og *organisatoriske foranstaltninger […] til at hindre forskelsbehandling/diskrimination af fysiske personer.*" Spørgsmålet er, hvilke foranstaltninger det skal være? Ordlyden i betragtningen giver som bekendt ikke noget entydigt svar, men en af hovedfigurerne bag ICO's AI-vejledninger

[179] ICO: *Data minimisation and privacy-preserving techniques in AI systems* (2019)
[180] McCombe: *Intro to future selection methods for Data Science* (2019)

Reuben Binns[181] foreslår forskellige tekniske foranstaltninger til at mitigere diskriminationsrisici i AI-modeller (*mitigerende foranstaltninger*).[182]

Den *første* mitigerende foranstaltning til at sikre fairness – eller mere præcist *informational fairness* (afsnit 4.5.2) – kan være at *tilføje* eller *fjerne* data om en under- eller overrepræsenterede gruppe. Eksempelvis var der som nævnt i afsnit 3.2.1.1 tale om diskrimination over for en kvindelige population, der var underrepræsenteret i træningsdataene. Her kan man enten tilføje flere data om kvinders låneansøgninger eller fjerne data om mænds. Dette er en god måde at afbalancere forholdene på – og samtidig styrke informational fairness i art. 5, stk. 1, litra a. En *anden* mitigerende foranstaltning er at *træne separate modeller* (f.eks. én for mænd og en anden for kvinder) og designe dem til at performe så godt som muligt på hver gruppe. En *tredje* mitigerende foranstaltning er, at de(n) dataansvarlige kunne *ændre* på dataene. Det kan gøres i de situationer, hvor træningsdata afspejler en tidligere diskrimination, som uddybet i afsnit 3.2.1.1., hvilket kan være med til at mindske sandsynligheden for reproduktion af de samme diskriminationsmønstre. Hvis man skal konkretisere yderligere, kan ændring af data gøres ved at fjerne *visse* "examples" fra træningsdata, som man forventer vil være et resultat af diskrimination. De datasæt der kan indgå i disse examples kan være noget så simpelt som labelled data. Supervised learning er som bekendt trænet på et datasæt indeholdende labelled data (afsnit 3.2.1). Ved at ændre på disse slags data (hvor man som nævnt har en mistanke om resultat af diskrimination) i træningsdatasættet, undgår man i vidt omfang, at træningsdata afspejler tidligere diskrimination. En *fjerde* mitigerende foranstaltning er at ændre på læringsprocessen, dvs. *ændre på selve måden* modellen lærer fra de data, udvikleren "fodrer" algoritmen med. Ergo kan det ændre på selve måden AI-robotten tænker på (dvs. dens "hjerne"), hvorfor man kan mindske sandsynligheden for, at den af egen drift behandler personoplysninger, der kan være i strid med art. 5, stk. 1, litra b (afsnit 4.2). Endelig er en *femte* mitigerende foranstaltning at *tilpasse modellen* efter man har trænet modellen op under træningsfasen i udviklings- og træningsfasen.

Reuben Binns understreger imidlertid vigtigheden af, at problemer i relation til diskrimination er et *bredt* fænomen, og at indførelse af mitigerende foranstaltninger ikke uden videre kan "fikse" alting ved et AI-system. Derfor er det fejlagtigt at tro, at man automatisk bliver compliant med eksempelvis ikke-diskriminationslovgivningen ved at indføre disse slags foranstaltninger. Når det er sagt, kan de

[181] https://www.cs.ox.ac.uk/people/reuben.binns/ (01.04.2021)
[182] ICO: *Ensuring lawfulness, fairness, and transparency in AI systems webinar* (2020)

tekniske mitigerende foranstaltninger alligevel være særdeles brugbare, afhængig af den specifikke kontekst for de(n) dataansvarlige.[183]

*5.1.1.3 Hvilke **accuracy measures** skal til for at sikre **accuracy**?*

GDPR nævner statistical accuracy i sammenhæng med profilering og automatiske afgørelser i betragtning 71, 6. pkt. Her bør de(n) dataansvarlige "*anvende passende matematiske og statistiske procedurer til profileringen*" af datasubjekterne som en del af de tekniske foranstaltninger. Her skal det understreges, at det ikke er tilsynsmyndighedernes opgave at bestemme selve måden, AI-systemer skal bygges på. Deres opgave er til gengæld at forstå, hvor *accurate* (afsnit 4.5.3) deres AI-systemer er, og hvilken *påvirkning* det har på datasubjekterne.[184] De(n) dataansvarlige bør derfor indføre passende *statistical accuracy measures*[185], mens de bygger og implementerer deres AI-systemer i henholdsvis udviklings- og træningsfasen samt anvendelsesfasen. Disse statistical accuracy measures bør afspejle balancen mellem to forskellige typer af *errors*, som er: *false positive* ["Type I"-error] og *false negative* ["Type II"-error]. Et eksempel på førstnævnte Type I-error er e-mails, som ukorrekt bliver klassificeret som spam, selvom de rent faktisk er originale mails (altså ikke-spam). Sidstnævnte Type II-error kan til gengæld være e-mails klassificeret som originale mails, selvom de er spam.

Ovennævnte errors kan gøres endnu mere virkelighedsnært ved at forestille sig et CV-filtreringssystem, hvis funktion er at vælge kvalificerede kandidater til en jobsamtale. Hvis systemet producerer et *false positive,* vil en *ukvalificeret* kandidat blive inviteret til en jobsamtale, hvilket bliver omkostningstungt for medarbejderen (f.eks. HR-chefen) og jobansøgeren, fordi det koster tid. Den tid kunne have været brugt bedre på en anden kandidat, der var mere egnet til jobbet. Modsat hvis CV-filtreringssystemet producerer et *false negative,* vil en *kvalificeret* kandidat på den ene side misse muligheden for at få jobbet og de(n) dataansvarlige vil på den anden side misse en god kandidat til jobbet. Eksemplet herfor viser, at der meget vel kan være vigtige forskelle mellem konsekvenserne af false positives og false negatives, når det kommer til de involverede datasubjekter.[186]

For at kunne anvende en statistical accuracy measure korrekt på en gruppe af datasubjekter, skal man vide, at det *ikke* er en *statisk målestok* – selvom den er målt på statiske test data. Selvom systemet er

[183] Ibid.
[184] ICO: *Accuracy of AI system outputs and performance measures* (2019)
[185] Ibid.
[186] ICO: *Guidance on the AI auditing framework* (2020), s. 49

statistical accurate om eksisterende populationsdata, kan populationen sagtens ændre sig løbende i forhold til forskellige typer af karakteristika som f.eks. adfærdsændringer mv. Derfor kan AI-systemet sandsynligvis blive *mindre* statistical accurate over tid, (hvilket i øvrigt ikke er særlig gunstigt for datasubjektet, jf. afsnit 4.5.3). Dette fænomen er i machine learning kendt som "*concept//model drift*"[187]. Det er således afgørende, at de(n) dataansvarlige regelmæssigt vurderer driften og genoptræningen af modellen. Hvis en model evaluerer CV'er som en del af en rekrutteringsøvelse, – og evaluering af kandidaternes kvalifikationer ændrer sig over tid hvert andet år – bør man kunne forudse behovet for at genoptræne modellens data på regelmæssig basis.

*5.1.1.4 Hvilke Expl**AI**nable **[XAI]-foranstaltninger** skal til for at sikre **explainability**?*

En [XAI]-foranstaltning er en metode til at reducere kompleksiteten[188] af AI-systemer og dermed til at efterleve kravet om explainability.[189] Der er forskellige måder at reducere denne kompleksitet og uigennemsigtighed på. Det britiske datatilsyn har på baggrund heraf foreslået seks [XAI]-foranstaltninger, der kan være relevante, når det kommer til at forklare AI både før og efter, at der træffes en afgørelse.[190] I afsnit 5.2 indgår fairness som et *ekstra* opmærksomhedspunkt i afvejningen mellem "accuracy vs explainability". På den baggrund er én af ICO's foranstaltninger særlig vigtig at fremhæve for at sikre opfyldelse af explainability. Denne [XAI]-foranstaltning er *fairness explanation*.

Fairness explanation handler om at hjælpe med at forklare datasubjekterne de ting, udvikleren gjorde (og fortsat vil gøre) for at sikre, at automatiske afgørelser er unbiased og fair. Ved at forklare hvordan man hindrer biases og diskrimination, får datasubjekterne indblik i, hvordan de bliver behandlet mere lige. Dette øger ikke kun fairness og explainability – men også datasubjekternes tillid.[191]

Flere af [XAI]-foranstaltningerne kan på hver sin måde indplaceres i andre bestemmelser inden for GDPR. I forlængelse af gennemsigtighedsprincippet i art. 5, stk. 1, litra a, påhviler det de(n) dataansvarlige efter art. 12, stk. 1, 1. pkt. at træffe *passende foranstaltninger* til at give enhver information i art. 13, stk. 2, litra f og art. 14, stk. 2, litra g[192] på en letforståelig måde over for datasubjektet. Disse art. 12-

[187] Ibid.
[188] Denne kompleksitet blev gennemgået i afsnit 4.5.4 men også afsnit 3.2.1
[189] Bestemmelserne indeholdt i begrebet explainability blev gennemgået i afsnit 4.5.4
[190] ICO: *Explaining decisions made with AI – Part 1* (2020), s. 20
[191] Ibid. s. 27*f*
[192] Afsnit 4.5.4

foranstaltninger kan søges opfyldt ved at implementere én eller flere [XAI]-foranstaltninger. Hvis det er relevant, kan de alternativt søges indpasset i de foranstaltninger, der følger af art. 22, stk. 3 og 4[193].

Ud fra ovennævnte betragtninger, kan man konstatere, at transparens er en forudsætning for, at datasubjektet kan gøre sine rettigheder gældende.[194] Explainability er til gengæld en forudsætning for accountability.[195] [XAI]-foranstaltninger må derfor siges at være vigtige tiltag for at overholde GDPR ved udvikling og anvendelse af AI, da de holder hånden under datasubjektet på tværs af forordningen.

5.1.2 Opsummering

Bestemmelserne i afsnit 4 er nu blevet gennemgået. Det drejer sig om: *behandlingsprincipperne* (art. 5, stk. 1, litra a, b, c og d), *automatiske afgørelser herunder profilering* (art. 22 og art. 4, nr. 4) og *oplysningspligten og indsigtsretten* (art. 13, stk. 2, litra f/art. 14, stk. 2, litra g/art. 15, stk. 1, litra h). Bestemmelserne i afsnit 5 er også blevet gennemgået, for så vidt angår *den røde tråd om ansvarlighed* (art. 5, stk. 2)[196] og *den risikobaserede tilgang til databeskyttelse* (art. 25)[197] med en *dokumentation* for alle de vurderinger, som den risikobaserede tilgang er udtryk for (art. 24).[198]

Alle udvalgte bestemmelser, som nævnt ovenfor, har vist sig at have stor juridisk betydning i forhold til at overholde kravene i GDPR ved udvikling og anvendelse af machine learning. Først og fremmest har teknologien givet anledning til store udfordringer i indsamlingsfasen. Herved viste afsnit 4.1, hvor uoverskueligt det er at se ind i GDPR's anvendelsesområde. Dernæst viste afsnit 4.2 og 4.3, hvor udfordrende det er at overholde helt grundlæggende behandlingsprincipper i art. 5, stk. 1, litra b og c. Selvom disse afsnit også kan gå ud over privacy, har analyserne vist, at de negative effekter heraf ikke er nær så katastrofale som dem, der blev skitseret i afsnit 4.5.2, 4.5.3 og 4.5.4. De tre sidstnævnte afsnit viste nemlig, hvor slemt unfairness, inaccuracy og manglende explainability kan gå ud over datasubjekterne. Det er derfor, at art. 5, stk. 1, litra a, og d samt gennemsigtighedsbestemmelserne fra afsnit 4.5.4, er de bestemmelser, som indgår i de trade-offs, der følger af afsnit 4.5. Det er imidlertid *ikke nok* at efterleve kravene om fairness, accuracy og explainability – disse skal *også* holdes op mod privacy i afsnit 4.5.1. Kravene kan sågar risikere at komme i *indbyrdes konflikt* med hinanden (f.eks. "accuracy vs explainability" i afsnit 5.2.1.3), hvilket kan problematisere tingene

[193] Afsnit 4.4
[194] Blume & Motzfeldt: *Databeskyttelse og udvikling af kunstig intelligens* (2020), s. 6
[195] Ibid. s. 8
[196] Korfits & Lotterup: *Databeskyttelsesforordningen og databeskyttelsesloven med kommentarer* (2020), s. 341
[197] Datatilsynet: *Behandlingssikkerhed* (2018), s. 26
[198] Ibid.

yderligere. De(n) dataansvarlige er imidlertid blevet præsenteret for en række tekniske foranstaltninger i afsnit 5.1.1.2, 5.1.1.3 og 5.1.1.4. med *afsæt i en risikobaseret tilgang* til udvikling og anvendelse af AI. Disse kan *både* bruges som hjælp til at løse mange af de udfordringer, som afsnit 4.5.2, 4.5.3, 4.5.4 (og 4.5.1) giver anledning til og bidrage til at foretage en *bedre* vurdering af trade-offs.

5.2 Vurdering af trade-offs

Et trade-off indebærer en alternativomkostning eller offeromkostning – på engelsk: *opportunity cost*.[199] Trade-offs giver derfor økonomisk mening at foretage en vurdering af i relation til AI-drevne løsninger. Opgaven er at kunne *balancere* de forskellige trade-offs på en passende måde, sådan så de ikke vil blive overtrådt i henhold til art. 5, stk. 2.

5.2.1 Hvor store fordele eller ulemper er der ved én eller flere trade-offs for de(n) dataansvarlige selv eller for datasubjekterne?

For at efterleve art. 5, stk. 2 kan de(n) dataansvarlige med fordel gøre brug af den i afsnit 5.3 nævnte *trade-off-model*. Den vil nu bruges til at vurdere og diskutere graden af fordele eller ulemper ([↑]; [↑↑]; [↓]; [↓↓]), som udviklingen og anvendelsen af machine learning har for de(n) dataansvarlige selv eller for datasubjekterne. Hele afsnit 5.2.1 bygger på de tidligere analyser i afsnit 3, 4 og 5. Brugen af trade-off-modellen sker dermed i forlængelse af de i afsnit 2 nævnte metoder og retskilder og inden for rammerne af GDPR. Det betyder, at der de facto er tale om flere kildehenvisninger, end dem der er eksplicit nævnt i afsnit 5.2.1.1 – 5.2.1.3. Dette gør trade-off-modellen både praktisk anvendelig og teoretisk orienteret.

*5.2.1.1 **Accuracy** vs **privacy***

Det kan virke som en ulempe for de(n) dataansvarlige at indhente relevante, nødvendige og proportionale personoplysninger i overensstemmelse med dataminimeringsprincippet (afsnit 4.3). Ikke desto mindre, kan flere data som udgangspunkt gøre AI-systemer mere præcise [$accuracy$ ↑]. Hvis der f.eks. er flere kunder inkluderet i træningsdataene, vil en forudsigelse af f.eks. fremtidige køb baseret på kundernes købshistorik, pege hen imod en større præcision end tilfældet ville være, hvis der var færre kunder inkluderet.[200] Dette er med til at efterleve art. 5, stk. 1, litra d (afsnit 4.5.3), hvilket må

[199] Parkin: *Economics* (2012), s. 30*ff*
[200] ICO: *Trade-offs* (2019)

betegnes som en fordel for de(n) dataansvarlige. Accuracy kan endda stige endnu mere ved at tilføje nye features til et eksisterende datasæt, hvilket kan øge relevansen i forhold til, hvad modellen forsøger at forudsige $[accuracy \uparrow\uparrow]$. Eksemplet herpå kan være en købshistorik suppleret med yderligere demografiske data.[201] Ulempen for datasubjekterne er imidlertid, at indsamlingen af flere personoplysninger kan have konsekvenser for deres privatliv $[privacy \downarrow]$, jf. afsnit 4.5.1. Ved at indsamle yderligere/supplerende PII, kan det endda føre til en større effekt og dermed øge de negative konsekvenser betragteligt (f.eks. ved at føre til biased eller diskriminerende outcomes) $[privacy \downarrow\downarrow]$. Husky vs Wolf i afsnit 4.5.3 er et skrækeksempel på dette. Alt i alt er det de(n) dataansvarliges opgave at kunne balancere ovennævnte trade-off $accuracy$ $[\uparrow (\uparrow)]$ vs $privacy$ $[\downarrow (\downarrow)]$ på en passende måde med det formål at øge privacy til fordel for datasubjekterne. Det kan de(n) dataansvarlige gøre ved at indføre en eller flere *mitigerende foranstaltninger* i afsnit 5.1.1.2 til at mitigere risiciene for biased eller diskriminerende outcomes (f.eks. ved at ændre på datene i Husky vs Wolf og/eller tilpasse modellen efter man har trænet modellen op). Antallet af de "passende" tekniske foranstaltninger vil naturligvis afhænge af *graden* af den negative indvirkning privacy har på datasubjekterne $[privacy \downarrow (\downarrow)]$, jf. afsnit 5.2.1.1.1. Bemærk i øvrigt at Husky vs Wolf ikke nødvendigvis er begrænset til de mitigerende foranstaltninger, men kan *også* gøre brug af *accuracy measures* i afsnit 5.1.1.3. Det skyldes, at Husky vs Wolf i virkeligheden indeholdt træningsdata, som blev labelled forkert, inden det førte til unfairness. Derfor kan de(n) dataansvarlige benytte accuracy measures til at forstå, hvor accurate Husky vs Wolf er, og hvilken påvirkning det har på datasubjekterne. Husky vs Wolf blev dermed brugt som eksempel til at illustrere, at *antallet* og *typen* af de tekniske foranstaltninger i høj grad afhænger af *konteksten*. Derfor er der tale om en *konkret* vurdering i forhold til den specifikke udvikling og anvendelse af machine learning og i den kontekst, den er implementeret i (se afsnit 5.3).

5.2.1.1.1 Hvor mange foranstaltninger bør de(n) dataansvarlige "skrue op" for?

Ovennævnte tekniske foranstaltninger kan derfor anses for at være en klar fordel for datasubjekterne, da det er med til at beskytte deres privacy. Spørgsmålet er så, om de(n) dataansvarlige synes, at det er det værd at beskytte datasubjekterne, når de efterfølgende finder ud af, hvilke konsekvenser det kan medføre for dem selv? De(n) dataansvarlige skal nemlig være opmærksom(me) på, at selvom foranstaltningerne på den ene side er med til at beskytte datasubjekterne, er disse tekniske redskaber på den anden side med til at *reducere præcisionen af de AI-outputs*, der fremkommer.[202] Dette kan

[201] Ibid.
[202] Ibid.

gå ud over accuracy $[accuracy \downarrow (\downarrow)]$ alt afhængig af, hvor meget der "skrues op" for de tekniske foranstaltninger. Så hvis trade-offet viser $accuracy$ $[\uparrow\uparrow]$ vs $privacy$ $[\downarrow\downarrow]$, kan implementering af *få* foranstaltninger øge privacy – men mindske accuracy – en smule, hvorefter konstellationen f.eks. nu vil vise $accuracy$ $[\uparrow]$ vs $privacy$ $[\downarrow]$. Alternativt kan de udlignes som følge af moderate antal foranstaltninger. Dog hvis *mange* tekniske foranstaltninger imidlertid blev indført, kan det medføre større ulemper for de(n) dataansvarlige $[accuracy \downarrow\downarrow]$, end hvis færre af slagsen blev indført. Det betyder i praksis, at ikke nok med at algoritmens funktionalitet bliver forringet ved at den får en lavere accuracy, det bliver også *omkostningstungt* at indføre for mange tekniske foranstaltninger. Derfor kan det udledes, at det ikke altid er en god idé eller en fordel, at implementere et betydeligt antal foranstaltninger (eller for den sags skyld forskellige typer af foranstaltninger), når beskyttelsen af datasubjekterne *kunne opnås med det, der var mindre*.

5.2.1.2 ***Fairness*** *vs* ***fairness***

Udover at accuracy vs privacy kan volde problemer, kan det samme siges at være gældende for fairness vs fairness. Endnu en ulempe for datasubjekterne – og for de(n) dataansvarlige – er, at mangel på data om f.eks. en minoritetspopulation kan skabe et unfair system $[fairness \downarrow]$.[203] Her er det de(n) dataansvarlige opgave at vende om på denne negative tendens og gøre systemet mere fair over for minoritetspopulationen samlet set ("***den samlede effekt***"). Dette fordrer, at de(n) dataansvarlige indsamler flere data om mennesker fra én eller flere minoritetsgrupper ud af den samlede minoritetspopulation. Indsamlingen af personoplysninger til disse minoritetsgrupper/stikprøver er med til at gøre systemet mere accurate $[accuracy \uparrow]$ på disse mindre grupper af mennesker. Dette skulle gerne medføre, at hele minoritetspopulationen bliver mere fair $[fairness \uparrow (\uparrow)]$ på et overordnet plan.

5.2.1.2.1 ***Privacy*** *vs* ***fairness***

For at gøre AI-systemet mere fair over for minoritetspopulationen, vil det dog være nødvendigt at indsamle data på *beskyttede karaktertræk* om disse forskellige minoritetsgrupper for at teste, om AI-systemet er diskriminerende (herefter betegnet som "***sideeffekten***"). Denne diskriminationstest var tilsvarende noget, som blev påkrævet sideløbende af banken i afsnit 4.5.3 og dermed, ifølge Art. 29-Gruppens retningslinjer, forventes af de(n) dataansvarlige at foretage. Når disse data skal indsamles for at udføre testen, står de(n) dataansvarlige over for en svær afvejning mellem *privacy* (ved ikke at

[203] Ibid.

indsamle disse karakteristika) og *fairness* (at benytte karakteristikaene til at teste systemet og gøre det mere fair). Her vil rådet til de(n) dataansvarlige være, at **sideeffekten** kun må begrænses i det omfang, hvor den **samlede effekt** $[fairness \uparrow (\uparrow)]$ som minimum er på linje med (men helst overstiger) **sideeffekten** $[privacy \downarrow]$ – dog uden at privacy formindskes for meget til f.eks. $[privacy \downarrow\downarrow]$.

Ovennævnte problematik burde Farmbrella også tage behørigt stilling til, set i lyset af den i afsnit 4.5.2 nævnte udfordring med at få forebygget diskrimination i AI-løsningen. Her var virksomhedens minoritetspopulation de facto kvinder og mænd i landbrugsbranchen. Minoritetsgruppen ville til gengæld bestå af kvinder, da de fleste direktører inden for landbrugsbranchen som bekendt er mænd, der udgør majoritetsgruppen.[204] Der er med andre ord tale om, at der mangler data om minoritetsgruppen af kvinder. Udvikleren fra Farmbrella skal derfor indsamle flere data om kvinderne med henblik på at gøre systemet mere accurate $[accuracy \uparrow]$ med det formål at gøre hele minoritetspopulationen mere fair $[fairness \uparrow (\uparrow)]$. Her vil ***sideeffekten*** være at indsamle data om beskyttede karaktertræk herunder køn og alder. Herefter vil opgaven være at begrænse ***sideeffekten***. Der findes forskellige måder, udvikleren kan gøre dette på. En af måderne er i virkeligheden at *diskriminere* mændene ved at give kvinderne én eller flere *fordele*. Hvis der f.eks er en kvinde med kompetencerne, må hun vises frem i lyset i stedet for en mand. Det kan godt være, at manden har mere erfaring end hende, men alene fordi hun er kvinde, vil hun give en bedre *dynamik* til gruppen. Det kan være tankevækkende direkte at diskriminere mændene på den måde, men alligevel kan det være med til at løse problemet, hvad angår den samlede diskrimination. Med andre ord: *diskrimination kan hindre diskrimination*. Det er ikke fair over for den mand, som måske var bedre end kvinden – men alligevel skal hun vælges, blot fordi hun er en kvinde. Ikke desto mindre skal det være noget, som Farmbrella *aktivt* skal "fodre" algoritmen med, sådan så algoritmen afspejler disse kriterier. Det er derfor vigtigt, at data bliver brugt korrekt på målgruppen, og at udvikleren har den fornødne brugsforståelse. Hvis denne tillige kan tænke etik[205] ind i det, vil det være et stort plus.

Kunsten er som sagt, at ***sideeffekten*** kun må begrænses i et omfang, hvor den **samlede effekt** som minimum er på linje med (men helst overstiger) **sideeffekten** – dog uden at privacy formindskes for meget. Det betyder i praksis, at Farmbrella ikke skal diskriminere mændene *for meget* ved at give kvinderne *alt for mange* fordele – fordi ellers ville privacy falde for meget til f.eks. $[privacy \downarrow\downarrow]$.

[204] Casestudiet om Farmbrella i afsnit 2.2 og bilag 2
[205] Mottelson: *Magtblinde it-specialister er en trussel mod demokratiet* (2016)

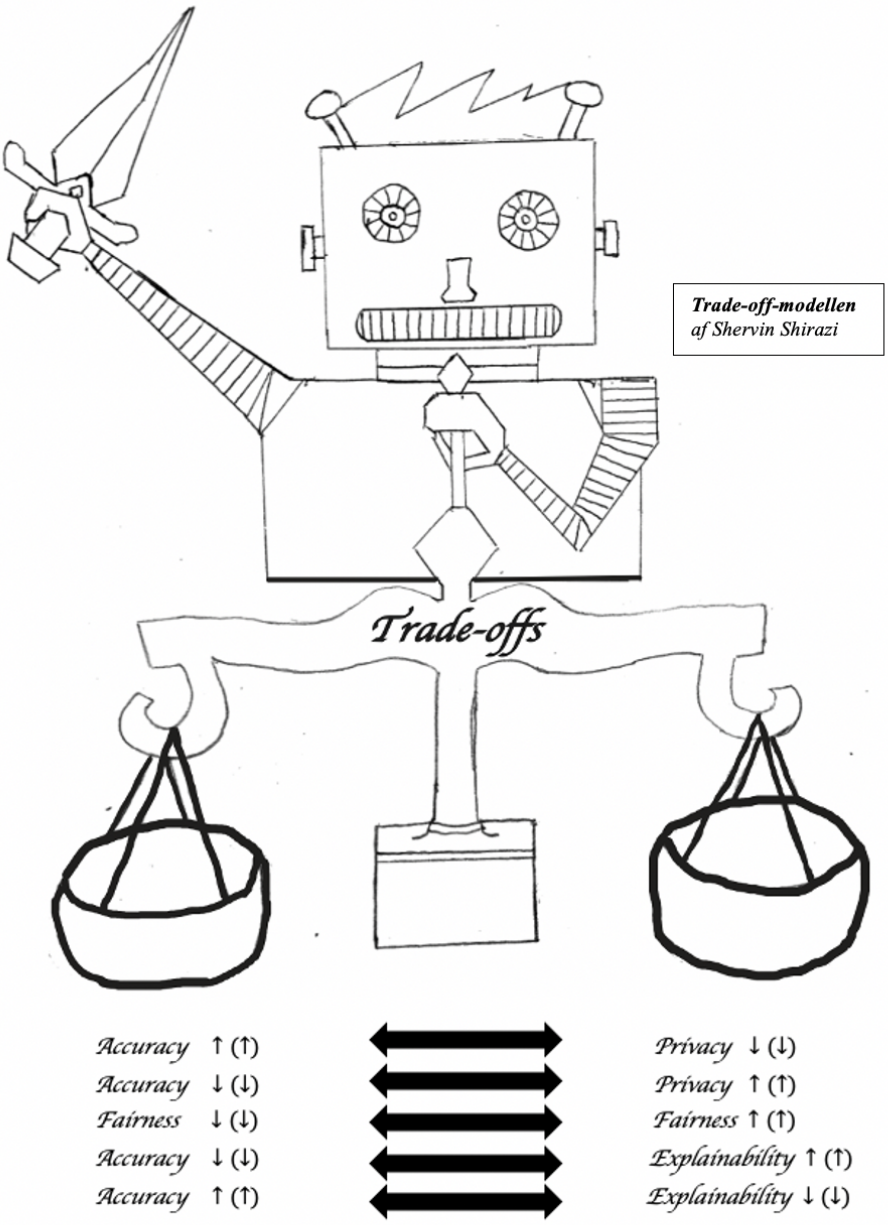
Hvis dette blev en realitet, ville det ikke havde været en forbedring, at den **samlede effekt** steg til f.eks. $[fairness \uparrow\uparrow]$. Det ønskværdige scenarie for Farmbrella er derimod, at den **samlede effekt** steg til $[fairness \uparrow\uparrow]$, selvom **sideeffekten** blev formindsket en smule til $[privacy \downarrow]$ ved at diskriminere mændene til en vis grad. På den måde kan Farmbrella afbalancere diskriminationsforholdene på den bedst mulige måde, hvilket *bidrager til ikke-diskrimination* for en hel minoritetspopulation i landbrugsbranchen – selvom nogle mænd af og til kan risikere at "mærke" en mindre grad af diskrimination. Denne vurdering af trade-offs via trade-off-modellen er derfor med til – i vidt omfang – at hindre unfair forskelsbehandling i forhold til begrebet *substantive fairness* i art. 5, stk. 1, litra a – selvom det gik lidt ud over *privacy*. Om end privacy bliver formindsket en smule, bør det antages, at de uafhængige tilsynsmyndigheder i alle dele af EU vil godkende denne grundige vurdering af trade-offs. Det skyldes, at der ikke umiddelbart på nuværende tidspunkt foreligger bedre alternativer for de(n) dataansvarlige i dette tilfælde.

*5.2.1.3 **Accuracy** vs **explainability***

Sidst men ikke mindst kan der opstå problemstillinger vedrørende accuracy vs explainability. Meget simple AI-systemer $[accuracy \downarrow\downarrow]$ kan være meget forklarbare $[explainablity \uparrow\uparrow]$. F.eks. ved hjælp af decisions trees, hvor forholdet mellem input og output (ofte) er nemme at forstå (afsnit 4.5.4). Det må alt andet lige betegnes som en fordel for datasubjektet. Omvendt vil det være en ulempe for de(n) dataansvarlige at have et inaccurate AI-system, hvorfor det er i dennes interesse at optimere på det mest muligt. Hvis det sker, vil meget komplekse AI-systemer $[accuracy \uparrow\uparrow]$ imidlertid gøre det svært at følge systemets bagvedliggende logik. F.eks. kan neural networks eller deep learning gøre det svært at vide, hvorfor den er kommet frem til outputtet $[explainablity \downarrow\downarrow]$. Det værst tænkelige scenarie er, hvis en automatisk AI-afgørelse om datasubjektet er truffet uden nogen form for forklaring eller information herom $[explainablity \downarrow\downarrow]$. Dette vil begrænse datasubjekternes autonomi og muligheder for selvbestemmelse, hvilket med stor sandsynlighed skaber et unfair system $[fairness \downarrow (\downarrow)]$.[206] Her kan man med fordel implementere én eller flere [XAI]-foranstaltninger. I dette tilfælde er *fairness explanation* fra afsnit 5.1.1.4 mest velgenet, da det både øger explainability og fairness uden at gå alt for meget på kompromis med accuracy. Det kan gøres ved at skrue op for fairness explanation i en sådan grad, at de(n) dataansvarlige får et moderat system $[accuracy \uparrow]$ samtidig med at explainability og fairness sikres eller øges på tilstrækkelig vis $[explainablity \uparrow]$ og $[fairness \uparrow]$.

[206] ICO: *Explaining decisions made with AI – Part 1* (2020), s. 13

5.3 Hvordan skal de(n) dataansvarlige løse og dokumentere de pågældende trade-offs?



Tabel 2: Denne gang ses ikke justitia men derimod en robot med et sværd i den ene hånd og en vægtskål i den anden. Sværdet skal symbolisere robottens magt, mens vægtskålen skal symbolisere de trade-offs robotten har fået integreret i dens algoritme. Det er de(n) dataansvarliges opgave at vejlede sin "robot" til at kunne balancere de forskellige trade-offs på hver side af vægtskålen på en passende måde. Venstre side af vægtskålen repræsenterer graden af de fordele/ulemper udviklingen og anvendelsen af machine learning har for de(n) dataansvarlige. Højre side repræsenterer graden af de fordele/ulemper udviklingen og anvendelsen af machine learning har for datasubjekterne. De(n) dataansvarlige skal også kunne påvise/dokumentere de afvejninger, robotten har foretaget.

De(n) dataansvarlige, som udvikler og anvender machine learning, bliver nødt til at *identificere* og *vurdere* trade-offs og finde en *passende balance* mellem de databeskyttelseskrav, der står i konflikt med hinanden. ICO har flere gange understreget, at håndteringen af trade-offs er et *nøgleområde* inden for afdækningen af de AI-specifikke risici, hvorfor de(n) dataansvarlige allerede har løst nogle af største udfordringer, hvis de kan afveje disse trade-offs på en passende måde, jf. art. 5, stk. 2.[207]

Som dataansvarlig kan man derfor løse og dokumentere de identificerede trade-offs ved at iagttage følgende procedure:

1. ***Identificering af trade-offs***
    a. *Hvilke trade-offs kan man identificere ved at udvikle og anvende et modelvalg inden for machine learning?* ***(afsnit 4.5.1 – 4.5.4)***
2. ***Løsning af trade-offs***
    a. *Har man "passende" tekniske og organisatoriske foranstaltninger til at løse de identificerede udfordringer i forbindelse med trade-offs?* ***(afsnit 5.1.1.2 – 5.1.1.4)***
3. ***Trade-off-modellen***
    a. *Vurdering af trade-offs:*
        i. *Hvor store fordele eller ulemper* ([↑]; [↑↑]; [↓]; [↓↓]) *er der ved én eller flere trade-offs for de(n) dataansvarlige selv eller for datasubjekterne?* ***(afsnit 5.2)***
    b. *Dokumentation af trade-offs:*
        i. *Beskriv og illustrér de afvejninger der er foretaget ved hjælp af trade-off-modellen. Vis det som dokumentation over for tilsynsmyndigheden (**afsnit 5.3**)*

Når det kommer til anvendelsen af de databeskyttelsesretlige regler på AI-baserede løsninger, fremhæver ICO, at der ikke er tale om en fast fremgangsmåde i forhold til at løse AI-udfordringerne på. Det betyder i praksis, at det ikke er meningen, at organisationer og offentlige myndigheder skal – eller kan – følge det samme hierarki, når udfordringerne skal løses. Det handler til gengæld om at finde den rette *balance* på tværs af en eller flere trade-offs (afsnit 5.2). Når først de(n) dataansvarlige har fundet denne balance, kan det efterfølgende bruges som uundværlig dokumentation over for tilsynsmyndighederne. Der er i øvrigt altid tale om en konkret vurdering i forhold til den specifikke udvikling og anvendelse af machine learning og i den kontekst, den er implementeret i[208].

[207] ICO: *Developing the ICO AI Auditing Framework: an update* (2019)
[208] Et eksempel herpå findes i afsnit 5.2.1.1

Ovennævnte synspunkt fra ICO falder i god tråd med, hvad GDPR rent faktisk sigter hen imod, når det handler om arbejdet med compliance. Den sigter nemlig i retning af, at der bør skabes indsigt i databehandlingens *karakter, omfang, sammenhæng og formål* og det er netop i denne konkrete kontekst, at risikoen for datasubjekterne skal vurderes.[209] Derudover skal det på ingen måder undervurderes, at disse ord/parametre udtrykkeligt indgår i art. 24 og 25's ordlyd.

Hvis det ikke er muligt at opnå et passende trade-off mellem to eller flere databeskyttelsesretlige krav i løbet af udviklings- og træningsfasen (afsnit 3.2.3), skal de(n) dataansvarlige være forberedt på at *standse* deployment af AI-systemet omgående.[210]

6 Konklusion

Det konkluderes, at måden, hvorpå man kan sikre overholdelse af GDPR ved udvikling og anvendelse af machine learning i relation til compliance er at følge den udvidede forenklede AI-livscyklus (illustreret i afsnit 3.2.3). Modellen skaber den røde tråd igennem hele afhandlingen og bidrager til at nå målet om at blive compliant med AI. Her er der i afsnit 4 blevet belyst nogle af de væsentligste udfordringer, der kan opstå i forbindelse med at blive compliant med AI. Derudover er der i afsnit 5 blevet belyst de løsninger og dokumentation, der er nødvendige for at sikre denne compliance.

Fase 1: indsamlingsfasen; først og fremmest er det de(n) dataansvarliges opgave at finde ud af, om machine learning i medfør af art. 2, stk. 1 indebærer en behandling af personoplysninger (4.1). Dette er ofte tilfældet, idet modellen typisk indeholder en større mængde PII. Udfordringen med machine learning er imidlertid evnen til at kunne sondre mellem personal versus non-personal data, da der kan være tale om mixed datasets. Dernæst skal formålet fastlægges (4.2). Grunden til, at formålet ofte først kan specificeres efter – og ikke før indsamlingen af personoplysninger – skyldes den fundamentale udfordring, at machine learning og formålsbestemthedskravet ikke fungerer sammen. Dette påvirker andre bestemmelser såsom art. 5, stk. 1, litra c negativt, hvorefter efterlevelse af kravene indeholdt i dataminimeringsprincippet skaber flere udfordringer (4.3). Hvis de(n) dataansvarlige i øvrigt agter at videresælge oplysninger i algoritmen, kan det være en udfordring ikke at videresælge dem til formål, der er uforeneligt med de(t) oprindelige (4.2).

[209] Olsen: *Håndbog i dataansvarlighed* (2020), s. 409
[210] ICO: *Trade-offs* (2019)

Fase 2: udviklings- og træningsfasen *og* ***fase 3: anvendelsesfasen;*** efterfølgende skal man identificere trade-offs, herunder om machine learning er fair (4.5.2), accurate (4.5.3) og explainable (4.5.4). Der foreligger mange udfordringer med at efterleve bestemmelserne, idet et modelvalg kan afføde AI-specifikke risici. Høj-risiko-faktorer *kan* være alvorlig diskrimination som følge af især 1) ubalancerede træningsdata, 2) træningsdata, som afspejler tidligere diskrimination og 3) profilering. Når det er sagt, har denne afhandling samtidig vist, at det er en sandhed med væsentlige modifikationer. Mennesker diskriminerer altså også – måske endda mere end AI-robotten. Fordelen ved at bruge AI er, at den f.eks. kan styre, hvad der må profileres og hvad der ikke må profileres. Dermed kan man sikre, at AI-løsninger bliver mindre diskriminerende eller mindre biased end de havde været, hvis et menneske havde lavet dem. Dette er en central pointe, fordi en udvikler kan – ved vurderingen af trade-offs (5.2) – gå ind og fjerne, tilføje eller ændre på disse variable og dermed få et andet slags output ud af det. Dette var også påkrævet af udvikleren fra Farmbrella ved aktivt at give kvinderne én eller flere fordele i algoritmen (5.2.1.2.1). IT-udviklere besidder således en evne til at kontrollere og ændre på afgørende beslutninger ved udvikling og anvendelse af machine learning.

Fase 4: forankringsfasen; for at løse udfordringerne i de tre første faser kan de(n) dataansvarlige implementere tekniske AI-foranstaltninger. Dataminimering kan sikres ved hjælp af dataminimeringsteknikker. F.eks. kan privacy-preserving methods i den sidste ende gøre identificerbarheden mere gennemsigtig og dermed øge evnen til at skille de pseudonymiserede og anonymiserede oplysninger ad i det blandede datasæt (5.1.1.1). Fairness kan sikres ved hjælp af tekniske foranstaltninger til at mitigere diskriminationsrisici. Dette kan f.eks. gøres ved tilføje eller fjerne data om en under- eller overrepræsenterede gruppe (5.1.1.2). Accuracy kan sikres ved hjælp af accuracy measures afspejlet i balancen mellem false positives og false negatives. F.eks. kan en produktion af false negative få konsekvenser for både en kvalificeret kandidat og for de(n) dataansvarlige selv (5.1.1.3). Endelig kan explainability sikres ved hjælp af en [XAI]-foranstaltning. F.eks. kan man reducere kompleksiteten af AI-systemer ved at implementere en fairness explanation (5.1.1.4).

Udover identificering af trade-offs i fase 2 og 3, er særligt de tre sidstnævnte foranstaltninger i fase 4 med til at øge kvaliteten af den samlede vurdering, der foretages af trade-offs (5.2) med afsæt i *trade-off-modellen* (5.3). Dette øger også kvaliteten af den *beskrivelse* og *illustration*, som skal bruges til at dokumentere afvejningen af trade-offs over for tilsynsmyndighederne i bl.a. EU's medlemsstater.

7 Perspektivering

EU-kommissionen har som sagt fremsat nyt forslag om regulering af AI. Med regelsættet taler den ledende næstformand Margrethe Vestager om en *europæisk tilgang* til AI.[211] Det indebærer til dels en omfavnelse af teknologien, men samtidig også et fast tag om de AI-specifikke risici, der er forbundet med at bruge teknologien. AI skal dermed bruges så meget som muligt, men det skal være på menneskets vilkår.[212] Fokus fastholdes dermed på fysiske personers rettigheder og frihedsrettigheder, og forslaget bygger på en risikobaseret tilgang til udvikling og anvendelse af AI (afsnit 2.1, 5.1, 5.1.1 og 5.1.2). Disse ting gør, at denne afhandling fortsat vil have relevans efter lovforslagets endelige vedtagelse, fordi fokusområdet er det samme. Den eneste nuance er blot, at de nye krav gøres mere specifikke ved at blive inddelt i risikable kategorier. Dette kunne dog være særdeles spændende at undersøge nærmere. Brud på regelsættet kan føre til bøder på op til 6% af den årlige globale omsætning, hvilket er højere end, hvad GDPR foreskriver (afsnit 1).

Litteraturliste

Bøger

- Andersen, Mads Bryde: *Ret & metode,* 1. udgave, 1. oplag, Gads Forlag, 2002
- Blume, Peter: *Den nye persondataret*, 2. udgave, Jurist- og Økonomforbundets Forlag, 2018
- Hansen, Lone L. & Werlauff, Erik: *Den juridiske metode – en introduktion*, 2. udgave, 1. oplag, Jurist- og Økonomforbundets Forlag, 2016
    - (Citeret: Hansen & Werlauff: *Den juridiske metode* (2016))
- Olsen, Birgitte Kofod m.fl.: *Eksponeret – Grænser for privatliv i en digital tid,* 1. udgave, 1. oplag, Gads Forlag, 2018
    - (Citeret: Olsen m.fl.: *Eksponeret* (2018))
- Korfits, Kristian & Lotterup, Anders: *Databeskyttelsesforordningen* og *databeskyttelsesloven med kommentarer*, 1. udgave, Jurist- og Økonomforbundets Forlag, 2020
- Motzfeldt, Hanne Marie: *Retssikkerheden bør følge med den automatiserede forvaltning*. Uddrag fra Olsen, Birgitte Kofod m.fl.: *Eksponeret – Grænser for privatliv i en digital tid,* 1. udgave, 1. oplag, Gads Forlag, 2018

[211] https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682 (21.04.2021)
[212] Den europæiske tilgang står i stærk kontrast til den kinesiske, der i vid udstrækning bruger teknologien til overvågningsformål. Mozur: *Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras* (2018)

    - (Citeret: Motzfeldt: *Retssikkerheden bør følge med den automatiserede forvaltning* (2018))
- Olsen, Birgitte Kofod: *Håndbog i dataansvarlighed*, 1. udgave, Djøf, 2020
- Parkin, Michael: *Economics*, 10. udgave, Pearson, 2012
- Revsbech, Karsten m.fl.: *Forvaltningsret – almindelige emner*, 6. udgave, 1. oplag, Jurist- og Økonomforbundets Forlag, 2016
- Sørensen, Karsten Engsig m.fl.: *EU-retten*, 6. udgave, 1. oplag, Jurist- og Økonomforbundets Forlag, 2014
- Taulli, Tom*: Artificial Intelligence Basics: A Non-Technical Introduction*, 1. udgave, Apress, 2019

Artikler

- Almeida, Fernando m.fl.: *Strengths and Limitations of Qualitative and Quantitative Research Methods*, European Journal of Education Studies, 3(9), 369-387, 2017
- Blume, Peter & Motzfeldt, Hanne Marie: *Databeskyttelse og udvikling af kunstig intelligens – svømmer databeskyttelsesretten over sine egne bredder?* Revision & Regnskabsvæsen, RR.2020.10.0014, 2020
    - (Citeret: Blume & Motzfeldt: *Databeskyttelse og udvikling af kunstig intelligens* (2020))
- Brkan, Maja: *Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond,* International Journal of Law and Information Technology, 27, 91–121, 2019
- Castañón, Jorge: *Machine Learning Methods that Every Data Scientist Should Know,* Towards Data Science, 2019
- Datalogisk Institut: *Computer fortæller om man dør af COVID-19*, Københavns Universitet, 2021
- Kjær, Sebastian: *Ugens Startup: Med avanceret ansigtsanalyse vil Justface Retail give fysiske butikker data-superkræfter,* TechSavvy, 2021
- Kunckel, Charlotte m.fl.: *Kunstig intelligens, GDPR og andre juridiske udfordringer,* Ret & Indsigt, 2018
- McCombe, Madeline: *Intro to future selection methods for Data Science*, Towards Data Science, 2019

- Mottelson, Aske: *Magtblinde it-specialister er en trussel mod demokratiet*, Information, 2016
- Mozur, Paul: *Inside China's Dystopian Dreams: A.I., Shame and Lots of* Cameras, The New York Times, 2018
- Nielsen, Mads: *Kunstig intelligens styrker kræftbehandling*, Kræftens Bekæmpelse, 2018
- Palmer, Cathryne & Bolderston, Amanda: *A Brief Introduction to Qualitative Research*, The Canadian Journal of Medical Radiation Technology, 37(1), 16-19, 2006
- Soper, Taylor: *How Olympic athletes use machine learning and data analysis to reach peak performance levels*, GeekWire, 2016

Love, betænkninger mv.

- Den Europæiske Menneskerettighedskonvention, vedtaget i 1950 – EMRK
- Den Europæiske Unions Charter om grundlæggende rettigheder, Den Europæiske Unions Tidende (2010/C 83/02) – EU's Charter
- Europa-Parlamentets og Rådets forordning (EU) 2018/1807 af 14. november 2018 om en ramme for fri udveksling af andre data end personoplysninger i Den Europæiske Union, Den europæiske Unions Tidende (L 303/59) – FFD regulation
- Europa-Parlamentets og Rådets forordning (EU) 2016/679 af 27. april 2016 om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger og om ophævelse af direktiv 95/46/EF (generel forordning om databeskyttelse) (2016/679) – Databeskyttelsesforordningen
- Europa-Parlamentets og Rådets direktiv 95/46/EF af 24/10/1995 om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger – Databeskyttelsesdirektivet
- Justitsministeriet: Databeskyttelsesforordningen - og de retlige rammer for dansk lovgivning, Betænkning nr. 1565, del I, bind 1, 24. maj 2017
    - (Citeret: Bet. 1565/2017)
- Lov nr. 502 af 23. maj 2018, Lov om supplerende bestemmelser til forordningen om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger – Databeskyttelsesloven
- Lov nr. 429 af 31. maj 2000, Lov om behandling af personoplysninger – Persondataloven
- Lov nr. 688 af 8. juni 2018, Lov om forbud mod forskelsbehandling på grund af handicap – Lov om forbud mod handicapdiskrimination

- Lovbekendtgørelse nr. 645 af 8. juni 2011, Bekendtgørelse af lov om ligebehandling af mænd og kvinder med hensyn til beskæftigelse mv. – Ligebehandlingsloven
- Lovbekendtgørelse nr. 438 af 16. maj 2012, Bekendtgørelse af lov om etnisk ligebehandling – Lov om etnisk ligebehandling
- Traktaten om Den Europæiske Unions Funktionsmåde, Den Europæiske Unions Tidende (2012/C 326/01) – TEUF

Vejledninger og retningslinjer og rapporter

- Agencia Española de Protección de Datos (AEPD): *Audit Requirements for Personal Data Processing Activities involving AI*, januar 2021
- Article 29 Data Protection Working Party: *Opinion 03/2013 on purpose limitation*, WP 203, 2. april 2013
    - (Citeret: WP203 (2013))
- Article 29 Data Protection Working Party: *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, WP 251, rev.01, 6. februar 2018
    - (Citeret: WP251 (2018))
- Baxter, Palema & Jack, Susan: *Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers*, The Qualitative Report, 13(4), 544-559, 1. december 2008
- Council of Europe & European Union Agency for Fundamental Rights: *Handbook on European data protection law*, april 2018
- Datatilsynet: *Behandlingssikkerhed - Databeskyttelse gennem design og standardindstillinger*, juni 2018
    - (Citeret: Datatilsynet: *Behandlingssikkerhed*, (2018))
- Datatilsynet: *Vejledning - Samtykke*, maj 2021
- Det norske datatilsyn: *Kunstig intelligens og personvern*, januar 2018
- European-Parliament: *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*, Scientific Foresight Unit (STOA) & Panel for the Future of Science and Technology, juni 2020
    - (Citeret: EU-Parliament: *The impact of the GDPR on artificial intelligence* (2020))
- Future of Privacy Forum (FPF): *The Privacy Expert's Guide To AI and Machine Learning*, oktober 2018

- Information Commissioner's Office (ICO): *Accuracy of AI system outputs and performance measures*, 2. maj 2019
- Information Commissioner's Office (ICO): *An overview of the Auditing Framework for Artificial Intelligence and its core components*, 26. marts 2019
- Information Commisioner's Office (ICO): *Big data, artificial intelligence, machine learning and data protection*, marts 2017
- Information Commissioner's Office (ICO): *Data minimisation and privacy-preserving techniques in AI systems*, 21. august 2019
- Information Commissioner's Office (ICO): *Data Protection Impact Assessments and AI*, 23. oktober 2019
- Information Commissioner's Office (ICO): *Developing the ICO AI Auditing Framework: an update*, 4. juli 2019
- Information Commissioner's Office (ICO): *Enabling access, erasure, and rectification rights in AI*, 15. oktober 2019
- Information Commisioner's Office (ICO): *Explaining decisions made with AI – Draft guidance for consultation – Part 1: The basics of explaining AI*, The Alan Turing Institute, 20. maj 2020
- Information Commisioner's Office (ICO): *Guidance on AI and data protection*, 30. juli 2020
- Information Commisioner's Office (ICO): *Guidance on the AI auditing framework – Draft guidance for consultation*, 19. februar 2020
    - (Citeret: ICO: *Guidance on the AI auditing framework* (2020))
- Information Commisioner's Office (ICO): *Human bias and discrimination in AI systems*, 25. juni, 2019
- Information Commissioner's Office (ICO): *Known security risks exacerbated by AI*, 23. maj 2019
- Information Commissioner's Office (ICO): *Project explain: Interim report*, juni 2019
- Information Commissioner's Office (ICO): *Trade-offs*, 25. juli 2019
- The European Data Protection Board (EDPB): *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default*, 13. november 2019
    - (Citeret: EDPB: *Guidelines 4/2019 on art. 25 DPbDD* (2019))

Links

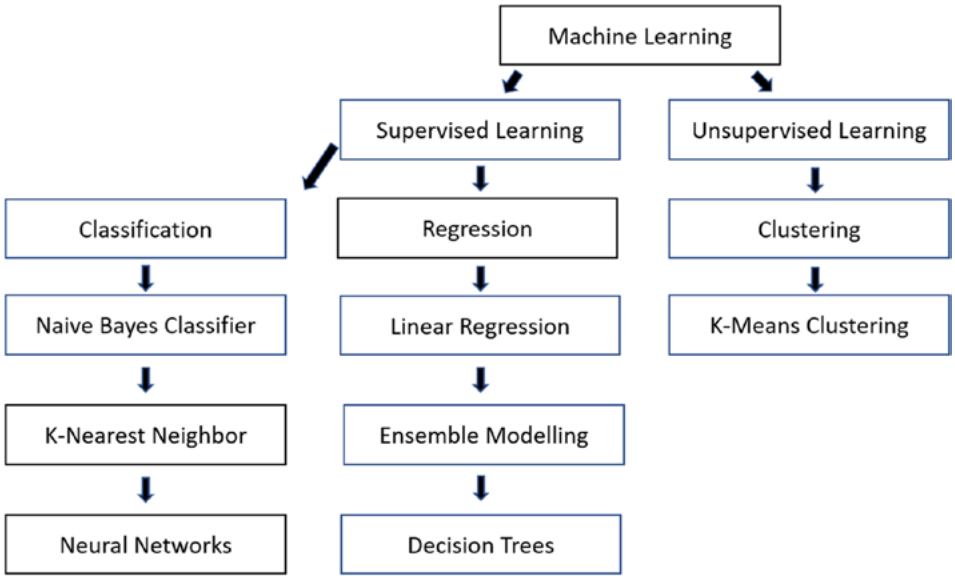
- https://videncenter.kl.dk/viden-og-vaerktoejer/informationssikkerhed-og-gdpr/juridisk-ai-vaerktoejskasse/
  - (senest besøgt 25.02.2021)
- https://www.carve.dk/2019/09/09/no-personal-data-no-gdpr-right-yes-but/
  - (senest besøgt 27.02.2021)
- https://www.regeringen.dk/nyheder/2019/national-strategi-for-kunstig-intelligens/
  - (senest besøgt 01.03.2021)
- https://www.digitaltrends.com/cool-tech/machine-learning-v-for-victory-terrorist-identification/
  - (senest besøgt 01.03.2021)
- https://www.datatilsynet.dk/om-datatilsynet
  - (senest besøgt 10.03.2021)
- https://edpb.europa.eu/our-work-tools/article-29-working-party_da
  - (senest besøgt 20.03.21)
- https://edpb.europa.eu/sites/edpb/files/files/news/endorsement_of_wp29_documents.pdf
  - (senest besøgt 22.03.2021)
- https://www.kmd.dk/indsigter/kunstig-intelligens-skal-kunne-forklare-sine-anbefalinger
  - (senest besøgt 26.03.2021)
- https://www.cs.ox.ac.uk/people/reuben.binns/
  - (senest besøgt 01.04.2021)
- https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682
  - (senest besøgt 21.04.2021)
- https://farmbrella.dk
  - (senest besøgt 24.04.2021)
- https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence
  - (senest besøgt 14.05.2021)
- ISO/IEC 2382-28:1995(en), Information technology — Vocabulary — Part 28: Artificial intelligence — Basic concepts and expert systems
  - (senest besøgt 14.05.2021)

- https://manoa.hawaii.edu/exploringourfluidearth/physical/world-ocean/map-distortion/practices-science-precision-vs-accuracy
    - (senest besøgt 20.05.2021)
- https://theengineeringofconsciousexperience.com/understanding-artificial-intelligence-and-its-future-neil-nie-tedxdeerfield/
    - (senest besøgt 23.05.2021)
        - (Citeret: Neil Nie: *Understanding Artificial Intelligence and Its Future*, TEDxDeerfield (2017))
- https://ico.org.uk/about-the-ico/news-and-events/events-and-webinars/ensuring-lawfulness-fairness-and-transparency-in-ai-systems-webinar/
    - (senest besøgt 23.05.2021)
        - (Citeret: ICO: *Ensuring lawfulness, fairness, and transparency in AI systems webinar* (2020))
- https://www.alphagomovie.com
    - (senest besøgt 23.05.2021)
        - (Citeret: DeepMind: *AlphaGo – The Movie | Full Documentary* (2020))
- https://www.zetland.dk/historie/sekdrB5l-moV7PnW6-fbf2b
    - (senest besøgt 26.05.2021)

Øvrigt

- Datatilsynet: *Årsberetning 2020*, publiceret 7. maj 2021
- Europa-Kommissionen: *Hvidbog om kunstig intelligens – en europæisk tilgang til ekspertise og tillid*, februar 2020
- Finansministeriet og Erhvervsministeriet: *National strategi for kunstig intelligens*, marts 2019
- Kommunal- og moderniseringsdepartementet: *Nasjonal strategi for kunstig intelligens*, januar 2020
- Kommunernes Landsforening (KL) og Kammeradvokaten: *Præsentation af juridisk værktøjskasse for ansvarlig AI. Håndtering af de databeskyttelsesretlige krav og risici ved udvikling og anvendelse af AI-løsninger i kommunerne*, juni 2020
- Madsen, Lars Henrik Gam: *Retsdogmatisk forskning*, 1. januar 2021

Bilag 1 – Framework for machine learning algorithms



Kilde: Taulli: *Artificial Intelligence Basics: A Non-Technical Introduction* (2019), s. 55

Bilag 2 – Specialesamarbejde

Agrosektoren oplever et stort generationsskifte, hvor ejere og selskabsformen vil blive skiftet ud. I 2019 var 36% af ejerne ud af de 24.000 danske landbrugsbedrifter 66 år eller ældre, og der er et bredt ønske om at omdanne nuværende IS-selskaber til AS og hertil kræves oprettelse af en professionel bestyrelse. Farmbrella ser derfor, at nu er det rette tidspunkt at udvikle et AI-værktøj, der kan hjælpe denne målgruppe med at sammensætte den rigtige bestyrelse. Der er et stærkt behov for en såkaldt "AI Board Recruiter", der ved hjælp af machine learning skal sammensætte den rigtige bestyrelse ud fra de præferencer, den enkelte virksomhed har. Den type personoplysninger der behandles af algoritmen, er i øjeblikket almindelige art. 6-oplysninger. AI-løsningen bruger for det meste unsupervised learning, da Farmbrella ikke rigtig har nogen historiske data til at finde sammenhænge i, hvad der gør en bestyrelse god. Lige nu bliver det baseret på kompetencer. Virksomheden er som nævnt i afsnit 2.2 nøje udvalgt, da de befinder sig på et stadie, hvor der skal tænkes compliance i udviklingen og anvendelsen af deres AI-løsning. Om ikke andet kommer AI Board Recruiteren i hvert fald til at være en væsentlig motor i målet om at øge antallet af abonnenter med 4000 over de næste 3 år, svarende til en øget omsætning på 14,7 mio. kroner. Endvidere har Farmbrella brug for at vide, hvilke

udfordringer det indebærer og hvilke løsninger og dokumentation, der kræves fra dem. Derfor har udvikleren – der har specialiseret sig inden for machine learning – indvilliget i at indgå et samarbejde.

Kilde: Bilag 2 er udarbejdet på baggrund af et kvalitativt casestudie med Fambrella (afsnit 2.2)

Bilag 3 – Begrebsoversigt

Accuracy vs precision	Accuracy refererer til, hvor tæt en måling er på den sande værdi eller den accepterede værdi. Præcision referer til, hvor tæt en måling er på hinanden indbyrdes. Det er muligt at være præcis uden at være accurate. Omvendt er det også muligt at være accurate uden at være præcis.[213] Herved bemærkes, at begreberne defineres forskelligt inden for økonomisk terminologi, mens præcision er indeholdt i begrebet accuracy i art. 5, stk. 1, litra d inden for juridisk terminologi
Artificial intelligence [AI]	Kunstig intelligens [KI] gør computere i stand til at lære fra erfaringer, hvilket ofte involverer en behandling af personoplysninger ved brug af sofistikerede algoritmer
Biases	Tendensen til at over- eller underestimere værdien af én eller flere parametre
Black box	Man kender til input og output, men ikke hvad der sker i inde i udregningen (man får således ikke mellemregninger med)
Clustering	En form for unsupervised learning som anvender unlabbeled data og bruger algoritmer til at lave grupperinger og sammenhænge i data
Compliance	Efterlevelse af regler
Decision trees	Beslutningstræer er et modelvalg, der er baseret på en arbejdsprocedure med ”ja/nej”-beslutninger. Man starter på øverste niveau af et træ, hvorefter man gradvist arbejder sig nedad indtil det nederste niveau er nået – og et output gives
Deep learning	Dyb læring er en type af AI som bruger neural networks til at efterligne visse processer i hjernen. Ordet ”dyb” henviser til et vis antal skjulte lag i det neurale netværk, hvilket styrker AI-modellen til at lære
Explainability	Evnen til at kunne forklare AI, når der træffes en automatisk afgørelse (i modsætning til en ”black box”). Explainabilty er med andre ord en proces til f.eks. at forstå underliggende årsager bag modelvalgene bedre
Fairness	Rimelighed/retfærdighed

[213] https://manoa.hawaii.edu/exploringourfluidearth/physical/world-ocean/map-distortion/practices-science-precision-vs-accuracy (20.05.2021)

Features	En feature er en kolonne af data, som indeholder bestemte træk eller karakteristika af et objekt, (data)subjekt eller lignende
Inference	Machine learning inference refererer til en inferensfase, som er den fase, hvor man "sætter AI-modellen i produktion" under deployment
Labelled data	Kategoriserede data, der ikke har en numerisk værdi men i stedet giver sig udryk i tekst (f.eks. ved at beskrive en race eller et køn)
Learning algorithm	Læringsalgoritme, der ofte involverer typer af korrelationer mellem data
Machine learning	Maskinlæring, hvor et system lærer og forbedrer sig ved at behandle data i stort omfang, uden at systemet skal pogrammeres manuelt
Mitigation measures	Mitigerende foranstaltninger, som er foranstaltninger til at neddæmpe/afbøde risici
Neural networks	Neurale netværk er sofistikerede AI-modeller som efterligner visse processer i hjernen. Et neural netværk har forskellige lag, der forsøger at finde unikke mønstre, der involverer analyser af flere forskelige lag (multiple layers)
Output/outcome	AI-modellen giver brugeren et resultat
Privacy	Privatlivsbeskyttelse/menneskerettigheder
Proxy variables	Proxy variable er variabler, som korrelerer med andre variable
Random data	Tilfældige data
Supervised learning	Vejledt læring, der er en AI-model, som bruger labelled data
Target	Et mål/noget man stiler efter
Trade-offs	Opportunity cost (alternativomkostning/offeromkosnting)
Training input/training set	Træningsdata, der indsættes som input til at skabe en algoritme
Unsupervised learning	Ikke-vejledt læring, der er en AI-model, som bruger unlabelled data
User input	Brugeren giver noget input til AI-modellen

Kilde: Begrebsoversigten i bilag 3 er udarbejdet på baggrund af kildehenvisninger i brødteksten og Taulli: *Artificial Intelligence Basics: A Non-Technical Introduction* (2019), s. 179-184