Åpenhet og rettferdighet i algoritmens tidsalder

Hvordan prinsippene om åpenhet og rettferdighet i personvernforordningen art. 5 (1) a) skal forstås ved bruk av kunstig intelligens som beslutningsstøtte

Kandidatnummer: 60

Antall ord: 14 966

JUS399 Masteroppgave
Det juridiske fakultet

UNIVERSITETET I BERGEN

10.12.2021

Innholdsfortegnelse

1 Innledning

1.1 Problemstillingen og dens aktualitet

Oppgavens tema er åpenhets- og rettferdighetsprinsippet i General Data Protection Regulation (heretter «personvernforordningen» eller «forordningen»).[1] Prinsippene fremgår av personvernforordningen art. 5 (1) a), som fastslår at «personopplysninger skal (…) behandles på en (…) rettferdig og åpen måte».[2] Problemstillingen er hvordan prinsippene skal forstås ved bruk av kunstig intelligens (KI) som beslutningsstøtte.

Den raske teknologiske utviklingen har satt personvernet under stadig større press, ettersom utviklingen har muliggjort å samle inn, dele og sammenstille store mengder personopplysninger.[3] Utviklingen førte til utarbeidelsen av personvernforordningen, som trådte i kraft mai 2018. Formålet med personvernforordningen er å verne fysiske personer i forbindelse med behandling av deres personopplysninger, samt å tilrettelegge for fri utveksling av personopplysninger innenfor EØS-området, jf. art. 1 (1). Formålene skal sikres gjennom enhetlig regulering av personvern i EØS-området.

KI er ikke nevnt i personvernforordningen. Forordningen er imidlertid utformet teknologinøytralt, som innebærer at den gjelder behandling av personopplysninger uansett hva slags teknologi som benyttes. Personvernforordningen omfatter dermed også nye typer teknologi, som KI. For å sikre at forordningen favner bredt, blant annet med hensyn til teknologinøytralitet, er ordlyden abstrakt formulert.[4] Det kan derfor være vanskelig å vite hvordan bestemmelsene skal forstås ved bruk av konkret teknologi, i spesifikke situasjoner.[5] Oppgavens formål er derfor å finne ut av hvordan åpenhet og rettferdighet, som sentrale, men abstrakte, prinsipper i personvernforordningen, skal forstås ved bruk av KI når det benyttes som beslutningsstøtte.

[1] Europaparlamentets og Rådets forordning (EU) 2016/679 av 27. april 2016 om vern av fysiske personer i forbindelse med behandling av personopplysninger og om fri utveksling av slike opplysninger samt om oppheving av direktiv 95/46/EF [personvernforordningen].
[2] Den engelske versjonen av art. 5 (1) a) er: «Personal data shall be (…) processed (…) fairly and in a transparent manner».
[3] Personvernforordningens fortalepunkt 6.
[4] Dag Wiese Schartum, *Personvernforordningen – en lærebok*, Fagbokforlaget, 2020, s. 29.
[5] Schartum (2020), s. 29.

Utviklingen av KI utfordrer måten man ser på personvern. Mens KI fortusetter bruk av store mengder data, krever personvernet bruk av så lite data som mulig, jf. personvernforordningen art. 5 (1) c). Med nok personopplysninger og en algoritme kan alvorlig sykdom avdekkes tidligere og med høyere presisjon enn leger klarer, og kriminalitet forhindres.[6] Prisen å betale er svakere personvern. Dette reiser spørsmål om hvordan samfunnet kan dra nytte av de positive mulighetene KI gir, uten at det går på bekostning av personvernet. To forhold ved KI som gjennomgående trekkes frem som problematisk for opprettholdelsde av personvernet og personvernforordningen, er manglende transparens og urettferdighet. Dette medfører at KI, som kapittel 4-6 viser, utfordrer prinsippene om åpenhet og rettferdighet i personvernforordningen art. 5 (1) a).

Manglende transparens knytter seg til at det er vanskelig å forklare hvordan KI fungerer. Problematikken oppstår på grunn av måten KI utvikles på, og omtales gjerne som «the black box problem» eller på norsk «den sorte boksen».[7] En av hovedfordelene med kunstig intelligente systemer er at de kan prosessere langt større mengder data enn mennesker. Dette gjør at KI kan se sammenhenger mennesker ikke ser, som kan bidra til at riktigere beslutninger tas raskere. Utfordringen er at det tar mennesker uoverkommelig lang tid å regne ut logikken bak utregningene til KI-systemene, og selv utviklerne får et problem når logikken bak KI skal forklares. Manglende evne til å forklare hvordan KI fungerer, vil, som oppgavens kapittel 4 viser, være i strid med åpenhetsprinsippet i personvernforordningen art. 5 (1) a).

Algoritmer gir ikke alltid rettferdige svar. KI trenes ofte på store mengder data, som den bruker for å bygge opp en mal for hvordan den skal løse ukjente problemstillinger. Dersom dataen KIen trenes på inneholder skjevheter, for eksempel dersom dataen inneholder diskriminering, vil den lære av skjevhetene og ende med å komme med diskriminerende utfall.[8] Problematikken omtales gjerne som «machine bias», på norsk skjevheter eller forutinntatthet. Det norske språket har ikke et enkeltstående dekkende begrep for machine bias, og det engelske begrepet brukes derfor i oppgaven. Som drøftelsen i kapittel 5.10. viser,

[6] Cade Metz, "A.I. Shows Promise Assisting Physicians", *The New York Times*, 11.02.2019, https://www.nytimes.com/2019/02/11/health/artificial-intelligence-medical-diagnosis.html (lest 07.12.2021); Lisa Quest, Anthony Charrie, Lucas du Croo de Jong og Subas Roy, "The Risks and Benefits of Usin AI to Detect Crime", *Harvard Business Review*, 09.08.2018, https://hbr.org/2018/08/the-risks-and-benefits-of-using-ai-to-detect-crime (lest 07.12.2021)
[7] Datatilsynet, *Kunstig intelligens og personvern*, 2018, s. 12.
[8] Se kapittel 2 og 6.2.

strider KI som inneholder machine bias mot rettferdighetsprinsippet i personvernforordningen art. 5 (1) a).

Machine bias, kombinert med manglende evne til å forkare hvordan KI kommer frem til svarene den gjør, får det til å virke nærmest umulig å overholde personvernforordningens art. 5 ved bruk av KI. Det er ikke bare de registrerte som har interesse av at KI er i tråd med kravene i personvernforordningen. Etterlevelse av reglene bygger tillit, som gjør folk villige til å dele data, som man igjen er avhengig av for å utvikle gode løsninger.[9] Alle involverte har en gjensidig interesse av å utvikle KI i tråd med personvernforordningen. Det er nettopp å kombinere personvern med utvikling av KI som er målet for prosjektene i Datatilsynets KI-sandkasse, noe som understreker problemstillingens aktualitet.[10]

For å konkretisere drøftelsen av åpenhets- og rettferdighetsprinsippet vil oppgavens avsluttende del fokusere på to eksempler på bruk av KI som beslutningsstøtte, som begge er i strid med åpenhets- og rettferdighetsprinsippet i personvernforordningen art. 5 (1) a). Eksemplene viser hvordan det paradoksalt nok er fordelen med KI, at den kan behandle store mengder opplysninger og nyttiggjøre seg av disse på en måte mennesker ikke klarer, som kan medføre at KI ikke kan benyttes i tråd med personvernregelverket.

1.2 Rettskilder og metodiske utfordringer

Oppgaven tar utgangspunkt i personvernforordningen. Hjemmelsgrunnlaget for personvernforordningen finnes i EUs primærrett, nærmere bestemt i Den europeiske unions pakt om grunnleggende rettigheter[11] art. 8 nr. 1 og i traktaten om Den europeiske unions virkeområde (TEUV) [12] art. 16 nr. 1, der retten til personvern er nedfelt.[13] Personvernforordningen utgjør en del av EUs sekundærlovgivning, som vil si at forordningen er en konkretisering av de politiske mål som er nedfelt i EUs primærrett.[14]

[9] Heidi Sævold, «Skal hjelpe norske selskaper å sikre trygg bruk av kunstig intelligens: Nå ønsker de innspill om konkrete temaer», *digi.no*, 12.10.2020, https://www.digi.no/artikler/skal-hjelpe-norske-selskaper-a-sikre-trygg-bruk-av-kunstig-intelligens-na-onsker-de-innspill-om-konkrete-temaer-br/500742 (lest 27.11.2021).
[10] Datatilsynets nettsider, «Sandkassesiden», https://www.datatilsynet.no/regelverk-og-verktoy/sandkasse-for-kunstig-intelligens/ (lest 23.11.2021).
[11] Charter of Fundamental Rights of the European Union 2000/C 364/01.
[12] Consolidated version of the The Treaty on the Function of the European Union (TFEU) 2012/C 326/01.
[13] Fortalepunkt 1.
[14] Odd Stemsrud, *EØS-rett i et nøtteskall*, Gyldendal, 2016, s. 69.

Personvernforordningen trådte i kraft i 2018, og er en relativt ny rettsakt. Rettskildebildet er derfor noe tynt, spesielt når det kommer til autoritative rettskilder. Personvernforordningen ble vedtatt den 27. april 2016, og erstattet det tidligere personverndirektivet 95/46/EF (personverndirektivet) da den trådte i kraft i 2018.[15] Ettersom personvernforordningen er en forordning, gjelder den direkte og er bindende i sin helhet, ord for ord, i alle EUs medlemsland, jf. TEUV art. 288 (2). Forordninger benyttes av EU på rettsområder der det er behov for en ensartet tolkning og presisering i alle land.[16] Det var nettopp enhetlighet som var målet med personvernforordningen.[17]

EU-lovgivning får ikke direkte virkning i EFTA-landene. Norge har likevel gjennom EØS-avtalen forpliktet seg til å gjennomføre store deler av EU-retten i norsk rett, og håndheve den på samme måte som EU-statene gjør.[18] For at lovgivning fra EU skal få virkning i EFTA-landene, må rettsaktene først innlemmes i EØS-avtalen. Forordninger skal etter EØS-avtalen art. 7 gjennomføres i nasjonal rett «som sådan», altså ord for ord. Ettersom Norge har et dualistisk system, må EU-rettsaktene også inkorporeres i norsk rett før de får virkning i Norge.[19] Personvernforordningen er innlemmet i EØS-avtalen, og gjennomført i norsk rett gjennom ny personopplysningslov som trådte i kraft 20. juli 2018.[20] Personvernforordningen gjelder som norsk lov etter § 1.

All materiell EØS-rett skal i tråd med homogenitetsprinsippets presumpsjon tolkes og anvendes på samme måte som i EU.[21] EU-retten særpreges av at den er autonom, hvilket innebærer at den etablerer sine egne løsninger og tolkningslære.[22] Også for EU-retten skal det tas utgangspunkt i rettsaktenes ordlyd, men ordlyden tillegges mindre vekt enn etter norsk rettskildelæren. Ettersom det er over 20 ulike språkversjoner, som alle er offisielle, er det

[15] Europaparlaments- og rådsdirektiv 95/46/EF av 24. oktober 1995 om beskyttelse av fysiske personer i forbindelse med behandling av personopplysninger og om fri utveksling av slike opplysninger.
[16] Fredrik Sejersted, Finn Arnesen, Ole-Andreas Rognstad, Sten Foyn, Olav Kolstad, *EØS-rett*, 3. utgave, Universitetsforlaget, 2011 s. 53.
[17] Fortalepunkt 9 og 10.
[18] Avtale om Det europeiske økonomiske samarbeidsområde av 2. mai 1992 (EØS-avtalen), AVT-1992-05-02-1; Sejersted mfl., (2011), s. 21.
[19] Sejersted mfl. (2011), s. 87.
[20] EØS-avtalen, vedlegg XI nr. 5e; Lov av 15. Juni 2018 nr. 38 om behandling av personopplysninger (personopplysningsloven).
[21] Sejersted mfl. (2011), s. 22.
[22] Ibid., s. 44.

viktigere å komme frem til meningen med bestemmelsene, heller enn å fintolke formuleringene.[23]

Et annet særtrekk ved tolkning av EU-rettskildene er at de skal tolkes dynamisk, i lys av utviklingen som har skjedd på tidspunktet for anvendelsen av bestemmelsene.[24] For personvernforordningen betyr dette at bestemmelsene må tolkes i tråd med den teknologiske utviklingen som skjer. Et virkemiddel for å sikre dynamisk tolkning er forordningens teknologinøytralitet. I EU-retten legges det også stor vekt på generelle prinsipper, eksempelvis åpenhets- og rettferdighetsprinsippet i personvernforordningen art. 5 (1) a), som anvendes aktivt i den rettslige argumentasjonen.[25]

Ved tolkningen av rettskildene er formålet sentralt. Fortalene til rettsaktene er sentrale kilder for å kartlegge rettsaktenes overordnede og konkrete formål. Fortalene er et utslag av kravet om at alle rettsakter skal begrunnes i TEUV art. 296. Fortalene er ikke bindende, men kan ha stor betydning som veiledning til hvordan artiklene skal forstås, og som tolkningsbakgrunn.[26] Rettskildemessig har fortalene noe lik funksjon som forarbeider etter norsk rettskildelære.[27] Ettersom det foreligger få andre tungtveiende rettskilder, vil fortalen til personvernforordningen («fortalen») være en sentral tolkningsfaktor.

Praksis fra EU-domstolen er en annen sentral kilde. EU-domstolen har enerett på en autoritativ tolkning av EU-retten, og øvrige EU-institusjoner, nasjonale myndigheter og domstoler plikter å legge EU-domstolens tolkninger til grunn.[28] Ettersom personvernforordningen er en forholdsvis ny rettsakt, foreligger det få avgjørelser fra EU-domstolen angående tolkningen av bestemmelsene. Jeg har derfor ikke funnet uttalelser fra EU-domstolen som gir hensiktsmessige bidrag til oppgaven.

Det finnes noe juridisk teori om åpenhets- og rettferdighetsprinsippet i personvernforordningen art. 5 (1) a). Selv om juridisk litteratur har begrenset rettskildemessig vekt, vil oppgaven vise til juridisk teori som tolkningsbidrag.

[23] Ibid., s. 45.
[24] Sejersted mfl. (2011), s. 45.
[25] L.c.
[26] Eva Jarbekk og Simen Sommerfeldt, *Personvern og personvernforordningen i praksis*, Cappelen Damm, 2019, s. 36.
[27] Sejersted mfl. (2011), s. 53.
[28] Ibid., s. 55.

Personvernrådet (European Data Protection Board) er opprettet i tråd med personvernforordningen art. 68 og 70. Personvernrådet er et uavhengig organ, som kommer med uttalelser og retningslinjer. Retningslinjene utgjør viktige rettskilder som skal sørge for enhetlig tolkning og anvendelse av personvernforordningen. Ettersom Personvernrådet består av representanter fra nasjonale datatilsyn og EUs eget datatilsyn (European Data Protection Supervisor), gir Personvernrådets retningslinjer og uttalelser uttrykk for en felles forståelse av reglene i personvernforordningen, på tvers av landegrensene.[29]

Personvernrådet er en videreføring av Artikkel 29-gruppen (Article 29 Working Party), som fylte samme rolle for personverndirektivet. Uttalelser fra Artikkel 29-gruppen kan fortsatt være relevante der de gjelder videreførte bestemmelser fra personverndirektivet. Personvernrådet har også vedtatt enkelte av Artikkel 29-gruppens veiledninger som fortsatt gjeldende.[30] Uttalelsene og retningslinjene til Personvernrådet er ikke rettslig bindende. Ettersom Personvernrådets kompetanse er direkte hjemlet i personvernforordningen, er den faktiske rettskildemessige vekten stor. Det skal derfor gode grunner til for å se bort ifra Personvernrådets retningslinjer. Sett sammen med mangelen på tungtveiende, autoritative rettskilder, utgjør retningslinjer og uttalelser fra Personvernrådet i praksis en sentralt tolkningsfaktor for personvernforordningen.

I oppgaven tas det hensyn til uttalelser fra EUs ekspertgruppe på KI (Artificial Intelligense High Level Expert Group, heretter KI-ekspertgruppen). EUs ekspertgrupper er satt ned av EU-kommisjonen. Ekspertgruppenes oppgave er å legge grunnlaget for regelverk eller strategier for EU-kommisjonen. For eksempel var de syv etiske retningslinjene foreslått av KI-ekspertgruppen veiledende for kommisjonens foreslåtte Artificial Intelligence Act (heretter AIA).[31] Uttalelsene fra ekspertgruppene er kun veiledende, og innholdet utgjør ikke rettskilder før de eventuelt vedtas av EU-kommisjonen som rettsakter. Uttalelsene har derfor, i likhet med juridisk teori, ikke større vekt enn det argumentenes kvalitet tilsier. Ekspertgruppenes uttalelser har likevel et skinn av demokratisk legitimitet, ettersom ekspertgruppene er oppnevnt av Kommisjonen som den eneste EU-institusjonen som legger

[29] Datatilsynets nettsider, «Det europeiske personvernrådet», oppdatert 19.07.2020c, https://www.datatilsynet.no/regelverk-og-verktoy/internasjonalt/personvernradet/ (lest 06.10.2021).
[30] Personvernrådet, *Endorsement 1/2018 of Working Party Article 29 Documents*, 2018.
[31] EU-kommisjonens nettsider, "High-level expert group on artificial intelligence", https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai (lest 09.11.2021); Proposal for a Regualtion of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (heretter forkortet AIA for Artificial Intelligence Act).

frem lover. Det må også vektelegges at det er eksperter på feltet som uttaler seg i veiledningene.

I oppgaven vil det også bli tatt hensyn til et varslet vedtak fra det norske Datatilsynet i det som omtales som «IB-saken».[32] Vedtak fra nasjonale datatilsyn er ikke bindende for andre nasjonale datatilsyn eller EU-organer. Det varslede vedtaket gir uttrykk for hvordan det norske Datatilsynet tolker personvernforordningen, og kan få tilsvarende vekt som forvaltningspraksis i Norge.

1.3 Avgrensninger

Med hensyn til innleveringsfristen avgrenses det til kilder som er publisert før 01.12.2021.

1.3.1 Automatiserte avgjørelser og profilering

Oppgavens problemstilling er hvordan åpenhets- og rettferdighetsprinsippet i personvernforordningen art. 5 (1) begrenser bruken av KI som beslutningsstøtte. Problemstillingen avgrenser dermed mot avgjørelser som utelukkende er basert på automatisert behandling, som nevnt i personvernforordningen art. 22 (1). Etter art. 22 (1) har den registrerte rett til ikke å være gjenstand for en utelukkende automatisert avgjørelse, som har rettsvirkning, eller annen tilsvarende betydelig påvirkning, for vedkommende. I oppgaven er det altså snakk om behandlingstilfeller der et menneske har det siste ordet i beslutningen, og KI brukes som støtte for mennesket som avgjør saken.

Profilering er definert som «enhver form for automatisert behandling av personopplysninger som innebærer å bruke personopplysninger for å vurdere visse personlige aspekter knyttet til en fysisk person, særlig for å analysere eller forutsi aspekter som gjelder nevnte fysiske persons arbeidsprestasjoner, økonomiske situasjon, helse, personlige preferanser, interesser, pålitelighet, atferd, plassering eller bevegelser», jf. personvernforordningen art. 4 (4). I art. 22 (1) er profilering nevnt som en form for utelukkende automatisert avgjørelse. At profilering er en «automatisert» behandling betyr at profilering kan benyttes for å fatte en utelukkende automatisert avgjørelse, men profilering kan også være et av flere elementer i en helt eller delvis automatisert avgjørelse. Noen profileringstilfeller vil dermed falle innunder

[32] Datatilsynet, sak 20/03087 (IB-saken), 2020a.

personvernforordningen art. 22 (1), og andre ikke. Oppgaven avgrenser derfor ikke mot profileringstilfeller. Et eksempel der profilering brukes i en delvis automatisert avgjørelse er tilfeller der en kunde fyller ut et skjema for å få lån i en bank. Opplysningene analyseres av en algoritme som svarer på hvor sannsynlig det er at kunden vil tilbakebetale lånet. En bankansatt avgjør så om kunden skal innvilges lån, blant annet på bakgrunn av svaret fra algoritmen.

Som forklaringen av KI i kapittel 2 viser, sammenfaller definisjonen av profilering i personvernforordningen art. 4 (4) i stor grad med måten maskinlærings-KI fungerer. Behandlingen KI gjør av personopplysninger er automatisert i den forstand at man mater KIen med personopplysninger som KIen analyserer og gir en vurdering av, uten menneskelig innblanding i prosessen. En avgrensning mot profileringstilfeller ville i stor grad utelukket behandling av KI.

1.3.2 Øvrige prinsipper i personvernforordningen art. 5

Personvernforordningen bygger på en rekke prinsipper og formål som kommer til uttrykk både gjennom fortalen og de enkelte artiklene. Personvernforordningen art. 5 lister opp flere prinsipper som skal følges ved all behandling av personopplysninger. I oppgaven avgrenses det til rettferdighets- og åpenhetsprinsippene i art. 5 (1) a). Lovlighetsprinsippet, som også er nevnt i art. 5 (1) a), skal ikke behandles fordi prinsippet ikke reiser like prinsipielt interessante spørsmål som åpenhet og rettferdighet. Heller ikke de øvrige prinsippene i art. 5 (1) b)-f) skal behandles. Prinsippene gjennomgås likevel kort i oppgavens kapittel 3, for å vise at også de påvirker hvordan KI kan brukes som beslutningsstøtte. Av hensyn til oppgavens omfang er utgangspunktet for avklaringen av prinsippenes innhold at behandling skjer ved bruk av KI. Avklaringen sier likevel noe om innholdet av prinsippene også generelt.

1.4 Den videre fremstillingen

I oppgaven utforsker jeg hvordan åpenhets- og rettferdighetsprinsippet i personvernforordningen art. 5 (1) a) skal forstås når KI brukes som beslutningsstøtte. For å drøfte prinsippene i lys av KI, fokuserer kapittel 2 på å definere hvordan KI skal forstås i denne oppgaven, og hvordan KI utvikles. I kapittel 3 redegjøres det for sentrale begreper og

aktører i personvernforordningen, prinsippene i art. 5 (1) sin betydning i rettskildebildet og de øvrige prinsippene i art. 5 (1).

Kapittel 4 og 5 fokuserer på å utrede hvordan åpenhets- og rettferdighetskapittelet skal forstås i lys av KI. I kapittel 4 drøftes åpenhetsprinsippets innhold med fokus på kravet til informasjonens innhold og formkravene til informasjonen. Åpenhetsprinsippet vil også ses opp mot problematikken rundt den svarte boksen. I kapittel 5 drøftes rettferdighetsprinsippets innhold. Rettferdighetsprinsippet ses deretter opp mot problematikken rundt machine bias. Prinsippene har noen overlappende sider, og drøftelsene vil til en viss grad I kapittel 6 presenteres IB-saken og beslutningsstøtte i domstolene, som to eksempler på bruk av KI som beslutningsstøtte, og hvordan åpenhets- og rettferdighetsprinsippet har slått ut i de tilfellene. Avslutningsvis reflekterer jeg over funnene i oppgaven og veien videre for personvern og KI.

2 Kunstig intelligens

2.1 Definisjon av kunstig intelligens

For kunne å drøfte hvordan åpenhet og rettferdighet påvirker bruken av KI, må KI defineres. Oppgaven forsøker ikke å definere KI utover å klargjøre de avgrensninger som er nødvendige for at problemstillingen oppgaven skal svare på blir tilstrekkelig klar. I tillegg er det viktig med en viss oversikt over hvordan KI er bygget opp og fungerer, for å kunne forstå de rettslige problemene slike systemer skaper.

Det finnes ingen omforent, entydig definisjon på hva KI er, eller hva som må til for at et system skal kunne defineres som KI.[33] Avhengig av hvilket fagområde man tilhører kan ulike definisjoner benyttes, og definisjonen utvikler seg gjerne i takt med den teknologiske utviklingen.[34]

I *Nasjonal strategi for kunstig intelligens*, som har tatt utgangspunkt i EUs KI-ekspertgruppes definisjon,[35] defineres KI som:

> «Kunstig intelligente systemer utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål. Enkelte KI-systemer kan også tilpasse seg gjennom å analysere og ta hensyn til hvordan tidligere handlinger har påvirket omgivelsene.»[36]

Definisjonen er vid og teknologinøytral, og rommer mange former for KI. En vid definisjon kan være hensiktsmessig ettersom man ikke trenger å redefinere KI i takt med at KI-teknologien utvikler seg. KI kan nemlig utvikles gjennom ulike tilnærminger og teknikker, se kapittel 2.

I AIA art. 3 (1) er det foreslått å definere KI ut ifra utviklingsmetoden som er brukt. EU-kommisjonens foreslåtte definisjon er dermed ikke teknologinøytral. De ulike

[33] Christian Bendiksen og Eirik Norman Hansen, *Når juss møter AI*, Gyldendal, 2019, s. 11.
[34] Kommunal og moderniseringsdepartementet, *Nasjonal strategi for kunstig intelligens*, 2020, s. 9.
[35] Independent High Level Expert Group set up by the European Commission, *A definition of AI: Main capabilities and disciplines*, 2019a, s. 1.
[36] Kommunal og moderniseringsdepartementet (2020), s. 9.

utviklingsmetodene er listet opp i regelverkets Anneks 1 og er maskinlæringstilnærminger, logikk- og kunnskapsbaserte tilnærminger og statistiske tilnærminger.[37]

De to vanligste utviklingsmetodene for KI er regelbaserte systemer og maskinlæring. Regelbaserte systemer utformes av mennesker, og systemet følger et fast regelsett. Dette innebærer at det er relativt enkelt å forstå hvordan algoritmene fungerer og den underliggende logikken bak KIens konklusjon kan forklares.[38] Dette medfører imidlertid at regelbaserte systemer ikke fungerer optimalt når systemet, eller problemet det skal løse, når en viss kompleksitet, for eksempel i skjønnsmessige avgjørelser. KI bygget opp på regelbaserte systemer får dermed et begrenset bruksområde.[39] Oppgaven avgrenser derfor mot KI utviklet med regelbaserte systemer.

I oppgaven avgrenses det til KI utviklet etter maskinlæringsteknikken, som forklares nærmere i kapittel 2.2, fordi det de siste årene har blitt den vanligste måten å utvikle KI som benyttes til skjønnsmessige avgjørelser og som kan gi skjeve utfall det er vanskelig å forklare. Det er derfor KI utviklet med maskinlæringsteknikk som har størst risiko for å være i strid med åpenhets- og rettferdighetsprinsippet i personvernforordningen art. 5 (1) a). Definisjonen ligger også innenfor definisjonen foreslått i AIA, og oppgaven er fortsatt relevant dersom regelverket vedtas slik det er foreslått. Oppgavens drøftelser og konklusjoner kan også være relevante for KI som ikke er utviklet gjennom maskinlæringsteknikken.

2.2 Utvikling av kunstig intelligens – maskinlæring og dyplæring

Maskinlæringssystemer er, i motsetning til regelbaserte systemer, ikke bygget på forhåndsskrevne regler laget av mennesker. Det særegne ved maskinlæringssystemer er at de lærer fra erfaringer og kan handle uten å bli spesifikt programmerte til akkurat den handlingen.[40] Når man utvikler maskinlæringssystemer gir man et system av algoritmer, som er opprettet av utviklere, tilgang på data (input) som systemet på egenhånd analyserer og trekker ut erfaringer fra, basert på mønstre og sammenhenger i dataene. Algoritmene lærer fra dataene, og forbedrer seg selv ved at parameterne i algoritmene justeres ettersom den ser flere

[37] AIA, Anneks 1, a)-c).
[38] NOU 2020:11, *Den tredje statsmakt – Domstolene i endring*, s. 254
[39] Ibid., s. 254
[40] Bendiksen og Hansen (2019), s. 19.

læringseksempler, gjerne fra historisk data. Det er flere faktorer som kan påvirke hvor treffsikker KIen blir. Resultatene kan bli mer treffsikre jo mer mer data algoritmene lærer fra, men treffsikkerheten avhenger også av dataenes kvalitet og hvordan dataene er bygget opp. Gjennom prøving og feiling blir systemet mer og mer nøyaktig, og denne prosessen kalles å trene algoritmen.[41] Når en algoritme er tilstrekkelig trent, klarer den å prosessere ukjent informasjon og komme med et eget, forhåpentligvis korrekt, svar (output). Dette innebærer at mennesker ikke har kontroll over prosessen fra KIen gis informasjon og frem til svaret er gitt.

Det finnes ulike typer maskinlæring. En grovinndeling kan gjøres mellom veiledet læring, ikke-veiledet læring og mellomløsningen forsterket læring.[42] Ved veiledet maskinlæring gis algoritmene tilgang på merket data og ønsket utfall av dataen.[43] Algoritmene finner sammenhenger mellom dataene og det ønskede utfallet. Etter nok læring kan algoritmene gi svar på ukjente data. Ved ikke-veiledet læring får systemet tilgang på ustrukturerte data og utfallet. Systemet må selv finne mønstre i dataene, uten retningslinjer for hva den skal lete etter.[44] Dette kan for eksempel være faktumet i en rettssak, og utfallet av rettssaken.

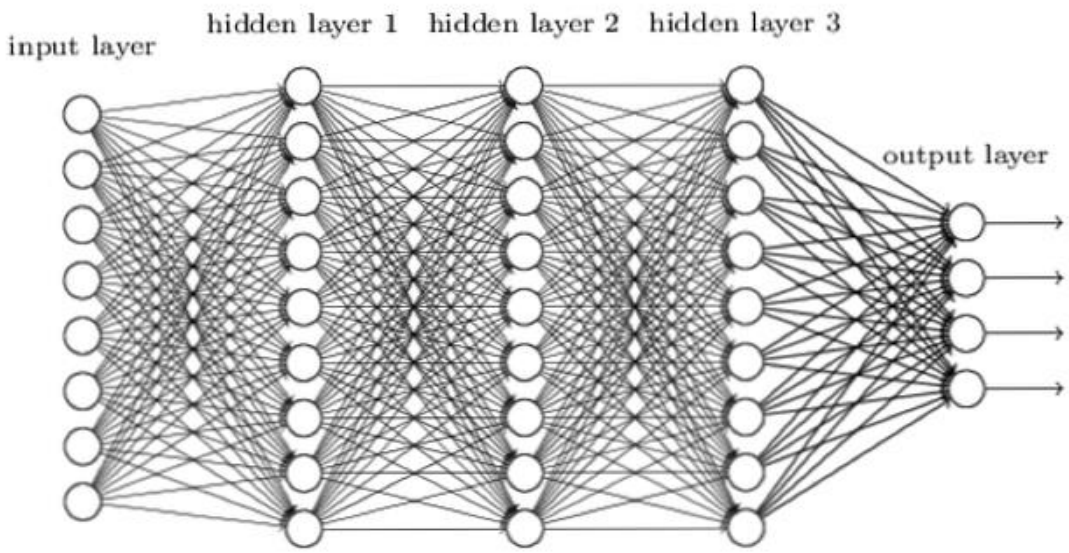
En siste kategori er forsterket læring. Da får algoritmene et mål de skal oppnå, ulike handlinger de kan utføre og tilbakemeldinger på handlingene de velger. Systemet prøver ulike handlinger og tilpasser valgene den tar etter hva som gir best resultat.[45] For mennesker kan det være umulig å se de små justeringene i handlingsalternativene som vil gi best resultat. Det vil videre ikke skilles mellom KI utviklet med veiledet, ikke-veiledet og forsterkende læringsmodeller, da alle har samme særegne evne til å endre output basert på erfaringer opparbeidet under treningen.[46]

Når algoritmer trenes på store mengder data, dannes nevrale nettverk. Nettverkene etterligner nervesystemet i den biologiske hjernen, og består av sammenhenger mellom ulike faktorer.[47] Illustrasjonen i figur 1 viser hvordan de nevrale nettverkene er bygget opp med ett inputlag, flere skjulte lag og ett outputlag. Algoritmene er i stand til å se sammenhenger som mennesker ikke klarer å fange opp på grunn av de store datamengdene systemet kan

[41] NOU 2020:11 s. 254
[42] Bendiksen og Hansen (2019), s. 19.
[43] L.c.
[44] L.c.
[45] L.c.
[46] L.c.
[47] Ibid., s. 18.

prosessere. Jo mer data KIen får trene algoritmene på, jo flere lag med nettverk kan dannes, og jo flere lag med nettverk som dannes, jo «dypere» blir nettverkene.

Utviklerne bestemmer antall lag med nettverk på forhånd, men algoritmen justerer selv på vekten mellom parameterne. Når nettverkene blir dype nok, blir modellen så komplekst sammensatt at mennesker ikke klarer å omfavne og forstå hvilke tilpasninger algoritmen gjør. Det ville tatt mennesker uoverkommelig lang tid å regne ut matematikken bak et stort nevralt nettverk. Dette gjør det umulig selv for utviklerne av KIen å bryte opp beslutningsprosessen eller resonnere seg frem til hvorfor systemet løste oppgaven som den gjorde. Det er kompleksiteten av modellen, som blant annet kommer av de dype nevrale nettverkene, som danner grunnlaget for den svarte boks' problem; justeringen algoritmen gjør i de skjulte lagene medfører at beslutningsprosessen til KIen er like lite transparent som en svart boks. Dette setter åpenhetsprinsippet på spissen, med hensyn til at man skal vite hvordan personopplysningene om seg selv brukes, se oppgavens kapittel 4.



Figur 1: Illustrasjon av dype nevrale nettverk med inputlag, tre skjulte lag og outputlag.[48]

Mens maskinlærings-KI er attraktiv for å ta beslutninger på en langt mer effektiv måte, der man tar hensyn til langt flere faktorer og sammenhenger enn det mennesker klarer, er den ikke feilfri. Feilene stammer gjerne fra at dataene som brukes for å trene KIen ikke er gode nok, eksempelvis data som inneholder urettmessige skjevheter – machine bias. Machine bias kan defineres som systematiske og gjentatte feilslutninger i KI-systemer som medfører

[48] Michael Nielsen, *Neural Networks and Deep Learning*, kapittel 5, www.neuralnetworksanddeeplearning.com (lest 11.10.2021).

urettmessige utfall, eksempelvis at en gruppe mennesker prioriteres over en annen. Dette setter rettferdighetsprinsippet på spissen, da behandling av personopplysninger ikke må usaklig forskjellsbehandle, se oppgavens kapittel 5.

3 Kort om personvernforordningen

3.1 Sentrale begreper og aktører i personvernforordningen

Som nevnt er formålet med personvernforordningen fysiske personers personvern og fri flyt av personopplysninger i EØS. Personvernforordningen gjelder all behandling av **personopplysninger**. Med personopplysninger menes alle opplysninger om en identifisert eller identifiserbar fysisk person, jf. personvernforordningen art. 4 (1). Begrepet rommer et bredt spekter av opplysninger, så lenge de kan knyttes til en enkeltperson. Den det behandles opplysninger kalles **«den registrerte»**, jf. art. 4 (1). Med **behandling** menes enhver «operasjon» som gjøres med personopplysninger, f.eks. organisering og strukturering av opplysninger jf. personvernforordningen art. 4 (2). Dette omfatter tilfeller der KI mates med personopplysninger som sammenstilles, analyseres og konkluderes på.

Når personopplysninger behandles oppstår det en rekke plikter for den som behandler opplysningene, for eksempel må prinsippene i personvernforordningen art. 5 følges, jf. art. 5 (2). Det oppstår også flere rettigheter for den registrerte, eksempelvis rett til informasjon etter art. 12-14. Den som er ansvarlig for å følge opp pliktene i personvernforordningen kalles «**behandlingsansvarlig**», og behandlingsansvarlig er den som bestemmer formålet med behandlingen av personopplysningene og hvilke midler som skal benyttes i behandlingen, jf. art. 4 (7). Det er eksempelvis den behandlingsansvarlige som bestemmer at KI skal benyttes for å behandle personopplysninger. Det er vanlig å sette ut behandling av personopplysninger til andre, og den som behandler personopplysninger på vegne av behandlingsansvarlig kalles **«databehandler»**, jf. art. 4 (8). Dette vil være tilfellet der en bank (behandlingsansvarlig) sender opplysninger om sine kunder (de registrerte) til et inkassobyrå (databehandler) som benytter seg av KI for å beregne betalingsevnen til kundene (behandlingsaktivitet). De sentrale aktørene i behandling av personopplysninger er altså den registrerte, behandlingsansvarlig og eventuelt databehandler.

3.2 Prinsipper for behandling av personopplysninger – artikkel 5

3.2.1 Prinsippenes betydning i rettskildebildet

Prinsippene i personvernforordningen art. 5 (1) tilsvarer delvis prinsippene i tidligere personverndirektiv art. 6 og personopplysningsloven 2000 § 11. Likevel var ikke alle prinsippene angitt i tidligere regelverk, i hvert fall ikke eksplisitt. Prinsippene om lovlighet, formålsbegrensning, dataminimering, korrekte opplysninger og lagringsbegrensning var oppgitt i personverndirektivet art. 6 (1) a)-e). Prinsippene om åpenhet og rettferdighet hadde ikke direkte hjemmel. Prinsippet om åpenhet kom likevel til uttrykk gjennom personverndirektivet art. 10, og rettferdig behandling blir nevnt i fortalen til personverndirektivet punkt 28 og 38. Det fulgte også av personverndirektivet art. 6 (1) a) at personopplysninger skulle behandles på «rimelig (…)vis». Betydningen av rimelig overlapper delvis med rettferdig, men ordlyden er ny. At åpenhets- og rettferdighetsprinsippet nå er knesatt i en egen artikkel omtalt som «grunnloven» i den nye forordningen, reiser spørsmål om prinsippenes innhold og betydning.[49]

Prinsippene i art. 5 legger premissene for hva som er lovlig behandling av personopplysninger. De utgjør i seg selv rettslige krav som må følges av alle behandlingsansvarlige og databehandlere.[50] Eventuelle unntak fra prinsippene må ha hjemmel.[51] I tillegg er prinsippene sentrale tolkningsfaktorer for de øvrige bestemmelsene i forordningen og i nasjonal særregulering om behandling av personopplysninger.[52] Ved spørsmål om hvordan en annen bestemmelse i personvernforordningen skal tolkes, kan man se til prinsippene i art. 5 (1). Personvernforordningens art. 5 er på mange måter ankeret man kan vende tilbake til i de mange uavklarte situasjonene innen personvern, og bruken av KI er en av de.

[49] Jarbekk og Sommerfeldt (2019), s. 46
[50] Skullerud m.fl. (2018), s. 74.
[51] L.c.
[52] L.c.

3.2.2 Øvrige prinsipper i personvernforordningen artikkel 5

De øvrige prinsippene i personvernforordningen art. 5 og hvordan de kan komme i konflikt med KI vil kort berøres her.

Lovlighet i art. 5 (1) a)

Lovlighet nevnes sammen med åpenhet og rettferdighet i art. 5 (1) a). Prinsippet innebærer at kravene som følger av personvernforordningen, herunder rettslig grunnlag etter art. 6 og 9, samt at behandling av personopplysninger må være i samsvar med EU-retten forøvrig.[53] Dersom AIA blir vedtatt, innebærer lovlighetsprinsippet at behandling av personopplysninger må samsvare med denne rettsakten.

Øvrige prinsipper i art. 5 (1) b)-f) og (2)

Etter art. 5 (1) b) skal personopplysninger samles inn for «spesifikke, uttrykkelig angitte og berettigede formål og ikke viderebehandles på en måte som er uforenlig med disse formålene (...)». Formålet må være konkret angitt, behandlingsansvarlig kan ikke «se an» hva personopplysningene skal brukes til. Den registrerte må også forstå formålet for å kunne vurdere om personopplysningene behandles i tråd med formålene. Når KI utvikles viser det seg ofte at modellene er nyttige på andre bruksområder enn tiltenkt. I tillegg kan KI ved hjelp av informasjonsmønstre vise mer informasjon enn det som var formålet. Det kan bli fristende for behandlingsansvarlig å benytte seg av de nye bruksområdene og all informasjonen. Dette vil imidlertid bryte med prinsippet om formålsbegrensning.

Prinsippet om dataminimering i art. 5 (1) c) innebærer at databehandler skal samle inn så få personopplysninger som mulig for å oppnå formålet. I fortalepunkt 39 uttales det at opplysninger bare skal samles inn så lenge formålet med rimelighet ikke kan oppnås på annen måte. Prinsippet står i sterk kontrast til måten KI utvikles, der målet er å bruke så mye data som mulig for å utvikle de best mulige algoritmene.

Videre skal personopplysningene som behandles være korrekte og oppdaterte etter art. 5 (1) d). De skal heller ikke lagres lenger enn nødvendig etter art. 5 (1) e). At personopplysningene skal være korrekte blir problematisk dersom KIen inneholder skjevheter som gir feilaktige slutninger om en person. For eksempel kan KI som er trent på personopplysninger fra en

[53] Skullerud m.fl. (2018), s. 75; Schartum (2020), s. 88-89.

bestemt gruppe mennesker gi uriktige resultater når de brukes på en annen gruppe mennesker.[54]

Etter prinsippet om integritet og konfidensialitet skal personopplysninger behandles på en måte som sikrer tilstrekkelig sikkerhet for personopplysningene, jf. art. 5 (1) f). Behovet for sikkerhet stiger når man behandler større mengder personopplysninger eller gjennomfører kompliserte operasjoner, som er tilfellet for KI.

Den behandlingsansvarlige har ansvar for å dokumentere at prinsippene etterleves, jf. art. 5 (2). Kravet er nærmere konkretisert i personvernforordningen art. 24 om at «den behandlingsansvarlige gjennom egnede tekniske og organisatoriske tiltak [skal] sikre og påvise at behandlingen utføres i samsvar med denne forordning». Kravet reiser blant annet spørsmål om hvordan behandlingsansvarlig skal sikre og dokumentere rettferdighet i et system der man ikke vet hvordan algoritmene fungerer.

[54] Joy Buolamwini og Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", *Conference on fairness, accountability and transparency, Proceedings of Machine Learning Research* vol. 81 januar 2018, s. 77-91, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf (lest 07.12.2021).

4 Åpenhetsprinsippet

Dette kapittelet beskriver åpenhetsprinsippet i personvernforordningen art. 5 (1) a) . Åpenhetsprinsippet handler om å informere den registrerte. Drøftelsen er delt inn etter hvilke krav åpenhetsprinsippet stiller til informasjonens innhold og form, i henhold til ordlyden i forordningens art. 5 (1) a), fortalen, veiledning fra Personvernrådet og EUs ekspertgruppe på KI, og juridisk litteratur.

4.1 Overordnede hensyn

Ordlyden av «åpen» behandling i art. 5 (1) a) nedsetter prinsippet om transparens, som er et av de viktigste prinsippene i forordningen. Kravet om åpenhet er viktig for at den registrerte skal kunne forstå hva som skjer med sine opplysninger, og åpenhetsprinsippet må tolkes med dette som formål.[55] Den registrerte har krav på innsyn i og lett tilgjengelig og forståelig informasjon om behandlingen som finner sted.[56] Kravet til åpenhet er altså todelt; det må gis informasjon om selve behandlingsaktiviteten (innholdskrav), og den registrerte må forstå informasjonen (formkrav). Det er behandlingsansvarliges ansvar at informasjonen er forståelig og tilgjengelig for den registrerte, jf. personvernforordningen art. 5 (2), jf. (1) a).

Åpenhet er i presisert i personvernforordningen kapittel III, og innebærer blant annet at den registrerte skal ha kontroll på egne opplysninger, bli informert om behandlingen og forstå behandlingen av personopplysningene, jf. art. 13 og 14, på en forståelig måte, jf. art. 12. Åpenhet er essensielt for å skape tillit til behandlingsprosessen og fungerer som et verktøy så den registrerte kan vurdere om behandlingen tilfredsstiller personvernforordningens krav. Åpenhet er utgjør en forutsetning for at den registrerte kan utøve sine rettigheter etter personvernforordningen. For å oppnå dette er ikke informasjonskravene i personvernforordningen art. 12-14 uttømmende. Informasjonskravet må tilpasses behandlingaktiviteten.

[55] Jarbekk og Sommerfeldt (2019), s. 46-47.
[56] L.c.

4.2 Informasjonens innhold

Etter personvernforordningen art. 13 (2) f) og 14 (2) g) skal «forekomsten av automatiserte avgjørelser, herunder profilering, som nevnt i artikkel 22 nr. 1 og 4, og, i det minste i nevnte tilfeller, relevant informasjon om den underliggende logikken samt om betydningen og de forventede konsekvensene av en slik behandling» opplyses om.[57] Oppgaven avgrenser mot art. 22 (1)-tilfeller. Informasjonskravet i art. 13 (2) f) kommer derfor ikke direkte til anvendelse. Ordlyden av «i det minste i nevnte tilfeller» åpner imidlertid for at kravene også gjelder annen behandling enn art. 22 (1)-tilfellene. Av Artikkel 29-gruppens veiledning om automatiserte avgjørelser og profilering fremgår at det er «good practice» å følge informasjonskravene i art. 13 (2) f) ved all form for profilering.[58] Schartum legger, uten nærmere forklaring, til grunn at profilering likestilles med helt automatiserte avgjørelser i art. 13 (2) f), selv om profilering ikke innebærer en avgjørelse.[59]

Dersom informasjonskravet i art. 13 (2) f gjelder all profilering, vil det fortsatt ikke omfatte all bruk av KI. Ut ifra det overordnede formålet om åpenhet i art. 5 (1) a), kan ikke den registrerte uten informasjonen nevnt i art. 13 (2) f) forstå hvordan personopplysningene behandles, eller vurdere om behandlingen er rettferdig. For å sikre de overordnede hensynene bak personvern må informasjonskravene angitt i art. 13 (2) f) gjelde for bruk av KI som beslutningsstøtte. Ordlyden i art. 13 (2) f) brukes derfor som utgangspunkt i den videre drøftelsen.

Behandlingsansvarlig har dermed opplysningsplikt om at KI benyttes som beslutningsstøtte. Opplysningsplikt om bruk av KI er også lagt til grunn av EUs KI-ekspertgruppe i deres uttalelser om åpenhet i sine etiske retningslinjer for bruk av KI.[60] Ett av KI-ekspertgruppens syv prinsipper for etisk bruk av er at KI må være «transparent». Prinsippet er oversatt til «gjennomsiktig» i *Nasjonal strategi for kunstig intelligens*.[61] Selv om den norske

[57] Art. 13 og 14 angir ganske samsvarende informasjonskrav avhengig av om personopplysningene samles inn fra den registrerte eller ikke. Ordlyden i art. 13 (2) f) og 14 (2) g) er identisk, det henvises derfor kun til 13 (2) f).
[58] Artikkel 29-gruppen, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, revidert og tilsluttet 06.02.2018b.
[59] Schartum (2020), s. 171.
[60] Independent High Level Expert Group on Artificial Intelligence set up by the European Commission (2019b), *Ethics guidelines for trustworthy AI*, s. 18; Retningslinjene er basert på EUs charter om grunnleggende rettigheter og internasjonal menneskerettighetslovgivning, og har bunnet ut i syv prinsipper for etisk og ansvarlig utvikling av KI (KI-ekspertgruppen (2019b) s. 14.) Disse syv prinsippene er lagt til grunn av den norske regjeringen for ansvarlig utvikling og bruk av KI i Norge (Kommunal- og moderniseringsdepartementet (2020), s. 58).
[61] KI-ekspertgruppen (2019b), s. 14; Kommunal- og moderniseringsdepartementet (2020), s. 59.

oversettelsen av «transparent» er ulik i personvernforordningen og *Nasjonal strategi for kunstig intelligens*, samsvarer den engelske ordlyden i personvernforordningen og de etiske retningslinjene. KI-ekspertgruppens uttalelser er derfor relevante for tolkningen av hvordan åpenhetsprinsippet i personvernforordningen art. 5 (1) a), skal forstås ved bruk av KI. Der KI-ekspertgruppens uttalelser gjennomgås, brukes begrepet «gjennomsiktighetsprinsippet» for å klargjøre at det er prinsippet i de etiske retningslinjene, og ikke åpenhetsprinsippet i personvernforordningen som omtales.

Etter art. 13 (2) f) må også «den underliggende logikken» opplyses om. Hvis de registrerte ikke forstår beslutningsprosessen, kan de ikke håndheve sine rettighete. Kravet underbygges av KI-ekspertgruppens uttalelser om forklarbarhet som en fasett ved gjennomsiktighetsprinsippet.[62] Forklarbarhet innebærer muligheten til å forklare både den tekniske prosessen og de menneskelige beslutninger som tas i tilknytning til KIen. Den tekniske forklarbarheten innebærer at beslutningene som tas av KIen kan forstås, og logikken følges av mennesker.[63] I kapittel 4.4 om den sorte boksen forklares det hvorfor kravet kan være vanskelig å oppfylle. Kravet innebærer at behandlingsansvarlig må forklare rasjonale bak, eller kriteriene for avgjørelsen.[64] At informasjonen skal være «relevant», tilsier at det gis informasjon av betydning for å forstå grunnlaget for beslutningen.[65] Hvor grundig forklaringen må være tilpasses proporsjonalt etter innvirkning behandlingsaktiviteten har på de registrerte.

Etter art. 13 (2) f) skal den registrerte opplyses om forventede konsekvenser. Dette følges opp av fortalepunkt 39 om at registrerte bør gjøres oppmerksomme på «risikoer». En potensiell risiko ved KI er machine bias, som innebærer at behandlingen ikke er rettferdig, se kapittel 5.10. Åpenhetsprinsippet og rettferdighetsprinsippet henger tett sammen. Åpenhet er i stor grad en forutsetning for rettferdighet, og den registrerte må ha tilstrekkelig informasjon til å bedømme om personopplysningene er gjenstand for rettferdig behandling.[66]

Åpenhet bidrar til å sikre forutberegnelighet for den registrerte. Artikkel 29-gruppen understreker at den registrerte på forhånd skal kunne bedømme omfanget og konsekvensene av behandlingen på en måte som gjør at de ikke senere blir overrasket over måten

[62] KI-ekspertgruppen (2019b), s. 18.
[63] L.c.
[64] Artikkel 29-gruppen (2018b), s. 25.
[65] L.c.
[66] Jarbekk og og Sommerfeldt (2019), s. 47; Skullerud mfl. (2018), s 75 og 121.

personopplysningene er behandlet på.[67] Skullerud m.fl. tolker åpenhetsprinsippet slik at behandlingen må være forutsigbar for den registrerte, så vedkommende kan innrette seg etter behandlingen.[68] Tolkningene henger sammen med fortalepunkt 39 om at den registrerte bør ha informasjon om potensielle risikoer, og er i tråd med åpenhetsprinsippets formål om at den registrerte skal forstå hvordan personopplysningene behandles. Spesielt ved komplekse og tekniske behandlingsaktiviteter, som behandling med KI som beslutningsstøtte, er det vanskelig for den registrerte å forutse påvirkningen behandlingsaktivitetene vil ha. Dette tilsier et skjerpet informasjonskrav.[69]

Etter fortalepunkt 39 skal behandlingsansvarlig gi «ytterligere informasjon for å sikre en rettferdig og åpen behandling». Det samme fremgår av fortalepunkt 60 og 71, med den presisering at «de særlige omstendighetene rundt behandlingen av personopplysningene og sammenhengen den skjer i» skal hensyntas. Behandlingsansvarlig må altså foreta en konkret vurdering av hvert brukstilfelle, slik at informasjonens innhold gir den registrerte den informasjonen som er nødvendig for å sikre rettferdig og åpen behandling. Derfor er det ikke nødvendigvis tilstrekkelig med standardisert informasjon. Dette ble også lagt til grunn av det norske Datatilsynet i IB-saken som gjennomgås i kapittel 6.1.[70] Der KI benyttes må informasjonen eksempelvis tilpasses behandlingsaktivitetens kompleksitet, hvor inngripende behandlingen er for den registrertes interesser og rettigheter og kunnskapen de registrerte har om KI.

KI-ekspertgruppen trekker også frem etterprøvbarhet som en fasett av gjennomsiktighetsprinsippet.[71] Etterprøvbarhet innebærer at datasettene og prosessene som leder opp til beslutningen skal bli så godt dokumentert som mulig for å sikre etterprøvbarhet og øke gjennomsiktigheten. Når etterprøvbarhet sikres, legger det også til rette for mulighet til å revidere algoritmen, samt forklarbarhet.

[67] Artikkel 29-gruppen, *Guidelines on transparency under Regulatiopn 2016/679*, revidert og tilsluttet 11.04.2018a, avsnitt 10.
[68] Skullerud mfl. (2018), s. 75.
[69] Artikkel 29-gruppen (2018a), avsnitt 10.
[70] Datatilsynet, sak 20/03087 (IB-saken) dokumentnummer 14, 07.08.2020b, s. 5.
[71] KI-ekspertgruppen (2019b), s. 18.

4.3 Forståelig informasjon

Personvernforordningen art. 12 gir nærmere vilkår for åpen behandling, og bestemmelsen må leses i lys av art. 5 (1) a). Kravene til behandlingsansvarlig i art. 12 kan deles inn i tre hovedelementer; 1) behandlingsansvarlig skal gi informasjon som gjør den registrerte i stand til å ta stilling til om personopplysningene er gjenstand for rettferdig behandling,[72] 2) behandlingsansvarlig skal sørge for at informasjonen er forståelig for den registrerte gjennom et «klart og enkelt språk», jf. art. 12 (1), 3) behandlingsansvarlig skal, gjennom å sørge for åpenhet rundt behandlingen, legge til rette for at de registrerte kan utøve de rettighetene den har etter personvernforordningen, jf. art. 12 (2).[73] Hovedfokuset i kapittelet er det andre hovedelementet.

Informasjonen skal gis på en «kortfattet, åpen, forståelig og lett tilgjengelig måte og på et klart og enkelt språk», jf. Art. 12 (1). Ordlyden tilsier at åpenhetsprinsippet innebærer krav til forklaringsevnen til behandlingsansvarlige. Ikke bare skal behandlingsansvarlig frembringe informasjonen, men informasjonen skal gis med et klart og enkelt språk, på en kortfattet måte som er forståelig og lett tilgjengelig for den registrerte. Fortalepunkt 39 gjentar kravene som følger av art. 12 med hensyn til lettfattelighet, enkelhet og klarhet. I fortalepunkt 58 presiseres det at der den registrerte er et barn, skal informasjonen gis på en måte barnet lett kan forstå.

Etter Artikkel 29-gruppens veiledning for åpenhetsprinsippet innebærer forståelighetskravet at den gjennomsnittlige registrerte må forstå informasjonen.[74] Dette forutsetter at behandlingsansvarlig har kunnskap om de registrerte, og kan tilpasse informasjonen etter mottakerens kunnskapsnivå. Også KI-ekspertgruppen uttaler at forklaringen må tilpasses målgruppen og dennes kompetanse.[75]

Fortalepunkt 60 og 71 om at særlige omstendigheter rundt og sammenhengen behandlingen skjer i skal tas hensyn til gjelder også informasjonens formkrav. Ifølge KI-ekspertgruppen må kommunikasjonen tilpasses den aktuelle bruken, herunder KI-systemets nøyaktighet og dets begrensninger.[76] Uttalelsene tilsier at behandlingsansvarlig må sørge for at den gjennomsnittlige registrerte forstår informasjonen, samtidig som forklaringen tilpasses

[72] Behandlet i kapittel 4.2.
[73] Skullerud, m.fl. (2018), s. 120-121.
[74] Artikkel 29-gruppen (2018a), avsnitt 9.
[75] KI-ekspertgruppen (2019b), s. 18.
[76] L.c.

mottakeren og eventuelle omstendigheter ved behandlingen. Dette innebærer at selv om den gjennomsnittlige registrerte forstår informasjonen, må det treffes ekstra tiltak dersom mottakeren trenger tilpasset informasjon, eller situasjonen eller KIen som benyttes gjør det nødvendig med ytterligere forklaring.

4.4 Den svarte boksen

Som nevnt i kapittel 2 er det utfordrende å forklare hvordan algoritmene fungerer. Navnet «den svarte boksen» spiller på at beslutningsprosessen til KIen ikke er transparent , men svart og uklar. Man vet ikke hvordan resultatet KIen leverer er blitt produsert.[77] Når man benytter maskinlæring for å utvikle KI benytter man en modell bygget på algoritmer, og det er denne modellen man kan få problemer med å forstå og forklare hvordan fungerer. Problemet er gjerne tilknyttet de dype nevrale nettverk som dannes i læringsprosessen til KIen, se figur 1, uten at dype nevrale nettverkene avgrenser problemstillingen rundt den svarte boksen. Forklaringsproblemene kan eksempelvis gjelde hvilke inputdata og/eller hvilken kombinasjon av inputdata som har vært viktigst.

Personvern handler i stor grad om å ivareta fysiske personers kontroll over bruk av egne data. Et viktig grep for å sikre hensynet er åpenhet.[78] Hensynet er vanskelig å ivareta gitt problematikken med den svarte boksen. Et tiltak for å sikre kontroll er åpne kildekoder. Kildekoder er imidlertid ikke forståelig for de fleste, selv om det kan gi offentligheten mulighet til å etterprøve KIen. Selv om tiltaket kan tilfredsstille informasjonens innholdskrav, oppfyller det ikke formkravet om forståelig informasjon.

Hvor vanskelig det er å forstå KIens beslutningsprosess avhenger blant annet av størrelsen på det nevrale nettverket. Størrelsen på nettverket avhenger blant annet av hvor mange inputverdier man opererer med og hvordan de ulike lagene er koblet sammen.[79] Jo flere inputverdier man har, jo fler faktorer kan KIen ta hensyn til, og jo flere lag den har, jo fler sammenhenger ser den. Det er muligheten for mange inputverdier og evnen til å se sammenhenger som medfører at KI kan ta riktigere og raskere beslutninger enn mennesker.

[77] Datatilsynet (2018), s. 12.
[78] Ibid., s. 18.
[79] Ibid., s. 13.

Det er paradoksalt at det er fordelene med KI som gjør at bruken kan være i strid med lovverket.

4.5 Oppsummert om åpenhetsprinsippet

Kravet til åpen behandling ved bruk av KI kan oppsummeres slik:

- Behandlingen må skje på en transparent måte
- Prinsippet har to sider: det må gis informasjon om selve behandlingsaktiviteten (innholdskrav), og den registrerte må enkelt kunne forstå denne informasjonen (formkrav).
- Kravet til informasjonens innhold:
    - Informasjonen som skal gis etter personvernforordningen art. 12-14 ikke er uttømmende,
    - Informasjon skal gjøre den registrerte i stand til å ta stilling til om personopplysningene er gjenstand for rettferdig behandling,
    - Den registrerte må informeres om at deres personopplysninger behandles av KI, herunder om profilering foretas,
    - Informasjonen skal legge til rette for at de registrerte kan utøve rettighetene sine etter personvernforordningen,
    - Den registrerte skal ha informasjon om eventuelle konsekvenser og hva slags påvirkning behandlingen har på den registrerte,
    - Den registrerte skal på forhånd kunne bedømme omfanget og konsekvensene av behandlingen slik at vedkommende ikke senere blir overrasket over måten personopplysningene er behandlet på,
    - Informasjonens innhold må tilpasses til mottakeren av informasjonen, det enkelte brukstilfellet, omstendighetene rundt og sammenhengen behandlingen skjer i,
    - Underliggende logikk skal opplyses om slik at beslutningene kan forstås,
    - Hvor grundig beslutningsprosessen må forklares tilpasses proporsjonalt etter innvirkning behandlingsaktiviteten har på livet til de registrerte,
- Formkravene til informasjonen:
    - Informasjonen må være gitt med et klart og enkelt språk, på en kortfattet måte, som gjør den forståelig og lett tilgjengelig for den registrerte,

    - Den gjennomsnittlige registrerte må kunne forstå informasjonen,
    - Forklaringen må tilpasses omstendighetene rundt og sammenhengen behandlingen skjer i, til hvem den skal gis til, herunder om det er et barn, og vedkommende sin kompetanse på området,
- Behandlingsansvarlig skal, gjennom en åpen behandling, legge til rette for at de registrerte kan utøve sine rettigheter,
- Prosessen skal være etterprøvbar i den forstand at datasettene og de prosessene som leder opp til beslutningen skal bli så godt dokumentert som mulig,
- Det stilles ekstra høye krav til informasjonens form og innhold når behandlingsansvarlig benytter KI.

5 Rettferdighetsprinsippet

Ordet «rettferdig» i personvernforordningen art. 5 (1) a) tilsier at behandlingen må være rimelig og ligge innenfor de moralske og etiske normene som følger av lover og regler, men også av samfunnets rettferdighetsoppfatning. Utover dette er rettferdighet et vidt begrep med et bredt anvendelsesområde. I dette kapittelet redegjøres det for rettferdighetsprinsippet. Kapittelet er strukturert etter de ulike aspektene ved rettferdighetsprinsippet; åpenhet, lovlighet, etikk, forbud mot usaklig forskjellsbehandling, behandlingens kontekst, mulige negative konsekvenser, rimelige forventninger, asymmetriske maktforhold og rettferdig oppbygging av algoritmen. Kildene som benyttes er ordlyden av lovteksten i art. 5 (1) a), fortalen til personvernforordningen, veiledninger fra Personvernrådet og EUs ekspertgruppe på KI og juridisk litteratur.

5.1 Åpenhet som forutsetning for rettferdighet

Åpenhet og rettferdighet henger tett sammen. Dette har blant annet resultert i at prinsippene omtales samlet flere steder i fortalen til personvernforordningen. Av fortalepunkt 39 fremgår det at det kreves tydelighet med hensyn til behandlingen. Den registrerte har blant annet rett til «informasjon for å sikre en rettferdig og åpen behandling». Det samme fremgår av fortalenpunkt 60. Fortalepunktene understreker kravet om tilstrekkelig informasjon for å sikre rettferdig behandling. Sammenhengen er naturlig da det er umulig for den registrerte å vite om behandlingen er rettferdig, dersom vedkommende ikke har informasjon om at behandlingen skjer, på hvilken måte, til hvilket formål, i hvilket omfang, etc. Åpenhet er dermed en forutsetning for rettferdighet.

Personvernrådets veiledning om innebygget personvern lister opp en rekke nøkkelelementer for å sikre rettferdighet gjennom innebygget personvern. Blant dem er at den registrerte skal informeres om hvordan behandlingen av personopplysninger fungerer når profileringsalgoritmer brukes.[80] Uttalelsen i veiledningen om innebygget personvern samsvarer med innholdskravet til informasjon for å sikre åpenhet, se oppgavens kapittel 4.2.

[80] Personvernrådet, *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default*, tilsluttet 13.11.2019b, avsn. 70.

At enkelte av kravene for å sikre åpenhet også gjelder for rettferdighet, understreker sammenhengen mellom de to prinsippene.

Jarbekk og Sommerfeldt har i sin bok *Personvern og personvernforordningen i praksis* oppsummert rettferdighetsprinsippet i en kravssjekkliste for behandlingsansvarlige.[81] Sjekklisten fremhever at de registrerte ikke må misledes eller lures når personopplysningene innhentes. Informasjonen som gis kan klart nok ikke være villedende dersom den registrerte skal kunne vurdere om behandlingen har vært rettferdig. Sammenhengen mellom åpenhet og rettferdighet, og kravet sammenhengen stiller til informasjon, tilsier at det skal gis korrekt og fullstendig informasjon som gjør den registrerte i stand til å vurdere om behandlingen har vært rettferdig.

5.2 Lovlighet og rettferdighet

I fortalepunkt 45 omtales tilfeller der behandling følger av en rettslig forpliktelse for den behandlingsansvarlige, behandling er nødvendig i allmennhetens interesse eller skjer for å utøve offentlig myndighet etter art. 6 (1) c) og e). Slik behandling bør ha rettslig grunnlag i unionsretten eller medlemslandenes nasjonalrett som bør presisere blant annet «denne forordnings allmenne vilkår for lovlig behandling av personopplysninger (...) og andre tiltak for å sikre lovlig og rettferdig behandling».[82] Lovlighetsprinsippet og rettferdighetsprinsippet henger altså sammen; behandling vil ikke være rettferdig så lenge den ikke er lovlig.

Lover er i de fleste tilfeller uttrykk for det som regnes som rettferdig. Lovlighetsaspektet av rettferdighetsprinsippet er dermed i stor grad sikret gjennom kravet om lovlighet i art. 5 (1) a). Det finnes imidlertid tilfeller der noe som er lovlig, ikke er rettferdig. Barnetrygdskandalen i Nederland, der KI avdekket feil i barnetrygdsaker, er et eksempel. Feilene kunne bestå av en glemt signatur, og ende i så store tilbakebetalingskrav at mange familier gikk konkurs.[83] Selv om signering er et lovlig krav, er det urettferdig å kreve tilbakebetalt all mottatt barnetrygd på det grunnlaget. Lovlighetsprinsippet faller utenfor oppgavens rammer, og kartlegges ikke nærmere.

[81] Jarbekk og Sommerfeldt (2019), s. 48.
[82] Fortalepunkt 45.
[83] Gabriel Geiger, "How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud", *vice.com*, 01.03.2021, https://www.vice.com/en/article/jgq35d/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud (lest 09.12.2021)

5.3 Etikk og moral

Ordlyden av «rettferdig» i art. 5 (1) a) tilsier at behandlingen må være moralsk og etisk i tråd med lover, regler og samfunnets rettferdighetsoppfatning. Rettferdighet er altså ikke bare et rettslig krav, men også et sterkt ønske generelt i samfunnet. Hva som er rettferdig reiser politiske, filosofiske og etiske spørsmål, og argumenter fra ulike fagområder glir inn i hverandre. De fleste har en klar oppfatning av om de har blitt behandlet urettferdig i konkrete saker, men det mer konkrete innholdet av uttrykket kan være vanskeligere å beskrive. Ordlydstolkningen tilsier at rettferdighetsprinsippets innhold varierer etter kulturen og samfunnet de involverte partene tilhører, samt at innholdet vil utvikle seg over tid.

En av grunntankene i sosialdemokratiet er ideen om likeverd.[84] Av dette følger at i et sosialdemokrati som Norge, er det rettferdig å forskjellsbehandle for at befolkningen skal få like muligheter, heller enn at rettferdighet er å behandle alle likt.[85] Den sosialdemokratiske rettferdighetstankegangen er fremmed for kapitalistiske land som USA, men også for mange europeiske land. I relasjon til KI forutsetter sosialdemokratisk tankegang at algoritmene tar hensyn til ulikheter og utøver skjønn. Også kulturelle forskjeller mellom konservative land og liberale land, som Norge, påvirker rettferdighetstankegangen. I konservative land kan det anses rettferdig å benytte KI som med høy sikkerhet kan fastslå en persons seksuelle legning i en ansettelsesprosess. Dette ville blitt sett på som urettferdig i Norge.

Terrorhendelsen 09.11.2001 og Covid 19-pandemien er eksempler på hvordan rettferdighetsoppfatning kan endres over tid. De to hendelsene medførte at overvåkning, og dermed oppgivelse av personopplysninger, i større grad anses som rettferdig, med de gode formålene å hindre terror og å hindre spredning av sykdom.[86] Politiske prioriteringer kan også påvirke hva som er en rettferdig avgjørelse i et offentlig vedtak om eksempelvis innvandring eller folketrygd, avhengig av politiske målsetninger.

At innholdet i rettferdighetsprinsippet utvikles over tid, er i tråd med EU-rettens dynamiske karakter. Ulike tolkninger avhengig av kulturell tilhørighet vil på den andre siden undergrave

[84] Dag Einar Thorsen, «Sosialdemokrati» i Store norske leksikon, 28.12.2020, https://snl.no/sosialdemokrati, (lest 04.12.2021).
[85] Ibid.
[86] ACLU sine nettsider, «Surveillance Under the Patriot Act" https://www.aclu.org/issues/national-security/privacy-and-surveillance/surveillance-under-patriot-act (lest 09.12.2021); Deborah Brown & Amos Toh, "Technology is Enabling Surveillance, Inequality During the Pandemic", *hrw.org*, https://www.hrw.org/news/2021/03/04/technology-enabling-surveillance-inequality-during-pandemic

formålet om en ensartet personvernlovgivning i EØS-området. Det må derfor søkes etter tydeligere retningslinjer for rettferdighetsprinsippets innhold i andre kilder enn ordlyden selv, for å sikre en ensartet praktisering av rettferdighetsprinsippet.

Også av Personvernrådets veiledning om innebygget personvern følger det at behandlingen av personopplysninger må være etisk forsvarlig.[87] Det presieseres at etisk forsvarlig behandling blant annet innebærer at behandlingsansvarlig tar hensyn til de større innvirkningene på den registrertes rettigheter og integritet.[88] Behandlingsansvarlig må altså ta hensyn til de større ringvirkningene behandlingen kan ha for den registrerte, utover sin egen behandlingsaktivitet og formålet med denne. Det er altså ikke tilstrekkelig at den aktuelle behandlingen behandlingsansvarlig gjennomfører oppfyller kravene etter personvernforordningen, dersom behandlingen kan skade den registrerte på andre måter. Et eksempel på dette er den nevnte IB-saken, der elever ved IB-skoler som en konsekvens av Covid 19-pandemien fikk avgangskarakterene sine påvirket av en skjev algoritme.[89] Behandlingsansvarlig hadde ikke tatt hensyn til at avgangskarakterene påvirket elevenes studie- og jobbmuligheter i ettertid.[90]

5.4 Usaklig forskjellsbehandling og diskriminering

Usaklig forskjellsbehandling og diskriminering er tydelige tilfeller av urettferdig behandling. Også Personvernrådet tolker rettferdighetsprinsippet dithen at behandlingen ikke kan diskriminere.[91] Fortalepunkt 71 gir uttrykk for at der det skjer profilering, innebærer rettferdighet at personopplysninger burde bli sikret på en måte som tar hensyn til potensiell risiko for interessene og rettighetene til den registrerte. Selv om fortalepunkt 71 gjelder profileringstilfeller, er innholdet relevant for tolkningen av rettferdig bruk av alle typer KI.

Fortalepunkt 71 tilsier at ved profilering må behandlingsansvarlig kartlegge den registrertes interesser og rettigheter, og kontrollere at disse ivaretas. Risiko for en persons rettigheter og friheter, som referert til i fortalepunkt 71, kan etter fortalepunkt 75 oppstå når behandlingen medfører diskriminering, økonomisk tap eller særlig økonomisk eller sosial ulempe, eller der personlige aspekter blir evaluert. Der det foreligger risiko for diskriminering, må personopplysningene sikres med hensyn til risikoen, ifølge fortalepunkt 71 og 75. Videre

[87] Personvernrådet (2019b), avsn. 70.
[88] L.c.
[89] Datatilsynet (2020b).
[90] Ibid., s. 8.
[91] Personvernrådet (2019b), avsn. 70.

følger det av fortalepunkt 71 at behandlingsansvarlig må hindre forskjellsbehandling på bakgrunn av de særlige kategoriene personopplysninger i personvernforordningen art. 9 (1).

At usaklig forskjellsbehandling særlig er trukket frem i fortalepunkt 71 og 75, selv om det uansett ville inngått som en del av den registrertes «interesser og rettigheter», gir behandlingsansvarlig en ekstra forpliktelse til å kontrollere at behandlingen ikke gir negative utslag for den registrerte i form av usaklig forskjellsbehandling. For KI vil dette blant annet innebære et ansvar for å sikre at inputdataene KI bygges opp på ikke er skjeve og at KIen ikke har skjeve utfall.

5.5 Omstendighetene rundt og sammenhengen behandlingen skjer i

Hva som er rettferdig avhenger av omstendighetene rundt og sammenhengene behandlingen skjer i. Av fortalepunkt 39 og 60 fremgår det at behandlingsansvarlig bør gi informasjon som er nødvendig for å sikre en rettferdig behandling. I punkt 60 legges også til at det skal «tas hensyn til de særlige omstendighetene rundt behandlingen av personopplysningene og sammenhengen den skjer i» når informasjonen gis. For at rettferdighetsprinsippet skal ivaretas tilsier dette at det ikke nødvendigvis er tilstrekkelig med standardisert informasjon til de registrerte. Behandlingsansvarlig må gi informasjon om eventuelle spesielle forhold ved behandlingen. Spesielle forhold ved behandlingen kan eksempelvis være at den utføres ved hjelp av KI, at noen av de registrerte har dårlige språkevner eller at maktbalansen mellom behandlingsansvarlig og den registrerte er skjev. Disse omstendighetene må tas hensyn til når behandlingsansvarlig vurderer hva slags informasjon som skal gis. Dette aspektet har en klar kobling til åpenhetsprinsippet.

Også fortalepunkt 71 presiserer at særlige omstendigheter og sammenhengen personopplysningene behandles i må hensyntas. I vurderingen er den registrertes interesser, rettigheter og forhindring av usaklig forskjellsbehandling sentralt. Rettferdig behandling forutsetter at det foretas en konkret vurdering av den registrertes interesser og rettigheter. I vurderingen av hvor konkret vurderingen må være vil det blant annet være relevant å se på hvor stor påvirkning behandlingen har på den registrerte. Det må eksempelvis kreves en mer konkret vurdering der KI brukes i en ansettelsesprosess enn der KI brukes for å foreslå kjøp i en nettbutikk.

5.6 Vurdering av og informasjon om mulige negative konsekvenser

Av fortalepunkt 39 fremgår det at de registrerte bør gjøres oppmerksomme på «risikoer, regler, garantier og rettigheter i forbindelse med behandling av personopplysninger». At de registrerte bør gjøres oppmerksom på «risikoer» tilsier at rettferdig behandling forutsetter at de registrerte i forkant av behandlingen skal informeres om mulige negative konsekvenser eller implikasjoner av behandlingen. Uten den informasjonen kan ikke den registrerte ta informerte valg om hvordan vedkommendes personopplysninger brukes. Videre tilsier det at den registrerte bør ha informasjon om «regler, garantier og rettigheter» og hvordan de kan utøve sine rettigheter at behandlingen ikke vil være rettferdig med mindre behandlingsansvarlig gir den registrerte de nødvendige verktøy for å ivareta sine rettigheter.

I veileder om behandling med grunnlag i kontraktsoppfyllelse etter personvernforordningen art. 6 (1) b) tolker Personvernrådet rettferdighetsprinsippet som at behandlingsansvarlig må vurdere mulige negative konsekvenser for den registrerte.[92] Etter fortalepunkt 39 bør den registrerte gjøres oppmerksom på mulige negative konsekvenser ved behandlingen. Dette forutsetter at behandlingsansvarlig har vurdert hvilke mulige negative konsekvenser behandlingen kan medføre for den registrerte. Tolkningen til Personvernrådet finner støtte i fortalen, samt hos Jarbekk og Sommerfeldt.[93] Personvernrådet retter fokuset mot vurderingen behandlingsansvarlig gjør, og fortalepunkt 39 mot informasjonen til den registrerte. Veiledningen og fortalen kan ikke forstås som at så lenge risikoer er vurdert og de registrerte informert, kan hvilke som helst risikoer tillates. Behandlingsansvarlig må også iverksette tiltak for å minimere risikoene der de negative konsekvensene er uakseptable.

5.7 Rimelige forventninger

I veiledning om behandling med grunnlag i kontraktsoppfyllelse uttaler Personvernrådet at rettferdighetsprinsippet innebærer «recognizing the reasonable expectations of the data subjects".[94] At behandlingen bør samsvare med den registrertes forventninger uttales også i

[92] Personvernrådet, *Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects*, 2. versjon. 08.10.2019a, avsn. 12.
[93] Jarbekk og Sommerfeldt (2019), s. 48.
[94] Personvernrådet (2019a), avsn. 12.

Personvernrådets retningslinjer om innebygget personvern.[95] Uttalelsene understreker at forutberegnelighet er et grunnleggende prinsipp for den reigstrerte. Dersom behandlingen ligger utenfor den registrertes rimelige forventninger, er den ikke rettferdig. Selv om det ligger innenfor formålet med innsamlingen etter personvernforordningen art. 5 (1) b), kan en behanlding være urettferdig. For eksempel kunne ikke elevene i IB-saken forvente at personopplysningene ble behandlet av en algoritme, selv om det lå innenfor formålet opplysningene var samlet inn for (å sette standpunktkarakterer).[96]

Skullerud m.fl. tolker rettferdighetsprinsippet slik at sammenhengen mellom innsamling av opplysninger og det formål de skal brukes til må fremstå som «rimelig for den registrerte». Tolkningen er noe smalere enn Personvernrådets med hensyn til hva den registrerte må finne rimelig, men faller innunder Personvernrådets tolkning.[97] Jarbekk og Sommerfeldt har, i likhet med Personvernrådet, lagt til grunn at personopplysninger generelt skal behandles innenfor de registrertes rimelige forventninger.[98] Tolkningen til Personvernrådet og Jarbekk og Sommerfeldt om at den registrertes rimelige forventninger ikke bare gjelder formålet, men den totale behandlingen, legges til grunn.

I Skullerud m.fl. tolkes rettferdighetsprinsippet som at det må være forståelig og naturlig for den registrerte at det er de aktuelle personopplysningene som behandles for det bestemte formålet.[99] Forutberegnelighetshensyn tilsier at behandlingsansvarlig ikke kan behandle flere eller andre personopplysninger enn den registrerte forventer ut ifra formålet, selv om andre opplysninger kan bidra til å oppnå formålet.

5.8 Asymmetriske maktforhold

Personvernrådet uttaler i sin veiledning om behandling med grunnlag i kontraktsoppfyllelse at behandlingsansvarlig må ta hensyn til aktuelle konsekvenser av et skjevt maktforhold mellom den registrerte og behandlingsansvarlig.[100] Også i veiledningen om innebygget personvern uttaler Personvernrådet at asymmetriske maktforhold har betydning for behandlingens

[95] Personvernrådet (2019b), avsn. 70.
[96] Datatilsynet (2020b), s. 7.
[97] Personvernrådet (2019a), avsn. 12.
[98] Jarbekk og Sommerfeldt (2019), s. 48.
[99] Skullerud m.fl. (2018), s. 75.
[100] Personvernrådet (2019a), avsn. 12.

rettferdighet.[101] Asymmetriske maktforhold bør så langt som mulig unngås, og det skal uansett iverksettes egnede tiltak som kan dempe ubalansen.

Skjevheter oppstår eksempelvis i arbeidsforhold, der arbeidstakeren er prisgitt arbeidsgiveren for å ha en inntekt. Det kan også være et skjevt forhold der private forbrukere inngår en avtale om bruk av tjenester som leveres av en stor internettleverandør. Urettferdig behandling kan da innebære at internettleverandøren utnytter sin posisjon overfor forbrukeren for å tilegne seg mer informasjon enn det som er nødvendig og naturlig for avtalens art.

5.9 Rettferdige algoritmer

Et nøkkelelement for rettferdighet som Personvernrådet lister opp i veiledning om innebygget personvern er at algoritmene må være rettferdige og det må skje kvalifisert menneskelig innblanding for å hindre machine bias.[102] Selv om behandlingsansvarlig bruker KI rettferdig, vil ikke behandlingen av personopplysningene være rettferdig dersom algoritmen inneholder skjevheter som medfører urettferdig behandling. Hvordan algoritmer kan være urettferdig utdypes i kapittel 5.10. Personvernrådet påpeker at rettferdige algoritmer kan sikres gjennom kvalifisert menneskelig innblanding for å hindre machine bias.[103] Machine bias utdypes i kapittel 5.10.

Ett av de syv prinsippene som oppstilles i EUs KI-ekspertgruppe sine etiske retningslinjer er «diversity, non-discrimination and fairness», som er omtalt som «inkludering, mangfold og likebehandling» i Nasjonal strategi for kunstig intelligens.[104] Prinsippet er tett linket opp til «fairness» som ett av fire fundamentale prinsipper fra EUs charter om grunnleggende rettigheter og internasjonal menneskerettighetslovgivning.[105] KI-ekspertgruppens uttalelser kan si noe om hva rettferdighet, og dermed også rettferdighetsprinsippet i personvernforordningen art. 5 (1) a), innebærer for bruk av KI.

Prinsippet innebærer blant annet at urettferdige skjevheter i datasettene som brukes til trening av KIen må unngås.[106] Slike skjevheter kan føre til uønsket diskriminering av grupper

[101] Personvernrådet (2019b), avsn. 70.
[102] L.c.
[103] L.c.
[104] KI-ekspertgruppen (2019b), s. 14; Kommunal- og moderniseringsdepartementet (2020), s. 59.
[105] KI-ekspertgruppen (2019b), s. 11-12 og 18.
[106] Ibid., s. 18.

mennesker, som potensielt kan forverre både forutinntatthet og marginalisering.[107] Urettferdighet kan også oppstå gjennom bevisst utnyttelse av skjevheter eller måten algoritmene utvikles på.[108] Videre forutsetter rettferdighet at KI-teknologien må være tilgjengelig for alle, uavhengig av kjønn, alder, kompetanse og funksjonshemninger.[109] KI-ekspertgruppen anbefaler at brukere som blir påvirket av systemet inkluderes i utviklingen av KIen, både før og etter at KIen er tatt i bruk.[110] At algoritmen må være rettferdig henger tett sammen med de andre aspektene av rettferdighetsprinsippet.

5.10 Machine bias

En sentral problemstilling innen bruken av KI, er machine bias. I dette kapittelet presenteres dette begrepet, for å vise hvordan bruk av KI kan utfordre rettferdighetsprinsippet.

KI som beslutningsstøtte har flere fordeler sammenlignet med mennesker, spesielt når det kommer til KIens evne til å prosessere store mengder informasjon raskt. Samtidig har det vist seg at KI kan ha innebygde skjevheter, som kan være vanskelig både å oppdage, og å rette opp.[111] Disse skjevhetene omtales som «machine bias», og leder til forskjellsbehandling, og i noen tilfeller diskriminerende utfall. Machine bias kan for eksempel oppstå på grunn av skjevheter i datagrunnlaget som algoritmene trenes på. Som beskrevet i kapittel 2 utvikles KI ved å trene algoritmer på store mengder data. Dersom læringseksemplene KIen trenes på inneholder skjevheter, smitter skjevhetene over på algoritmene. Skjevheter kan for eksempel oppstå hvis datagrunnlaget er for tynt, inneholder dikriminering, ikke er objektivt eller ikke er representativ for gruppen KIen skal benyttes på. KIen er ikke bedre enn dataen den bygges på.

Et eksempel på KI som inneholder machine bias er risikovurderingsverktøyet COMPAS.[112] Brukstilfellet utdypes i kapittel 6.2. COMPAS benyttes i amerikanske domstoler og veileder dommere i vurderingen av om siktede i straffesaker skal varetektsfengsles. De historiske dataene som risikovurderingsverktøyet er trent på, inneholder diskriminerende utfall i form av at mørkhudede feilaktig varetektsfengsles hyppigere enn hvite. Dette medfører at KIen

[107] KI-ekspertgruppen (2019b), s. 18.
[108] L.c.
[109] L.c.
[110] Ibid. s. 19.
[111] Europarådet: Committee of Experts on Internet Intermediaries, *Algorithms and Human Rights*, DGI (2017)12, 2018, s. 26.
[112] Julia Angwin, Jeff Larson, Surya Mattu, & Lauren Kirchner. "Machine Bias", *ProPublica.no*, 23.05.2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (lest: 17.11.2021)

konkluderer med at hyppigere varetektsfengsling av mørkhudede er riktig måte å løse oppgaven på. En KI som inneholder machine bias er, etter det vi har sett over, ikke i tråd med rettferdighetsprinsippet.

At det er praktisk talt umulig for mennesker å regne ut matematikken bak algoritmene gjør skjevheter svært vanskelig å oppdage. Ut ifra problematikken rundt machine bias og diskriminering har Europarådet påpekt forskjellen mellom direkte og indirekte diskriminering som kan hjelpe med hensyn til å avgjøre hvorvidt en algoritme er rettferdig.[113] Direkte diskriminering er tilfeller der beslutningstakeren direkte baserer beslutningen på urettmessige kriterier som hudfarge eller kjønn.[114] Direkte diskriminering skjer gjerne underbevisst hos mennesker, ettersom det er vanskelig å unngå å registrere noens kjønn eller hudfarge. Faktorer som kjønn og hudfarge kan imidlertid ekskluderes fra datasettet som danner grunnlaget for kunstig intelligens. Dette kan forhindre direkte diskriminering i KI som benyttes som beslutningsstøtte.

Indirekte diskriminering er vanskeligere både å oppdage og å korrigere enn direkte diskriminering. Indirekte diskriminering er tilfeller der visse karakteristikker eller faktorer opptrer hyppigere i en gruppe som det er ulovlig å diskriminere.[115] Dette er tilfellet der algoritmer er basert på skjeve data; beslutningen er i seg selv ikke bygget på direkte diskriminerende faktorer, men blir diskriminerende på grunn av skjevheter i datagrunnlaget. Problemet er at selv om ulovlige diskrimineringsfaktorer utelates fra datasettet kan algoritmen utvikle «information proxies», informasjonskombinasjoner som med høy grad av sikkerhet kan gi den samme informasjonen som om faktoren var inkludert. Så lenge det finnes skjevheter i datasettet, vil KIen finne andre korrelasjoner mellom informasjonen og konklusjonen som vil gi tilnærmet samme diskriminerende utslag som ved direkte diskriminering.[116] Dette er mulig blant annet på grunn av de dype nevrale nettverkene som bygges basert på mønstre i dataen algoritmen er trent på. Dette var tilfellet da Apple lanserte sitt kredittkort, der KIen fortsatte å diskriminere kvinner også etter at kjønn ble fjernet som faktor fra datasettet.[117]

[113] Europarådet (2018), s. 27.
[114] L.c.
[115] Ibid., s. 28.
[116] L.c.
[117] Will Knight, “The Apple Card Didn't "See" Gender - and That's the Problem”, *wired.com*, 19.11.2019, https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/ (lest 19.10.2021).

Ekskludering av faktorer det er urettmessig å basere beslutningen på kan gjøre det enda vanskeligere å oppdage og rette opp eventuelle skjevheter. Ser man ikke faktoren, kan man heller ikke se mønsteret. I motsetning til mennesker klarer ikke KI å se at beslutningene gir urimelige resultater eller å motvirke diskriminering på eget initiativ. KI har ikke etiske prinsipper eller målsetninger å korrigere beslutningene sine etter. Ettersom algoritmer utviklet med maskinlæringsteknikker lærer mens de brukes, forsterker de sine egne skjevheter. Dette styrker deres potensiale til å skape ekkokamre som forsterker allerede eksisterende diskriminering.

Ettersom rettferdighetsprinsippet i personvernforordningen art. 5 (1) a) forutsetter at det ikke skjer diskriminering og usaklig forskjellsbehandling, må behandlingsansvarlig iverksette tiltak som forhindrer dette. Det kan være å inkludere inputfaktorer det er urettmessig å vektlegge i vurderingen. Formålet er å justere algoritmens vekting i retning av å ikke diskriminere mot disse faktorene, eller for å kontrollere algoritmene opp mot de urettmessige faktorene for å se om det foreligger diskriminering. Dette kan bli problematisk ettersom faktorer det er urettmessig å vektelegge ofte samsvarer med personopplysninger det i utgangspunktet er forbudt å behandle, jf. personvernforordningen art. 9 (1). Det er inntatt i AIA art. 10 (5) at ved bruk av høyrisiko KI er det lovlig å behandle særlige kategorier personopplysninger etter personvernforordningen art. 9 (1), dersom det er strengt nødvendig for å overvåke, avdekke og korrigere skjevheter. Dersom forslaget blir vedtatt, kan det bidra til å avdekke machine bias.

Risikoen for machine bias krever høy grad av bevissthet fra brukeren av KIen. Dette kan til en viss grad sikres gjennom opplæring i hva KIen egner seg til å si noe om og ikke, hvilke feil den kan gjøre og hvor nøyaktig den er. Selv om en bruker av KI som beslutningsstøtteverktøy er både forsiktig og bevisst, kan menneskets psykologi komme i veien. Som Daniel Kahnemans studier viser, lar mennesker selv irrelevante faktorer påvirke deres dømmekraft, en effekt som kalles *ankring*.[118]

Ankring har vist seg å påvirke selv dommere i deres beslutninger.[119] Dette er blant annet vist i en studie der tyske dommere, etter å ha fått fremlagt et faktum som beskrev et tilfelle av butikknasking, skulle ta stilling til hvor mange måneder fengsel vedkommende skulle få.[120]

[118] Daniel Kahneman, *Thinking Fast and Slow*, Penguin Books, 2011, s. 119-128.
[119] Kahneman (2011), s. 125-126.
[120] L.c.

Før de tok beslutningen, trillet dommerne en terning de ble fortalt landet på et helt tilfeldig tall. Terningene var imidlertid innstilt til å lande på henholdsvis tre eller ni. Dommerne som trillet et tretall, dømte i gjennomsnitt naskeren til fem måneder fengsel. Dommerne som trillet ni på terningen dømte i gjennomsnitt naskeren til åtte måneders fengsel. Konklusjonen på studiet var at dommerne ble påvirket av det tilfeldige tallet terningen viste når de skulle dømme.[121] Ankringseffekten påvirker også muligheten til å forklare de menneskelige beslutningene som tas i tilknytning til KIen, som er et krav etter åpenhetsprinsippet, se kapittel 4.2.

5.11 Oppsummert om rettferdighetsprinsippet

Kravet til rettferdig behandling kan etter gjennomgangen over oppsummeres slik:

- Behandlingen må være lovlig og i tråd med regelverket,
- Behandlingen må være åpen i form og innhold, for at den registrerte skal få kunne vurdere om behandlingen er rettferdig,
- Behandlingen må være i tråd med moralske og etiske prinsipper og normer,
- Den registrertes rettigheter og interesser må ivaretas av behandlingsansvarlig under den aktuelle behandlingen, men også i et større perspektiv,
- Behandlingen må ikke medføre usaklig forskjellsbehandling,
- Behandlingen og vurderingen av hvilken informasjon som skal gis må ta hensyn til de særlige omstendighetene og sammenhengen behandlingen skjer i,
- Behandlingsansvarlig skal vurdere og ta hensyn til mulige negative konsekvenser for den registrerte, og vedkommende skal opplyses om disse,
- Behandlingen må skje i tråd med den registrertes rimelige forventninger, og typen personopplysninger som behandles må være naturlig og forståelig for den registrerte,
- Behandlingen må ta hensyn til skjeve maktforhold og innvirkningen det kan ha på behandlingen,
- Algoritmene må i seg selv være rettferdige.

[121] Kahneman (2011), s. 126.

6 Eksempel: Åpen og rettferdig bruk av kunstig intelligens

I dette kapittelet gjennomgås IB-saken og beslutningsstøtte i domstolene som eksempler der KI brukes som beslutningsstøtte. Med bakgrunn i drøftelsen av åpenhets- og rettferdighetsprinsippet i kapittel 4 og 5, analyseres det hvordan prinsippene slår ut i sakene.

6.1 IB-saken

6.1.1 Faktum

Det norske datatilsynet varslet i august 2020 vedtak mot International Baccalaureate Organization (IBO). Bakgrunnen var at IBO, på grunn av manglende eksamensavvikling som følge av Covid-19 pandemien, anvendte en modell for å fastsette elevenes standpunktkarakterer. Modellen baserte avgjørelsen på elevens resultater fra større skriftlige arbeider, faglærerens foreslåtte karakter og historisk data fra skolen elevene gikk på, basert på tidligere elevers resultater. IBO opplyste hverken i forkant av karaktersettingen eller etter anmodning fra Datatilsynet om hvordan de ulike faktorene var vektet eller de endelige karakterene ble regnet ut. Datatilsynet uttrykte i varselet at kravet til åpenhet og rettferdighet var brutt. Datatilsynet avsluttet saken før det var fattet vedtak på grunn av manglende jurisdiksjon.[122] IBO reviderte likevel algoritmene, og fjernet historisk data som faktor.[123] Flere elever fikk fortsatt store avvik mellom karakter på større skriftlige arbeider, lærerens vurdering og sluttkarakteren, se tabell 1. Dette kan tyde på at eksterne faktorer fortsatt ble benyttet som faktor i utregningen til algoritmen. Drøftelsen i delkapittel 6.1.2. retter seg primært mot den første karaktersettingen.

[122] Datatilsynet, «Lukker IB-saken», 16.07.2021, https://www.datatilsynet.no/aktuelt/aktuelle-nyheter-2021/lukker-ib-saken/ (lest 11.10.2021). IBO hadde på tidspunktet for karaktersettingen hovedetablering i Storbritannia, som da fortsatt var en del av EØS. Hovedregelen er at det kun er datatilsynet i landet et selskap har hovedetablering som kan fatte vedtak etter personvernforordningen, jf. personvernforordningen art. 55 (1). Det norske Datatilsynet konkluderte med at det var sannsynliggjort at IBO hadde hovedetablering i Storbritannia, og avsluttet derfor saken før endelig vedtak var truffet. Dersom Storbritannias utmelding av EU hadde skjedd før karaktersettingen, kunne det norske Datatilsynet kanskje ha truffet vedtak i saken.

[123] IBO sine nettsider, "IB update on May 2020 Diploma Programme and Career-related Programme results", 17.08.2020, https://www.ibo.org/news/news-about-the-ib/update-m20-dp-cp-results/?fbclid=IwAR0s07X4C47WdqsuLJlSFJX__H7s7l_fNjFrwwSqicuIxekKtJioO21A7v0 (lest 25.11.2021).

6.1.2 Åpenhet og rettferdighet

En viktig del av rettferdighetsprinsippet er at behandlingen må være i tråd med den registrertes rimelige forventninger. Dette er et sentralt poeng i IB-saken, som også Datatilsynet trakk frem i deres varslede vedtak.[124] Vanligvis blir karakterer satt på bakgrunn av elevens egne akademiske prestasjoner, ettersom karakterene skal gjenspeile elevens evner i et fag. Det er dette elevene rimeligvis forventer. IBOs modell forutsatte imidlertid at elevene forutså og aksepterte at karakterene deres ble påvirket av en faktor utenfor deres kontroll og uten sammenheng med deres akademiske prestasjoner; andres akademiske prestasjoner. Behandlingen er ikke i tråd med elevenes rimelige forventninger, og den er derfor i strid med rettferdighetsprinsippet i personvernforordningen art. 5 (1) a).

Forholdet mellom elevene og skolen er asymmetrisk, da elevene er prisgitt skolen og ledelsen i sin skolegang. Etter rettferdighetsprinsippet skal skjeve maktforhold tas hensyn til. Det fremkommer ikke at dette er hensyntatt i behandlingen. Et mulig tiltak kunne vært å gjennomføre en personvernkonsekvensvurdering etter personvernforordningen art. 35 (1) der man innhentet synspunkt på den planlagte behandlingsaktiviteten fra elevene, jf. art. 35 (9). IBO burde også gjennomført en personvernkonsekvensvurdering ettersom de tok ibruk ny teknologi. Personvernkonsekvensvurderingen kunne avdekket bekymringer og tilrettelagt for en rettferdig behandling der det skjeve maktforholdet ble hensyntatt. Manglende hensyn til og tiltak mot det skjeve maktforholdet bryter med rettferdighetsprinsippet i art. 5 (1) a).

Modellen kunne videre føre til usaklig forskjellsbehandling, som også påpekes av Datatilsynet.[125] Vurderingsmodellen ble tilpasset den enkelte skole, og hvilken skole de registrerte tilhørte påvirket utfallet. Dette uten å ta hensyn til om den enkelte elevs akademiske prestasjoner var sammenlignbare med skolens og fagets historiske gjennomsnitt. Hvorvidt de historiske dataene hadde skjevheter som var i elevenes disfavør var avhengig av flere faktorer, som skolens geografiske plassering. Elever fra områder med dårlige resultater risikerte å bli satt ned i karakter, kun fordi de gikk på skolen i det området.[126] Modellen kunne dermed forskjellsbehandle ulike sosioøkonomiske grupper, som henger sammen med geografisk bosetting. At IBO sikret at det ikke var store avvik i gjennomsnittskaraktern på den

[124] Datatilsynet (2020b), s. 7.
[125] L.c.
[126] Larry Hardesty, "Study finds gender and skin-type bias in commercial artificial-intelligence systems", *MIT News*, 11.02.2018, https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212 (lest 06.12.2021).

enkelte skole avhjelper ikke problemet. Tynt datagrunnlag som skyldes få avholdte eksamener ved en relativt ny skole kan også være mulig kilde for skjevheter. Risikoen for usaklig forskjellsbehandlig av elevene ved IB-skolene strider med rettferdighetsprinsippet i personvernforordningen art. 5 (1) a).

Videre hadde ikke IB-skolene tatt tilstrekkelig hensyn til elevenes interesser og rettigheter i et større perspektiv enn den aktuelle behandlingsaktiviteten. Karakterer påvirker hvilke studie- og jobbmuligheter elevene får. Feilaktig karaktersetting kan ha store konsekvenser, og behandlingen gi negative ringvirkninger for elevenes interesser og rettigheter. Dette momentet ble også vektlagt av Datatilsynet.[127] I tillegg er en særlig omstendighet ved behandlingen at den gjaldt opplysninger om unge mennesker, som har lagt planer om og søkt på studier og jobber ut ifra forespeilede karakterer basert på egne akademiske prestasjoner. IBO har heller ikke tatt hensyn til den usikre livssituasjonen til elevene som en særlig omstendighet. Ettersom det ikke er tatt hensyn til de negative konsekvensene behandlingen kunne ha for elevene på lengre sikt, samt manglende henyn til omstendighetene rundt og sammenhengen behandling skjedde i, er behandlingen i strid med rettferdighetsprinsippet i art. 5 (1) a).

Innholdet av åpenhetsprinsippet må vurderes fra sak til sak, der omstendighetene rundt og sammenhengen behandlingen skjer i spiller inn på kravene som stilles. Det stilles høye krav til informasjonens form og innhold ved bruk av KI. Ettersom karakterer er avgjørende for hvilke utdannings- og arbeidsmuligheter elevene får, kan behandlingsaktiviteten påvirke resten av elevenes liv. Dette er faktorer som tilsier at ekstra tiltak burde vært iverksatt for å sikre åpenhet. Likevel var behandlingen foretatt av en profileringsalgoritme som hverken elevene eller offentligheten hadde tilgang på logikken bak. I brev fra elevenes advokat, sendt til Datatilsynet og IBO den 03.09.2021, ble flere eksempler der sluttkarakteren avvek fra studentens egne prestasjoner vist, uten logisk forklaring på utregningen, se tabell 1.[128] Det var ikke mulig å forstå eller resonnere seg frem til hvorfor systemet konkluderte som det gjorde, som også er et krav etter åpenhetsprinsippet. På denne bakgrunn bryter modellen med

[127] Datatilsynet (2020b), s. 8.
[128] Føyen Advokatfirma v/Arve Føyen, brev til Datatilsynet i sak 20/03087 (IB-saken), 03.09.2020, s. 3, tilgjengelig på: http://nrkbeta.no/wp-content/uploads/2020/09/2020-09-03-Letter-from-FT-IBO-result-anomalies1377198.1.pdf (lest 25.11.2021).

prinsippet om åpenhet i personvernforordningen art. 5 (1) a), noe Datatilsynet også konkluderte med.

Example #	Student	IBO grade July 2020	PG	IA	IBO grade August 2020	Comments
9	Subject	3	4	6	3	This is an exceptional example. The July as well as August 2020 grades are both set lower than the student's own performance (IA and PG). This is a clear indicator that external factors beyond the student's own performance (PG and IA) are still at play.
Student 5						
10	Subject	4	6	4	4	This example illustrates the lack of transparency. IBO has not exposed any information on how final grade is set when there is a discrepancy of 2 or more between PG and IA.

Tabell 1: Oversikt over to elevers karakterer.. Eksempel 9 viser at sluttkarakteren er lavere enn elevens karakter på større skriftlige arbeider (IA) og karakteren satt av læreren (PG). I eksempel 10 tilsvarer elevens sluttkarakter med IA, men avviker to karakter fra PG.[129]

Andre mangler ved åpenheten er at elevnee ikke fikk informasjon om eventuelle konsekvenser ved behandlingen. Dette aspektet av åpenhetsprinsippet bidrar til at kravet om at de registrerte kan bedømme om behandlingen var rettferdig oppfylles. Det er tydelig at elevene ikke hadde fått nok informasjon til å bedømme omfanget og konsekvensene av behandlingen, da de ble overrasket over måten personopplysningene ble behandlet. Behandlingen bryter med åpenhetsprinsippet i art. 5 (1) a) på disse punktene.

At Datatilsynet ba om innsyn i utregningene uten å få det, kan tyde på manglende mulighet for etterprøvbarhet. Heller ikke da algortimene ble revidert ble utregninene lagt frem. Selv om det ble gitt noe informasjon om utregningene ved andre karakterfastsettelse, kan ikke alle karakterene forklares ut ifra informasjonen, se tabell 1. Manglende etterprøvbarhet for offentligheten og de registrerte, strider med åpenhetsprinsippet i art. 5 (1) a).

Behandlingen bryter med rettferdighetsprinsippet i art. 5 (1) a) da den ikke samsvarte med de registrertes forventninger, asymmetriske maktforhold ikke var hensyntatt, usaklig

[129] Føyen Advokatfirma (2020), s. 3.

forskjellsbehandling kunne foreligge og elevenes interesser og rettigheter i et større perspektiv, ewm6 særlige omstendigheter rundt og sammenhengen behandlingen skjedde i, ikke var hensyntatt. Behandligen bryter med åpenhetsprinsippet i art. 5 (1) a) da det ikke ble iverksatt tiltak for å sikre åpenhet, til tross for bruk av ny teknologi og innvirkningen på de registrerte, det ikke var mulig å følge logikken bak KIens beslutninger, det ikke var tilstrekkelig informasjon til å bedømme om behandlingen var rettferdig og omfanget og konsekvensene av den og manglende mulighet for å etterprøve.

6.2 Beslutningsstøtte i domstolene

6.2.1 Faktum

Domstolskommisjonen, som har i oppgave å utrede domstolenes organisering og uavhengighet, publiserte sin andre delrapport i 2020.[130] Rapporten inneholdt forslag til domstolenes fremtidige organisering for å ivareta effektivitet og kvalitet. KI ble i rapporten beskrevet som et verktøy som kan forbedre domstolenes beslutningskvalitet og sikre likebehandling.[131] Samtidig påpekte rapporten at rettssikkerheten og domstolenes uavhengighet må ivaretas ved innføring av KI. Blant annet ble viktigheten av å forhindre diskriminering, sikre upartiskhet og en tilstrekkelig åpen beslutningsprosess nevnt.[132] En av de KI-baserte løsningene kommisjonen foreslo er risikovurderingsverktøy som beslutningsstøtte for dommere. Risikovurderingsverktøy er allerede i bruk i domstolene i en rekke stater i USA. Personvernforordningen gjelder ikke i USA, men brukstilfellet er egnet til å si noe om hvordan åpenhets- og rettferdighetsprinsippet kan slå ut dersom domstolskommisjonens forslag gjennomføres i norske domstoler. Vurderingen av hvordan åpenhets- og rettferdighetsprinsippet kan slå ut vil derfor ta utgangspunkt i et av de mest brukte risikovurderingsverktøyene i USA.[133]

ProPublica, en uavhengig, ideell organisasjon som produserer undersøkende journalistikk i offentlighetens interesse, gjennomførte en studie på beslutningsstøtteverktøyet COMPAS som

[130] Regjeringens nettsider, «Domstolkommisjonen», https://www.regjeringen.no/no/dep/jd/org/styre-rad-og-utval/innstillinger/innstillinger-fra-utvalg/innstillinger-levert-i-2020/domstolkommisjonen/id2565732/ (lest 11.10.2021)
[131] NOU 2020:11 s. 258.
[132] L.c.
[133] Angwin m.fl. (2016).

brukes i flere amerikanske stater.[134] Verktøyet gir siktede i straffesaker en risikoscore mellom 1 og 10. Risikoscoren er ment å si noe om hvor sannsynlig det er at siktede møter opp til rettsmøter og risikoen for gjentagelsesfare. Scoren gis basert på informasjon som er samlet inn gjennom et spørsmålsark.[135] Risikoscoren presenteres for dommeren, som står fritt til å bruke risikovurderingen i spørsmålet om hvorvidt den siktede skal varetektsfengsles.

6.2.2 Åpenhet og rettferdighet

Rettferdighetsprinsippet innebærer en rett til å ikke bli diskriminert. ProPublicas studie konkluderte med at algoritmene bak COMPAS er diskriminerende overfor mørkhudede.[136] Studien viste at nesten dobbelt så mange mørkhudede sammenlignet med hvite fikk en høy risikoscore, uten at dette stemte. For de som fikk en lav risikoscore for gjentagelsesfare, viste det seg at nesten halvparten av de hvite begikk en ny kriminell handling, mens i underkant av 30 prosent av mørkhudede gjorde det samme.[137] COMPAS ble utviklet basert på historisk data, og opptøyene i USA det siste året, med blant annet Black Lives Matter-bevegelsen, viser at dataene inneholder store skjevheter med hensyn til hudfarge.

Firmaet som eier COMPAS er uenige i konklusjonen til ProPublica om diskriminering, da hudfarge eller etnisitet ikke er en del av inputverdiene.[138] Som vist i drøftelsen om direkte og indirekte diskriminering i kapittel 5.10., kan diskriminering oppstå selv om faktorene det er ulovlig å diskriminere mot ikke er inkludert. Det er svært vanskelig å kunne gi en risikoscore uten å inkludere faktorer som har korrelasjon til hudfarge. KIen er trent på skjeve data, og studien viser at den diskriminerer.[139] Ettersom algortimene diskriminerer er den i strid med rettferdighetsprinsippet i art. 5 (1) a).

Tilfellet fra USA viser at Norge må utvise forsiktighet med å ta ibruk KI som er trent på data fra andre land, da disse ikke nødvendigvis er i etisk og moralsk overenstemmelse med norske verdier. Dataene som KI utviklet i andre land er trent på er heller ikke nødvendigvis representative for Norge, som er et annet grunnlag for machine bias. Hva som er etisk og moralsk på EU-nivå er vanskelig å avklare, med ulike rettferdighetsidealer i de europeiske

[134] ProPublicas nettsider, "About Us", https://www.propublica.org/about/ (lest 25.11.2021); Angwin m.fl. (2016).
[135] Angwin m.fl. (2016).
[136] Ibid.
[137] Ibid.
[138] Ibid.
[139] Ibid

landene, eksempelvis i konservative sammenlignet med liberale land. Ulik rettferdighetsvurdering kan skade personvernforordningenens formål om enhetlig regulering, tolkning og praktisering. Dette blir imidlertid opp til EU-domstolen å tolke og EU-kommisjonen å løse. Rettferdighetsprinsippet i art. 5 (1) a) kan altså slå ut på ulike måter avhengig av hvordan Norge implementerer risikovurderingsverktøy.

En løsning er å utvikle algoritmer trent på norske data. Norge har imidlertid en historie med systematisk diskriminering av minoriteter som samer, som kan gi utslag på algoritmene. Domstolskommisjonen fant også skjevheter basert på domstolenes geografiske plassering og dommernes erfaring.[140] Algoritmer trent på norske data kan også gi diskriminerende utslag i strid med rettferdighetsprinsippet.

En annen problemstilling er hvordan dommere kan bruke beslutningsstøtten på en rettferdig måte. Kahnemans studie på ankringseffekten, omtalt i kapittel 5.10., kan sammenlignes med beslutningsstøtte i domstolene. Ved bruk av risikovurderingsverktøy vil dommeren få presentert et tall i forkant av beslutningen. Tallet vil til forskjell fra studien ikke være tilfeldig, men regnet ut basert på opplysninger om den aktuelle personen, og er derfor mer legitimt å vektlegge enn terningens øyne var i studien. Risikoen for machine bias tilsier likevel at risikovurderingen burde tillegges begrenset vekt. Studien at en ren oppfordring om å utvise forsiktighet og bevissthet rundt risikoen for machine bias ikke er nok når underbevisstheten spiller inn på hvor mye vekt risikovurderingen får. At et menneske har det siste ordet, hjelper ikke det når KIen har det første ankrende ordet.

Problematikken viser viktigheten av opplæring for å sikre rettferdig bruk av algoritmene. Det er ikke gjort studier på hvordan risikovurderingsverktøy påvirker dommeres vurderinger. En studie kan bidra til å gi den registrerte nødvnedig informasjon for å vurdere om behandlingen er rettferdig, at dommeren kan ta hensyn til ankring som en særlig omstendighet, i best mulig grad sikre rettferdig bruk og vurdering av KIen og vurderingen av mulige negative konsekvenser for den registrerte.

Den registrertes rimelige forventninger spiller også inn. Siktede kan forvente å bli vurdert ut ifra egne handlinger og sakskomplekset i spørsmålet om varetektsfengsling, jf. vilkårene i straffeprosessloven §§ 171-173, jf. § 184. Ettersom risikovurderingsverktøyet er bygget opp

[140] NOU 2019:17, 8.2.5.

på historisk data, vil siktede til en viss grad vurderes ut ifra andres tidligere handlinger, uten at KIen nødvendigvis klarer å ta hensyn til individuelle ulikheter. Dette er ikke i tråd med siktedes forventninger, da man ikke skal dømmes for andres handlinger. Samtidig påvirkes dommere av egne erfaringer i sine beslutninger, og problemet eksisterer også hos menneskelige beslutningstakere.[141]

Et annet moment som viser at behandlingen er utenfor den registrertes rimelige forventninger er spørsmålene de siktede svarer på. ProPublica fikk tilgang til noen av spørsmålene som stilles ved bruk av COMPAS, blant annet om foreldrene til siktede var i fengsel og hvor mange venner og bekjente siktede har som ulovlig bruker narkotika. Selv om det kan være gode statistiske grunner bak spørsmålene, gir dette også inntrykk av at siktede risikerer varetektsfengsling på grunn av andres handlinger enn sine egne. En nærmere vurdering av hvorvidt bruk av historisk data og spørsmålene som stilles medfører at man riskierer varetektsfengsling for andres handlinger, faller utenfor oppgavens rammer. Likevel tyder momentene på at behandlingen ikke er i tråd med de registrertes rimelige forventninger, og rettferdighetsprinsippet i art. 5 (1) a) blir brutt.

Et siste moment innenfor rettferdighetsvurderingen er det asymmetriske forholdet mellom siktede og domstolene. Dette er noe avhjulpet gjennom at siktede i flere tilfeller har krav på offentlig oppnevnt forsvarer. Dette er imidlertid et tiltak som alltid er til stede, og den økte risikoen KI presenterer tilsier at det bør treffes ytterligere tiltak for å utjevne maktbalansen.

I tråd med åpenhetsprinsippet skal den registrerte få nok informasjon til å vurdere om vedkommende er utsatt for rettferdig behandling. Ifølge ProPublica får ikke de registrerte informasjon om hvordan utregningene skjer, eller hvilken vekt ulike momenter er tildelt. Hverken de registrerte eller det offentlige har hatt tilgang på logikken bak eller mulighet til å etterprøve risikovurderingene. ProPublica fikk etter forespørsel tilgang til de overordnede faktorene utregningen av gjentagelsesfaren består av, men ikke utregningen, med opphavsrett som begrunnelse. At større researchprosjekter ProPublicas studie avdekker diskriminering, er ikke det tilstrekkelig åpent. Man kan ikke være avhengig av rettighetsorganisasjoner, som det er få av i Norge innen personvern. Systemets begrensninger og egenskaper må opplyses om til den registrerte og dommerne som skal bruke risikoscoren. De registrerte ser ikke ut til å være opplyst om risikoen for machine bias, og det foreligger liten informasjon om opplæring av

[141] NOU 2020:11, s. 255.

dommerne. Innholdet av informasjonen som er gitt tilfredsstiller ikke innholdskravene i åpenhetsprinsippet i art. 5 (1) a).

Selv om eieren av COMPAS ikke er konfrontert med eller innrømmer det, kan manglende forklaring av utregningene bak COMPAS være en følge av den svarte boksen, omtalt i kapittel 4.4. Samtidig kan en dommers sinn, og dermed også beslutning, være like ugjennomtrengelig som en svart boks. Dersom en dommer direkte diskriminerer, kan det skje fordi det er umulig å ikke legge merke til siktedes hudfarge eller kjønn, ikke fordi dommeren bevisst ønsker å diskriminere. Skjevhetene i de historiske dataen stammer tross alt fra virkeligheten. En forskjell er at KI vil «svare» ærlig på spørsmålet om det foreligger diskriminering ved en undersøkelse. Mennesker kan nekte for diskriminering, uten mulighet for å overprøve hva som har skjedd i vedkommendes hjerne. En tankegang om at menneskelige beslutninger også har svakheter kan ikke unnskylde at KI også har det. Menneskelige svakheter må heller brukes som en mulighet til å se hvor ny teknologi kan forbedre dagens løsninger. Så lenge det iverksettes tiltak for å overprøve KIen, kan det være enklere å oppdage og rette opp diskriminering enn i et menneskelig sinn.

Konklusjonen er at dersom COMPAS blir tatt ibruk i Norge slik systemet brukes i USA i dag, strider behandlingen med både åpenhets- og rettferdighetsprinsippet. Dersom risikovurderingsverktøy skal tas i bruk i norske domstoler, må faktorene som er nevnt her tas hensyn til, slik at bruken er i overensstemmelse med åpenhets- og rettferdighetsprinsippet i personvernforordningen art. 5 (1) a).

7 Avsluttende refleksjoner

Innledningsvis ble det paradoksale ved at KIs fordel også er dens ulempe i møte med personvernregelverket nevnt. Personvernet er imidlertid ikke en absolutt rettighet, og fordelene KI kan gi for samfunnet må vektes mot personvern.[142] Drøftelsen i kapitlene 4-6 viser hvordan også åpenhetsprinsippet og rettferdighetsprinsippet må avveies mot hverandre. Et mindre komplisert system med algoritmer kan være enklere å forklare, men vil ikke nødvendigvis være like nøyaktig, ettersom det ikke kan ta hensyn til like mange faktorer og sammenhenger. Motsatt vil et grundig og nøyaktig system som tar hensyn til en lang rekke faktorer kunne ta gode avgjørelser, men det vil være vanskelig å forklare beslutningsprosessen. Det må derfor gjøres avveininger mellom å forbedre et systems forklarbarhet (som kan gjøre det mindre nøyaktig) eller å gjøre systemet mer nøyaktig (på bekostning av forklarbarheten).[143]

Mange håpet å få retningslinjer på avveiningene mellom utvikling og bruk av KI og personvern da EU-kommisjonen la frem AIA, men ble skuffet da forslaget ble lagt frem. Mens personvernforordningen har to hovedformål, å fremme EUs indre marked og å sikre personvernet til EUs innbyggere, har AIA ett hovedformål: å sikre funksjonen til EUs indre marked.[144] Rettsakten er derfor i hovedsak hjemlet i TEUV art. 114 om EUs indre marked, og ikke i en av bestemmelsene i TEUV del 1, kapittel II om fysiske personers rettigheter, slik personvernforordningen er. AIA er derfor ikke en rettighetsakt for fysiske personer, men fungerer heller som en bransjestandard for utviklere og brukere av KI. Reguleringen kan likevel bidra til et vern for fysiske personer i møte med KI.

Rettsakten er basert på en risikotilnærming, med streng regulering av høyrisiko KI, og langt mildere regulering for KI med medium og lav risiko. Enkelte typer KI er klassifiserte som uakseptable risikomessig, og er ikke tillatt. Hva som regnes som høyrisiko KI er basert på erfaringer. Blant annet er KI-systemer som skal brukes for å vurdere elever i undervisningssituasjoner, som i IB-saken, rangert med høy risiko.[145] Dette innebærer blant annet krav til evaluering av skjevheter i datasettene og informasjon til brukerne, se art. 10 (2) f) og 13. Kravene i AIA er mer konkrete enn i personvernforordningen, men svarer langt ifra

[142] Fortalepunkt 4.
[143] KI-ekspertgruppen (2019b), s. 18.
[144] AIA, fortalepunkt 1, første setning.
[145] AIA, Anneks II punkt (3) b).

på alle spørsmål som oppstår i forbindelse med åpen og rettferdig bruk av KI. Også denne forordningen bærer preg av abstrakte normer som skal regulere konkrete brukstilfeller.

En tenkning som kan være hensiktsmessig å ta med fra personvernreguleringen er innebygget personvern, som er et sentralt krav i personvernforordningen. Kravet innebærer at personvern tas hensyn til i alle utviklingsfaser av et system eller en løsning. Det skal sikre at behandlingen oppfyller personvernforordningens krav og ivaretar individets rettigheter gjennom hele prosessen.[146] Selv om KI kan kontrolleres i ettertid, er det mest hensiktsmessig å bygge inn personvern og andre etiske hensyn fra start.[147] I tillegg må mekanismer for å kontrollere at KIen fortsetter å overholde kravene i personvernforordningen underveis sikres. En algoritme kan bli skjev etter hvert, selv om den ikke er det fra starten av.

KI er basert på data, som er en refleksjon av historien. KI tar ikke valg som er etiske, men matematiske. Så lenge fortiden dveler i algoritmene vil ikke KI gjøre fremskritt, men gjenskape verden slik den eksisterer. Man kan ikke endre verden med det samme tankesettet som skapte den. Personvernforordningen står foreløpig som en festningsmur mot misbruk av personopplysninger, som gir individer rett til tilgang, kontroll og forutberegnelighet med hvordan personopplysningene brukes. Vi forstår enda ikke hva våre data i kombinasjon med KI kan brukes til, men åpenhet og rettferdighet blir viktige rettesnorer for videre utvikling av både KI og regelverk.

[146] Kommunal- og moderniseringsdepartementet (2020), s. 60.
[147] L.c.

8 Litteraturliste

8.1 Norske lover og forarbeider

Personopplysningsloven	Lov av 15. Juni 2018 nr. 38 om behandling av personopplysninger (personopplysningsloven)
Personopplysningsloven [2000]	Lov av 14. april 2000 nr. 31 om behandling av personopplysninger (personopplysningsloven) (opphevet)
Straffeprosessloven	Lov av 22. mai 1981 nr. 25 om rettergangsmåten i straffesaker (straffeprosessloven)
NOU 2020:11	Den tredje statsmakt – Domstolene i endring
NOU 2019:17	Domstolstruktur

8.2 EU-rettsakter og avtaler

TEUV	Konsolidert versjon av Traktaten om Den europeiske unions virkeområde (TEUV), Consolidated version of the Treaty on the Functioning of the European Union (TFEU), 26.10.2012, 2012/C 326/01.
EUs charter om grunnleggende rettigheter	Charter on the Fundamental Rights of the European Union, 18.12.2000, 2000/C 364/01.
Personvernforordningen	Europaparlamentets- og Rådsforordning (EU) 2016/679 av 27. April 2016 om vern av fysiske personer i forbindelse med behandling av personopplysninger og om fri utveksling av slike opplysninger samt om oppheving av direktiv 95/46/EF (generell personvernforordning).

Personverndirektivet	Europaparlaments- og Rådsdirektiv 95/46/EF av 24. oktober 1995 om beskyttelse av fysiske personer i forbindelse med behandling av personopplysninger og om fri utveksling av slike opplysninger.
EØS-avtalen	Avtale om Det europeiske økonomiske samarbeidsområde av 2. mai 1992, AVT-1992-05-02-1.

8.3 Veiledninger, retningslinjer, rapporter og offentlige dokumenter fra internasjonale organer og institusjoner

Personvernrådet (2018)	European Data Protection Bord, *Endorsement 1/2018 of Working Party Article 29 Documents*, 2018, https://edpb.europa.eu/sites/default/files/files/news/endorsement_of_wp29_documents_en_0.pdf (lest: 15.10.2021).
Europarådet (2018)	Council of Europe: The Committee of Experts on Internet Intermediaries**,** *Algorithms and Human Rights*, DGI (2017)12, 2018.
Artikkel 29-gruppen (2018a)	Article 29 Data Protection Working Party, *Guidelines on transparency under Regulation 2016/679*, 29.11.2017, revidert og tilsluttet av Personvernrådet 11.04.2018.
Artikkel 29-gruppen (2018b)	Article 29 Data Protection Working Party, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*, 03.10.2017, revidert og tilsluttet av Personvernrådet 06.02.2018.

Personvernrådet (2019a) European Data Protection Bord, *Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects*, 2. versjon, 08.10.2019.

Personvernrådet (2019b) European Data Protection Bord, *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default*, 13.11.2019.

KI-ekspertgruppen (2019a) Independent High Level Expert Group set up by the European Commission *A definition of AI: Main capabilities and disciplines*, 2019.

KI-ekspertgruppen (2019b) Independent High Level Expert Group on Artificial Intelligence set up by the European Commission, *Ethics guidelines for trustworthy AI*, 2019.

Artificial Intelligence Act (AIA) Proposal for a Regualtion of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final.

8.4 Uttalelser, rapporter og andre dokumenter fra norske organer og institusjoner

Datatilsynet (2021) Aktuelle nyheter 2021, «*Lukker IB-saken*», 16.07.2021. Tilgjengelig på: https://www.datatilsynet.no/aktuelt/aktuelle-nyheter-2021/lukker-ib-saken/ (lest 11.10.2021).

Datatilsynet (2020a) Sak 20/03087 (IB-saken).

Datatilsynet (2020b) Sak 20/03087 (IB-saken), saksdokument 14, 07.08.2020. Tilgjengelig på: https://www.datatilsynet.no/contentassets/04df776f85f64562945f1d261b4add1b/advance-notification-of-order-to-rectify-unfairly-processed-and-incorrect-personal-data.pdf (lest 09.10.2021)

Datatilsynet (2018) *Kunstig intelligens og personvern*, publisert januar 2018. Tilgjengelig på: https://www.datatilsynet.no/globalassets/global/dokumenter-pdfer-skjema-ol/rettigheter-og-plikter/rapporter/rapport-om-ki-og-personvern.pdf (lest 15.11.2021).

Kommunal- og moderniseringsdepartementet (2020) Kommunal og moderniseringsdepartementet, *Nasjonal strategi for kunstig intelligens*, publisert 14.01.2020. Tilgjengelig på: https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf (lest 11.10.2021)

8.5 Litteratur

Bendiksen & Hansen (2019) Bendiksen, Christian & Hansen, Eirik Norman, *Når juss møter AI*, Gydendal 2019.

Jarbekk & Sommerfeldt (2019) Jarbekk, Eva, & Sommerfeldt, Simen, *Personvern og personvernforordningen i praksis*, Cappelen Damm 2019.

Kahneman (2011) Kahneman, Daniel, *Thinking Fast and Slow*, Penguin Books 2011.

Schartum (2020) Schartum, Dag Wiese, *Personvernforordningen – en lærebok*, Fagbokforlaget 2020.

Sejersted m.fl. (2011) Sejersted, Fredrik, Arnesen, Finn, Rognstad, Ole-Andreas, Foyn, Sten & Kolstad, Olav, *EØS-rett*, 3. utgave, Universitetsforlaget 2011.

Skullerud, m.fl. (2018) Skullerud, Åste Marie Bergseng, Rønnevik, Cecilie, Skorstad, Jørgen, & Pellerud, Marius Engh, *Personopplysningsloven og personvernforordningen (GDPR) : Kommentarutgave*, Universitetsforlaget 2018.

Stemsrud (2016) Stemsrud, Odd, *EØS-rett i et nøtteskall*, Gyldendal 2016.

8.6 Rapporter og artikler publisert på nett, internettsider og brev

ACLU sine nettsider, «Surveillance Under the Patriot Act" https://www.aclu.org/issues/national-security/privacy-and-surveillance/surveillance-under-patriot-act (lest 09.12.2021).

Angwin, Julia, Larson, Jeff, Mattu, Surya, & Kirchner, Lauren, "Machine Bias", *ProPublica*.com, 23.05.2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (lest 12.10.21).

Brown, Deborah & Toh, Amos, "Technology is Enabling Surveillance, Inequality During the Pandemic", *hrw.org*, https://www.hrw.org/news/2021/03/04/technology-enabling-surveillance-inequality-during-pandemic (lest 09.12.2021).

Buolamwini, Joy og Gebru, Timnit, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", *Conference on fairness, accountability and*

transparency, Proceedings of Machine Learning Research vol. 81 januar 2018, s. 77-91, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf (lest 07.12.2021).

Datatilsynets nettsider, «Sandkassesiden», https://www.datatilsynet.no/regelverk-og-verktoy/sandkasse-for-kunstig-intelligens/ (lest 23.11.2021).

Datatilsynets nettsider, Regelverk og verktøy, *«Det europeiske personvernrådet (Personvernrådet)»*, sist endret 19.07.2020, https://www.datatilsynet.no/regelverk-og-verktoy/internasjonalt/personvernradet/ (lest 06.10.2021).

EU-kommisjonens nettsider, "High-level expert group on artificial intelligence", sist endret 27.09.2021, https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.(lest 11.10.2021).

Føyen Advokatfirma v/Arve Føyen, brev til Datatilsynet i sak 20/03087 (IB-saken), 03.09.2020, tilgjengelig på: http://nrkbeta.no/wp-content/uploads/2020/09/2020-09-03-Letter-from-FT-IBO-result-anomalies1377198.1.pdf (lest 25.11.2021).

Geiger, Gabriel, "How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud", *vice.com*, 01.03.2021, https://www.vice.com/en/article/jgq35d/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud (lest 09.12.2021).

Hardesty, Larry, "Study finds gender and skin-type bias in commercial artificial-intelligence systems", *MIT News*, 11.02.2018, https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212 (lest 06.12.2021).

IBO sine nettsider, "IB update on May 2020 Diploma Programme and Career-related Programme results", 17.08.2020, https://www.ibo.org/news/news-about-the-ib/update-m20-dp-cp-results/?fbclid=IwAR0s07X4C47WdqsuLJlSFJX__H7s7l_fNjFrwwSqicuIxekKtJioO21A7v0 (lest 25.11.2021).

Knight, Will, "The Apple Card Didn't "See" Gender - and That's the Problem", 19.11.2019, https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/ (lest 19.10.2021).

Metz, Cade, "A.I. Shows Promise Assisting Physicians", *The New York Times*, 11.02.2019, https://www.nytimes.com/2019/02/11/health/artificial-intelligence-medical-diagnosis.html (lest 07.12.2021).

ProPublicas nettsider, "About Us", https://www.propublica.org/about/ (lest 25.11.2021).

Regjeringens nettsider, «Domstolkommisjonen», https://www.regjeringen.no/no/dep/jd/org/styre-rad-og-utval/innstillinger/innstillinger-fra-utvalg/innstillinger-levert-i-2020/domstolkommisjonen/id2565732/ (lest 11.10.2021).

Sævold, Heidi, «Skal hjelpe norske selskaper å sikre trygg bruk av kunstig intelligens: Nå ønsker de innspill om konkrete temaer», *digi.no*, 12.10.2020, https://www.digi.no/artikler/skal-hjelpe-norske-selskaper-a-sikre-trygg-bruk-av-kunstig-intelligens-na-onsker-de-innspill-om-konkrete-temaer-br/500742 (lest 27.11.2021).

Thorsen, Dag Einar, «Sosialdemokrati» i Store norske leksikon, 28.12.2020, https://snl.no/sosialdemokrati, (lest 04.12.2021).

Quest, Lisa, Charrie, Anthony, du Croo de Jong, Lucas og Roy, Subas, "The Risks and Benefits of Usin AI to Detect Crime", *Harvard Business Review*, 09.08.2018, https://hbr.org/2018/08/the-risks-and-benefits-of-using-ai-to-detect-crime (lest 07.12.2021).

8.7 Tabeller og figurer

Illustrasjon av dype nevrale nettverk, 2. 15: Michael Nielsen, *Neural Networks and Deep Learning*, kapittel 5, tilgjengelig på www.neuralnetworksanddeeplearning.com (lest 11.10.2021).

Tabell med eksempler på karakterutregning, s. 46: Føyen Advokatfirma v/Arve Føyen, brev til Datatilsynet i sak 20/03087 (IB-saken), 03.09.2020, http://nrkbeta.no/wp-content/uploads/2020/09/2020-09-03-Letter-from-FT-IBO-result-anomalies1377198.1.pdf (lest 25.11.2021).