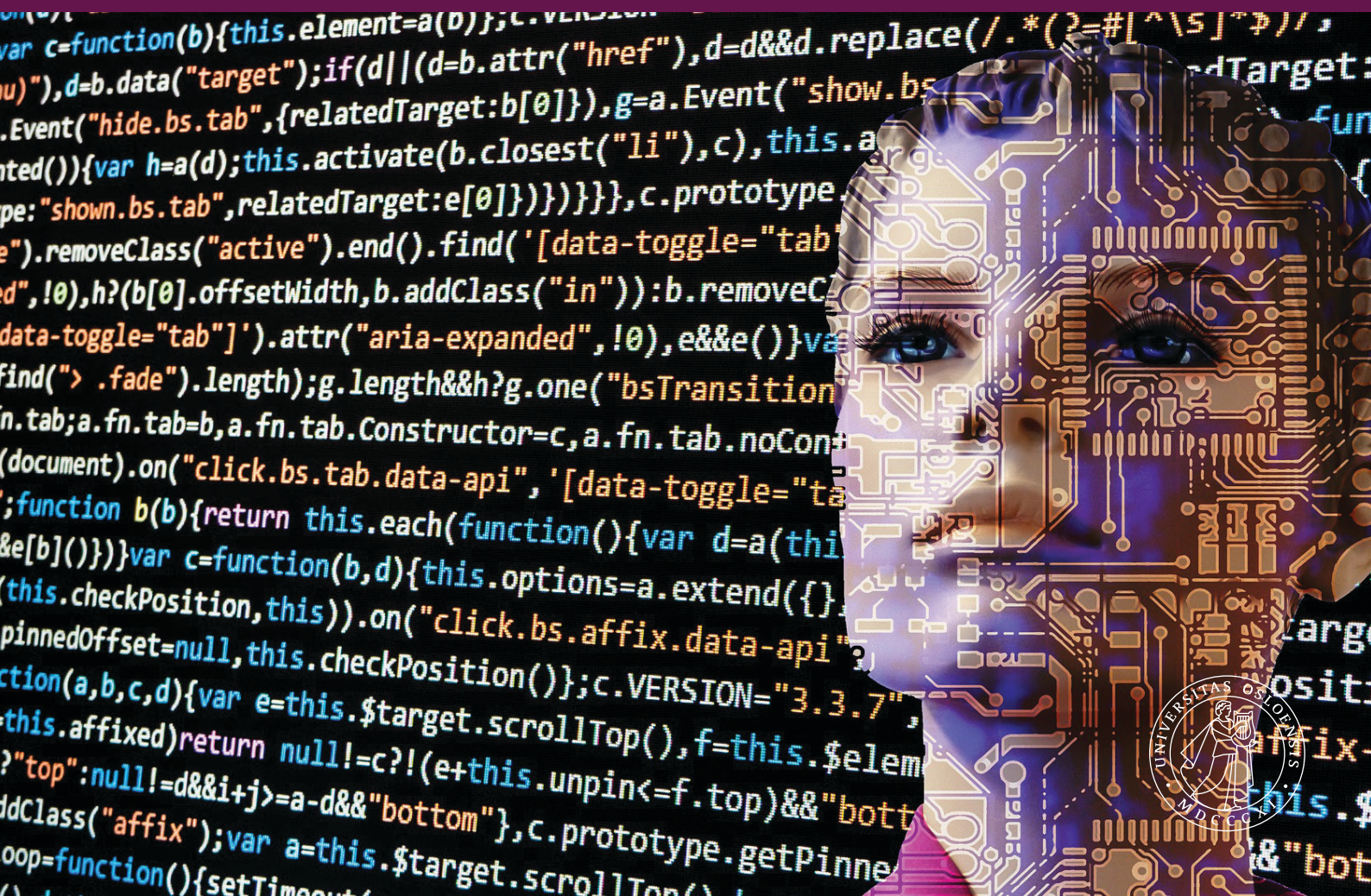


Borgernes rettigheter ved utvikling og bruk av kunstig intelligente systemer i offentlig forvaltning



Til Arbeids- og velferdsdirektoratet ved Seksjon for informasjonsforvaltning

Ved mandat gitt i faget lovgivningslære våren 2022 ble det oppnevnt et utvalg til å utrede gjennomføring av europeiske harmoniserte regler om kunstig intelligens i norsk forvaltning.

Utvalget legger med dette fram sin utredning og forslag til nye regeltekster. Utredningen er enstemmig.

Oslo, 8. juni 2022

Viktorija Østby

Clarissa Lage Dahl

Anna Pettersdatter Melkild

Innhold

1	Innledning	5
1.1	Om mandatet	5
1.2	Utvalgets sammensetning.....	5
1.3	Presiseringer og avgrensninger	5
1.4	Sentrale begreper	6
2	Bakgrunnen for lovforslaget	7
2.1	Kunstig intelligens – gevinster og utfordringer	7
2.1.1	Hva er kunstig intelligens?	7
2.1.2	Gevinster ved bruk av kunstig intelligens i velferdsforvaltningen	11
2.1.3	Utfordringer ved bruk av kunstig intelligens i velferdsforvaltningen.....	13
2.2	EUs forslag til forordning om harmoniserte regler for kunstig intelligens	16
2.2.1	Den europeiske tilnærmingen til kunstig intelligens.....	17
2.2.2	Et sammendrag av innholdet i EUs forslag til forordning om kunstig intelligens.....	20
2.3	Utviklingen på området	24
2.3.1	I Norge.....	24
2.3.2	I øvrige land.....	29
2.4	EØS-relevans.....	30
2.4.1	EØS-avtalens saklige virkeområde	31
2.4.2	EØS-avtalens geografiske virkeområde.....	32
2.4.3	Bør reglene innlemmes?	32
2.5	Behovet for lovendringer som følge av bruk av kunstig intelligente systemer i forvaltningen	35
2.5.1	En ny lov om kunstig intelligens	35
2.5.2	Endringer i forvaltningsloven	35
3	Overordnet om dagens rettskilddebilde på forvaltningsrettens område	41
3.1	Relevant lovgivning.....	41
3.1.1	Overordnede rettsnormer.....	41
3.1.2	Internasjonale forpliktelser og føringer	43
3.1.3	Nasjonal lovgivning	44
3.2	Forvaltningsrettslige prinsipper og regler	46
3.2.1	Forvaltningsrettslige prinsipper	46
3.2.2	Grunnprinsippenes utslag i forvaltningsloven.....	49
4	Grunnleggende prinsipper for utvikling og bruk av kunstig intelligente systemer	52
4.1	Transparens; forklarbarhet, etterprøvbarehet, åpenhet	52
4.1.1	Forklarbarhet	53
4.1.2	Etterprøvbarehet	56
4.1.3	Åpenhet.....	58
4.1.4	Målestokk og terskel for oppfyllelse av kravene om transparens.....	59
4.2	Forsvarlig forvaltning av data	60
4.2.1	Bakgrunn	60
4.2.2	Hva er dataforvaltning?.....	60
4.2.3	Hvorfor er dataforvaltning viktig?.....	63
4.2.4	Hva innebærer prinsippet om forsvarlig forvaltning av data?	64
4.2.5	Prinsippet føringer for norsk velferdsforvaltning	68

4.3	Teknisk robusthet og sikkerhet	69
4.3.1	Bakgrunn	69
4.3.2	Nøyaktighet.....	71
4.3.3	Robusthet.....	72
4.3.4	Cybersikkerhet	73
4.3.5	Pålitelighet og reproduserbarhet.....	73
4.4	Rettferdige, riktige og ikke-diskriminerende avgjørelser	74
4.4.1	Bakgrunn	74
4.4.2	Nærmere om problemstillingen.....	75
4.4.3	Hva er «rettferdige» avgjørelser?	75
4.4.4	Hva er «riktige» avgjørelser?	77
4.4.5	Hva er «ikke-diskriminerende» avgjørelser?.....	78
4.5	Personvern	83
4.5.1	Nærmere om personvernforordningens krav ved automatiserte avgjørelser	84
4.6	Individuell begrunnelse.....	87
4.7	Borgernes livskvalitet og velferd	88
4.7.1	Bakgrunn	88
4.7.2	Bærekraftig og miljøvennlig	90
4.7.3	Sosial påvirkning.....	92
4.7.4	Samfunn og demokrati.....	94
4.7.5	Ivaretagelse av borgernes livskvalitet og velferd	95
5	<i>Forvaltningens utvikling og bruk av kunstig intelligente systemer i lys av grunnleggende prinsipper</i>	96
5.1	Hvilke krav til utvikling og bruk bør stilles i lov og forskrift?	96
5.1.1	Lovfesting av grunnleggende prinsipper og operasjonaliseringen av disse	96
5.1.2	Utvikling av kunstig intelligente systemer	96
5.1.3	Krav til systemvedtaks innhold	97
5.2	Hva bør være konsekvensene av feil i systemet og uriktige systemvedtak?	98
5.2.1	«Effective redress». Et effektivt rettsmiddel	98
5.2.2	«Accountability». Ansvarlighet	99
5.2.3	Behovet for særskilte reguleringer om klage, omgjøring og retting av systemvedtak	100
5.2.4	Rettslig systeminteresse?.....	100
5.3	Hva vil skille saksbehandlingen ved bruk av kunstig intelligente systemer fra tradisjonell saksbehandling?.....	101
5.3.1	Effektiv saksgang	101
5.3.2	Systemoffentlighet og veiledningsplikt	102
5.4	Tilsyn og kontroll	102
5.4.1	Bakgrunn	102
5.4.2	Kontroll i enkeltsaker	103
5.4.3	Kontroll på systemnivå.....	105
6	<i>Administrative og økonomiske konsekvenser</i>	111
7	<i>Lovforslag.....</i>	112
8	<i>Merknader til bestemmelsene</i>	116
9	<i>Kilder.....</i>	125
9.1	Norske rettskilder	125
9.2	Internasjonale rettskilder.....	128

9.3	Litteratur	130
9.4	Annet.....	133

1 Innledning

1.1 Om mandatet

Oppdragsgiver er Arbeid- og velferdsforvaltningen (NAV) Seksjon for informasjonsforvaltning.

Utvalget har fått følgende mandat:

Studentgruppen skal vurdere om eventuelle europeiske regler om harmonisering av regler om kunstig intelligens er EØS-relevante. Det vil si om dette er regler som faller innenfor EØS-avtalens saklige virkeområde, og om Norge bør innlemme reglene i EØS-avtalen.

Studentgruppen skal vurdere hvordan de europeiske reglene om harmonisering av regler om kunstig intelligens kan gjennomføres på en hensiktsmessig måte i norsk rett. Gruppen skal vurdere om gjeldende regler må endres og foreslå forslag til nye regler. Gruppen skal ta stilling til hva som er en tilstrekkelig gjennomføringsmåte.

Studentgruppen skal særlig vurdere hvilken betydning EUs regler har for velferdsstatlig digitalisering og bruk av kunstig intelligens og særlig for regelverket som er relevant for Arbeids- og velferdsdirektoratet, herunder NAV-loven og folketryktdloven.

Studentgruppen kan foreslå en ny lov om bruk av kunstig intelligens i forvaltningen eller velferdsforvaltningen, eller foreslå endringer i eksisterende lovgivning.

1.2 Utvalgets sammensetning

Utvalget består av tre studenter i rettsvitenskap ved Universitetet i Oslo:

Viktorija Østby

Clarissa Lage Dahl

Anna Pettersdatter Melkild

1.3 Presiseringer og avgrensninger

Hovedproblemstillingen er hvordan EUs forslag til forordning om harmonisering av regler om kunstig intelligens (COM(2021) 206) kan gjennomføres i norsk rett. Mandatet er svært vidt og fordrer nærmere avgrensning.

Utvalget har valgt å fokusere på utformingen av regler om kunstig intelligens på forvaltningsrettens område, med særlig fokus på velferdsstatlig digitalisering og bruk av kunstig intelligens. Utvalget vil kort behandle EØS-relevansen til de europeiske reglene om harmonisering av kunstig intelligens.

Utvalget foreslår å implementere et nytt kapittel i forvaltningsloven om bruk av kunstig intelligens. I arbeidet har vi sammenholdt grunnprinsippene i de europeiske reglene om harmonisering av kunstig intelligens med forvaltningsrettens grunnleggende prinsipper, overordnede rettsnormer og annen relevant lovgivning.

1.4 Sentrale begreper

Selv om flere begreper vil bli definert fortløpende i den videre fremstillingen, ser vi det som hensiktsmessig med en kort redegjørelse for utredningens mest sentrale begreper her innledningsvis. Begrepet kunstig intelligens redegjøres det for i punkt 2. Med kunstig intelligente systemer (KI-systemer) sikter vi til systemer som anvender kunstig intelligens.

For utvalget har begrepene «rettslig systemavgjørelse» og «systemvedtak» stått særlig sentralt. Med rettslige systemavgjørelser menes en avgjørelse med rettslig innhold som treffes som ledd i utviklingen av kunstig intelligente systemer. Med systemvedtak mener vi enkeltvedtak generert av et kunstig intelligent system.

Utvalget vil også benytte seg av begrepene «input» og «output». Førstnevnte betegner den informasjonen som gis et system, mens sistnevnte sikter til det resultatet det kunstig intelligente systemet produserer.

Det er også grunn til å si noen ord om begrepet *bias*. Forslaget til forordningen bruker 'bias' som et samlebegrep for ulike utilstrekkeligheter ved et kunstig intelligent system, som gjør at systemet gir uønskede resultater. På norsk oversettes bias til forutinntatthet eller partiskhet, «skjevhet i vitenskapelig undersøkelse eller resultat, især pga. mangelfull systematikk i innsamlingen av data».¹ Nasjonal strategi for kunstig intelligens bruker begrepet om utvalgsskjevhet.² På engelsk defineres begrepet som «the fact of preferring a particular subject or thing» eller «the fact that information is not correct because of the method used in collecting or presenting it».³ En tredje definisjon er «the inclination or prejudice of a decision made by an AI system which is for or against one person or group, especially in a way considered to be unfair».⁴ Utvalget mener at definisjonen av bias som forutinntatthet, partiskhet og skjevhet kan bli for snever. I denne utredningen vil begrepet brukes om tilbøyeligheten som et kunstig intelligent system har til å systematisk treffe avgjørelser som avviker fra det som er korrekt, og som setter et individ eller en gruppe mennesker i en urettferdig posisjon på grunn av mangler ved datagrunnlaget.

Fremstillingen vil omhandle flere ulike aktører i møte med bruken av kunstig intelligente systemer. Aktørene defineres som følger:

- Bruker: Enhver fysisk eller juridisk person, offentlig myndighet eller enhver virksomhet eller annet organ, som anvender et KI-system under sin myndighet, såfremt systemet ikke benyttes som ledd i en personlig ikke-ervervsmessig aktivitet.⁵
- Tilbyder: En fysisk eller juridisk person, en offentlig myndighet, en virksomhet eller et annet organ, som utvikler eller får utviklet et KI-system med det formål å omsette systemet eller benytte det under eget navn og varemerke, enten mot betaling eller gratis.⁶

¹ Det Norske Akademis Ordbok, s.v. «bias». <https://naob.no/ordbok/bias>.

² Nasjonal strategi for kunstig intelligens.

³ Cambridge Dictionary, s.v. «bias». <https://dictionary.cambridge.org/dictionary/english/bias>.

⁴ Ntoutsis (2020) s. 4.

⁵ COM(2021) 206 art. 3 nr. 4.

⁶ COM(2021) 206 art. 3 nr. 2.

- Importør: Enhver fysisk eller juridisk person etablert i EU/EØS, som omsetter eller benytter et KI-system, som bruker et navn eller varemerke til en fysisk eller juridisk person etablert utenfor EU/EØS.⁷
- Distributør: Enhver fysisk eller juridisk person i omsetningskjeden utover tilbyderen og importøren, som gjør et KI-system tilgjengelig på markedet i EU/EØS uten å ha innflytelse på systemets egenskaper.⁸

2 Bakgrunnen for lovforslaget

2.1 Kunstig intelligens – gevinster og utfordringer

2.1.1 Hva er kunstig intelligens?

En redegjørelse for hva som ligger i begrepet «kunstig intelligens» følger av JOU 2020: 5 *Kunstig intelligens – muligheter og risikoer i velferdsforvaltningen*. I denne utredningen vil utvalget ta utgangspunkt i EUs definisjon av kunstig intelligens, slik den kommer til uttrykk i forordningsforslaget, og vurdere begrepets innhold og betydning på grunnlag av denne definisjonen.

2.1.1.1 Forordningsforslagets definisjon

EUs ekspertgruppe for kunstig intelligens foreslo å definere kunstig intelligens som *systemer som utviser intelligent adferd ved å analysere sine omgivelser og utføre handlinger med en viss grad av autonomi for å oppnå spesifikke mål*.⁹ I tillegg utviklet Joint Research Centre (JRC) en operasjonell definisjon av kunstig intelligens, basert på en kartlegging av alle betydninger av kunstig intelligens fra et politisk, forskningsmessig og industrielt perspektiv.¹⁰ EU-kommisjonen kom imidlertid til at begrepet «kunstig intelligent system» av rettssikkerhetshensyn burde ha en klarere definisjon, ettersom det vil være avgjørende for om reglene i forordningsforslaget vil komme til anvendelse. Kommisjonens definisjon er i stor grad basert på en definisjon som allerede er brukt av OECD.¹¹

Forordningsforslaget definerer et kunstig intelligent system på følgende måte:

‘artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;

Et kunstig intelligent system er følgelig en *programvare* som er utviklet ved hjelp av en eller flere av teknikkene eller tilnærmingene som er opplistet i forslagens vedlegg I:

ANNEX I
ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES
referred to in Article 3, point 1

⁷ COM(2021) 206 art. 3 nr. 6

⁸ COM(2021) 206 art. 3 nr. 7

⁹ High-Level Expert Group on Artificial Intelligence (2018).

¹⁰ Magiega, Tambiama (2022).

¹¹ OECD (2019).

- (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
- (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
- (c) Statistical approaches, Bayesian estimation, search and optimization methods.

EU-kommisjonen skal holde listen oppdatert i tråd med utviklingen i markedet og teknologien, og foreta endringer ved å vedta delegerede rettsakter.¹²

Bokstav (a) refererer til *maskinlæring*, som er en spesialisering innenfor eller underkategori av kunstig intelligens som fagfelt eller informasjonsteknologi. Ved maskinlæring må et kunstig intelligent system *selv* finne ut hvordan det skal løse et problem som ikke nøyaktig kan spesifiseres, eller hvor problemløsningen ikke kan beskrives ved hjelp av symbolske regler for resonnementer (type «hvis X, så Y»)¹³. Systemet «lærer» å finne mønstre i store datamengder, i stedet for å bli programmert til å utføre prosesser på en bestemt måte. Jo mer data systemet får å jobbe med, og jo mer det får «øve seg» på å finne mønstre, desto mer nøyaktige blir resultatene.

Det er tre hovedformer for maskinlæring: *veiledet læring* («supervised learning») *ikke-veiledet læring* («unsupervised learning») og *forsterket læring* («reinforcement learning»). Ved veiledet læring gir vi systemet eksempler på problemet og løsningen, slik at systemet kan lære å generalisere fra disse eksemplene og identifisere løsningene i andre situasjoner. Hvis vi lar systemet analysere mange bilder av hunder og forteller at bildene forestiller en hund, vil systemet (forhåpentligvis) klare å identifisere en hund også når det blir forelagt nye bilder. Det er imidlertid viktig at systemet har fått øve seg på et tilstrekkelig *representativt og mangfoldig* datagrunnlag. Hvis vi utelukkende benytter bilder av små, hvite hunder, vil systemet ikke ha forutsetninger for å klare å klassifisere en schæfer som en hund. Datakvalitet har derfor stor betydning ved veiledet maskinlæring.

Ved ikke-veiledet læring lar vi systemet selv finne mønstre og sammenhenger i et datasett, for eksempel ved å sortere en gruppe mennesker etter trekk som systemet finner felles for disse. En slik tilnærming kan oppdage mønstre og systemer som tidligere ikke var kjent for mennesker.¹⁴ Ikke-veiledet læring kan brukes til å trene opp *dyplæringsalgoritmer*, som anvender en datastruktur av *nevrale nettverk* inspirert av den menneskelige hjernen.¹⁵ Enkelt forklart vil et slikt system kunne få input fra flere ulike kilder, og foreta analyser på flere nivåer for å komme frem til løsningen. Det kan for eksempel lære seg at hunder kjennetegnes ved at de har en snute, og ikke en nese.

Forsterket læring går ut på at systemet får menneskelige tilbakemeldinger på om løsningen som systemet kommer frem til er god eller dårlig. Systemet lærer å finne den beste strategien for komme til løsningen ved å prøve, feile og bli korrigert underveis.¹⁶ Målet er at systemet skal få flest mulig gode tilbakemeldinger. Denne tilnærmingen brukes ofte ved opplæring av et system som skal gi anbefalinger, for eksempel i forbindelse med markedsføring.

¹² COM(2021) 206 fortalepunkt 6.

¹³ High-Level Expert Group on Artificial Intelligence (2018).

¹⁴ Teknologirådet (2018).

¹⁵ Store Norske Leksikon (2021).

¹⁶ Teknologirådet (2018).

Det er viktig å være bevisst på at bruk av maskinlæringsteknikker innebærer en risiko for uriktige resultater. Det vil alltid være en avviksprosent, selv om den kan være veldig liten. En god indikator på hvor «riktige» resultater man kan forvente å få fra et system er systemets *nøyaktighet*, som er et mål på prosentandelen riktige resultater i forhold til uriktige.¹⁷

Bokstav (b) refererer til kunstig intelligente systemer som anvender *logikk- og kunnskapsbaserte tilnærminger* («logic- and knowledge-based approaches»), det vil si symbolske (menneskelig lesbare) representasjoner av problemstillinger og logikk. Det kan ses på som en form for kompleks programmering, hvor vi gir systemet informasjon og forteller hvordan systemet skal behandle denne informasjonen. Systemet kan da prosessere store mengder menneskelig kunnskap på en effektiv måte.

Slike tilnærminger forutsetter at systemet har tilgang på informasjon i form av kunnskap. Informasjonen må formuleres eller *representeres* på en måte som systemet klarer å tolke («knowledge representation»). En samling av slik «ferdigtolket» kunnskap kalles for en *kunnskapsbase* («knowledge base»). Systemet prosesserer kunnskapen ved hjelp av matematiske algoritmer, det vil si instruksjoner for hvordan kunnskapen skal behandles ved hjelp av matematisk logikk. Systemets *slutningsmotor* («inference engine») anvender disse logiske algoritmene på kunnskapen i kunnskapsbasen, og utleder («deduces») ny informasjon. Systemer som typisk benytter denne formen for kunstig intelligens er *ekspertsystemer* («expert systems»), som for eksempel kan brukes i helsevesenet som hjelpemiddel for å stille diagnoser.¹⁸

Bokstav (c) refererer til *statistiske tilnærminger* («statistical approaches»), Bayesiansk estimering («Bayesian estimation»), søke- og optimaliseringsmetoder («search and optimization methods»). Etter utvalgets forståelse gjelder dette punktet kunstig intelligente systemer som brukes til å forstå data, hva som skjer når data endrer seg og optimalisere søkefunksjoner for å finne frem til den beste løsningen.

Definisjonen har blitt kritisert for å være for vid, og at den går utover det som normalt vil bli ansett som «intelligent».¹⁹ Det har blitt anført at alternativet i ets bokstav b i prinsippet vil omfatte et bredt spekter av programvare, da det meste av eksisterende programvare er «logikk- eller kunnskapsbasert» («logic- or knowledge based»)²⁰ At definisjonen også inkluderer teknikker som beskrevet i ets bokstav c, kan i praksis innebære at programvare som bruker en vanlig lineær regresjonsmodell vil falle innenfor begrepet «kunstig intelligens».²¹

EU-kommisjonens åpne høring rundt forordningsforslaget fikk 304 høringssvar – et langt høyere antall enn hva som er normalt tilfelle for andre teknologirelaterte reguleringer. Teknologiselskaper og bransjeorganisasjoner har hatt et sterkt ønske om å påvirke reglens utforming. Flere medlemsland og aktører har forespurt mer klarhet rundt regelverket, spesielt rundt definisjonen av «kunstig intelligens» og klassifikasjonen av forbudte og høyrisiko kunstig intelligente systemer.²²

¹⁷ High-Level Expert Group on Artificial Intelligence (2018) s. 5.

¹⁸ Store Norske Leksikon (2021).

¹⁹ Clarke (2021).

²⁰ Ibid.

²¹ Barczentewicz (2021).

²² Clarke (2021).

I et nyere dokument – en «Precedency Compromise Text» fra det slovenske formannskapet i EU – har definisjonen av «kunstig intelligens» blitt snevret inn ytterligere.²³ Den reviderte definisjonen ser slik ut:

‘artificial intelligence system’ (AI system) means a system that

- (i) receives machine and/or human-based data and inputs,
- (ii) infers how to achieve a given set of human-defined objectives using learning, reasoning or modelling implemented with the techniques and approaches listed in Annex I, and
- (iii) generates outputs in the form of content (generative AI systems), predictions, recommendations or decisions, which influence the environments it interacts with;

I tillegg er det foreslått endringer i reguleringens fortalepunkt 6, hvor det nå fremheves at definisjonen av kunstig intelligens bør være basert på egenskaper som skiller kunstig intelligente systemer fra «klassisk programvare og programmering». Det er også lagt til at systemet skal kunne oppnå et gitt menneskedefinert mål ved å *lære, resonnerer eller modellere*. Disse endringene skal bringe klarhet rundt det juridiske begrepet «kunstig intelligens», og bidra til å unngå at tradisjonell programvare som normalt ikke anses for å være kunstig intelligens, inkluderes i virkeområdet til den foreslåtte reguleringen.²⁴

Den reviderte definisjonen gir ikke uttrykk for Europarådets endelige posisjon, men vil igjen diskuteres med Europaparlamentet og EU-kommisjonen (også kjent som «trilogforhandlinger» på EU-sjargong).²⁵ Den enorme interessen rundt formuleringene tilsier imidlertid at lovens virkeområde ønskes redusert.

Et forslag er å definere kunstig intelligente systemer ut fra deres *egenskaper* eller *effekter*, i stedet for *teknikker*. Denne tilnærmingen begrunnes med at en for teknisk definisjon fort risikerer å bli utdatert med den teknologiske utviklingen. Definisjonen til EUs ekspertgruppe for kunstig intelligens gir i større grad uttrykk for en slik tilnærming.²⁶ I forlengelse av dette er det også foreslått å definere *høyrisiko* kunstig intelligente systemer ut fra sannsynligheten for at systemet kan påføre noen skade.²⁷

En for omfattende definisjon av kunstig intelligens vil, som det er argumentert for, medføre risiko for at også programvare som i prinsippet *ikke* er det vi anser som «kunstig intelligent» vil omfattes av reguleringene i forordningsforslaget. Dette kan redusere insentivene til å utvikle ny teknologi som anvender kunstig intelligens, fordi virksomheter ikke ønsker eller ikke har råd til å bli pålagt det ansvaret som følger av regelverket. Dette gjelder kanskje særlig for små og mellomstore bedrifter. Det vil koste penger å vurdere om systemet samsvarer med de kravene som stilles, og det vil koste enda mer å la være dersom kravene ikke blir oppfylt. Regelverket kan følgelig bli en hindring for utviklingen og bruken av nyttige verktøy, som bør være allment tilgjengelige i *Europas digitale tiår*.

Et argument mot å innskrenke definisjonen er at det vil stille for mange krav til et system før det kan klassifiseres som et kunstig intelligent system. Det vil medføre en risiko for at systemer som ikke er tilstrekkelig kvalifiserte for den snevre definisjonen ikke blir fanget opp av regelverket, selv om de i realiteten anvender kunstig intelligens. Det har blitt argumentert for at ønsket om en mer innskrenket definisjon er et forsøk på å redusere kostnadene for større

²³ 2021/0106(COD).

²⁴ Ibid. s. 3 og 8.

²⁵ Kayser-Bril (2021).

²⁶ Clarke (2021).

²⁷ Ibid.

virksomheter og myndigheter, som vil påløpe ved slik regulering av kunstig intelligente systemer.²⁸

En utfordring i forbindelse med definisjonen av kunstig intelligens er at regler som kan virke fornuftige for lovgivere og rettsanvendere, ikke nødvendigvis stemmer overens med den *tekniske realiteten* ved programutvikling. Det reiser spørsmålet om hva forordningsforslaget søker å regulere. Er målet å regulere programvare- og teknologisk utvikling? Eller er hensikten å regulere *risikoen* ved bestemte former for kunstig intelligens? Dersom formålet er å regulere risikoen, vil det være en fordel om definisjonene tar utgangspunkt i systemets egenskaper og virkninger på sine omgivelser.²⁹

2.1.1.2 Definisjonen i den nasjonale strategien for kunstig intelligens

Den nasjonale strategien for kunstig intelligens tar utgangspunkt i EUs ekspertgruppes definisjonen av kunstig intelligens, og definerer kunstig intelligens på følgende måte:

«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.»³⁰

2.1.2 Gevinster ved bruk av kunstig intelligens i velferdsforvaltningen

Flere og mer komplekse arbeidsoppgaver fordrer en *digital transformasjon* i forvaltningen. Begrepet «digital transformasjon» innebærer en endringsprosess og «et redesign av virksomheten på alle nivå».³¹ Virksomheten må endre hvordan den utfører sine oppgaver, tilbyr tjenester, jobber mer effektivt eller skaper nye tjenester. Digital transformasjon handler ikke utelukkende om digitalisering, men om å løse oppgaver på nye måter.³²

Samtidig ligger det i begrepet at en slik transformasjon må bygge på *digital teknologi*. Teknologien er i stadig endring og utvikling, og et viktig element i en digital transformasjon er å utnytte mulighetene ved ny teknologi til å muliggjøre nye tjenester og løsninger. Det bør også være rom for å eksperimentere og lære av erfaringer. Teknologiske systemer må tilpasses sine omgivelser og brukere. Forbedring er en løpende prosess.³³

Et viktig element i en digital transformasjon er *utnyttelse av data*. Norsk offentlig sektor har enorme mengder informasjon, som må betraktes og behandles som en ressurs. Informasjonen kan brukes til å utvikle bedre tjenester og løse samfunnsoppdrag på ny måter.³⁴

Et annet viktig element er *brukerne*. Digital transformasjon tvinger brukerne til å endre sin måte å interagere med offentlig sektor på. Samtidig forventer brukerne at også offentlig sektor følger den teknologiske utviklingen, slik at offentlige tjenester er konkurransedyktige i møte med kommersielle tjenester.³⁵

²⁸ Bryson (2022).

²⁹ Barczentewicz (2021).

³⁰ Kommunal- og moderniseringsdepartementet (2020) pkt. 1.1.

³¹ Digitaliseringsdirektoratet (u.å.).

³² Ibid.

³³ Ibid.

³⁴ Ibid.

³⁵ Ibid.

Den åpenbare fordelene ved kunstig intelligente systemer er at de kan imitere menneskelige handlinger og resonneringer på en mer systematisk og effektiv måte. Kunstig intelligente systemer har kapasitet til å prosessere mye data på kort tid, og kan bruke sekunder på å finne informasjon som et menneske ville ha brukt langt lenger tid på å søke etter. Ved hjelp av kunstig intelligens kan offentlig sektor utnytte potensialet i den tilgjengelige informasjonen, og anvende data som en ressurs. Samtidig kan offentlige tjenester forbedres, fornyes og utvikles i takt med teknologien, og gi brukerne en fullverdig brukeropplevelse.

I dag oppsøker de fleste av oss offentlige tjenester *ved behov*. Da er det viktig at tjenestene er lett tilgjengelige, enkle å bruke og effektive. Her har digitaliseringen spilt en viktig rolle. Samtidig er det mange utfordringer som kunne ha vært unngått dersom de ble oppdaget og fulgt opp tidligere. Problemet er at det er vanskelig å forutsi når en borger vil ha sitt neste møte med offentlig sektor, og det er heller ikke kapasitet eller ressurser til å operere med slike forutsigelser for hele den norske befolkningen.

Ved hjelp av maskinlæring kan mange tjenester gå fra å være «situasjonsorienterte» og *reaktive* til å være *proaktive*. Kunstig intelligente systemer som bruker maskinlæring kan forutse potensielle hendelser eller utfordringer, snarere enn å reagere når de allerede har oppstått. Maskinlæring brukes nettopp til å lage prediksjoner. Man tar utgangspunkt i den informasjonen man har tilgjengelig (i form av data), og bruker den til å generere informasjon man *ikke* har.³⁶ Slike prediksjonssystemer kan brukes til å for eksempel foreta risikovurderinger for sykdom, vurdere hvorvidt en person vil være i stand til å komme tilbake i arbeid etter langvarig sykefravær eller til å persontilpasse undervisning, helsetjenester og andre offentlige tjenester.³⁷

En studie fra 2021, *Using Machine Learning to Create an Early Warning System for Welfare Recipients*,³⁸ viser at maskinlæring kan hjelpe med å identifisere «high-risk individuals». Studien fant at kunstig intelligente systemer kan forutsi hvem som vil få trygdeytelser i en lengre periode med (mer enn) 20 % større sikkerhet enn andre metoder. Dette kan blant annet forklares med at et kunstig intelligent system kan ta i betraktning utrolig mange forskjellige faktorer som kan påvirke sannsynligheten (rundt 1800). Kunstig intelligens kan på den måten komplementere menneskelig behandling.

I stedet for å introdusere algoritmer som *erstatning* for menneskelig behandling, bør algoritmene brukes som tillegg til menneskelig behandling. Algoritmene kan brukes til å (1) forutsi forekomsten av problemet, og (2) sortere/prioritere («triage») hvilke grupper som kan få en automatisk behandling av en algoritme, og hvilke som må behandles av mennesker. Denne bredere formuleringen av automatiseringsproblemstillingen fremhever viktigheten av å nøyaktig estimere tilbøyeligheten til både mennesker og algoritmer til å gjøre feil/treffe feil beslutninger, og å bruke slike estimater i et optimaliseringsrammeverk for å fordele innsats effektivt.³⁹

³⁶ Teknologirådet (2018) s. 8.

³⁷ Ibid. s. 9.

³⁸ Sansone (2021).

³⁹ Raghu (2019).

2.1.3 utfordringer ved bruk av kunstig intelligens i velferdsforvaltningen

2.1.3.1 Det digitale gapet

Global Risks Report fra 2021, en analytisk rapport med fokus på risiko for og konsekvenser av økende forskjeller og fragmentering av samfunnet, fremhevet «barriers to digital inclusivity» som en «kritisk risiko» for hele verden, både på kort og på lang sikt.⁴⁰ Covid-19 har skutt fart i den *fjerde industrielle revolusjonen*, og verdensbefolkningen har blitt introdusert for en ny hverdag med elektronisk handel, nettundervisning, helsehjelp på mobilen og hjemmekontor. Disse endringene har endret måten vi samhandler og kommuniserer på, og det er ingen grunn til å tro at utviklingen ikke vil fortsette selv om pandemien nå er forbi. Mange samfunn har stor nytte av disse fleksible løsningene. Samtidig er det en risiko for at eksisterende forskjeller i samfunnet blir enda mer synlige.

Selv om mange land har kommet langt i digitaliseringsprosessen, er det fremdeles mange land som ikke klarer å henge med. I Europa havner Romania nederst på listen ifølge DESI-indeksen,⁴¹ etterfulgt av Bulgaria og Hellas. Disse landene scorer under gjennomsnittet på menneskelig kapital, digitale forbindelser, integrering av digital teknologi og digitale offentlige tjenester. I Romania har kun 31 % av befolkningen mellom 16 og 74 år grunnleggende digitale ferdigheter; i Bulgaria utgjør denne andelen av befolkningen kun 29 %. I Bulgaria har bare 59 % av alle husholdninger bredbåndstilkobling. Et utdatert juridisk rammeverk står i veien for en digital offentlig forvaltning, og kun 36 % av internettbrukerne benytter seg av elektroniske forvaltningstjenester. I Hellas har utviklingen gått i riktig retning, og 51 % av befolkningen mellom 16 og 74 år har grunnleggende digitale ferdigheter. Selv om det er en høyere andel enn i både Romania og Bulgaria, mangler nesten halvparten av den greske befolkningen digital kompetanse. Danmark og Finland scorer høyest når det kommer til digitalisering, og begge landene er godt over gjennomsnittet i EU når det gjelder befolkningsandel med grunnleggende digitale ferdigheter – henholdsvis 70 og 76 %.⁴² I Norge er andelen tilsvarende 83 %.⁴³

Tallene illustrerer *det digitale gapet* i EU. Store deler av befolkningen har i realiteten ikke muligheter til å sørge for livets opphold i et digitalt marked. Sentrale utfordringer er økt digital avhengighet og automatisering, informasjonsundertrykkelse og manipulasjon samt mangelfulle rettslige reguleringer.

Digitalisering kan sies å være en forutsetning for automatisering ved hjelp av kunstig intelligens. Når store deler av befolkningen i et land mangler grunnleggende digitale ferdigheter, vil det ikke bidra til effektivisering av velferdstjenestene å implementere en teknologisk løsning som brukerne enten ikke vil ha tilgang til eller ikke vil forstå. Det vil heller ikke fungere å benytte seg av slike løsninger internt i forvaltningen; manglende kompetanse hos de ansatte vil kunne svekke rettssikkerhetsgarantiene når ingen i realiteten vet hvordan et kunstig intelligent system tar beslutninger. Selv om en stor andel av den norske befolkningen har tilstrekkelig digital kompetanse, er det viktig å være bevisst på utfordringene som den resterende delen av befolkningen opplever ved digitalisering og automatisering av velferdsforvaltningen.

⁴⁰ World Economic Forum (2021).

⁴¹ *The Digital Economy and Society Index*, se European Commission (u.å.).

⁴² Tallene er hentet fra rapportene til de enkelte landene, se European Commission (2022).

⁴³ Tall hentet fra rapporten fra 2021, se European Commission (2022).

2.1.3.2 Brukernes inntrykk av digitale tjenester og kunstig intelligens

En undersøkelse fra 2019 viser at årsaken til at enkelte brukere av offentlige tjenester i Norge velger bort digitale løsninger til fordel for analog kontakt (telefon, besøk o.l.) ikke er mangel på tilgang til internett eller digitale enheter.⁴⁴ Det skyldes heller emosjonell motstand mot å bruke digitale løsninger eller misnøye med brukervennligheten.

En utfordring som blir pekt ut er behandling av mer omfattende eller komplekse problemstillinger. Mange opplever fremdeles at digital informasjon ikke er tilstrekkelig for vurdering av konkrete saksforhold.⁴⁵ Samtidig er det ikke alltid mulig å opplyse om alle sider av saken på en digital plattform.⁴⁶ Det er ikke utenkelig at brukerne vil ha tilsvarende holdninger overfor kunstig intelligente systemer. Selv om et kunstig intelligent system klarer å ta i betraktning mange ulike faktorer, kan det komme til et annet resultat enn et menneske fordi det ikke klarer å ta rimelighetshensyn eller gjøre etiske vurderinger.

Beslutningsprosessen kan fort oppleves kynisk og svekke tilliten til forvaltningen.

Andre utfordringer ved bruk av digitale løsninger knytter seg til *kompleksiteten* av løsningene og *manglende forståelse* for disse.⁴⁷ Brukernes erfaringer og tilbakemeldinger tilsier at det er et behov for innsikt og forklarbarhet når det kommer til slike løsninger. Problemet er at kunstig intelligente systemer er enda vanskeligere å redegjøre for enn mange av de eksisterende digitale løsningene. Opplevelsen av beslutningsprosessene i offentlig sektor som en «black box» kan ytterligere bidra til å svekke tilliten blant borgerne.

Det er dokumentert skepsis til kunstig intelligens blant befolkningen. En del mener at konsekvensene av kunstig intelligens først og fremst vil bli negative.⁴⁸ De færreste ønsker å la kunstig intelligente systemer ta avgjørelser som har større betydning for dem, for eksempel avgjørelser som påvirker deres rettigheter.⁴⁹

Samtidig stiller både lovverket og brukerne krav til at forvaltningen må være *effektiv*. Det offentliges plikt til å beskytte borgerne fra potensielle uheldige virkninger av kunstig intelligens står følgelig på bekostning av fristelsen til å øke effektiviteten i forvaltningen.⁵⁰ En effektiv saksbehandling som i liten grad involverer mennesker vil kreve færre ressurser. Det kan imidlertid øke antallet feilaktige eller uhensiktsmessige avgjørelser, og føre til enda flere klagesaker og begjæringer om omgjøring. Det må tas i betraktning når man skal ta stilling til om «en enkel saksbehandling vil være mer kostnadseffektiv enn en mer omhyggelig behandling».⁵¹

2.1.3.3 Datamengde

Maskinlæringsystemer er avhengige av store mengder informasjon fra den virkelige verden i form av data. I 2020 nådde den globale produserte datamengden nye høyder som følge av en økt etterspørsel i forbindelse med pandemien. Den globale mengden data forventes å øke til over 180 zettabytes innen 2025. Det er imidlertid kun en mindre prosentandel av denne

⁴⁴ Korsgaard (2019) s. 8.

⁴⁵ Ibid. s. 11.

⁴⁶ Ibid. s. 12.

⁴⁷ Ibid. s. 25.

⁴⁸ Tennøe (2017).

⁴⁹ Fremtind (2021). Undersøkelsen gjaldt nordmenns holdninger til kunstig intelligens i bank- og forsikrings spørsmål, men er like fullt illustrerende for poenget.

⁵⁰ Kuziemski (2020).

⁵¹ NOU 2019: 5 s. 140.

datamengden som faktisk blir lagret; kun 2 % av datamengden som ble produsert i 2020 ble lagret og beholdt til 2021.⁵² Samtidig økes lagringskapasiteten, og det blir billigere å lagre data – i enkelte tilfeller billigere enn å slette data.⁵³

Den økte bevisstheten rundt betraktningen av data som en ressurs har de siste årene fått virksomheter til å organisere egne data bedre. En annen faktor som har bidratt til dette er det økte fokuset på vern av personopplysninger. Nye regelverk for behandling av slike opplysninger setter grenser for både mengden opplysninger som kan samles inn og tidsrommet for lagring av opplysningene. En utfordring ved bruken av kunstig intelligente systemer er å håndtere datamengden som er nødvendig for å utvikle og anvende systemene på en forsvarlig måte. Dette er både viktig for at virksomheten skal kunne utnytte potensialet som ligger i data, og for å overholde personvernregler og andre krav som stilles til behandling av data, for eksempel prinsippet om kun én gang. En forutsetning for å kunne føre kontroll med datamengdene er at virksomheten har den nødvendige kompetansen, og greier å sette av tilstrekkelig tid og ressurser.

2.1.3.4 Risiko for diskriminering

Bruk av kunstig intelligens kan medføre ulike former for forskjellsbehandling og diskriminering. I 2019 avdekket et forskerteam at en algoritme som ble brukt på over 200 millioner mennesker i amerikanske sykehus i stor grad favoriserte lyse pasienter over mørkhudede.⁵⁴ Algoritmen skulle forutsi hvilke pasienter som kunne ha behov for mer medisinsk hjelp. Slike og lignende prediksjonsalgoritmer brukes ofte i helsesektoren for å identifisere og hjelpe pasienter med komplekse helseutfordringer. Etnisk opprinnelse var ikke en variabel som ble brukt i denne algoritmen, men det ble brukt en annen variabel som var nært relatert, nemlig historikk over helseutgifter. Hensikten var at disse utgiftene skulle gi en oversikt over omfanget av en persons behov for helsetjenester. Algoritmen tok imidlertid ikke hensyn til at mørkhudede mennesker ikke har like god tilgang på helsetjenester som hvite mennesker, og at dette medfører at det generelt brukes mindre penger på helsetjenester til mørkhudede enn til hvite. Selv om mørkhudede i utgangspunktet hadde de samme helseutfordringene som hvite, ble de ikke prioritert av algoritmen på grunn av at helseutgiftene ikke var like høye. Forskerteamet pekte på at valget av praktiske, tilsynelatende effektive proxyer for virkelige omstendigheter kan være en viktig kilde til algoritmisk bias og diskriminering.

Et annet bemerkelsesverdig eksempel på algoritmisk diskriminering er COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)-algoritmen som ble brukt i amerikanske rettssystemer for å forutsi sannsynligheten for tilbakefall blant tiltalte.⁵⁵ På grunn av data som ble brukt, modellen som ble valgt og hele prosessen med å utvikle algoritmen, endte algoritmen med å forutsi dobbelt så mange falske positive for tilbakefall for mørkhudede tiltalte (45 %) enn for hvite tiltalte (23 %). Hvite tiltalte som begikk nye lovbrudd i løpet av de neste to årene, ble feilaktig ansett for å utgjøre en lav risiko nesten dobbelt så mange ganger (28 %) enn mørkhudede tiltalte (48 %).

Kunstig intelligente systemer kan bli diskriminerende selv om de utvikles på et tilsynelatende representativt grunnlag. Dette forklares gjerne med at systemene «overtar» menneskelige holdninger og fordommer som finnes i samfunnet, og gjør dem mer synlige. Et eksempel er

⁵² Statista Research Department (2020).

⁵³ Fertik (2015).

⁵⁴ Obermeyer (2019).

⁵⁵ Larson (2016).

Google Translate som oversetter «han er sykepleier» til «hun er sykepleier» som følge av at historiske data om kjønn og yrke er matet inn i algoritmen.⁵⁶

2.1.3.5 Regulatoriske utfordringer

De regulatoriske utfordringene ved utvikling av kunstig intelligens kan deles inn i tre kategorier. På makronivå fordrer slik teknologi etablering av nye rettigheter og plikter for enkeltindivider. Kunstig intelligens vil påvirke mennesker på en helt ny måte, som eksisterende regulatoriske instrumenter ikke har tatt høyde for. Det er viktig at menneskers grunnleggende rettigheter og plikter sikres parallelt med den raske teknologiske utviklingen. Dette pålegger lovgivere et ansvar for å vurdere både de tilsiktede og utilsiktede konsekvensene som kunstig intelligens kan ha for borgerne og samfunnet, vurdert i forhold til eksisterende verdier i forvaltningen som rettferdighet, åpenhet og ikke-diskriminering. På et mesonivå nødvendiggjør innfasingen av kunstig intelligens utvikling av dynamiske måter å vurdere kunstig intelligente systemer på. Systemene må måles, overvåkes og evalueres. Det forutsetter retningslinjer for evaluering av input, output, informasjonsforvaltning, resultater og virkninger. Dette utfordrer allerede etablerte retningslinjer i forvaltningen, som for eksempel eksisterende retningslinjer for digitalisering og risikovurderinger. På mikronivå oppstår det en ny form for spenning mellom legitimiteten av algoritmiske beslutninger og rettighetene til borgerne som blir gjenstander for slike beslutninger.⁵⁷

I tillegg kan det oppstå såkalte «agenda-flaskehals», hvor det blir lagt for mange strategier i forhold til hvor mange som faktisk kan gjennomføres i praksis. Det er ikke uvanlig at politikere som står bak slike agendaer har mål og ambisjoner som ikke er helt forenlige med realiteten for de ordinære offentlige virksomhetene og byråkratiet.

2.2 EUs forslag til forordning om harmoniserte regler for kunstig intelligens

Europaparlamentet og -rådets *forslag til forordning om fastsettelse av et harmonisert regelverk om kunstig intelligens* (KI-forordningen) er et forslag til et risikobasert regelverk for kunstig intelligens.⁵⁸ Forslaget ble lagt frem 21. april 2021, samtidig med en EU-strategi og -handlingsplan for kunstig intelligens. Det bygger på EUs koordinerte plan for kunstig intelligens fra 2018 (som nå er revidert),⁵⁹ og hvitboken om kunstig intelligens.⁶⁰ Forslaget er hjemlet i artikkel 114 i TEUV, som gjelder harmonisering av EU-medlemsstatenes lovgivning rundt det indre markedes funksjon.⁶¹

Både offentlige og private aktører, i og utenfor EU, vil omfattes av det nye regelverket, såfremt det kunstig intelligente systemet omsettes på unionsmarkedet eller anvendes slik at det påvirker mennesker som befinner seg innenfor EUs grenser. Regelverket kan gjelde for leverandører og utviklere av kunstig intelligente systemer, så vel som tilbydere⁶²

⁵⁶ Kayser-Bril (2020).

⁵⁷ Kuziemski (2020).

⁵⁸ Europalov (u.å.).

⁵⁹ COM(2018) 795 (plan fra 2018); COM(2021) 205 (revidert plan).

⁶⁰ COM(2020) 65.

⁶¹ Regjeringen (2021).

⁶² European Commission (2021).

For å sikre overholdelse av regelverket, skal hver medlemsstat utpeke én eller flere nasjonale kompetente myndigheter til å føre tilsyn med at regelverket blir implementert og overholdt. I tillegg skal medlemsstatene utpeke en nasjonal tilsynsmyndighet som det offisielle offentlige kontaktpunktet, som også vil representere landet i *European Artificial Intelligence Board (EAIB)*.

I tillegg til representanter fra de ulike medlemslandene, vil EAIB bestå av representanter fra det europeiske datatilsynet (EDPS) og EU-kommisjonen. Styret skal tilrettelegge for en harmonisert og effektiv implementering av regelverket. Styret vil gi anbefalinger og uttalelser til EU-kommisjonen vedrørende høyrisiko kunstig intelligente systemer, og om andre relevante forhold. Det vil også fungere som et kompetanse- og konsultasjonssenter for nasjonale myndigheter. På nasjonalt nivå vil EAIB være likestilt med nasjonal tilsynsmyndighet, tilsvarende tilsynsmekanismen etter personvernforordningen.

Overtredelser av regelverket skal ilegges effektive sanksjoner, herunder administrative bøter. Følgende bøtesatser foreslås i forordningsforslaget: Opptil 30 millioner euro eller 6 % av den totale årlige omsetningen i foregående regnskapsår ved forbudt praksis eller manglende overholdelse av datakrav foreslås; Opptil 20 millioner euro eller 4 % av den totale årlige omsetningen i foregående regnskapsår for manglende overholdelse av de øvrige kravene eller forpliktelsene; Opptil 10 millioner euro eller 2 % av total årlig omsetning i foregående regnskapsår ved levering av uriktig, ufullstendig eller villedende informasjon etter forespørsel fra kontrollorganer og nasjonale kompetente myndigheter.

EU-kommisjonen vil utarbeide retningslinjer for fastsettelse av administrative bøter. Dette vil sikre harmonisert praksis innenfor EU/EØS. Også EU-institusjoner og -organer vil være underlagt reglene og risikere sanksjoner ved brudd. Dette begrunnes med at disse bør gå foran som et godt eksempel for andre aktører. Det europeiske datatilsynet (EDPS) vil ha myndighet til å ilegges slike institusjoner og organer administrative bøter.⁶³

2.2.1 Den europeiske tilnærmingen til kunstig intelligens

EU-kommisjonen har siden 2014 arbeidet med å lette utviklingen av en datadrevet økonomi i Europa.⁶⁴ Dette har blant annet resultert i EUs personvernforordning (også omtalt som GDPR – the General Data Protection Regulation), som ble formelt vedtatt i EU i 2016, og gjennomført i Norge ved personopplysningsloven i 2018. EU har også vedtatt en regulering om fri flyt av andre opplysninger enn personopplysninger,⁶⁵ forordning om cybersikkerhet⁶⁶ og direktiv om åpne data og viderebruk av informasjon fra offentlig sektor.⁶⁷

I 2018 presenterte EU-kommisjonen sin første *strategi for kunstig intelligens*, og ga sin tilslutning til en koordinert plan med EUs medlemsstater for å fremme utviklingen og bruken

⁶³ Ibid.

⁶⁴ European Commission (u.å.). Kilden benytter begrepet «data-agile economy», som kan oversettes til «datasmidig økonomi». I Norge har imidlertid begrepet «datadrevet økonomi» blitt benyttet, og derfor blir begrepet også brukt her.

⁶⁵ Forordning 2018/1807/EU.

⁶⁶ Forordning 2019/881/EU.

⁶⁷ Direktiv 2019/1024/EU.

av kunstig intelligens i Europa.⁶⁸ I tillegg ble det nedsatt en ekspertgruppe for kunstig intelligens.⁶⁹

I 2019 presenterte ekspertgruppen de *etiske retningslinjene for pålitelig kunstig intelligens*. Ifølge disse retningslinjene skal pålitelig kunstig intelligens være *lovlig, etisk og robust*. I «lovlig» ligger at teknologien skal være i overensstemmelse med alle relevante lover og reguleringer. I «etisk» ligger at teknologien skal respektere etiske prinsipper og verdier. I «robust» ligger et krav til teknologisk robusthet, samt robusthet overfor teknologiens sosiale omgivelser.

Retningslinjene inneholder syv nøkkelkrav som kunstig intelligente systemer må oppfylle for å anses som pålitelige:

Menneskelig intervensjon og tilsyn: Kunstig intelligente systemer bør styrke mennesker ved å tillate dem å treffe informerte beslutninger og fremme deres grunnleggende rettigheter. Samtidig er det viktig å sikre gode tilsynsmekanismer, for eksempel ved at et menneske fører tilsyn med systemet eller har full kontroll over det.

Teknisk robusthet og sikkerhet: Kunstig intelligente systemer må være motstandsdyktige og sikre. Det må være trygt å anvende slike systemer. Det er derfor viktig å ha en plan i tilfelle noe går galt. I tillegg må systemene være nøyaktige og reproducerbare, for å minimere og forhindre utilsiktede negative virkninger av systemet.

Personvern og datastyring: Kunstig intelligente systemer må respektere personvernet og beskyttelse av data. I tillegg er det viktig å ha tilstrekkelige datastyringsmekanismer, som tar hensyn til dataens kvalitet og integritet, og sikrer at kun berettigede personer får tilgang til slik data.

Transparens: Data-, system-, og forretningsmodellene for kunstig intelligente systemer bør være transparente. Transparens kan oppnås ved at systemets funksjoner etterprøves. I tillegg bør kunstig intelligente systemer og beslutninger truffet av disse kunne forklares på en måte tilpasset den/dem som blir berørt. Mennesker må gjøres klare over at de samhandler med et kunstig intelligent system, og informeres om systemets muligheter og begrensninger.

Mangfold, ikke-diskriminering og rettferdighet: Urettferdig behandling av mennesker må unngås, da det kan ha flere negative virkninger, fra marginalisering av sårbare grupper til forverring av fordommer og diskriminering. For å fremme mangfold bør kunstig intelligente systemer være tilgjengelige for alle, uavhengig av funksjonshemninger eller andre forutsetninger, og involvere relevante interessenter gjennom hele systemets livssyklus.

Samfunnsmessig og miljømessig velvære: Kunstig intelligente systemer bør komme alle mennesker til gode, inkludert fremtidige generasjoner. Det må derfor sikres at de er bærekraftige og miljøvennlige. Systemene må ta hensyn til miljøet, inkludert andre levende vesener. Det må vurderes hvilke sosiale og samfunnsmessige virkninger kunstig intelligente systemer har på miljøet.

Ansvarlighet: Det bør iverksettes tiltak for å sikre ansvar for kunstig intelligente systemer. *Auditabilitet*, som gjør det mulig å vurdere algoritmer, data og designprosesser, spiller en viktig rolle på dette punktet. Dette gjelder særlig for

⁶⁸ European Commission (2018a).

⁶⁹ European Commission (2018b).

systemer innenfor kritiske områder. I tillegg er det viktig å sikre at eventuelle skader eller tap blir kompensert i tilstrekkelig grad.

I 2019 la også EU-kommisjonens nåværende president, Ursula von der Leyen, frem sitt politiske program for perioden 2019—2024. Programmet inneholder seks hovedprioriteringer, herunder den *digitale agendaen* for et Europa rustet for den digitale tidsalder («A Europe fit for the digital age»). Her ble kunstig intelligens fremhevet som en teknologi som i særlig stor grad endrer verden vi lever i; både når det gjelder vår hverdag, og verdensøkonomien- og samfunnet som helhet. Videre ble det satt et mål om at Europa skal være ledende i den globale digitale utviklingen. Europa skal oppnå *teknologisk suverenitet*, ved å bli mindre avhengig av kritisk teknologi og digitale tjenester fra resten av verden. Samtidig skal investeringer i europeisk kompetanse og europeiske næringer sette globale standarder for *den nye generasjonen teknologi*.

Ny lovgivning for en koordinert europeisk tilnærming til menneskelige og etiske implikasjoner av kunstig intelligens ble lovet fremmet i løpet av von der Leyens første 100 dager i embetet. Denne lovgivningen skulle også vurdere hvordan «big data» kunne brukes til å skape innovasjon og verdi for samfunnet og næringslivet.⁷⁰

I 2020 presenterte EU-kommisjonen sin strategi for EUs digitale fremtid, som en del av den digitale agendaen for 2019—2024.⁷¹ Kommisjonens tre hovedfokusområder i denne perioden vil være (1) å tilrettelegge for teknologi som fungerer *for mennesker*, (2) å skape en rettferdig og konkurransedyktig digital økonomi og (3) skape et åpent, demokratisk og bærekraftig samfunn.

I tillegg publiserte EU-kommisjonen sin *hvitbok* om kunstig intelligens, med alternativer for å realisere en pålitelig og sikker utvikling av kunstig intelligens i Europa, med respekt for EU-borgernes verdier og rettigheter. Hvitboken består av to hovedelementer: et *politisk rammeverk* for å samordne innsatsen på europeisk og nasjonalt nivå og skape et «økosystem av fortrefelighet», og *retningslinjer for fremtidige lovverk* om kunstig intelligens i Europa for å skape et «økosystem av tillit». Med hvitboken fulgte i tillegg en europeisk strategi for *data*. Den angir en rekke politiske tiltak for å nå målet om en attraktiv, sikker og dynamisk europeisk *data-økonomi*.⁷²

I 2021 presenterte EU-kommisjonen sin visjon for Europas digitale transformasjon innen 2030; *Europas digitale tiår*. I tillegg ble det lansert et såkalt «digitalt kompass», som bygger på den digitale strategien fra 2020. Kompasset har fire hovedpunkter: ferdigheter, sikker og bærekraftig digital infrastruktur, digital transformasjon av bedrifter og digitalisering av offentlig sektor. Formålet med kompasset er å gjøre EUs ambisjoner for 2030 om til *konkrete mål*, blant annet at: minst 80 % av alle voksne bør ha grunnleggende digitale ferdigheter; alle befolkede områder skal dekkes av 5G; $\frac{3}{4}$ selskaper skal bruke sky-tjenester, *big data* og kunstig intelligens i sin virksomhet; alle viktige offentlige tjenester skal være tilgjengelige online; 80 % av EUs befolkning bør bruke en elektronisk identifikasjonsløsning.

Samme år kom EU-kommisjonen med en meddelelse om å *fremme en europeisk tilnærming til kunstig intelligens*,⁷³ i tillegg til en ny koordinert plan for kunstig intelligens og et

⁷⁰ von der Leyen (2018) og Regjeringen (2019).

⁷¹ European Commission (2019).

⁷² COM(2020) 65.

⁷³ COM(2020) 65.

forordningsforslag til en europeisk harmonisering av regler om kunstig intelligens,⁷⁴ til sammen omtalt som «kunstig intelligens-pakken».

Den nye koordinerte planen er en gjennomgang av planen fra 2018, som la grunnlaget for politisk koordinering av kunstig intelligens og oppmuntret medlemslandene til å utvikle nasjonale strategier. Den teknologiske, økonomiske og politiske utviklingen nødvendiggjorde en revisjon av planen, særlig med hensyn til miljø og helse. Den nye planen beskriver politiske endringer og investeringer som medlemslandene må gjennomføre for å utvikle det som kalles en «menneskesentrert, bærekraftig, sikker, inklusiv og pålitelig kunstig intelligens». Strategien for å nå målet skal nås ved hjelp av følgende handlinger: *akselerere* investeringer i teknologier som bruker kunstig intelligens, *handle* ut fra strategier og programmer for kunstig intelligens og *tilpasse* politikken rundt kunstig intelligens.⁷⁵

Forordningsforslaget er verdens første rettslige rammeverk som regulerer bruken av kunstig intelligens. Det bygger på en risikobasert tilnærming, hvor kunstig intelligente systemer blir klassifisert i ulike kategorier på grunnlag av *hvilken risiko de utgjør* for sikkerheten, livet og rettighetene til enkeltmennesker.

2.2.2 Et sammendrag av innholdet i EUs forslag til forordning om kunstig intelligens

Forslaget skal sikre at borgerne har *tillit til kunstig intelligens*. Mange kunstig intelligente systemer som brukes i dag innebærer begrenset til ingen risiko, og bidrar til å løse ulike problemstillinger i samfunnet. Enkelte kunstig intelligente systemer kan imidlertid utgjøre en stor risiko for borgerne og samfunnet. Det er først og fremst slik risiko forslaget til forordningen tar sikte på å regulere og redusere, for å unngå uheldige og uønskede utfall.

Et spørsmål som naturlig oppstår i denne sammenhengen, er *hva som menes med risiko*. Forslaget til forordningen knytter risikobegrepet til graden av negativ innvirkning som kunstig intelligente systemer kan ha på menneskers helse, sikkerhet og grunnleggende rettigheter. For eksempel kan det være vanskelig å klarlegge *hvorfor* et kunstig intelligent system har kommet til en bestemt konklusjon, eller utført en bestemt handling. Når man ikke vet hvorfor systemet fungerer slik det gjør, kan man heller ikke være sikker på at resultatet er *riktig*. I tilfeller hvor resultatet generert ved hjelp av et kunstig intelligent system har innvirkning på menneskers helse, sikkerhet og rettigheter, vil det være særlig problematisk dersom resultatet ikke kan forklares eller begrunnes. Denne problemstillingen gjør seg kanskje spesielt gjeldende ved bruk av kunstig intelligente systemer i offentlig sektor, av myndigheter og etater som har ansvar for å forvalte borgernes rettigheter og plikter. I slike tilfeller vil risikoen ved anvendelsen av kunstig intelligente systemer være at borgerne kan miste tilgang til goder som de egentlig har rett på, eller bli pålagt plikter som de egentlig ikke skal ha.

Forslaget til forordningen søker å adressere slik risiko som oppstår ved anvendelse av kunstig intelligente systemer i samfunnet. Forslaget bygger på en risikobasert tilnærming, med inndeling i *risikokategorier*; fra minimal eller ingen risiko, til uakseptabel risiko. Jo høyere risiko et kunstig intelligent system utgjør, desto strengere reguleringer gjelder for anvendelsen av systemet. Mange av de foreslåtte reglene gjelder for *høyrisiko* kunstig intelligente systemer. Disse anses for å potensielt utgjøre en stor trussel mot enkeltmennesker og samfunnet som helhet.

⁷⁴ COM(2021) 205 og COM(2021) 206.

⁷⁵ COM(2021) 205.

Risikokategoriene deles inn i (1) uakseptabel risiko, (2) høyrisiko, (3) begrenset risiko og (4) minimal eller ingen risiko. Forslaget gir anvisning på en metode for å identifisere og definere høyrisiko kunstig intelligente systemer. Grunnlaget for risikoklassifikasjonen er systemets *tiltenkte formål* («intended purpose»), i tråd med EUs eksisterende produktsikkerhetsregelverk. Det betyr at hvilken risikogruppe et kunstig intelligent system faller inn under, avhenger av hvilken *funksjon* systemet har og det *spesifikke formålet* som systemet anvendes for.⁷⁶

2.2.2.1 Kunstig intelligente systemer med uakseptabel risiko

Kunstig intelligente systemer som utgjør en *uakseptabel risiko* for helsen, sikkerheten og rettighetene til enkeltmennesker, foreslås forbudt. Dette er systemer som anses å utgjøre en uakseptabel trussel fordi anvendelsen av dem strider mot grunnleggende rettigheter og verdier i EU, for eksempel menneskerettighetene. En liste over disse systemene finnes i forslagets artikkel 5.

Dette gjelder blant annet systemer som kan brukes til å manipulere personer utenfor deres bevissthet eller utnytte sårbare grupper (for eksempel barn eller funksjonshemmede) ved å påvirke oppførselen deres, systemer som gjør det mulig for myndighetene å drive med *social scoring*, samt biometriske identifikasjonssystemer i det offentlige rom (masseovervåkning) til bruk for rettshåndhevelse.

2.2.2.2 Høyrisiko kunstig intelligente systemer

Forslagets del III inneholder særskilte regler for kunstig intelligente systemer som klassifiseres som *høyrisiko*. Dette er systemer som har potensial til å vesentlig påvirke menneskers helse, sikkerhet eller grunnleggende rettigheter. Slike systemer vil være lovlig å bruke, men foreslås underlagt omfattende reguleringer.

Hvilke kunstig intelligente systemer som klassifiseres som høyrisiko, beror – i tråd med klassifikasjonsmetodikken i forslaget – på systemets tiltenkte formål. I tillegg er det relevant å vurdere hvor mange personer som potensielt kan bli berørt ved anvendelsen av systemet, hvilken betydning systemets utfall vil ha, hvor reversible eventuelle skader er, samt i hvilken grad eksisterende unionsregler effektivt kan forhindre eller vesentlig minske risikoene.

En spesifisering av hvilke systemer som anses for å være *høyrisiko* finnes i forslagets artikkel 6 og 7, med videre henvisninger til forslagets III. Her finnes en liste over kritiske områder hvor anvendelse av kunstig intelligente systemer innebærer en høy risiko. Når det gjelder kunstig intelligente systemer som kan brukes i velferdsforvaltningen, er det sentralt å trekke frem at følgende systemer klassifiseres som *høyrisiko*:

«5. Access to and enjoyment of essential private services and public services and benefits:

- (a) AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services;»

For at et høyrisiko kunstig intelligent system skal være lovlig å bruke, må det overholde de pålagte kravene i forslagets del III kapittel 2. Kravene skal sikre at menneskers grunnleggende rettigheter respekteres i møte med kunstig intelligente systemer, og bidra til å skape tillit til slike systemer. Kravene knytter seg til datakvalitet, teknisk dokumentasjon og loggføring, transparens, menneskelig tilsyn og robusthet, nøyaktighet og sikkerhet. Hvis kravene ikke overholdes, vil nasjonale myndigheter kunne få tilgang til informasjon som er nødvendig for å

⁷⁶ European Commission (2021).

undersøke om bruken av det kunstig intelligente systemet er i overensstemmelse med regelverket.

Listen med høyrisiko KI-systemer i III omfatter både løsninger som allerede finnes, og som er under utvikling. Det er foreslått at Kommisjonen skal kunne utvide listen innenfor bestemte områder og ved bruk av en nærmere fastsatt vurderingsmetode. Dette skal sikre at listen kan oppdateres parallelt med utviklingen av kunstig intelligens. I tillegg skal det etableres en database over høyrisikosystemer i regi av Europakommisjonen.⁷⁷

Før et høyrisiko kunstig intelligent system kan plasseres på markedet eller tas i bruk, må leverandøren foreta en samsvarsvurdering av systemet. Det innebærer at leverandøren må demonstrere at systemet overholder de pålagte kravene. Hvis systemet endres eller anvendes til et annet formål, vil man måtte utføre samsvarsvurderingen på nytt.

I tillegg må leverandører av høyrisiko kunstig intelligente systemer implementere kvalitets- og risikostyring for å sikre overholdelse av regelverkets krav og minimere risikoen for brukere og berørte personer, selv etter at systemet har blitt plassert på markedet. Tilbydere av slike systemer skal også ha mulighet til å rapportere om alvorlige hendelser eller brudd på grunnleggende rettighetsforpliktelser som de får kjennskap til.

Som nevnt, er det åpent for at EU-kommisjonen kan endre listen over høyrisiko kunstig intelligente systemer. Definisjonen er følgelig verken endelig eller uttømmende. Høyrisiko-klassifikasjonen kan synes å være en åpen standard innenfor forordningsforslagets rammer, som kan endre seg i tråd med endringer i samfunnet og teknologien, og som må tolkes av domstolene og EU-domstolen. En ulempe med en slik dynamisk standard er at det kan oppstå uenigheter rundt hvilke systemer som til enhver tid skal klassifiseres som høyrisiko.

2.2.2.3 Kunstig intelligente systemer med begrenset risiko

Kunstig intelligente systemer som utgjør en *begrenset risiko*, er underlagt visse *transparensforpliktelser*. Dette er typisk systemer som brukes til å samhandle med andre mennesker, for eksempel *chatbots*. Slike systemer må være innrettet på en slik måte at personen blir informert om at vedkommende samhandler med et kunstig intelligent system, og ikke et annet menneske. Vedkommende skal være i stand til å treffe en informert beslutning om å fortsette eller avbryte samhandlingen med systemet.

Denne såkalte *informasjonsplikten* gjelder ikke hvis det klart fremgår av sammenhengen rundt bruken at man samhandler med et kunstig intelligent system. Plikten gjelder heller ikke for lovregulerte kunstig intelligente systemer som har til formål å avsløre, lokalisere, identifisere eller reise tiltale etter en straffbar handling, hvis ikke det gjelder systemer som er offentlig tilgjengelige og etablert for å rapportere straffbare handlinger.⁷⁸

2.2.2.4 Kunstig intelligente systemer med minimal eller ingen risiko

De fleste kunstig intelligente systemer som er i bruk i dag utgjør minimal eller ingen risiko for enkeltmennesker eller samfunnet som helhet. Dette er systemer som ikke omfattes av noen av de tre øvrige kategoriene. Disse systemene er ikke underlagt noen særskilte forpliktelser.

Leverandører og brukere av slike systemer kan selv velge om de ønsker å følge retningslinjene for pålitelig kunstig intelligens, og eventuelt egne etiske retningslinjer.

⁷⁷ Regjeringen (2021).

⁷⁸ Regjeringen (2021).

2.2.2.5 Øvrige reguleringer

Før et høyrisiko kunstig intelligent system kan plasseres på markedet, må følgende trinn være utført:

1. Systemet må være utviklet fortrinnsvis ved hjelp av interne ex ante konsekvensutredning for kunstig intelligens («Artificial Intelligence Impact Assessments») og etiske retningslinjer («Codes of Conduct»), overvåket av tverrfaglige team.
2. Systemet må gjennomgå en godkjent samsvarsvurdering («conformity assessment») og kontinuerlig overholde kravene til kunstig intelligente systemer som følger av forordningsforslaget, i løpet av systemets livssyklus. En slik dynamisk prosess sikrer tilsyn og validering. Hvis det blir gjort endringer i systemet, må dette trinnet gjentas.
3. Systemet må registreres i en dedikert EU-database.
4. En samsvarserklæring («declaration of conformity») må være signert, og systemet må være CE-merket («Conformité Européenne mark»).

I tillegg pålegger forordningsforslaget både EU- og nasjonale myndigheter, tilbydere og sluttbrukere ulike plikter, som bidrar til kontinuerlig overvåking av systemet fra ulike «nivåer».

Mennesker har rett til å bli informert om at de interagerer med et kunstig intelligent system og ikke et annet menneske. Forordningsforslaget introduserer derfor spesifikke transparensforpliktelser for både brukere og tilbydere av kunstig intelligente systemer.

Høyrisiko kunstig intelligente systemer skal overholde EU-standarder, inkludert sikkerhets- og samsvarsvurderinger. Slike vurderinger stiller krav om at produkter og tjenester som omsettes innenfor EU oppfyller de tekniske, juridiske og etiske standardene som gjenspeiler EUs kjerneverdier. Hvis et kunstig intelligent system anses kvalifisert i henhold til slike standarder, vil det bli CE-merket og følgelig få tillatelse til å plasseres på EU-markedet. CE-merket fungerer som en sikkerhetssertifisering.⁷⁹

Det kan være verdt å merke seg at disse vurderingene også må foretas for trenings-, testing- og valideringsdata. Slik data må være representativ, av høy kvalitet, riktig merket og feilfri for å unngå bias og sikre ikke-diskriminerende resultater. Følgelig vil også input-data være underlagt de høye standardene for pålitelig kunstig intelligens.⁸⁰

Forordningsforslaget vil ha et *ekstraterritorielt omfang*, ved at den også vil forplikte virksomheter *utenfor* EU. Forslaget er følgelig enda et eksempel på fenomenet *Brüssel-effekten*, som refererer til EUs unilaterale makt til å regulere globale markeder. Vilkåret er at systemet har virkning innenfor EU, uavhengig av hvor leverandøren eller brukeren er etablert.

Det er den fysiske eller juridiske personen som tilgjengeliggjør det kunstig intelligente systemet på EU-markedet som er pålagt ansvar i form av å være tilbyder etter forordningsforslaget. Dette gjelder uavhengig av om denne fysiske eller juridiske personen selv har designet eller utviklet systemet, og uavhengig av om vedkommende er etablert innenfor eller utenfor EU. Leverandører utenfor EU vil også være underlagt

⁷⁹ COM(2021) 206 art. 46.

⁸⁰ Kop (2021).

forordningsforslagets krav hvis det endelige resultatet som det kunstig intelligente systemet produserer, anvendes i EU.⁸¹

2.3 Utviklingen på området

2.3.1 I Norge

2.3.1.1 Utvikling

Norge er blant de ledende landene når det gjelder digitalisering i Europa, ifølge DESI-indeksen,⁸² som måler digital modenhet og utvikling i EU/EØS-land. I analysen fra 2021 ble Norge rangert som nummer fem, etter Danmark, Finland, Sverige og Nederland.⁸³ Resultatene er særlig gode når det gjelder digitalisering av offentlige tjenester.⁸⁴

Den digitale utviklingen i Norge begynte for alvor i 2011. EU hadde nylig utarbeidet sin 2020-strategi,⁸⁵ og utpekt informasjons- og kommunikasjonsteknologi (IKT) som ett av de viktigste virkemidlene for å løfte Europa ut av den økonomiske krisen og legge grunnlaget for bærekraftig økonomisk vekst.⁸⁶ Selv om Norges økonomiske situasjon var noe annerledes enn den i EU-landene, var IKT ikke et mindre relevant tema innenfor våre grenser. Den økende bevisstheten rundt den økonomiske og demokratiske verdien av slik teknologi begynte å gi uttrykk for et behov for en mer helhetlig og strategisk nasjonal tilnærming. I tillegg ble også norske markeder og utvikling påvirket av europeiske prioriteringer og valg.

For å følge opp EUs strategi, meddelte det daværende Fornyings-, administrasjons- og kirkedepartementet at det, i samarbeid med andre departementer, ville utarbeide en nasjonal IKT-strategi eller en «digital agenda» for Norge. Denne ville reflektere den europeiske digitale agendaen.⁸⁷ Allerede da lå Norge godt an til å hente ut gevinsten fra IKT i form av produktivitetsvekst og verdiskaping, både i næringslivet og i offentlig sektor.

Den digitale agendaen for Norge ble presentert i 2012.⁸⁸ Agendaen hadde et langsiktig perspektiv, frem mot 2020. Formålet var å legge frem regjeringens politikk for hvordan Norge som samfunn skulle utnytte mulighetene for verdiskaping og innovasjon ved hjelp av IKT og internett. Den teknologiske revolusjonen over de foregående 20 årene hadde lagt grunnlag for den nye *digitale revolusjonen*. IKT var ikke lenger en støttefunksjon, men en *kjernefunksjon*. Virksomheter som tidligere leverte varer og tjenester ved hjelp av fysiske eller analoge løsninger, begynte å gå over til heldigitale systemer. Privat sektor var tidlig ute med denne omleggingen. Tjenester i offentlig sektor var ikke i like stor grad påvirket av digitaliseringen, men også offentlige aktører måtte belage seg på å gjennomgå tilsvarende endringer for å

⁸¹ Ipek (2022); COM(2021) 206 art. 2.

⁸² *The Digital Economy and Society Index*.

⁸³ European Commission (2021).

⁸⁴ Osloregionens Europakontor (2021).

⁸⁵ COM(2010) 2020.

⁸⁶ Prop. 1 S (2011–2012).

⁸⁷ Ibid.

⁸⁸ Meld. St. 23 (2012–2013).

tilpasse seg nye nettbaserte tjenester og forretningsmodeller. Spørsmålet var ikke *om* samfunnet ville digitaliseres, men *når*.⁸⁹

Norge hadde en rekke fordeler i møte med den digitale revolusjonen, herunder en kompetent og omstillingsvillig offentlig sektor og innbyggere med høy grad av tillit til myndighetene. Det ble også fremhevet at den norske befolkningen var opptatt av ny teknologi.⁹⁰ Utviklingen og implementeringen av digitale tjenester i samfunnet ville skje mer eller mindre uavhengig av hva staten la til rette for. En aktiv politikk for å ta steget fra et analogt til et digitalt samfunn ville imidlertid legge til rette for en hensiktsmessig digitalisering av samfunnet. Et sentralt mål var at Norges befolkning skulle være på nett, ha digital kompetanse og bruke digitale tjenester på en trygg og sikker måte. Et annet sentralt mål var at offentlig sektor skulle etterspørre digitale tjenester og løsningen. Godt personvern, høy sikkerhet og tillit til IKT-systemer skulle skape grunnlaget for IKT-basert verdiskaping.⁹¹

Det ble trukket frem flere grunnleggende prinsipper i den digitale agendaen fra 2012 for hvordan staten skulle opptre på enkeltområder, for IKT-politikken og for valg mellom ulike virkemidler: staten skulle i sin politikk legge til rette for ny virksomhet, og ikke virke hemmende for ny innovasjon i eksisterende virksomheter; reguleringer skulle være teknologinøytrale; åpent tilgjengelige, internasjonale standarder skulle være utgangspunktet; staten skulle være varsom i sin regulering på IKT-området for å ikke hindre utvikling og innovasjon; reguleringen skulle være effektiv, og ikke gå lenger enn nødvendig; offentlig sektor skulle gå foran som et eksempel og bidra til å realisere gevinstene ved nye teknologiske løsninger; og staten skulle arbeide for et åpent og ikke-diskriminerende internett.⁹²

I 2014 trådte eForvaltningsforskriften og endringer i forvaltningsloven i kraft, og digital kommunikasjon ble hovedregelen innenfor forvaltningslovens virkeområde.⁹³ Flere digitaliseringsprosjekter ble gjennomført i løpet av de neste årene, blant annet A-ordningen og digital postkasse.⁹⁴

I 2015 presenterte regjeringen en ny digital agenda for Norge,⁹⁵ hvor hovedmålsettingene var (1) en brukerrettet og effektiv offentlig forvaltning og (2) verdiskaping og deltakelse for alle. En hovedprioritering var å skape sammenhengende og helhetlige offentlige tjenester, hvor forvaltningen kunne gjenbruke informasjon i stedet for å spørre brukerne om informasjonen på nytt. Videre skulle Norge være en del av det digitale indre marked i Europa, og det skulle legges til rette for samvirke med europeiske løsninger for effektiv digitalisering av offentlig sektor. Til slutt skulle personvern og informasjonssikkerhet være en integrert del av utviklingen og bruken av IKT; behandling av personopplysninger skulle baseres på gode forholdsmessighetsvurderinger med utgangspunkt i behandlingsformålet, og informasjonssikkerhet skulle ivaretas ved en risikobasert tilnærming med utgangspunkt i trussel- og sårbarhetsvurderinger.⁹⁶

⁸⁹ Ibid.

⁹⁰ Ibid.

⁹¹ Ibid.

⁹² Ibid.

⁹³ Sanner (2014).

⁹⁴ Kommunal- og distriktsdepartementet (2014).

⁹⁵ Meld. St. 27 (2015–2016).

⁹⁶ Ibid.

I 2019 gikk statlig og kommunal sektor sammen om en digitaliseringsstrategi for offentlig sektor. Strategien følger opp den digitale agendaen fra 2015, og definerer felles mål og innsatsområder for digitaliseringsarbeidet frem mot 2025.⁹⁷ For å realisere strategien, er det utarbeidet en *handlingsplan med innsatsområder*, hvor hvert innsatsområde inneholder flere *initiativer*.⁹⁸ Det er totalt åtte innsatsområder: sammenhengende tjenester med brukeren i sentrum; økt deling av data og verdiskaping; klart og digitaliseringsvennlig regelverk; felles økosystem for nasjonal digital samhandling; styring og samordning for mer sammenhengende offentlig sektor; styrket samarbeid med privat sektor; økt digital kompetanse i offentlig sektor; og digital sikkerhet.

Brukeren i sentrum: Offentlige tjenester skal oppleves sammenhengende og helhetlige av brukerne, uavhengig av hvilke virksomheter som tilbyr dem. Dette krever samarbeid på tvers av forvaltningsnivåer og sektorer. Et sentralt virkemiddel ved arbeidet på dette innsatsområdet er inndelingen av borgernes viktigste møter med offentlig sektor i løpet av livet i *livshendelser*.⁹⁹ Det er etablert en behovskatalog basert på behovene i de ulike livshendelsene, med formål om å realisere disse. Det arbeides med å utvikle felles prinsipper for god brukskvalitet. I tillegg er det laget en konseptskisse for en *digital assistent*, som blant annet skal bidra til enklere tilgang til egne data og persontilpasset informasjon.¹⁰⁰

Økt deling av data: Forvaltningen skal dele data når den kan, og skjerme data når den må. Åpne data skal gjøres tilgjengelig for viderebruk – til utvikling av nye tjenester og verdiskaping. I tillegg skal brukere av offentlige tjenester slippe å oppgi informasjon som allerede er innhentet flere ganger. Det har blitt opprettet et nasjonalt ressurscenter for deling av data, en nasjonal verktøykasse for deling av data og en datafabrikk.¹⁰¹

Digitaliseringsvennlig regelverk: Hindringer i regelverket som bremser digitaliseringen skal fjernes. Hindringene skyldes oftest kompleksitet, lav kompetanse og usikkerhet rundt regelverket. Nasjonalt ressurscenter for deling av data tar imot innspill rundt regelverksutfordringer. Det har også blitt publisert en veileder for digitaliseringsvennlig regelverk og klart lovspråk.¹⁰²

Felles digitalt økosystem: Eksisterende plattformer for nasjonal samhandling og utvikling av tjenester skal utnyttes bedre i et felles økosystem. Det er blant annet signert en samarbeidsavtale om digitalisering for å koordinere statlige og kommunale interesser.¹⁰³

Styring og samordning: Gevinstene ved digitalisering skal uthentes på en mer systematisk måte. Det er også viktig å styrke samordningen på tvers av sektorer, og mellom statlig og kommunal sektor. I 2020 fikk Skate-organet (organet for styring og

⁹⁷ Kommunal- og distriktsdepartementet (2019).

⁹⁸ Digitaliseringsdirektoratet (u.å.a).

⁹⁹ Digitaliseringsdirektoratet (u.å.b).

¹⁰⁰ Digitaliseringsdirektoratet (u.å.c).

¹⁰¹ Digitaliseringsdirektoratet (u.å.d).

¹⁰² Digitaliseringsdirektoratet (u.å.e).

¹⁰³ Digitaliseringsdirektoratet (u.å.f).

koordinering av tjenester for e-forvaltning) et nytt mandat, og skal ha en aktiv rolle i utviklingen, gjennomføringen og oppfølgingen av digitaliseringsstrategien.¹⁰⁴

Samarbeid med privat sektor: Et digitalt samarbeid med næringslivet og frivillige organisasjoner kan gi grunnlag for nye innovative tjenester. Programmet StartOff, som skal koble oppdragsgivere med gründervirksomheter, ble lansert i 2021, og initierte 12 prosjekter det første året.¹⁰⁵

Økt digital kompetanse: Digital kompetanse er en kritisk faktor for offentlige virksomheters evne til digital transformasjon. Det er planer om å utarbeide en strategi for digital kompetanse i offentlig sektor.¹⁰⁶

Digital sikkerhet: De digitale løsningene som tilbys av det offentlige skal være trygge og sikre. Forvaltningen skal bli tilstrekkelig moden når det gjelder informasjonssikkerhet, og bli bedre rustet til å følge endringene i trusselbildet.¹⁰⁷

I 2020 la regjeringen frem den nasjonale strategien for kunstig intelligens.¹⁰⁸ Strategien var forventet i lys av utviklingen på området de siste årene, utviklingen i EU og den nasjonale satsingen på digitalisering. Her beskrives blant annet hva som ligger i begrepet «kunstig intelligens», hvor kunstig intelligens kan anvendes, og hvordan kunstig intelligens kan anvendes for å bidra til de ulike nasjonale målsettingene. Strategien er ikke begrenset til noen tidsperiode. Dette begrunnes med at kunstig intelligens er et område i stadig utvikling, og det kan derfor være behov for å revurdere og revidere strategien i takt med teknologi- og samfunnsutviklingen.

Den digitale transformasjonen som Norge har gjennomgått det seneste tiåret har ytterligere bekreftet at vi har gode forutsetninger for å lykkes med ny teknologi, inkludert kunstig intelligens. Offentlig sektor har evnet å tilpasse seg nye løsninger, og har kapasitet og kompetanse til å utvikle seg videre. I tillegg har vi en god infrastruktur og mye data i offentlig sektor. Sammen ligger forholdene til rette for at forvaltningen kan implementere og anvende kunstig intelligens for å løse viktige oppgaver på en bedre og mer effektiv måte.

Samtidig innebærer utvikling av ny teknologi at det oppstår nye problemstillinger. Det kan være at eksisterende normer og regler utfordrer utviklingen. For eksempel er det norske samfunnet bygget på tillit og respekt for grunnleggende verdier som menneskerettigheter og personvern – som kan legge begrensninger på hvor langt vi kan gå ved utviklingen og bruken av kunstig intelligente systemer. Det kan også være at problemstillingene som oppstår er såpass nye at vi enda ikke har de nødvendige midlene for å løse dem. Å anvende kunstig intelligens i praksis har ikke vært et aktuelt tema før nylig. Derfor har vi heller ikke regler som regulerer slik teknologi. Den nasjonale strategien for kunstig intelligens legger føringer for utvikling og bruk av kunstig intelligens i Norge, som skal bidra til at våre grunnleggende verdier og rettigheter blir respektert i møte med den nye teknologien. Strategien fremhever også viktige prinsipper som kan gi et utgangspunkt ved vurderingen av nye problemstillinger som oppstår ved utvikling og bruk av kunstig intelligens.

¹⁰⁴ Digitaliseringsdirektoratet (u.å.g).

¹⁰⁵ Digitaliseringsdirektoratet (u.å.h).

¹⁰⁶ Digitaliseringsdirektoratet (u.å.i).

¹⁰⁷ Digitaliseringsdirektoratet (u.å.j).

¹⁰⁸ Kommunal- og distriktsdepartementet (2020).

De nasjonale føringene:¹⁰⁹

- kunstig intelligens som utvikles og brukes i Norge skal bygge på etiske prinsipper, og respektere menneskerettighetene og demokratiet
- forskning, utvikling og bruk av kunstig intelligens i Norge skal bidra til ansvarlig og pålitelig kunstig intelligens
- utvikling og bruk av kunstig intelligens i Norge skal ivareta den enkeltes integritet og personvern
- digital sikkerhet skal bygges inn i utvikling, drift og forvaltning av løsninger som bruker kunstig intelligens
- tilsynsmyndigheter skal føre kontroll med at systemer basert på kunstig intelligens på sitt tilsynsområde opererer innenfor prinsippene for ansvarlig og pålitelig bruk av kunstig intelligens

Norge deltar i programmet for et digitalt Europa (DIGITAL) i perioden 2021–2027.¹¹⁰ Dette er et investerings- og kapasitetsbyggingsprogram for digital omstilling og bruk av innovativ digital teknologi i samfunnet og næringslivet. Ett av satsingsområdene i programmet er data og kunstig intelligens, som må ses i sammenheng med EUs datastrategi. Målet er blant annet å tilrettelegge for deling av data, samt fremme testing og implementering av kunstig intelligente systemer i samfunnet.¹¹¹

Norge har hatt en aktiv rolle i EUs aktiviteter rundt kunstig intelligens, og var blant annet med på utarbeidelsen av EUs koordinerte plan for kunstig intelligens i 2018. De nåværende etiske prinsippene for utvikling og bruk av kunstig intelligens i den nasjonale strategien er basert på EUs ekspertgruppes prinsipper, som ble presentert i 2019.¹¹²

2.3.1.2 Norges posisjon til forslaget

Norge har gitt uttrykk for en positiv holdning til forordningsforslaget. Det er gitt støtte til den brede definisjonen av kunstig intelligens, på bakgrunn av den norske praksisen med teknologinøytral lovgivning. Det gis også støtte til den risikobaserte tilnærmingen til reguleringen.

Det er foreslått mindre endringer i formuleringene rundt de forbudte kunstig intelligente systemene. Når det gjelder høyrisiko kunstig intelligente systemer, er det gitt tilbakemeldinger på flere punkter. Det gjelder blant annet et ønske om konkrete retningslinjer for vurderingen av hvorvidt et system klassifiseres som høyrisiko eller ikke og utvikling av harmoniserte standarder for krav til høyrisiko kunstig intelligente systemer. Det ønskes også ytterligere presisering av kvalitetskravene til datasett, og strengere krav til transparens ved bruk av kunstig intelligente systemer i forbindelse med rettsanvendelse.

Det er også spørsmål om forordningsforslagets forhold til annen lovgivning, herunder personvernforordningen, forbrukerretten og arbeidsretten.

¹⁰⁹ Ibid.

¹¹⁰ Bjornsson (u.å.).

¹¹¹ Digitaliseringsdirektoratet (u.å.k).

¹¹² Osloregionens Europakontor (2021).

2.3.2 I øvrige land

Vest-europeiske land har topplasseringene i rangeringen *Oxford Insights Government AI Readiness Index 2021*; 10/20 av topp 20 land, og 6 av topp 10 rangerte land. Storbritannia er høyest rangert – med en score på 81,25 – etterfulgt av Finland, Nederland, Sverige, Frankrike, og Danmark. Norge havner like under Danmark, med en score på 76,14.

De høye scorene skyldes i stor grad den politiske viljen blant myndighetene i disse landene til å utnytte fordelene ved kunstig intelligens, sterke nasjonale økonomier som er innrettet for digitale forretningsmodeller og befolkninger med avanserte digitale ferdigheter. De fleste landene på listen har enten allerede etablert sin nasjonale strategi for kunstig intelligens, eller er i ferd med å utvikle en. Nasjonale strategier er viktige virkemidler for å kartlegge muligheter og utfordringer som myndighetene står overfor som følge av utviklingen av kunstig intelligens. Slike strategier kan også gi myndighetene en enhetlig plan, slik at de kan levere gode løsninger på tvers av sektorer.¹¹³

I FN-målingen «E-Government Survey» fra 2020 ble Danmark ansett for å være ledende innenfor offentlig digitalisering.¹¹⁴ Danmark fikk sin nasjonale strategi for kunstig intelligens i 2019. Strategien har et sterkt fokus på en etisk tilnærming til teknologien.¹¹⁵ Strategien legger opp til at bruken av kunstig intelligens skal være til fordel for både borgere, virksomheter og offentlige tjenester. Samtidig slår den danske regjeringen fast at det fremdeles må være mennesker som skal treffe viktige beslutninger – ved hjelp av kunstig intelligens, og ikke omvendt. Strategien inneholder 24 initiativer som skal styrke utviklingen og anvendelsen av kunstig intelligens i Danmark.

Ett av initiativene retter seg mot *sikkerhetsaspektet* ved kunstig intelligens. Målet er å kartlegge sikkerhetsmessige risikoer som kan oppstå for myndigheter og virksomheter ved økt bruk av kunstig intelligente systemer. Det er utarbeidet en veiledning med konkrete tiltak som skal lette arbeidet med IT-sikkerhet og databeskyttelse i både offentlig og privat sektor, blant annet konkrete eksempler på kontraktsvilkår for sikkerhet som myndighetene kan stille til leverandører av kunstig intelligens.¹¹⁶

Et annet initiativ hadde som mål å skape juridisk klarhet rundt utviklingen og anvendelsen av kunstig intelligens. En interdepartemental/tverrdepartemental arbeidsgruppe fikk i oppgave å undersøke hvorvidt problemstillinger ved anvendelsen av kunstig intelligens kunne håndteres innenfor de daværende lovgivningsmessige rammene. Arbeidsgruppen skulle avdekke eventuelle behov for reguleringer rundt kunstig intelligens, på nasjonalt nivå og på EU-nivå. Arbeidsgruppens arbeid skulle også danne grunnlag for dialog med EU-kommisjonen om slike reguleringer.¹¹⁷

Ett av initiativene skal sikre transparens ved bruk av kunstig intelligens i offentlig sektor. Som en del av dette arbeidet ble det iverksatt et pilotprosjekt for utvikling og testing av strategier for hvordan offentlige myndigheter kunne sikre lovgivningens krav om rimelig, ansvarlig og transparent bruk av kunstig intelligens som grunnlag for beslutningstaking.¹¹⁸

¹¹³ Oxford Insights (2022).

¹¹⁴ Digitaliseringsstyrelsen (2020).

¹¹⁵ Regjeringen (2019).

¹¹⁶ Ibid. s. 29–30.

¹¹⁷ Ibid. s. 30.

¹¹⁸ L.c.

Det siste initiativet som er interessant å nevne her er initiativet til *signaturprosjekter*. Den danske regjeringen lanserte et investeringsfond for kommuner og regioner, med formål om å øke bruken av kunstig intelligens og utbredelsen av digitale velferdsløsninger. Det var imidlertid mangel på konkrete erfaringer med bruk av kunstig intelligens i velferdsforvaltningen. For å raskest mulig få slike erfaringer, bestemte regjeringen å gjennomføre en rekke signaturprosjekter. Signaturprosjektene skulle teste teknologien på områder i samfunnet hvor det var potensial for å øke kvaliteten og produktiviteten i offentlige kjerneoppgaver. Erfaringene skulle bidra til å utvikle løsninger som kunne brukes i hele landet. Signaturprosjektene skulle ta utgangspunkt i en offentlig driftsoppgave, hvor det i samarbeid mellom offentlig og privat sektor skulle utvikles en løsning ved hjelp av kunstig intelligens. Regjeringen valgte i første omgang følgende områder: helsesektoren, arbeids- og velferdsforvaltningen og «tverrgående saksbehandling». Felles for de tre områdene var et stort volum av oppgaver, og en mulighet for kunstig intelligens til å komme til nytte for et stort antall borgere.

Initiativet startet med 15 ulike signaturprosjekter.¹¹⁹ I 2021 ble det iverksatt nye 13 signaturprosjekter, blant annet for raskere saksbehandling for bevilgning av «kroppsbårne hjelpemidler»,¹²⁰ og digital inkludering og support ved talegjenkjenning.¹²¹ I 2022 har det blitt iverksatt ytterligere 12 signaturprosjekter.¹²² Det signaliserer at signaturprosjektene virker etter sin hensikt, og at det er et behov for å prøve teknologien i praksis.

Erfaringene fra signaturprosjektene har vist til konkrete utfordringer ved bruk av kunstig intelligens i offentlig sektor:¹²³

- Varierende datakvalitet, datamengde og mangel på felles datastandarder
- Juridiske utfordringer, blant annet når det gjelder personvernreglene
- Utfordringer med IT-infrastruktur, blant annet begrenset teknisk adgang til data og datadeling mellom ulike systemer
- Mangel på juridisk kompetanse og kunnskap om kunstig intelligens
- Usikkerhet rundt etiske aspekter ved kunstig intelligens, herunder ansvarlighet og transparens

2.4 EØS-relevans

Utvalget er bedt om å vurdere om den eventuelle EU forordningen om kunstig intelligens er EØS-relevant, herunder om regelen faller innenfor EØS-avtalens saklige virkeområde, og om Norge bør innlemme reglene i EØS-avtalen. EU-relevans er en forutsetning for at nye EU-rettsakter vurderes innlemmet i EØS-avtalen via ene.¹²⁴

Det finnes to klare indikasjoner på EØS-relevans. For det første vil den være relevant dersom rettsakten endrer eller opphever en rettsakt som allerede er innlemmet i EØS-avtalen.¹²⁵ I

¹¹⁹ Digitaliseringsstyrelsen (2019).

¹²⁰ Signaturprosjekt (2021a).

¹²¹ Signaturprosjekt (2021b).

¹²² Digitaliseringsstyrelsen (2021) og Digitaliseringsstyrelsen (2022).

¹²³ Digitaliseringsstyrelsen (2021).

¹²⁴ Baur (2016) s. 53.

¹²⁵ L.c.

forslaget til forordning om kunstig intelligens fastslår artikkel 75-82 endringer i flere slike rettsakter, herunder Skipsutstyrdirektivet 2014¹²⁶ og forordningen for sikkerhet i sivil luftfart.¹²⁷ Likevel kan neppe disse mindre endringene regnes for å klart fastsette EØS-relevans. Det indikerer likevel at deler av forslaget kan berøre EØS-relevante områder.

For det andre vil EØS-relevante rettsakter normalt være merket «with relevance to the EEA».¹²⁸ Dette er ikke tilfellet for foreliggende forslag, men som nevnt både av Juridisk fakultet ved Universitet i Oslo og Digdir i deres respektive innspill til det norske posisjonsnotatet¹²⁹ er det likevel sannsynlig at det endelige forordningen vil være EØS-relevant.

Vurdering som derfor må foretas er om forordningen «faller innenfor EØS-avtalens saklige og geografiske virkeområde, slik dette er definert i EØS-avtalens hoveddel, vedlegg og protokoller».¹³⁰ Vurderingen er til en viss grad skjønnsmessig idet de objektive og rettslige kriteriene som avtalen oppstiller ikke er helt presise. Denne skjønnsmessigheten gjør at vurderingen i realiteten i visse tilfeller også kan ha visse politisk sider. Disse vil vi i liten grad drøfte. Vurderingen av EØS-relevans vil etter dette bero på en analyse av rettsakten i lys av formålene til EØS og midlene til å oppnå disse formålene.¹³¹

En EØS-rettsakt kan også bare delvis være EØS-relevant. Da vil bare de EØS-relevante delene bli innlemmet i EØS-avtalen.¹³²

2.4.1 EØS-avtalens saklige virkeområde

EØS-avtalen artikkel 1 fastslår avtalens saklige virkeområde – å opprette et enhetlig europeisk økonomisk samarbeidsområde. Midlene for å oppnå dette formålet er etter annet ledd fri bevegelse av varer, personer, tjenester og kapital (bokstav a-d), opprettelse av et system som sikrer at konkurransen ikke vriss og at reglene overholdes likt internt i samarbeidet (bokstav e), samt nærmere samarbeid på områder som forskning og utvikling, miljø, utdannelse og sosialpolitikk (bokstav f).

Relevansbegrepet kan i denne sammenheng tolkes både snevert og vidt.¹³³ Relevans i snever forstand viser til rettsakter som angår de fire friheter og konkurransereglene etter Del II til V av EØS-avtalen.¹³⁴ De følger av punkt 2.1 i begrunnelsen til forslaget («Explanatory Memorandum») at rettsgrunnlaget er TEUV artikkel 114 som gjelder bestemmelser om vedtakelsen av virkemidler som skal sikre opprettelsen av og funksjonen til det indre markedet. Fortalen til forslaget viser videre i punkt (1) til at formålet er å forbedre det indre markedets funksjon ved å etablere homogene rettslig rammer særlig for utvikling, markedsføring og anvendelse av kunstig intelligens som er i overensstemmelse med Unionens

¹²⁶ Direktiv 2014/90/EU.

¹²⁷ Forordning (EF) nr. 300/2008.

¹²⁸ Baur (2016) s. 54.

¹²⁹ Henholdsvis UiO ved Senter for rettsinformatikk (2021) og Digdir [2021?].

¹³⁰ Meld. St. 5 (2012–2013) pkt. 5.3.1 jf. utredningsinstruksen pkt. 5.5 og DFØ *Veileder til utredningsinstruksen*.

¹³¹ Baur (2016) s. 54.

¹³² Jf. Meld. St. 5 (2012–2013) pkt. 5.3.1.

¹³³ L.c.

¹³⁴ Jf. Meld. St. 5 (2012–2013) pkt. 5.3.1 og Almestad (2016) s. 89.

verdier. Det er klart at hovedvekten av forslaget etter dette vil være innenfor EØS-avtalens saklige virkeområde.

Likevel, som påpekt av UiO, inneholder forslaget bestemmelser som regulerer problemstillinger som ikke nødvendigvis hører under EØS-avtalens saklige virkeområde¹³⁵, særlig dersom man vurderer relevans i snever forstand. Enkelte av bestemmelsene kan tenkes å være omfattet av avtalens saklige virkeområde tolket i vid forstand. Ved en slik tolkning omfattes også tiltak som faller utenfor de fire friheter, på de områdene som er nedfelt i artikkel 78 i EØS-avtalens del VI.¹³⁶ Særlig relevant for foreliggende forslag er punktene «forskning og teknologisk utvikling» og «informasjonstjenester», men også «forbrukervern» og «små og mellomstore bedrifter» kan ha betydning.

Dersom det etter denne vurderingen fortsatt skulle gjenstå bestemmelser som faller utenfor EØS-avtalens saklige virkeområde, vil disse ikke bli innlemmet i avtalen.

2.4.2 EØS-avtalens geografiske virkeområde

Det geografiske virkeområdet til EØS-avtalen er fastsatt i art. 126 og omfatter «Kongeriket Norges territorium». Begrepet territorium tolkes i norsk rett i henhold til fast praksis i folkeretten slik at avtalen omfatter Norges landterritorium, indre farvann og territorialfarvann, men ikke økonomiskone, kontinentalsokkel eller det åpne hav.¹³⁷ Svalbard er i henhold til protokoll 40 også unntatt fra virkeområdet. Siden forslaget til forordning søker å regulere det indre markedet for kunstig intelligent systemer jf. artikkel 2, er det klart at den vil falle innenfor EØS-avtalens geografiske virkeområde.

2.4.3 Bør reglene innlemmes?

Spørsmålet om Norge bør innlemme forordningen i EØS-avtalen er fra et juridisk standpunkt til en viss grad besvart positivt i det ovenstående. EU-rettsakter som er EØS-relevante skal som hovedregel innlemmes.

Samtidig har spørsmålet også en politisk side. Innlemmelse av et forslag som direkte berører det indre markedet slik dette gjør, vil vise Norges støtte til EØS-samarbeidet. Det vil også være i tråd med den politiske linjen som Norge har holdt på digitaliseringsområdet. Som tidligere nevnt trakk regjeringen fram i 2012 at utgangspunktet for IKT-politikken skulle være internasjonale standarder.¹³⁸ I 2015 ble det videre lagt vekt på Norges deltakelse i det digitale indre marked i Europa og viktigheten av tilrettelegging for samarbeid med europeiske løsninger for effektiv digitalisering.¹³⁹ Norges aktive rolle i EUs aktiviteter rundt kunstig intelligens¹⁴⁰ og pågående deltakelse i *The Digital Europe Programme (DIGITAL)*¹⁴¹ taler for

¹³⁵ UiO nevner for eksempel reglene om bruk av systemer til biometrisk fjernidentifikasjon i realtid på offentlige steder med henblikk på rettshåndhevelse jf. artikkel 5(2) og kravene om høyrisiko-KI-systemer for rettshåndheving og migrasjonsstyring, asylforvaltning og grensekontroll jf. artikkel 6 og annekset III. Se UiO ved Senter for rettsinformatikk [2021].

¹³⁶ Jf. Meld. St. 5 (2012–2013) pkt. 5.3.1.

¹³⁷ L.c.

¹³⁸ Se utredningens punkt 2.3 med henvisning til Meld. St. 23 (2012–2013).

¹³⁹ L.c.

¹⁴⁰ Se utredningens punkt 2.3.

¹⁴¹ The European Commission (u.å.). For mer informasjon om Norges deltakelse se utredningens punkt 2.3 med videre henvisninger.

en sterk politisk vilje til å støtte EUs utvikling på området, inkludert innlemmelse av en eventuell forordning om kunstig intelligens.

Likevel kan intensjon alene ikke tale for innlemmelse. Forordningsteksten og dens egnethet til å oppnå de tiltenkte målene og beskyttelse av verdiene som Norge og EU står for, er også av betydning for spørsmålet.

Europaparlamentets forskningstjeneste¹⁴², heretter EPRS eller forskningstjenesten, publiserte i januar 2022 en briefing¹⁴³ om forslaget til forordning om kunstig intelligens¹⁴⁴. Om tilbakemeldingene på forslaget oppsummerer EPRS at interessenter og eksperter generelt støtter forslaget, men det rettes samtidig kritikk mot visse deler av forslaget med krav om endringer.¹⁴⁵ Tilsvarende holdning følger av Norges posisjonsnotat til forslaget.¹⁴⁶

Kritikken mot forslaget har for det første vært rettet mot definisjonen av kunstig intelligens i artikkel 3 nr. 1. Felles for kritikken er frykten for juridisk usikkerhet, men der noen kritikere mener at den uklare ordlyden er for snever og derfor åpner for anvendelsen av systemer som potensielt kan gjøre skade på de verdiene som forordningen skal beskytte, mener andre at ordlyden er for vid og dermed vil føre til overregulering.¹⁴⁷ Ordlyden har i tillegg blitt kritisert for å ikke være tilstrekkelig teknologinøytral til å kunne sikre mot fremtidig utvikling på området.¹⁴⁸

Det tyske forbundsrådet skriver i sin beslutning at en vid definisjon i prinsippet virker hensiktsmessig idet noen enhetlig definisjon enda ikke finnes. Samtidig har den også en svakhet ved at nesten alle IT-programmer vil falle innenfor definisjonen, som følge av at de tar i bruk enkle sorterings- og søkealgoritmer som definert i vedlegg I til forslaget.¹⁴⁹ Til og med beregning av gjennomsnitt eller standardavvik samt konvensjonelle metoder for offisiell statistikk kan omfattes av definisjonen, og forbundsrådet ber derfor blant annet om at definisjonen begrenses til *selvlærende systemer* («aus sich selbst heraus lernende Systeme»)¹⁵⁰.

¹⁴² The European Parliamentary Research Service (EPRS) er Europaparlamentets interne forskningstjeneste og tenke tank. Organets formål er å gi Parlamentet og dets utvalg uavhengig, objektiv og autoritativ forskning og analytisk støtte. For mer informasjon se blant annet <https://epthinktank.eu/about/> og <https://www.europarl.europa.eu/at-your-service/da-stay-informed/research-and-analysis>.

¹⁴³ EPRS-briefinger tar sikte på å gi medlemmer av Europaparlamentet en systematisk analyse av alle vesentlige forslag til EU-lovgivning på alle stadier av lovgivningsprosessen. Briefingene inneholder en redegjørelse for formålet, innholdet og juridiske aspektene ved den foreslåtte lovgivningen, analyserer endringer som lovgivningen vil medføre, samt eventuell tidligere lovgivning og bakgrunnen. I tillegg gis det en oversikt over interessentenes synspunkter, samt meninger fra nasjonale parlamenter og de to rådgivende komiteene. For mer informasjon se: <https://epthinktank.eu/eu-legislation-in-progress/>.

¹⁴⁴ European Parliamentary Research Service (2022).

¹⁴⁵ Ibid s. 1 og s. 9.

¹⁴⁶ Kommunal- og moderniseringsdepartementet (2021).

¹⁴⁷ European Parliamentary Research Service (2022) s. 8-9.

¹⁴⁸ Ibid s. 9.

¹⁴⁹ Bundesrat Deutschland (2021) avsnitt 19.

¹⁵⁰ L.c.

Norge har i sitt posisjonsnotat ikke reist tilsvarende bekymring til definisjonen. I stedet uttrykkets det en støtte til bruken av en vid definisjon og muligheten for å tilpasse den gjennom endringer av vedlegg I med henvisning til den norske lovgivningstradisjonen.¹⁵¹

For det andre er det rettet kritikk mot den risikobaserte tilnærmingen til forslaget. De fleste interessentene støtter tilnærmingen¹⁵², herunder Norge¹⁵³ og Tyskland¹⁵⁴, men kritiserer av ulike grunner gjennomføringen. Enkelte advarer om at utformingen av forslaget ikke sikrer et tilstrekkelig høyt beskyttelsesnivå for de grunnleggende rettighetene til borgerne.¹⁵⁵ Det vises blant annet til at individer som blir berørt av AI-systemer og rettighetsorganisasjoner ikke vil ha noen klagerett, mulighet for rettshåndhevelse via domstolene eller oppreisningsmulighet.¹⁵⁶ EU oppfordres til å undersøke om eksisterende oppreisningsmekanismer i EU/EØS er tilstrekkelig tilpasset de utfordringene som KI representerer særlig overfor prinsippene transparens, sikkerhet og ansvarlighet¹⁵⁷, og eventuelt å sørge for eksplisitt å opprette slike mekanismer i forordningen.¹⁵⁸

I tillegg rettes særlig bekymring mot implementeringen av artikkel 5 (forbudt praksis) og artikkel 6 (regulerte høy-risiko praksiser) og at risikovurderingen i stor grad overlates til tilbyderens¹⁵⁹ av systemet fremfor av en uavhengig enhet før systemet distribueres.¹⁶⁰ Det tyske Forbundsrådet bemerker at forordningen i liten grad inneholder eksplisitte materielle krav til beskyttelse av grunnleggende rettigheter. Dette vurderes som bekymringsverdig fordi det sannsynliggjør at prosedyrene for kontroll av samsvar med regelverket, samt det offisielle tilsynet, vil være begrenset til en gjennomgang av tiltakene for kvalitetssikring etter forslaget kapittel 2.¹⁶¹ Norge uttrykker tilsvarende bekymring særlig med henvisning til vurderingen av høyrisiko KI-systemer etter artikkel 6 og anser det som potensielt nødvendig å opprette en prosedyre for å avgjøre grensetilfeller.¹⁶² Når de gjelder forbudet i artikkel 5 nr. 1 reiser både Norge og Tyskland bekymring over at det kun får anvendelse på faktisk og potensiell fysisk eller psykisk skade.¹⁶³ Tilsvarende manipulerende praksiser anvendes gjerne også til å utøve finansiell skade som indirekte kan ha psykiske konsekvenser.

En fjerde bekymring som reises av flere interessenter, er de potensielle virkningene av forslaget på investeringer og for små og mellomstore bedrifter (SMB). Interessenter viser til at kostnadene ved etterlevelse av forordningen og frykten for store bøter kan virke avskrekkende

¹⁵¹ Kommunal- og moderniseringsdepartementet (2021) pkt. 2.

¹⁵² European Parliamentary Research Service (2022) s. 8-9.

¹⁵³ Kommunal- og moderniseringsdepartementet (2021) pkt. 3.

¹⁵⁴ Bundesrat Deutschland (2021) avsnitt 3 og 13.

¹⁵⁵ European Parliamentary Research Service (2022) s. 9.

¹⁵⁶ Ibid s. 10.

¹⁵⁷ Kommunal- og moderniseringsdepartementet (2021) pkt. 6.

¹⁵⁸ European Parliamentary Research Service (2022) s. 10.

¹⁵⁹ Som definert i COM(2021) 206 art. 3 nr. 2.

¹⁶⁰ European Parliamentary Research Service (2022) s. 9-10.

¹⁶¹ Bundesrat Deutschland (2021) avsnitt 23.

¹⁶² Kommunal- og moderniseringsdepartementet (2021) pkt. 5.

¹⁶³ Bundesrat Deutschland (2021) avsnitt 20 og Kommunal- og moderniseringsdepartementet (2021) pkt. 4.

på SMBer fra utvikling av AI-systemer og dermed føre til å redusere investeringer i kunstig intelligens og kvele innovasjon på området.¹⁶⁴

De ovenstående eksemplene er ikke uttømmende, og det er viktig at Norge tar i betraktning alle de potensielle ulempene ved utformingen av forordningen. Siden forslaget ikke enda ikke er ferdig utformet og vedtatt, er det for tidlig å si noe om hvor tungtveiende betydning de nevnte utfordringene vil få, og om de i det hele tatt vil være relevante etter gjennomføring av endringer før eventuell vedtakelse av forordningen. Spørsmålet om hvorvidt Norge bør innlemme KI-forordningen kan etter dette vanskelig besvares på dette tidspunktet og en ny vurdering må foretas når det endelige forslaget er klart. Det er likevel et tungtveiende argument for innlemmelse at forslaget er det første av sitt slag i verden. Å vise sin støtte til et så ambisiøst forsøk på å regulere kunstig intelligens, kan være av stor politisk betydning selv om forslaget inneholder enkelte svakheter.

2.5 Behovet for lovendringer som følge av bruk av kunstig intelligente systemer i forvaltningen

2.5.1 En ny lov om kunstig intelligens

Utvalget mener som nevnt i punkt 2.4 at EUs forordning om harmonisering av kunstig intelligens høyst sannsynlig er EØS-relevant, og at det uansett er god grunn til å innlemme forordningen i EØS-avtalen. Vi ser det som hensiktsmessig med en egen lov om kunstig intelligens, som gjennomfører forordningen ved en inkorporasjonsbestemmelse i lovens § 1, sml. personopplysningslovens inkorporasjon av personvernforordningen i lovens § 1.

For en hensiktsmessig gjennomføring av forordningen, må all relevant eksisterende norsk lovgivning bringes i samsvar med de nye reglene. Av kapasitetshensyn vil vi ikke utforme en KI-lov. Utvalget vil i stedet fokusere på hva implementeringen av forordningen har å si på det forvaltningsrettslige området.

2.5.2 Endringer i forvaltningsloven

Utvalget mener EU-forordningen fordrer at det gjøres endringer i forvaltningsloven. Som vi allerede har vært inne på, ser vi en økt digitalisering og automatisering av forvaltningen, og det er et uttalt politisk mål at offentlig forvaltning skal ha en høy grad av automatisering.¹⁶⁵ Forvaltningsloven er nødt til å tilpasses de krav EU stiller til bruk av systemer som anvender kunstig intelligens, for at forvaltningsorganene skal kunne benytte seg av kunstig intelligente systemer i tråd våre internasjonale forpliktelser.

Forvaltningslovutvalget foreslår i NOU 2019: 5 tre paragrafer som gjelder digital forvaltning direkte. Det gis en hjemmel for helautomatiserte avgjørelser (§ 11), krav til offentlig dokumentasjon av det rettslige innholdet i automatiserte saksbehandlingssystemer (§ 12), og en hjemmel for nærmere regler om digital kommunikasjon mellom forvaltningsorganer og den enkelte (§ 13). Etter vår oppfatning er det imidlertid et behov for ytterligere regler om utvikling og bruk av kunstig intelligens i forvaltningen.

¹⁶⁴ European Parliamentary Research Service (2022) s. 9 og Kommunal- og moderniseringsdepartementet (2021) pkt. 7.

¹⁶⁵ Se blant annet Meld. St. 27 (2015–2016), Kommunal- og moderniseringsdepartementet (2019) og Kommunal- og moderniseringsdepartementet (2020).

Utgangspunktet for reguleringen bør etter utvalgets oppfatning være den rettslige statusen til utvikling og bruk av kunstig intelligente systemer. Forvaltningsloven knytter en rekke grunnleggende forvaltningsrettslige prinsipper og regler til beslutningstypene enkeltvedtak og forskrift, særlig til enkeltvedtak. Det vi har valgt å kalle «rettslige systemavgjørelser», vil etter dagens forvaltningslov falle mellom to stoler, idet slike avgjørelser innehar kjennetegn for både forskrift og enkeltvedtak.

Som nevnt innledningsvis vil vi med «rettslige systemavgjørelser» sikte til avgjørelser med rettslig innhold som treffes som ledd i utviklingen av kunstig intelligente systemer. En slik avgjørelse vil det blant annet være behov for der regelen som skal programmeres er uklar. For å kunne programmere en uklar regel, må det foretas en rettslig vurdering av hvilket tolkningsalternativ programkoden skal legge til grunn. Rettslige systemavgjørelser vil også være nødvendige der det for å gi en sammenhengende rettslig avklaring av et rettsspørsmål, er et behov for å utfylle eksisterende lovgivning. Automatisert rettsanvendelse fordrer et detaljert og sammenhengende regelverk.

I det følgende vil vi forsøke å anvende forvaltningslovens definisjoner på rettslige systemavgjørelser, og se nærmere på hvorfor dagens definisjoner ikke egner seg på denne beslutningstypen.

2.5.2.1 Forvaltningslovens definisjoner anvendt på rettslige systemavgjørelser

2.5.2.1.1 Vil rettslige systemavgjørelser være vedtak i forvaltningslovens forstand?

Spørsmålet er om utvikling og anvendelse av kunstig intelligente systemer til bruk i automatisert saksbehandling vil kunne karakteriseres som vedtak etter dagens forvaltningslov. Forvaltningsloven definerer et vedtak som en «avgjørelse som treffes under utøving av offentlig myndighet», og som er «bestemmende for rettigheter og plikter til private personer», jf. fvl. § 2 bokstav a.

Med «avgjørelse» siktes det til at det må foretas en avgrensning mot rådgivning, veiledning og informasjonsgivning. Ordlyden «bestemmende for rettigheter og plikter» tilsier at avgjørelsen må endre eller fastslå ens rettsstilling. Det er lagt til grunn i praksis at «rettigheter» ikke skal forstås i snever forstand som noe man har rettskrav på å få dersom bestemte vilkår er oppfylt, men at det også omfatter goder man får utdelt etter forvaltningens mer eller mindre frie skjønn. Videre må avgjørelsen gjelde «private personer». Betegnelsen må forstås vidt. Hensikten er i første rekke å trekke grensen mot forvaltningens interne avgjørelser.

Med vilkåret om at avgjørelsen må være truffet «under utøvelse av offentlig myndighet» tilsier ordlyden at det må være en avgjørelse særpreget for det offentlige. Dette vil typisk være der myndighetene utøver sin heteronome kompetanse til å nedlegge forbud og påbud, men også der den tildeler goder, som tjenester og tillatelser. Privatrettslige disposisjoner myndighetene foretar faller imidlertid utenfor.

Spørsmålet er om rettslige beslutninger som tas under utviklingen av kunstig intelligente systemer vil være en avgjørelse truffet med offentlig myndighetskompetanse som endrer en rettslig tilstand for privatpersoner.

I vår sammenheng er det på det rene at saksbehandling er en type virksomhet som særpreger det offentlige, slik at det vil snakk om offentlig myndighetsutøvelse. Det kan for eksempel være snakk om et system som skal være bestemmende for tildeling av goder og tjenester til befolkningen. Behandlingen av privatpersoners enkeltsaker gjennom den automatiserte saksbehandlingen vil også rette seg mot privatpersoner. Det er videre på det rene at beslutninger med rettslig innhold vil karakteriseres som en «avgjørelse». Spørsmålet er om

rettslige beslutninger som tas som ledd i utviklingen av kunstig intelligente systemer er «bestemmende for rettigheter eller plikter». Det avgjørende her vil være om programmeringen endrer en rettslig tilstand.

Hvorvidt programmeringen endrer en rettslig tilstand eller ei, vil naturlig nok avhenge av den konkrete rettsregelen som kodes. Dersom rettsregelen er klar og programmeringsvennlig – med andre ord der det er snakk om en «skal»-bestemmelse som det ikke er tvilsomt hvordan skal tolkes – vil ikke programmeringen medføre noen endring i rettstilstanden. Her vil det simpelthen være snakk om å en anvendelse av lovteksten, og ikke en «avgjørelse» av rettigheter og plikter.

Er det derimot snakk om en rettsregel som fordrer at det underveis i programmeringen tas valg om hvordan en rettsregel skal forstås, vil selve programmeringen endre en rettslig tilstand. Det vil her være en forskjell mellom de rettslige normene som fremgår av de tradisjonelle rettskildene, og de som fremgår av programmet. Programinstruksjonen vil da «innebære en rettslig presisering i forhold til det øvrige regelverk, eller representere regler som ikke direkte er basert på eksisterende skrevne regler».¹⁶⁶

Dette er i tråd med den generelle oppfatning i juridisk teori. Bing skriver at der det foreligger «avvik i datamaskinprogrammet, vil programmet være med på å bestemme rettigheter eller plikter til et ubestemt antall personer, og er dermed en forskrift i lovens forstand».¹⁶⁷ Han presiserer at et slikt avvik for eksempel kan foreligge «hvis man i programmet definerer vage uttrykk, omformulerer en vurdering til en kompleks, fast regel osv.». Frihagen hevder et dataprogram vil kunne være en forskrift, og med det også et vedtak, «der programmet snevrer inn de hensyn og kriterier regelverket ellers stiller opp».¹⁶⁸ En forutsetning er imidlertid at «programinstruksjonen er ment å være en bindende avgjørelse, og at den har vært gjenstand for rettslige vurderinger».¹⁶⁹

Konklusjonen er etter dette at en programkode som på grunnlag av rettslige vurderinger utgjør en endring av rettstilstanden sammenlignet med det øvrige regelverket, og som er ment å være rettslig bindende overfor befolkningen, må anses å være et vedtak. Det er grunn til å merke seg at de fleste lovbestemmelser, også de som tilsynelatende er enkelt utformet, som hovedregel vil måtte underlegges en tolkning for at vi skal komme frem til rettsregelen.¹⁷⁰ Utviklingen av systemer og programkode vil med det som regel være et vedtak i lovens forstand. Programkoding som har en normerende funksjon vil etter dette oppfylle vilkårene, mens der dette ikke er tilfellet, vil programmeringen være ren rettsanvendelse.

Det er interessant å merke seg at spørsmålet ikke har vært debattert på flere tiår.¹⁷¹ Mye har skjedd de siste tiårene innen programmering, særlig gjennom fremskritt innen maskinlæring. Spørsmålet er hvilken betydning det vil ha for vurderingen at algoritmen ikke kun er et kodet beslutningstre etter en regeldrevet modell, men en datadrevet modell som gjennom datamateriale har «lært» hvordan konkrete avveininger skal foretas.

Utgangspunktet for vurderingen vil måtte være det samme. Det springende punkt vil være om programvaren innebærer en endring sammenlignet med den øvrige rettstilstanden. Dersom

¹⁶⁶ Schartum (1989) s. 675.

¹⁶⁷ Bing (1977) s. 222.

¹⁶⁸ Frihagen (1986 II) s. 840.

¹⁶⁹ Schartum (1989) s. 675.

¹⁷⁰ Eng (2020). Forelesning rettsfilosofi.

¹⁷¹ Schartum (2013) s. 259.

algoritmen baserer seg på en modell matet med et aktuelt datagrunnlag, vil algoritmen gjenspeile den rettsstilstanden som dette datagrunnlaget baserer seg på. Algoritmen vil med det ikke utgjøre et «avvik» fra den øvrige rettsstilstanden, og følgelig ikke være en «avgjørelse» som er «bestemmende» for privatpersoners rettsstilstand etter fvl. § 2 bokstav a. En slik algoritme vil etter dette ikke være et vedtak. Samtidig kan man si at det foreligger en «avgjørelse» i bestemmelsen av hvilke data algoritmen skal mates med, og at dette i seg selv vil være et vedtak.

Dersom rettsstilstanden skulle endre seg som følge av en endring i formell lov, vil algoritmen måtte endres. Bildet kan imidlertid bli noe mer komplisert ved endringer som følge av andre forhold enn formell lov, for eksempel gjennom rettspraksis, endringer i annen lov, endringer i faktiske forhold som får betydning for tolkning o.l. Algoritmen vil i seg selv være konservativ. Den vil ikke «finne opp noe nytt» helt på egenhånd. Dersom algoritmen kun mates med behandlinger av saker den selv har foretatt, vil den ikke bringe noe nytt på banen. Dersom algoritmen også mates med dommer, kjennelser og andre beslutninger tatt av mennesker eller andre maskiner, vil eventuelle nyskapninger i rettsstilstanden ikke være en følge av algoritmen selv, men en «avgjørelse» tatt på annet hold. Igjen vil imidlertid hvilket datamateriale du gir til algoritmen være gjenstand for en rettslig vurdering.

Vi ser her at drøftelsene fort blir kompliserte og lite fruktbare. Det blir tydelig at forvaltningslovens vedtaksdefinisjon ikke er utformet med tanke på beslutninger tatt som ledd i utvikling av kunstig intelligente systemer til bruk i automatisert saksbehandling. Dette taler for at slike beslutninger ses som en egen beslutningstype etter loven.

Det samme er tilfellet ved vurderingen av hvorvidt *anvendelse* av kunstig intelligente systemer på et konkret faktum faller innunder vedtaksdefinisjonen. Ved kjøring av en algoritme på et konkret faktum vil resultatet genereres automatisk av maskinen. Det blir kunstig å legge til grunn at systemet «fatter avgjørelser». Samtidig vil en slik systembasert masseforvaltning utvilsomt være bestemmende for privatpersoners rettigheter og plikter, slik at behovet for forvaltningslovens rettssikkerhetsgarantier gjør seg like gjeldende som ved tradisjonell saksbehandling.

Bing påpekte allerede på 70-tallet at et dataprogram «fører til en viss utvisking av grensen mellom normgivning og rettsanvendelse», og at «vekselvirkning mellom norm og praksis» lett synes å gå tapt.¹⁷² Bing fremholdt at når «normen detaljbeskrives i datamaskinprogrammet, har normgiveren i minste detalj foreskrevet løsningen av alle enkeltvedtak». Dersom utfallet allerede er foreskrevet, vil det ikke foreligge noen «avgjørelse» ved bruken av programmet. Saken vil i realiteten være forhåndsavgjort. Ved bruk av en maskinlæringsalgoritme derimot, vil det være vanskelig å foreskrive resultatet på samme måte, fordi man ikke vet hvilke faktorer systemet kommer til å vektlegge.

Oppsummeringsvis vil de rettslige avgjørelsene som tas som ledd i utviklingen av kunstig intelligente systemer som den klare hovedregel falle innunder lovens vedtaksdefinisjon, såfremt det ikke er snakk om programmering av en lovregel som ikke behøver å underlegges tolkning. Bruk av systemet på et konkret faktum vil imidlertid falle utenfor lovens vedtaksdefinisjon. Dette vil etter alle solemerker også gjelde for et system basert på maskinlæring, for selv om ikke resultatet her er foreskrevet på samme måte som detaljbeskrevne programmer, oppleves det kunstig å legge til grunn at maskinen «fatter avgjørelser». Det er programkoden og eventuelt datagrunnlaget som eventuelt må angripes og underlegges rettslig kontroll.

¹⁷² Bing (1977) s. 222-223.

Som lagt til grunn i EU-forordningen, må høyrisiko-KI-systemer være gjenstand for menneskelig tilsyn («human oversight»). Dersom et menneske foretar en rettslig vurdering av hvorvidt algoritmens resultat er rettslig holdbart i hver enkeltavgjørelse, vil denne vurderingen i seg selv kunne utgjøre en «avgjørelse». Det er imidlertid grunn til å tenke seg at slikt tilsyn vil skje på et mer overordnet og generelt nivå. Dette vil vi gjøre nærmere rede for i del 5.4.

Etter utvalgets oppfatning taler den kompliserte og sprikende rettslige statusen til utviklingen og bruk av kunstig intelligente systemer, for at slike avgjørelser lovfestes som en egen beslutningstype.

2.5.2.1.2 Forskrift eller enkeltvedtak?

Et annet viktig poeng i denne sammenheng er at «rettslige systemavgjørelser» er vanskelig å kategorisere som enten forskrift eller enkeltvedtak.

Etter dagens lovverk kan et vedtak være enten en forskrift eller et enkeltvedtak. For å avgjøre hva slags vedtak det er snakk om, må det foretas en vurdering av hvem vedtaket retter seg mot. En forskrift er et vedtak som retter seg mot «et ubestemt antall eller en ubestemt krets av personer», jf. fvl. § 2 bokstav c. Forskriften skiller seg med det fra et enkeltvedtak, som kjennetegnes ved at vedtaket retter seg mot «en eller flere bestemte personer», jf. bokstav b.

En programkode vil kunne anvendes overfor et ubestemt antall personer, og ikke kun overfor bestemte. Dette tilsier at rettslige beslutninger i forbindelse med programmeringen faller innunder forskriftsdefinisjonen. Samtidig vil programkoden anvendes i enkeltsaker, ved at de kunstig intelligente systemene behandler konkrete saker, og genererer et resultat.

Beslutningene som tar som ledd i utviklingen av systemene, vil med det også være avgjørende i saker med direkte betydning for enkeltpersoner. Fra denne synsvinkelen minner de rettslige beslutningene mer om enkeltvedtak.

Schartum skriver: «Selv om programkoden inneholder implisitte enkeltvedtak i enhver sak, får ikke dette effekt i enkeltsaker før systemet faktisk er anvendt. Dette tidsaspektet gjør at parallellen til enkeltvedtakene ikke er helt treffende».¹⁷³ Det fremstår konstruert å tale om at man i utviklingen av en programkode treffer enkeltvedtak, når man på dette stadiet i prosessen ikke forholder seg til enkeltpersoner, men et ubestemt antall personer.

Det er mye som taler for at de rettssikkerhetsgarantier som parten tilkjennes ved enkeltvedtak, også må tilkjennes den som får sin sak avgjort av et kunstig intelligent system. Dersom borgeren ønsker å klage på avgjørelsen, og det er på det rene at systemet har anvendt et korrekt, konkret faktum, vil borgeren måtte angripe systemet og de avgjørelsene som er fattet i utviklingen av systemet. I vår sammenheng, med NAV som oppdragsgiver, er det grunn til å merke seg at det helt på det rene at saker om tilskudd og støtte tradisjonelt regnes som enkeltvedtak.¹⁷⁴ Dette er typiske «masseforvaltningssaker» det vil kunne være mulig å underlegge automatisert rettsanvendelse.

Hvilke regler som bør gjelde for klage og omgjøring ved automatisert rettsanvendelse vil vi komme nærmere tilbake til i et senere kapittel. Her er det tilstrekkelig å merke seg at det vil utgjøre en stor fare for borgernes rettssikkerhet dersom man av de nevnte grunner ikke skulle anse saker avgjort av kunstig intelligente systemer som enkeltvedtak. Det ville ikke vært en god løsning om slike rettssikkerhetsgarantier forsvant med bruken av algoritmer.

¹⁷³ Schartum (2013) s. 260.

¹⁷⁴ Ot.prp. nr. 3 (1976–77) s. 56.

Det er heller ikke treffende å tale om rettslige systemavgjørelser som instruksjer eller veiledninger. Instruksjer er interne beslutninger som fastlegger handlingsrommet innenfor lovene, ved å gi bindende direktiver til underordnede forvaltningsorganer om saksbehandling, lovtolkning og skjønnsutøving. Veiledninger er orienteringer og faglige råd om fremgangsmåter og handlemåter, som ikke er rettslige bindende.¹⁷⁵

Rettslige systemavgjørelser vil ikke rette seg mot et underordnet forvaltningsorgan, men fungere som algoritmiske instruksjoner til det kunstig intelligente systemet på et konkret nivå. Selv om man kan tale om at man «instruerer» systemet, vil ikke disse instruksjonene gi et handlingsrom, men ha direkte betydning. Systemet er heller ikke et underordnet forvaltningsorgan. Hovedargumentet imot å se rettslige systemavgjørelser som instruksjer er imidlertid at beslutningstypen bør underlegges viktige rettssikkerhetsgarantier. Det vil heller ikke være snakk om en «veiledning», da beslutningen vil ha en konkret og avgjørende betydning for borgerens rettstilstand.

2.5.2.2 Hvorfor egen beslutningstype og nytt kapittel?

Utvalget mener at det bør innføres en ny beslutningstype i forvaltningsloven, og at det bør lovfestes et nytt kapittel med særlige reguleringer tilknyttet denne beslutningstypen. En grunn i favør av en slik løsning er at grunnlaget for rettslige systemavgjørelser, som gjennomgått, ikke passer særlig godt med forvaltningslovens definisjoner av vedtak, forskrift og enkeltvedtak. Utvalget holder med Schartum i at ingen av forvaltningslovens saksbehandlingsregler kommer direkte til anvendelse på slike avgjørelser, men fordi «de er ledd i aktiviteter som ofte innebærer offentlig myndighetsutøvelse, har likevel generelle forvaltningsrettslige prinsipper klar relevans».¹⁷⁶ Dette tilsier at det bør formuleres regler spesielt tilpasset rettslige systemavgjørelser.

Bruk av kunstig intelligente systemer reiser videre en rekke nye problemstillinger, som det er behov for at lovgiver tar stilling til. Det er snakk om å gi regler for beslutninger som vil ha stor innvirkning på borgernes rettsstilling, og som kan gripe inn i deres grunnleggende menneskerettigheter. Som vi vil gjennomgå om litt, gjør en rekke grunnleggende prinsipper seg gjeldende ved bruk av kunstig intelligente systemer – slik som prinsipper om transparens, personvern, cybersikkerhet og dataforvaltning. Regulering vil være nødvendig for å søke å hindre at det skal oppstå rettslige feil og ufullstendigheter i systemet. Kun slik vil en overgang til automatisert rettsanvendelse ivareta borgerens grunnleggende menneskerettigheter og rettssikkerhet.

Schartum peker i sin høringsuttalelse til NOU 2019:5 på fem punkter som kjennetegner den nye digitale forvaltningen:¹⁷⁷

- høy grad av automatisering, dvs. massive og detaljerte forhåndsavgjørelser av rettsspørsmål, uttrykt som algoritmer som kan programmeres og kjøres i en maskin;
- økt betydning av «rettslig systemutvikling», dvs. av den beslutningsprosessen som leder frem til de algoritmene som er nevnt i forrige kulepunkt;
- saksbehandling uten saksbehandler, dvs. enkeltvedtak som ikke er basert på noen saksbehandleres forståelse av beslutningsgrunnlaget og tilknyttede rettsspørsmål;
- stor grad av integrering mellom informasjonssystemer innen offentlig forvaltning;

¹⁷⁵ NOU 2019: 5 s. 632.

¹⁷⁶ Schartum (2013) s. 260.

¹⁷⁷ Schartum (2020).

- stor grad av integrering mellom informasjonssystemer i offentlig forvaltning og privat sektor.

Ved å la rettslige systemavgjørelser utgjøre en egen beslutningstype, vil vi kunne gi særskilte lovregler for denne beslutningstypen. Vi ser det som hensiktsmessig med et eget kapittel som på en oversiktlig måte samler alle reglene for denne beslutningstypen, samt andre tilhørende regler, i stedet for å implementere reglene spredd utover i forvaltningslovens ulike deler. Dette er også i tråd med forvaltningens system av i dag, som samler alle reglene for forskrifter i kapittel VII om forskrifter, og lar kapittel IV, V og VI gjelde for enkeltvedtak.

3 Overordnet om dagens rettskildebilde på forvaltningsrettens område

Utvalget vil i det følgende gjøre rede for dagens rettskildebilde, relevant i tilknytning til velferdsstatlig digitalisering og bruk av kunstig intelligens. Dette innebærer en redegjørelse av eksisterende lovgivning på området, og grunnleggende prinsipper og verdier det er avgjørende å verne om ved utarbeidelsen av regler om bruk av kunstig intelligente systemer. Betydningen av EUs forslag til forordning om kunstig intelligens vil drøftes opp mot de regler, prinsipper og verdier vi her legger til grunn.

3.1 Relevant lovgivning

3.1.1 Overordnede rettsnormer

Med «overordnede rettsnormer» sikter vi til rettsregler og rettslige prinsipper med forrang for annen lovgivning. Utvikling og bruk av kunstig intelligente systemer bringer en rekke overordnede rettsnormer på banen, der retten til privatliv og forbudet mot diskriminering står særlig sentralt. Disse grunnleggende rettighetene og pliktene skal være hensyntatt av kommisjonen i utarbeidelsen av forslaget til forordning, som oppgir at et grunnleggende mål med forordningen er å sørge for at «AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values».¹⁷⁸

Risikoer og muligheter ved kunstig intelligens til bruk i forvaltningen knyttet til diskriminering og manglende personvern sto sentralt for både JOU 2020: 5 *Kunstig intelligens – Muligheter og risikoer i velferdsforvaltningen* og JOU 2020: 2 *Automatisk kontroll i Arbeids- og velferdsetaten – Grunnleggende rettigheter ved automatisk kontroll med borgernes etterlevelse av velferdslovgivningen*.

Fokuset i vår JOU er rettet mot EUs forslag til forordning om kunstig intelligens, og hvorvidt nasjonal lovgivning må endres i møte med denne. De overordnede rettsnormene utgjør et viktig bakteppe for forståelsen av EU-forordningen, i tillegg til at de etter vår rettsorden skal utgjøre selve utgangspunktet for lovgivningsprosessen. I det følgende vil vi redegjøre for hvorfor dette er tilfellet.

3.1.1.1 Grunnloven

I vår rettsorden er Grunnloven av 1814 den høyeste rettskilden, og dens stilling som «lex superior» innebærer at Grunnloven i konflikttilfeller går foran alminnelig lov. Grunnloven utgjør det rettslige fundamentet for statsdannelsen, i tillegg til å sette rammer for

¹⁷⁸ COM(2021) 206 punkt 1.1 I «Explanatory Memorandum».

lovgivningsmyndigheten og de andre statsmaktene.¹⁷⁹ Den lovgivende makt Stortinget er gitt etter Grl. § 49, jf. § 75 bokstav a, må utføres innenfor de skranker Grunnloven oppstiller. Ved utformingen av enhver lov, og med det også vår lovgivning om kunstig intelligens, vil prinsippene i Grunnloven stå helt sentralt.

Ved grunnlovsreformen i mai 2014 fikk Grunnloven en rekke nye rettighetsbestemmelser, som i stor grad bygger på den europeiske menneskerettskonvensjon (EMK), samt FNs konvensjon om sivile og politiske rettigheter (SP) og FNs konvensjon om økonomiske, sosiale og kulturelle rettigheter (ØSK). Retten til personvern og ikke-diskriminering finner vi i henholdsvis Grl. §§ 102 og 98 annet ledd. Ved implementering av bestemmelsen om personvern var menneskerettsutvalget særlig bevisst viktigheten av å bevare rettigheten i den teknologiske utviklingen.¹⁸⁰

Grl. § 110 første ledd første punktum gir borgerne en rett til sosial trygget gjennom å sikre at den som «ikke selv kan sørge for sitt livsopphold, har rett til støtte fra det offentlige». Et minimumskrav til de systemer som ivaretar sosial trygghet er at de «bør være lovfestet, og myndighetene må ta ansvar for effektiv administrasjon og overvåkning av systemet».¹⁸¹ Kravet vil stå sentralt ved regulering av bruk av KI-systemer i forvaltningen.

Grl. § 100 om yringsfrihet er også verdt å nevne i denne sammenheng. Bestemmelsen pålegger forvaltningen en informasjonsplikt. Myndighetene skal sikre borgerne en «rett til innsyn i statens og kommunenes dokumenter», jf. femte ledd første punktum, som gjerne omtales som myndighetenes «passive» informasjonsplikt. Den «aktive» informasjonsplikten følger av bestemmelsens sjette ledd, og innebærer at myndighetene plikter å «legge forholdene til rette for en åpen og opplyst offentlig samtale». Det er nødvendig å lage regelverk som gjør det mulig for forvaltningen å overholde denne plikten.¹⁸²

Grunnlovsbestemmelsene skal tolkes i lys av sine folkerettslige forbilder, og praksis fra de internasjonale håndhevingsorganene som forelå ved vedtakelsen av bestemmelsene vil være relevant.¹⁸³ Myndighetenes plikt til å respektere og sikre menneskerettighetene er slått fast i Grl. § 2 annet punktum, jf. «sikre demokratiet, rettsstaten og menneskerettighetene», og er videre knesatt i Grl. § 92. I enhver lovgivningsprosess skal de internasjonale menneskerettighetene utgjøre en ledetråd.

3.1.1.2 Menneskerettsloven

Flere av de internasjonale konvensjonene om menneskerettigheter er gjort til en del av norsk rett gjennom menneskerettsloven. Dette gjelder blant annet EMK, SP og ØSK, jf. mrl. § 2 nr. 1-3. I tillegg til å være relevante ved tolkningen av grunnlovsbestemmelsene som bygger på dem, utgjør med det konvensjonsrettighetene selvstendige skranker for lovgiver.

Konvensjonen går «foran bestemmelser i annen lovgivning» ved motstrid, jf. mrl. § 3. Regelsettet kommer etter dette i en mellomstilling mellom lov og Grunnlov, og i teorien omtales de gjerne som «semikonstitusjonelle»¹⁸⁴.

¹⁷⁹ Skoghøy (2018) s. 15.

¹⁸⁰ Dokument 16 (2011–2012) s. 176.

¹⁸¹ Dokument 16 (2011–2012) s. 238.

¹⁸² Som vi vil komme nærmere tilbake til kan dette være en utfordring ved regulering av kunstig intelligens.

¹⁸³ Rt. 2015 s. 93 (Maria) avsnitt 57.

¹⁸⁴ Skoghøy (2018) s. 48.

Et sentralt bakteppe ved utformingen av bestemmelser om kunstig intelligens vil være EMK art. 8 om retten til privatliv, samt EMK art. 14, som oppgir at ethvert inngrep i en konvensjonsrettighet må skje på en måte som ikke diskriminerer. Bestemmelsen er aksessorisk, og kan med det forstås som en integrert del av den enkelte rettighetsbestemmelse.¹⁸⁵

Retten til privatliv er også nedfelt i SP art. 17, og ØSK har bestemmelser om retten til likebehandling, jf. art. 2, og forbudet mot diskriminering, jf. art. 3. ØSK art. 9 gir en rett til sosial trygghet, og art. 11 nr. 1 skal sikre en rett til «tilfredsstillende levestandard».

Myndighetenes plikt til å sikre borgernes rettigheter, jf. blant annet EMK art. 1, innebærer at lovgiver ikke kan gi lover som på ulovlig måte griper inn i individenes rettigheter. Det er interessant å merke seg at sikringsplikten for lovgivers del også innebærer at lovgiver må sørge for at det foreligger et «adequate legal framework».¹⁸⁶ Rettslig regulering av utvikling og bruk av kunstig intelligente systemer kan tenkes å være et nødvendig middel i sikringen av borgernes grunnleggende rettigheter.

3.1.2 Internasjonale forpliktelser og føringer

Det dualistiske prinsipp innebærer at nasjonal rett og folkeretten utgjør adskilte rettssystemer, slik at Norges traktatforpliktelser først vil være en del av norsk rett etter at det er truffet et vedtak om det av kompetent lovgivningsmyndighet.¹⁸⁷ Dette kan enten skje ved at traktaten i seg selv gis status som norsk lov («inkorporasjon») gjennom en henvisning til traktaten i en norsk lov eller lovbestemmelse, eller ved at det gis lovbestemmelser som bygger på eller gjengir den folkerettslige forpliktelsen («transformasjon»). Menneskerettsloven § 2, som gjennomgått over, er et eksempel på en bestemmelse i norsk lov som inkorporerer internasjonal rett.

Det er ikke bare menneskerettsloven som inkorporer relevante internasjonale menneskerettigheter. FNs konvensjon om avskaffelse av alle former for rasediskriminering (ICERD) er inkorporert gjennom likestillings- og diskrimineringsloven § 5. Et annet eksempel er de forpliktelser Norge har påtatt seg gjennom EØS-avtalen og EU-direktiver. Norge er blant annet forpliktet til å følge EUs likebehandlingsdirektiv (2006/54/EF) gjennom dens EØS-relevans, og EUs rasediskrimineringsdirektiv (2000/43/EF), som Norge på tross av manglende EØS-relevans har forpliktet seg til politisk. EU-forordninger kan også representere bindende rettsakter for Norge, og i denne sammenheng er det relevant å nevne rådsforordning nr. 1408/71 om anvendelse av trygdeordninger, og nr. 574 /72 om koordinering av trygdeytelser i EØS. Utvalget legger som nevnt til grunn at EU-forordningen om kunstig intelligens vil være EØS-relevant.

Det følger av EFTA-domstolens praksis at de «fundamentale rettighetene» i EU, som uttrykt i EUs Charter om grunnleggende menneskerettigheter og EMK, også gjelder i EØS-retten.¹⁸⁸ Forvaltningsorganer som håndhever regler med EØS-opphav, vil etter dette måtte se hen til blant annet EUs Charter art. 41 om retten til god forvaltning.

Videre innebærer presumsjonsprinsippet at traktater Norge er rettslig forpliktet av gjennom avtaler med andre stater og internasjonale organisasjoner, kan bli relevante ved fastleggelsen

¹⁸⁵ *Marckx v. Belgium* avsnitt 32 (No. 6833/74).

¹⁸⁶ *Söderman v. Sweden* avsnitt 85 (No. 5786/08).

¹⁸⁷ Skoghøy (2018) s. 39.

¹⁸⁸ E-2/03 *Ásgeirsson*.

av norsk rett. Prinsippet innebærer at man skal legge til grunn at norsk rett er i samsvar med folkeretten, og så langt det er mulig tolke de norske bestemmelsene på en måte som ikke kolliderer med folkeretten. Det er også grunn til å minne om at Grl. § 92 pålegger statens myndigheter å respektere og sikre de «for Norge bindende traktater om menneskerettigheter». Et eksempel her er FNs konvensjon om rettighetene til mennesker med nedsatt funksjonsevne (CRPD). Konvensjonen er ratifisert av Norge, men ennå ikke inkorporert.¹⁸⁹

3.1.3 Nasjonal lovgivning

3.1.3.1 Forvaltningsloven

Forvaltningsloven er en lov med generelle regler om hvordan den offentlige forvaltning skal behandle saker. Loven regulerer særlig partenes rettigheter under saksbehandlingen. Et sentralt hensyn bak loven er å sikre partenes formelle rettssikkerhet ved forvaltningens behandling av saker med inngripende betydning for privatpersoner.¹⁹⁰

Som allerede argumentert for, mener vi det er nødvendig med et nytt kapittel i forvaltningsloven. Etter utvalgets oppfatning tar ikke Forvaltningslovutvalgets utredning om ny forvaltningslov¹⁹¹ tilstrekkelig hensyn til de problemstillingene som reiser seg som følge av ny teknologi, herunder teknologi som anvender kunstig intelligens. Det er et behov for å redegjøre nærmere for hvordan forvaltningen kan styrke og bevare individenes rettigheter ved bruk av kunstig intelligens og systemavgjørelser.¹⁹²

3.1.3.2 NAV-loven

Lov om arbeids- og velferdsforvaltningen (NAV-loven) skal «legge til rette for en effektiv arbeids- og velferdsforvaltning», jf. § 1 første ledd første punktum. Loven hjemler og regulerer virksomheten til Arbeids- og velferdsforvaltningen (NAV), og samordner bruken av arbeidsmarkedsloven, folketrygdloven, sosialtjenesteloven og andre velferdsmessige lover.

Lovens § 1 annet punktum slår fast at NAV skal «møte det enkelte menneske med respekt, bidra til sosial og økonomisk trygghet og fremme overgang til arbeid og aktiv virksomhet». Loven omhandler blant annet regler om behandling av personopplysninger, jf. § 4a, brukervedvirkning, jf. § 6 og taushetsplikt, jf. § 7.

3.1.3.3 Folketrygdloven

Lov om folketrygd regulerer folketrygden, som er et nasjonalt og sosialt forsikringssystem. Som det fremgår av lovens § 1-1 har loven tre hovedformål. For det første skal loven sikre økonomisk trygghet i ulike faser av livet, som for eksempel ved arbeidsløshet, sykdom og alderdom. For det andre skal folketrygden bidra til å utjevne inntekt og levekår gjennom livet, både for den enkelte og på gruppenivå. For det tredje skal loven bidra til «hjelp til selvhjelp», i den hensikt at den enkelte skal forsørge seg selv, og klare seg selv best mulig i det daglige.

3.1.3.4 Sosialtjenesteloven

Lov om sosiale tjenester i arbeids- og velferdsforvaltningen har som formål å bedre levekårene for vanskeligstilte, bidra til sosial og økonomisk trygghet, fremme likeverd og likestilling, og forebygge sosiale problemer, jf. § 1.

¹⁸⁹ Regjeringen har i Hurdalserklæringen fra 2021 slått fast at CRPD skal inkorporeres i norsk lov.

¹⁹⁰ Graver (2019) s. 299.

¹⁹¹ NOU 2019: 5

¹⁹² Se også Schartums kritikk av NOU 2019:5 i sin Høringsuttalelse til forslaget og artikkelen «Utilstrekkelig forslag om digital forvaltning i ny forvaltningslov» (2019).

3.1.3.5 Likestillings- og diskrimineringsloven

Lov om likestilling og forbud mot diskriminering skal «fremme likestilling og hindre diskriminering» på grunn av vesentlige forhold ved en person, slik som blant annet kjønn, etnisitet og religion, jf. § 1 første ledd. I annet ledd slås det fast at loven med «likestilling» mener «likeverd, like muligheter og like rettigheter», og det presiseres at dette forutsetter «tilgjengelighet og tilrettelegging».

3.1.3.6 Personopplysningsloven

Lov om behandling av personopplysninger gir generelle bestemmelser om behandling av personopplysninger, og loven gjennomfører personopplysningsforordningen i norsk rett. Med «personopplysning» menes opplysninger om en identifisert eller identifiserbar fysisk person, jf. personvernforordning art. 4 nr. 1, jf. personopplysningsloven § 1.

Loven kommer til anvendelse ved «helt eller delvis automatisert behandling av personopplysninger og ved ikke-automatisert behandling av personopplysninger som inngår i eller skal inngå i et register», jf. § 2 første ledd, som gjennomfører forordningens art. 2. Begrepet «automatisert» er ikke legaldefinert, men gjelder behandling som involverer en maskin, som med det i mindre eller ingen grad involverer menneskelig medvirkning. Bruk av systemer som anvender kunstig intelligens vil åpenbart være «automatiserte», slik at personopplysningsloven vil komme til anvendelse dersom det er snakk om personopplysninger.

3.1.3.7 Offentleglova

Lov om rett til innsyn i dokument i offentlig verksemd – offentliglova – omhandler som lovtittelen tilsier retten til innsyn i forvaltningens virksomhet. Et viktig hensyn bak loven er offentlighetsprinsippet, som innebærer at myndighetenes virksomhet skal foregå åpent, slik at dokumenter, opplysninger og argumenter tilgjengeliggjøres for enhver.¹⁹³

Offentlighetsprinsippet kommer til uttrykk i Grl. § 100 femte ledd, og offentliglova er ment å gjennomføre denne bestemmelsen. Offentliglova suppleres av en rekke andre regler som skal sikre offentlighetsprinsippet, som blant annet forvaltningslovens regler om partsoffentlighet for dokumenter, som gir sterkere retter for partene enn offentliglova gjør.

3.1.3.8 Arkivloven

Lov om arkiv har som formål å sikre arkiv som blant annet inneholder rettslig eller viktig forvaltningsmessig dokumentasjon. Poenget er å kunne ta vare på og gjøre tilgjengelig denne dokumentasjon for ettertiden, jf. formålsbestemmelsen i § 1. Loven gjelder som utgangspunkt for alle offentlige organer, jf. § 5. Lovens § 6 fastsetter en generell plikt til å oppbevare arkiver av en viss kvalitet, og operasjonaliserer etter dette formålsbestemmelsen.

Arkivlovutvalgets forslag til ny § 10 gir en mer presis angivelse av hva dokumenteres for virksomheter som benytter en «helt eller delvis automatisert rettsanvendelse». Som utvalget påpeker, er slik dokumentasjon nødvendig for en «åpen og gjennomsiktig offentlig forvaltning som er gjenstand for allmennhetens regler og kontroll».¹⁹⁴

¹⁹³ Frihagen (1972) s. 3.

¹⁹⁴ NOU 2019: 9 s. 36.

3.2 Forvaltningsrettslige prinsipper og regler

3.2.1 Forvaltningsrettslige prinsipper

Utvalget vil i det følgende gjennomgå hvilke grunnleggende forvaltningsrettslige prinsipper vi mener det er særlig sentralt å verne om ved utarbeidelsen av regler for automatiserte beslutningssystemer.

3.2.1.1 *Legalitet*

Legalitetsprinsippet innebærer at forvaltningen (i) må utlede sin kompetanse fra loven, (ii) opptre innenfor rammen av loven og (iii) ha hjemmel for sine inngrep i formell lov, jf. GrL § 113.¹⁹⁵ Legalitetsprinsippet må ses i sammenheng med maktfordelingslæren og folkesuverenitetsprinsippet. Forvaltningen kan ikke på egenhånd fastsette sin kompetanse, og den er nødt til å forholde seg til rettsregler av høyere rang både ved normering og faktiske handlinger, jf. tolkningsprinsippet «lex superior».

Videre innebærer legalitetsprinsippet at forvaltningen kun vil ha hjemmel i lov for et inngrep dersom loven oppfyller visse materielle minstekrav. Lovhjemmelen må være tilstrekkelig «tilgjengelig og presis».¹⁹⁶ Et viktig hensyn bak lovkravet er å sikre forutberegnelighet overfor borgeren. Videre gir det rom for parlamentarisk kontroll med Regjeringen gjennom forvaltningsapparatet, og domstolskontroll av et inngreps lovlighet.

I denne sammenheng er det et poeng å påpeke at for vide fullmakter til forvaltningen kan uthule kravet til en konkret hjemmel.¹⁹⁷ Det er imidlertid på det rene at skjønsmessige bestemmelser er en forutsetning for et fungerende lovverk. Forvaltningen må utstyres med fullmakter for å kunne løse oppgaver som ikke er løst på forhånd.

Demokratisk styring av forvaltningen – gjennom lovgivning og bevilgningsvedtak, men også den alminnelige instruksjons- og organisasjonsmyndigheten til regjering, fylkesting og kommune – er et grunnleggende prinsipp. Legalitetsprinsippet er således en viktig brikke i sikringen av forvaltningens demokratiske legitimitet.

3.2.1.2 *Rettsikkerhet*

Ordet «rettsikkerhet» kan brukes på flere forskjellige måter. Tradisjonelt betegner ordet at borgerne skulle ha sikkerhet for sine rettigheter. Et samfunn regulert av rettsregler håndhevet av myndighetene, i motsetning til et samfunn der den sterkeste rett rår, ville i dette perspektivet representere en «rettsstat», og slik ivareta «rettsikkerheten».¹⁹⁸

Begrepet tar i tillegg innover seg viktigheten av at den enkelte beskyttes mot inngrep fra det offentlige selv. Denne forståelsen av begrepet sammenfaller med det forvaltningsrettslige legalitetsprinsippet. Forvaltningen må i tillegg til å ha hjemmel for sine inngrep, følge loven. Dette kan ses som et generelt krav om rettsriktighet. Forvaltningen er pålagt å tilstå en rettighet eller yte et gode der dette følger av loven.¹⁹⁹

Legalitetsprinsippet og kravet om rettsriktighet setter i utgangspunktet ingen begrensninger for lovens innhold, men i senere års teori har begrepet «rettsikkerhet» blitt ansett å omfatte et

¹⁹⁵ Graver (2019) s. 79.

¹⁹⁶ HR-2016-1833-A avsnitt 15.

¹⁹⁷ NUT 1958: 3 s. 14, Graver (2019) s. 82.

¹⁹⁸ NOU 2019: 5 s. 138.

¹⁹⁹ Graver (2019) s. 113.

krav om veridiktighet, samt visse kvalitative minstekrav til forvaltningen.²⁰⁰ Begrepet kan etter dette også omfatte materielle skranker for lovgivningens innhold – skranker vi finner igjen i blant annet Grunnloven, menneskerettighetene, EØS-loven og det generelle prinsippet om forholdsmessighet.

Vi kan også snakke om prosessuell rettssikkerhet. Det siktes her til regler som skal sikre en behandling som er rettferdig, hensynfull og korrekt. Retten til kontradiksjon står i denne sammenheng svært sentralt, og den prosessuelle rettssikkerheten kan ses som et krav om forsvarlighet.²⁰¹ Forsvarlighetsprinsippet vil vi omtale nærmere nedenfor.

I NOU 2019: 5 legger utvalget til grunn at begrepet «rettssikkerhet» ofte brukes som en «samlebetegnelse for beskyttelse mot overgrep og vilkårlighet fra myndighetene, til å kunne forutberegne sin rettsstilling og forsvare sine interesser».²⁰² Begrepet kan også omfatte hensyn til «likhet og rettferdighet», og blir gjerne fremhevet med svake gruppers stilling i tankene. Det er en slik forståelse av «rettssikkerhet» vi vil legge til grunn i det følgende.

3.2.1.3 Forsvarlighet

Prinsippet om forsvarlighet kan sies å være både prosessuelt og materielt, i den forstand at det stilles krav til både den måten forvaltningen går frem på i saksbehandlingen, samt til forsvarligheten av sakenes innhold. For det sistnevnte kravet vil forholdsmessighet stå sentralt, i tillegg til krav om nøytralitet og saklighet. Skillet mellom forsvarlig saksbehandling og forsvarlige avgjørelser er imidlertid ikke skarpt, og vil gå ut på det samme.²⁰³

En sentral side ved forsvarlighetsprinsippet er kravet til kontradiksjon. Den som berøres av en offentlig avgjørelse, skal kunne øve innflytelse og få fremlegge saken sin. For å realisere dette må vedkommende få *varsel*, *innsynsrett* i avgjørelsen og en *uttalelsesrett*.²⁰⁴ Disse rettighetenes sikring i forvaltningsloven kommer vi nærmere tilbake til om litt.

Retten til kontradiksjon fremgår også av de prinsipper EU-domstolen har oppstilt for forvaltningen, og EU-charteret art. 41. EMD-praksis viser at kontradiktorisk forvaltningspraksis er en forutsetning for et inngrep skal kunne være «nødvendig i et demokratisk samfunn».

En forsvarlig saksbehandling innebærer videre at alle de nødvendige sidene av saken utredes, både de faktiske og de rettslige sidene. Forsvarlige utredninger vil sørge for at den overordnede politikk i større grad gjennomføres, og at individenes interesser og rettigheter ivaretas. Slik fremmes også forvaltningens legitimitet og borgernes tillit. Hensynet til å fatte avgjørelser innen rimelig tid tilsier imidlertid at det må være grenser for hvor grundig forvaltningen skal gå til verks.²⁰⁵ I samme retning kan hensynet til at det offentlige forvalter fellelskapets ressurser trekke.

Regler om for eksempel habilitet, utredning og klage anviser hva som vil være en forsvarlig saksbehandling, samtidig som reglene skal sikre forsvarlige avgjørelser. Forsvarlige avgjørelser er avgjørelser som er nøytrale, saklige og forholdsmessige. Dette innebærer at forvaltningen ikke kan legge vekt på hensyn som er irrelevante eller utenforliggende, eller la

²⁰⁰ Graver (2019) s. 114.

²⁰¹ Graver (2019) s. 113.

²⁰² NOU 2019: 5 s. 139.

²⁰³ Graver (2019) s. 102.

²⁰⁴ Rt. 1997 s. 1019 (Rettergangsbot).

²⁰⁵ Graver (2019) s. 108.

være å vurdere viktige hensyn.²⁰⁶ Læren omtales gjerne som myndighetsmisbrukslæren. Læren anses å omfatte forbudet mot å ta utenforliggende hensyn og forskjellsbehandle, vilkårlighet og grovt urimelige avgjørelser.²⁰⁷

Likhetsgrunnsetningen om at vi alle er like for loven, og ikke må utsettes for usaklig eller uforholdsmessig forskjellsbehandling, jf. Grl. § 98, står sterkt i forvaltningsretten.²⁰⁸ Et annet prinsipp som står sterkt, er forholdsmessighetsprinsippet. Dette prinsippet innebærer at forvaltningen ikke skal benytte fullmaktene sine til å gripe inn i individenes rettigheter og interesser i større utstrekning enn nødvendig.²⁰⁹ Forsvarlighetskravet kan også innebære at det oppstilles visse velferdsmessige minstemål, og Boe har argumentert for et generelt krav til en forsvarlig minstestandard.²¹⁰

3.2.1.4 Effektivitet

En overordnet begrunnelse for mål om effektivitet er å sikre at samfunnets ressurser anvendes på en god måte, og at samfunnsmessige mål nås. I tillegg til å begrunnes i samfunnshensyn, vil en effektiv saksbehandling være av betydning for den enkelte som er part i en sak.²¹¹

Begrepet «effektivitet» kan anvendes med ulikt innhold. Som regel brukes begrepet som et mål på saksbehandlingstiden, med andre ord hvor hurtig saksavviklingen har vært. Med effektivitet kan det imidlertid også siktes det til blant annet styringseffektivitet og kostnadseffektivitet. Styringseffektivitet vil si graden av målrealisering, altså om målet med loven nås gjennom forvaltningens vedtak truffet i medhold av den. For å måle kostnadseffektiviteten ser man på ressursbruken i forhold til målrealiseringen. En rask og enkel saksbehandling vil ikke være kostnadseffektiv dersom den gir feilaktige og u hensiktsmessige avgjøres, som påklages og kreves omgjort.²¹² Forvaltningen bør være «effektiv» i alle de betydningene av begrepet som nå er nevnt.

3.2.1.5 Tillit

Forvaltningslovutvalget omtaler «tillit» som «fravær av behov for å verifisere og bekrefte andres handlinger og bygger på en subjektiv opplevelse av at andre sannsynligvis vil handle på en måte som er god eller nyttig for oss, eller som i det minste ikke er skadelig for oss».²¹³ Dersom borgerne har høy tillit til forvaltningen, vil vi i større grad få en velfungerende forvaltning. Tillit gjør nemlig forvaltningen mer effektiv, blant annet ved å senke konfliktnivået og behovet for oppfølging og kontroll. Tillit er også en forutsetning for å gjennomføre langsiktige reformer, der resultatet ikke vil være synlig med det samme.

For å sikre tilliten til forvaltningen er forvaltningen nødt til å opptre i samsvar med de prinsipper vi allerede har gjennomgått. Forsvarlige avgjørelser, som er saklige og forholdsmessige, bidrar til forståelse og aksept fra borgernes side. En partisk og korrump forvaltning vil være et enormt skår i rettssikkerheten, og svekke forvaltningens effektivitet. Forvaltningslovutvalget fremhever viktigheten av en «åpen, tilgjengelig og forståelig»

²⁰⁶ Graver (2019) s. 120.

²⁰⁷ Graver (2019) s. 122.

²⁰⁸ Forvaltningslovutvalget foreslår en lovfestelse av saklighetskravet i lovens formål, se utkast § 1 (2).

²⁰⁹ Graver (2019) s. 137.

²¹⁰ Boe (1993) s. 715-719. Han begrunner kravet i Rt. 1990 s. 874 (Fusa).

²¹¹ NOU 2019: 5 s. 139.

²¹² NOU 2019: 5 s. 140.

²¹³ NOU 2019: 5 s. 141.

forvaltning, for «jo mer transparent forvaltningen er, desto mindre kan foregå i det skjulte».²¹⁴ Tillit inngis ved at borgeren opplever systemet som forutberegnelig. Det er også viktig at borgeren føler seg ivaretatt, ved at borgerens behov blir sett og imøtegått.²¹⁵

Tilliten kan svekkes, og dersom dette først skjer, kan det være vanskelig å bygge den opp igjen. Forvaltningen må hele tiden jobbe for å opptre på en tillitsfull måte, og fatte riktige og rettferdige avgjørelser. Som nevnt innledningsvis har det på det velferdsrettslige området funnet sted en rekke «social security scandals», og flere av disse knytter seg til uheldig bruk av teknologi i velferdsretten.²¹⁶ Den enorme verdien av høy tillit er følgelig helt sentralt å ha in mente ved utviklingen av velferdsrettslige reformer. Dette inkluderer en grundig vurdering av grunnleggende prinsipper og menneskerettigheter.

3.2.1.6 *Autonomi*

Automatiserte rettsprosesser vil gjennom kategorisering og sin proaktive rolle kunne utgjøre en trussel for den personlige autonomien. Borgerne har en selvbestemmelsesrett, som blant annet begrunner kravet om kontradiksjon. Ved å orienteres om sakens forhold, og ha adgang til å forsvare sine interesser, utøver borgeren sin autonomi.²¹⁷ Forvaltningen må ikke overkjøre borgeren på en måte som gjør at hun verken får forståelse for eller innflytelse over prosessen.²¹⁸ En slik «kafka-prosess» vil senke tilliten til systemet betraktelig, og utgjøre en trussel for rettssikkerheten.²¹⁹

3.2.2 Grunnprinsippenes utslag i forvaltningsloven

I det følgende vil utvalget se nærmere på hvordan disse gjennomgåtte grunnprinsippene operasjonaliseres i forvaltningsloven. Selv om det ikke alltid kommer uttrykkelig til uttrykk i loven, er en rekke av forvaltningslovens regler tuftet på grunnleggende prinsipper om blant annet forsvarlighet, rettssikkerhet, tillit og effektivitet.

3.2.2.1 *Veiledningsplikten*

Forvaltningens veiledningsplikt følger av forvaltningsloven § 11. Formålet med veiledningen er å gjøre parter og andre interesserte i stand til å «vareta sitt tarv i bestemte saker på best mulig måte», jf. første ledd annet punktum. Med «tarv» menes *interesser*.

Veiledningsplikten vil øke borgernes tillit til forvaltningen, og deres personlige autonomi. Kontradiksjon er også en viktig side av kravet til rettssikkerhet. Plikten vil også kunne bidra til at saken blir bedre belyst, noe som sikrer forsvarlige og materielt riktige avgjørelser. Veiledningsplikten kan også gjøre saksbehandlingen mer effektiv, for eksempel ved at flere skjemaer vil være riktig utfylt med korrekt informasjon, slik at man slipper feilaktige avgjørelser på bakgrunn av feil faktum.

Veiledningsplikten innebærer som utgangspunkt at forvaltningen må informere om lovreglene på området, og hvordan disse tolkes og praktiseres. Veiledningen kan også være mer praktisk rettet, for eksempel ved å informere om når en avgjørelse kan ventes, eller gjennom hjelp til

²¹⁴ NOU 2019: 5 s. 145. Transparens som grunnprinsipp i forbindelse med utviklingen av systemer som anvender kunstig intelligens vil vi komme tilbake til.

²¹⁵ Thorgersen (2017) s. 105.

²¹⁶ Robodebt, SyRI eller for Norges del NAV-skandalen.

²¹⁷ Eckhoff/Smith (2014) s. 65 flg.

²¹⁸ Bernt (2003) s. 44.

²¹⁹ Det Norske Akademis Ordbok, s.v. «Kafka-prosess». <https://naob.no/ordbok/Kafka-prosess>

utfylling av skjemaer og utarbeidelse av klager der dette er nødvendig, jf. forvaltningslovforskriften § 3. Som utgangspunkt aktualiseres veiledningsplikten først dersom noen tar kontakt med forvaltningen, men veiledningsplikten er mer vidtrekkende overfor parter i en forvaltningssak, jf. fvl. § 11 annet ledd første punktum.

Hvor omfattende veiledning forvaltningen plikter å gi vil avhenge av behovet til den som skal veiledes. En ressurs svak person vil gjerne ha behov for mer veiledning for å kunne «vareta sitt tarv». Omfanget av veiledningsplikten vil også bero på forvaltningens «situasjon og kapasitet», jf. § 11 første ledd tredje punktum. Det må foretas en samlet helhetsvurdering.

Veiledningsplikten på typisk masseforvaltningsområder vil som hovedregel oppfylles gjennom generelle opplysninger på organets nettsider, såfremt individuell veiledning gis ved behov, for eksempel overfor den som ikke vil eller kan benytte seg av digitale verktøy.²²⁰ I LB-2012-115118 la lagmannsretten til grunn at «tildeling av ytelser etter folketrygdloven er utpreget masseforvaltning, der det treffes et overveldende antall enkeltvedtak hvert år». Etter domstolens oppfatning måtte dette sette «klare rammer for hvor inngående veiledning den enkelte kan ha krav på å få».

3.2.2.2 Utrednings- og informasjonsplikt

Etter forvaltningsloven § 17 plikter forvaltningsorganet å forsikre seg om at «saken er så godt opplyst som mulig før vedtak treffes», jf. første ledd første punktum. Utredningsplikten må ses i sammenheng med det grunnleggende kravet om forsvarlighet. Ordlyden kan peke i retning av en omfattende utredningsplikt, men det er på det rene at bestemmelsen må tolkes i lys av et krav om kostnadseffektivitet og behandling innen rimelig tid. Hensynet til at forvaltningen er satt til å forvalte samfunnets ressurser, og at disse ikke er ubegrensede, får innvirkning på de krav som kan stilles til utredningen.

Dersom saken ikke er tilstrekkelig opplyst, vil dette fort få betydning for avgjørelsens gyldighet. Feilbedømmelse ved mangelfull opplysning er såpass nærliggende at det ofte vil være grunn til å statuere ugyldighet.

Etter § 17 annet ledd må forvaltningsorganet forelegge opplysninger for parten, som denne har rett til å gjøre seg kjent med etter §§ 18 og 19, dersom disse skulle dukke opp under saksforberedelsen. Det samme gjelder andre opplysninger av «vesentlig betydning», jf. tredje ledd. Informasjonsplikten må ses i sammenheng med den grunnleggende retten til kontradiksjon, som er en viktig forutsetning for prinsipper som rettssikkerhet, forsvarlighet, effektivitet og tillit.

3.2.2.3 Plikt til å forhåndsvarsle

Det følger av fvl. § 16 første ledd at forvaltningen «før vedtak treffes» skal varsle en part om vedtaket, og gi denne anledning til å uttale seg. Parten skal gis den informasjonen som er «påkrevd for at parten på forsvarlig måte» skal kunne ivareta sine interesser, jf. annet ledd. Det betyr at parten må gis den informasjon som er nødvendig for at hun skal forstå saken og eventuelt komme med innsigelser. Reglene om forhåndsvarsling er ikke absolutt, jf. tredje ledd, og kan blant unnlates dersom varsling vil medføre en uforholdsmessig stor ressursbruk.²²¹

Igen ser vi at retten til kontradiksjon får et klart utslag i lovens saksbehandlingsregler, og at forvaltningslovens regler er tuftet på viktige prinsipper om rettssikkerhet, forsvarlighet, effektivitet og tillit.

²²⁰ Stub (2021) § 11 note 3.

²²¹ NOU 2019: 5 s. 333.

3.2.2.4 Innsyn

Forvaltningsloven § 18 gir parten en rett til innsyn i saken både under saksforberedelsen og etter at det er truffet et vedtak i saken. På denne måten kan parten påse at forvaltningen baserer seg på korrekt informasjon, og forholder seg saklig og upartisk til denne. Den som ikke regnes som part, må kreve innsyn etter andre regler, særlig offentleglova.²²²

En åpen og tilgjengelig forvaltning er en forutsetning for borgernes tillit til systemet. Dersom forvaltningens virksomhet ikke er tilstrekkelig transparent, vil det heller ikke være mulig å benytte sentrale rettssikkerhetsgarantier som retten til klage og omgjøring – der kontradiksjon igjen står sentralt. Innsynsretten er imidlertid ikke absolutt, og loven oppstiller en rekke unntak, jf. §§ 18a-d og 19. Det gjøres for eksempel unntak fra opplysninger om andre personers helseforhold, jf. § 19 annet ledd, utfra personvern hensyn.

3.2.2.5 Partenes opplysningsplikt

Forvaltningens utredningsplikt begrenses til en viss grad av en forventning om at partene også har en plikt til å bidra til sakens opplysning. Denneplikten er ikke lovfestet i dagens forvaltningslov. Forvaltningslovutvalget foreslår å lovfeste partens opplysningsplikt i de saker der parten selv har tatt initiativet til saken.²²³

På folketrygdlovens område finner vi særlige bestemmelser om partens opplysningsplikt. Etter ftrl. § 21-3 har parten en opplysningsplikt ved et ytelseskrav. Parten skal gi forvaltningen tilstrekkelig informasjon til å vurdere om vedkommende har rett på ytelsen, og eventuelle endringer som kan føre til bortfall av eller endringer i ytelsen.

3.2.2.6 Begrunningsplikt

Det klare utgangspunktet er at forvaltningen plikter å begrunne enkeltvedtak, jf. fvl. § 24 første ledd første punktum. Begrunnelsen skal vise til rettsreglene avgjørelsen bygger på, og i den utstrekning det er «nødvendig for å sette parten i stand til å forstå vedtaket», skal reglens innhold og vedtakets problemstillinger gjengis, jf. § 25 første ledd. Etter annet og tredje ledd skal det også fremgå hvilke faktiske forhold vedtaket bygger på, og de hovedhensyn som har vært avgjørende.

En begrunnelse gir parten grunnlag for å vurdere om hun ønsker å påklage vedtaket eller kreve omgjøring. Dette sikrer kontradiksjon, autonomi og materielt riktige avgjørelser. Begrunnelsen gir også parten forståelse for hvorfor utfallet ble som det ble, slik at hun kan slå seg til ro med at avgjørelsen er fattet på et forsvarlig grunnlag. Dette inngir tillit til forvaltningen.

3.2.2.7 Taushetsplikt og unntak fra denne

Det følger av forvaltningsloven § 13 at forvaltningens tjenestepersoner har en «plikt til å hindre at andre får adgang eller kjennskap til» informasjon om «noens personlige forhold» eller forretningshemmeligheter, som vedkommende får tilgang til gjennom sin posisjon. Loven gir en rekke unntak fra taushetsplikten i § 13a flg. Taushetsplikten etter NAV-loven § 7 er strengere, ved at det gis færre godkjente unntak fra plikten.

Med digitalisering og økt samarbeid på tvers av organisasjonsstrukturer er det grunn til å merke seg at utgangspunktet er en taushetsplikt. Dette utgangspunktet er blant annet begrunnet i retten til personvern og privatliv, som er grunnlovsfestet i Grl. § 102. Som vi vil komme nærmere tilbake til kan krav om innsyn og transparens komme i konflikt med regler som skal beskytte personvernet. Samarbeid på tvers av organer kan gi bedre utredning og flere

²²² NOU 2019: 5 s. 341.

²²³ NOU 2019: 5 s. 330.

riktige avgjørelser, men kan også representere en trussel mot borgernes personvern og autonomi.

4 Grunnleggende prinsipper for utvikling og bruk av kunstig intelligente systemer

I dette kapittelet vil vi gjennomgå hvilke prinsipper som etter utvalgets oppfatning må være grunnleggende ved utvikling og bruk av kunstig intelligente systemer. Disse overordnede prinsippene vil representere grunnleggende krav til systemene, og inngå som en egen prinsippbestemmelse i utvalgets forslag. Prinsippene vil i tillegg utgjøre grunnlaget for utforming av regler om utvikling, bruk og innhold.

I det følgende vil vi se nærmere på transparens, personvern, dataforvaltning, cybersikkerhet, treffsikkerhet, begrunnelse og borgernes krav til livskvalitet og velferd.

4.1 Transparens; forklarbarhet, etterprøvbarehet, åpenhet

Et overordnet prinsipp for kunstig intelligente systemer og systemvedtak er at de må være transparente. Transparens har utgjort en kjerneverdi for EU i hele prosessen med å regulere kunstig intelligens²²⁴, og prinsippet kommer til uttrykk flere steder i forslaget til forordning om harmonisering av regler for kunstig intelligens.

Beaudouin et al. (2020) definerer transparens som at «information is made available as-is»²²⁵, med andre ord som tilgjengeliggjøring av teknisk informasjon om systemets innhold. KI-forordningen anvender etter vår oppfatning en bredere forståelse av begrepet, selv om det er noe uklart hvilken definisjon forordningen legger til grunn.²²⁶ I vårt lovforslag bruker vi derfor transparens som et paraplybegrep som omfatter kriteriene forklarbarhet og etterprøvbarehet, samt et krav om åpenhet rundt bruken av algoritmer.

For forståelsen av begrepenes innhold vil vi i det følgende se nærmere på de mest sentrale bestemmelsene om transparens i forordningsforslaget fra EU.

Art. 11

Technical documentation

1. The technical documentation of a high-risk AI system shall be drawn up before that system is placed on the market or put into service and shall be kept up-to-date.

Art. 12

Record-keeping

1. High-risk AI systems shall be designed and developed with capabilities enabling the automatic recording of events ('logs') while the high-risk AI systems is operating. Those logging capabilities shall conform to recognised standards or common specifications.
2. The logging capabilities shall ensure a level of traceability of the AI system's functioning throughout its lifecycle that is appropriate to the intended purpose of the system

²²⁴ [Ethics guidelines for trustworthy AI](#) (2018), [European Commission White Paper on AI](#) (2020), [European Parliament's Framework of ethical aspect of AI](#) (2020).

²²⁵ Beaudouin et al. (2020) s. 8-9.

²²⁶ Dette har forordningsforslaget fått kritikk for, se blant annet Kiseleva (2021). <https://europeanlawblog.eu/2021/07/29/making-ais-transparency-transparent-notes-on-the-eu-proposal-for-the-ai-act/>

Art. 13

Transparency and provision of information to users

1. High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured, with a view to achieving compliance with the relevant obligations of the user and of the provider set out in Chapter 3 of this Title.
2. High-risk AI systems shall be accompanied by instructions for use in an appropriate digital format or otherwise that include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users.

Art. 52

Transparency obligations for certain AI systems

1. Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use.

Som bestemmelsene gir uttrykk for, brukes begrepet transparenens både om «forklarbarhet» og ved «åpenhet» om bruken av kunstig intelligens overfor brukeren. Selv om forordningsforslaget artikkel 11 (Technical documentation) og artikkel 12 (Record-keeping)²²⁷ ikke inneholder begrepet transparenens, er det vi har valgt å kalle «etterprøvnbarhet» en viktig side av transparenensprinsippet. I det følgende vil vi se nærmere på begrepene forklarbarhet, etterprøvnbarhet og åpenhet.

4.1.1 Forklarbarhet

4.1.1.1 Hva er forklarbarhet?

En sentral side av prinsippet om transparenens er «forklarbarhet». Det spiller ingen rolle om systemets innhold er tilgjengelig «as-is», dersom det ikke er mulig for utviklere, brukere, myndighetene og offentligheten å forstå hvordan systemet fungerer.

Forklarbarhet innebærer at noe er mulig å *forklare*, altså at det er mulig å belyse gjennom mer detaljerte beskrivelser. Verbet «forklare» er i Det Norske Akademis Ordbok definert som å «gjøre tydelig; gjøre forståelig».²²⁸ At noe er «forklarbart» kan etter dette innebære at det må være mulig å gjøre forståelig gjennom en nærmere redegjørelse.

Forordningsforslaget art. 13 slår fast at høyrisiko KI-systemer skal designes og utvikles på en måte som belyser deres virkemåte. Brukerne må kunne være i stand til å «interpret the system's output and use it appropriately», jf. art. 13 (1). Det må med andre ord være mulig for brukerne å fortolke systemets output og anvende denne informasjonen korrekt.²²⁹

Fortalen avsnitt 47 presiserer at dette innebærer at «relevant documentation and instructions of use» med «concise and clear information, including in relation to possible risks to fundamental rights and discrimination, where appropriate» bør vedlegges ethvert høyrisiko

²²⁷ I den danske forordningsteksten: «Registrering».

²²⁸ Den Norske Akademis Ordbok, s.v. «forklare». <https://naob.no/ordbok/forklare>

²²⁹ I den danske forordningsteksten: «Høyrisiko-AI-systemer udformes og udvikles på en sådan måde, at deres drift er tilstrækkelig gennemsigtig til, at brugerne kan fortolke systemets output og anvende det korrekt».

KI-system.²³⁰ Art. 13 (2) oppgir at slike bruksanvisninger skal bestå av «concise, complete, correct and clear information that is relevant, accessible and comprehensible to users».²³¹

Forordningsforslaget stiller også krav til at KI-systemene skal være forklarbare overfor dem som fører menneskelig tilsyn med systemene. Menneskene som fører tilsyn må «be able to correctly interpret the high-risk AI system's output», jf. art. 14 (4) bokstav c, for slik å kunne avgjøre om det er nødvendig å «disregard, override or reverse the output»²³², jf. bokstav d. Systemene skal etter dette overvåkes av mennesker, som har evne til å se bort fra, tilsidesette eller omgjøre maskinens resultater. Dersom systemene ikke er forståelige for dem som skal utføre menneskelig tilsyn («human oversight»),²³³ vil kravet om menneskelig tilsyn i realiteten være uten betydning. Forklarbarhet er med det en forutsetning for menneskelig tilsyn.

Forordningsforslaget innebærer med dette at det skal være mulig for folk å forstå hvordan det aktuelle KI-systemet gjør beregninger og kommer frem til resultater. Det er imidlertid på det rene at enkelte KI-systemer vil være for komplekse til at den menneskelige fatteevne kan trenge til bunns i dem. Det er interessant å merke seg at KI-forordningen ikke virker å utelukke bruken av slike algoritmer for høyrisiko KI-systemer. I fortalens avsnitt 47 slås det fast at det kreves «a certain degree of transparency» for slike systemer, for å bøte på «the opacity that may make certain AI systems incomprehensible to or too complex for natural persons».²³⁴ Kravet er en «viss grad» av transparens, og ikke fullstendig transparens. Dette tilsier at «black-box»-algoritmer kan tas i bruk såfremt virkemåten er tilstrekkelig transparent.

4.1.1.2 Hvordan forklare?

Forskningsområdet for såkalt «forklarbar KI» («*explainable AI*») er stort og tverrfaglig, og rommer fagfelt som datavitenskap, statistikk, anvendt matematikk, informatikk, sosiologi, økonomi og juss.²³⁵ Bakgrunnen for det er at det er bred enighet om at forklarbarhet er essensielt for å sikre tillit til systemer som tar i bruk kunstig intelligens.

I litteraturen skilles det mellom globale og lokale forklaringer.²³⁶ Globale forklaringer er forklaringer av systemet som helhet, det kan si forklaringer av systemets funksjonalitet, for eksempel beslutningstrær, forhåndsdefinerte modeller, kriterier, klassifikasjonsstrukturer, datagrunnlag osv.²³⁷ Lokale forklaringer retter seg mot en spesifikk avgjørelse, altså en forklaring av forholdet mellom regelen og den konkrete regelanvendelsen. Dette vil omfatte begrunnelsen, årsakene og de konkrete omstendighetene som førte til det spesifikke resultatet, som for eksempel profilgruppe eller konkret vektning av ulike kriterier. Globale forklaringer kan gis både i forkant av (ex ante) en automatisert beslutning, eller i etterkant (ex post).

²³⁰ I den danske forordningsteksten: «... ledsages af relevant dokumentation og en brugsanvisning og indeholde kortfattede og klare oplysninger, blandt andet om mulige risici med hensyn til grundlæggende rettigheder og forskelsbehandling, hvis det er relevant».

²³¹ I den danske forordningsteksten: «... kortfattede, fuldstændige, korrekte og klare oplysninger, som er relevante, tilgængelige og forståelige for brugerne».

²³² I den danske forordningsteksten: «... se bort fra, tilsidesætte eller omgøre».

²³³ I den danske forordningsteksten er «human oversight» omtalt som «menneskeligt tilsyn».

²³⁴ I den danske forordningsteksten: «For at afhjælpe den uigennemsigtighet, der kan gøre visse AI-systemer uforståelige eller for komplekse for fysiske personer, bør det kræves en vis grad af gennemsigtighed».

²³⁵ Beaudouin (2020) s. 2.

²³⁶ Beaudouin (2020) s. 8.

²³⁷ Wachter (2017) s. 78.

Lokale forklaringer kan kun gis etter at den automatiserte beslutningen har funnet sted (ex post).

Beaudouin et. al (2020) mener globale forklaringer kan være brukerveiledning til hvordan modellen fungerer, herunder hvordan den lærer, hvilke treningsdata som er brukt og svakheter ved modellen.²³⁸ En slik brukerveiledning har en klar side til kravene om etterprøvnbarhet, som vi straks vil se nærmere på. Lokale forklaringer vil ha som hovedformål å gi brukeren svar som kan være til hjelp i den konkrete situasjonen brukeren befinner seg i, for eksempel knyttet til hvorfor søknaden ble avslått, utslagsgivende faktorer og endringsmuligheter.

Hvordan forklarbarhet kan måles, og hvilke krav som må stilles for at en forklaring for at den skal anses som tilstrekkelig, vil vi se nærmere på om litt.

4.1.1.3 *Hvorfor er forklarbarhet viktig?*

Forklarbarhet er en forutsetning for ivaretagelsen av grunnleggende demokratiske prinsipper om offentlighet, rettssikkerhet og demokratisk kontroll. Offentlighetsprinsippet gir som tidligere nevnt allmennheten en rett til innsyn i forvaltningens saksdokumenter. Prinsippet begrunnes i hensynet til et deliberativt og deltakende demokrati, kontrollhensyn for å avdekke vilkårlighet og urettferdighet, og rettssikkerhetshensyn, i form av å sikre respekten for legalitetsprinsippet og hindre myndighetsmisbruk.

Dersom forvaltningen skal anvende algoritmer som ledd i utføringen av forvaltningsmessige oppgaver, må borgerne ha mulighet til å forstå saksprosessen og kreve innsyn i forvaltningens systemer. Poenget må være at innsynskrav gir en mulighet til å gjøre en kjent med, påvirke og kontrollere forvaltningens virksomhet.²³⁹ Uten åpenhet og kunnskap om en prosess vil det være umulig å kritisere dens innhold. Et innsynskrav vil være av liten verdi dersom innsynet ikke gir et grunnlag for forståelse. Det er først gjennom forklarbare systemavgjørelser retten til informasjon og innsyn realiseres.

Behovet for demokratisk kontroll av forvaltningen kommer blant annet til uttrykk gjennom forvaltningens passive og aktive informasjonsplikt. Den passive informasjonsplikten begrunnes i offentlighetsprinsippet, som gir borgerne en «rett til innsyn i statens og kommunenes dokumenter», jf. Grunnloven § 100 femte ledd første punktum. Unntaket fra retten må fastsettes i lov og må begrunnes i «hensyn til personvern» eller «andre tungtveiende grunner», jf. annet punktum.

Offentleglova, som skal bidra til å realisere grunnlovsbestemmelsen, slår fast at det skal legges til rette for at «offentleg verksemd er open og gjennomiktig, for slik å styrkje informasjons- og ytringsfridommen, den demokratiske deltakinga, rettstryggleiken for den enkelte, tilliten til det offentlege og kontrollen frå ålmenta», jf. § 1. Etter lovens § 3 kan alle som utgangspunkt kreve innsyn i et offentlig organs dokumenter.

Hensynet bak offentlighetsprinsippet tilsier at informasjonen som gis ved innsyn bør være forståelige for brukerne. Dette er i tråd med bestemmelsen om klarspråk i den nye språklova, som presiserer at det offentlige «skal kommunisere på eit klart og korrekt språk som er tilpassa målgruppa», jf. språklova § 9. Forarbeidene til bestemmelsen presiserer at det med «språk» også siktes til «innhald, struktur og visuelle element».²⁴⁰

For å sikre en effektiv forvaltning, som ikke overleses av innsynskrav som kreves bearbeidet for å gjøre dem forståelige, er det behov for å utvikle standarder for forklaringer ved innsyn.

²³⁸ Beaudouin (2020) s. 13.

²³⁹ NOU 2003: 30 pkt 4.1.

²⁴⁰ Prop. 108 L (2019–2020) s. 119.

Av personvern hensyn kan det imidlertid være grunn til å forbeholde dyptgående innsynskrav, altså ettergåelse av modellen i sin helhet, til offentlige tilsyns- eller revisjonsmyndigheter. Dette taler for at innsynskrav bør besvares med globale og lokale forklaringer, og ikke gjennom en full tilgang til en rekonstruksjon av datamodellen med all brukt data tilgjengelig.

Forvaltningen har også en aktiv informasjonsplikt, som går ut på å «legge forholdene til rette for en åpen og opplyst offentlig samtale», jf. Grunnloven § 100 sjette ledd. Plikten er begrunnet i et formål om «bidra til kvalitet og mangfold i den offentlige debatt», og kan gjelde både det offentliges virksomhet eller privates virksomhet.²⁴¹ Den aktive informasjonsplikten er imidlertid ikke presisert i lovgivningen, og forvaltningslovutvalget vurderte det slik at veiledningsplikten til forvaltningen i stor grad ivaretar informasjonsbehovet.²⁴²

4.1.2 Etterprøvbarhet

4.1.2.1 Hva er etterprøvbarhet?

Begrepet «etterprøvbarhet» sikter til at noe er mulig å kontrollere eller få bekreftet i ettertid.²⁴³ Dette fordrer at det som skal etterprøves er blitt dokumentert, slik at man har en tilstrekkelig notoritet for å foreta en kontroll. KI-forordningen stiller flere krav til systemavgjørelsers etterprøvbarhet.

I tillegg til å stille krav om at en automatisert beslutning må kunne forklares overfor brukeren, stiller forslaget til forordning også krav om en mer «compliance-orientert» forklarbarhet.²⁴⁴ Art. 13 (1) angir at «an appropriate type and degree of transparency shall be ensured, with a view to achieving compliance with the relevant obligations of the user and of the provider»,²⁴⁵ slik disse er fastsatt i forordningens kapittel 3. Hva som vil være «tilstrekkelig transparent» etter forordningen må etter dette ses i sammenheng med de mer konkrete, operasjonaliserbare reglene i kapittel 3. Dette omfatter blant annet krav om innføring av et kvalitetsstyringssystem, arkivering av loggfiler og CE-merking.

Art. 13 (1) annet punktum må ses i sammenheng med kravet om teknisk dokumentasjon i art. 11. Etter denne bestemmelsen kreves det at all teknisk dokumentasjon om høyrisiko KI-systemer skal utarbeides i forkant av lanseringen av disse nye systemer, og etter dette holdes kontinuerlig oppdatert.

Etter art. 12 skal høyrisiko KI-systemer designes og utvikles på en slik måte at det er mulig å foreta en «automatic recording of events ('logs')»²⁴⁶ når programmet kjører, jf. (1). Loggene skal sikre at funksjonaliteten til systemene er så sporbare («level of traceability»)²⁴⁷ som må kreves sett hen til formålet med systemet, jf. (2). I fjerde ledd presiseres det at et minimumskrav til en logg for systemer klassifisert som høyrisiko er at man logger enhver

²⁴¹ NOU 1999: 27 s. 250.

²⁴² NOU 2019: 5 s. 245.

²⁴³ Det Norske Akademis Ordbok, s.v. «etterprøve». <https://naob.no/ordbok/etterpr%C3%B8ve>

²⁴⁴ Sovrano (2021) s. 131.

²⁴⁵ I den danske forordningsteksten: «Der sikres en passende type og grad af gennemsigtighed med henblik på opfyldelse af de relevante bruger- og udbyderforpligtelser».

²⁴⁶ I den danske forordningsteksten: «hændelser («logfiler») registreres automatisk».

²⁴⁷ I den danske forordningsteksten: «passende grad af sporbare af systemets funktion gennem hele dets livscyklus».

bruk av systemet, datagrunnlaget som er anvendt, dataen som er brukt som input og hvilke mennesker som har vært involvert i prosessen, jf. kravet i art. 14.

Art. 20 stiller krav om at tilbydere («providers») av høyrisiko systemer skal ta vare på automatisk genererte logger i en passende tidsperiode sett hen til formålet bak loggen. Tilbydere plikter etter art. 23 å gi offentlige myndigheter tilgang til loggene dersom det er i samsvar med lov eller avtale med brukeren. Forslaget til forordning stiller også krav til brukere og importører med flere.

I fortalen punkt 46 presiseres kravet til etterprøvbarehet. Det fremgår her at kravene til teknisk dokumentasjon og loggføring skal inneholde «general characteristics, capabilities and limitations of the system, algorithms, data, training, testing and validation processes used as well as documentation on the relevant risk management system».²⁴⁸ For at systemene skal være tilstrekkelig etterprøvbare må generell informasjon om systemet og utviklingen av systemet være tilgjengelig. Kravet om å ta vare på dokumentasjon om systemets risikostyringssystem («risk management system») innebærer at det må kunne være mulig å tilgjengelige vurderinger som er tatt og tas etter kravene til risikostyringssystemer for høyrisiko KI-systemer i art. 9.

Kriteriene forordningen oppstiller for etterprøvbarehet minner om kjennetegnene for globale forklaringer som presentert tidligere, og er viktige for generell kontroll og offentlighet rundt systemene. Muligheten for etterprøving av enkeltavgjørelser vil også være avhengig av tilstrekkelig dokumentasjon og loggføring, for slik å kunne avgjøre hvordan hver enkelt prediksjon har foregått.

Ved å stille krav om at systemene utvikles etter en viss teknisk standard, oppstiller man samtidig kriteriene for standardisering. Dette er inspirert av EUs New Legislative Framework (NLF), som legger ansvaret for konformitetsetterlevelsen av produktsikkerhet på produsenten ut fra en tanke om at produsentene er best egnet til nettopp det. Dette skiller seg fra områder der det offentlige selv foretar en sjekk av hvorvidt noe er i samsvar med de oppstilte krav.²⁴⁹

Forordningsforslaget kapittel IV regulerer standardiseringsprosessen for KI-systemer. Blant annet oppgis det i vedlegg IV (2) bokstav b at «the design specifications of the system, namely the general logic of the AI system and of the algorithms»²⁵⁰ skal være blant informasjonen som skal gis for å vise etterlevelse (compliance) med forordningen før KI-systemet gjøres tilgjengelig på markedet. Systemets konstruksjon, og det og algoritmenes generelle logikk, herunder begrunnelser og antakelser, er etter dette blant informasjonen som må kunne etterprøves. Ved å ha en slik kontroll ex ante vil man kunne minimere risikoen for at KI-systemet får uønskede konsekvenser.²⁵¹ Deretter kan et tilsynsorgan eller en annen tredjepart kontrollere at dokumentasjonen er korrekt, enten ex ante eller ex post systemet er lansert.

²⁴⁸ I den danske forordningsteksten: «systemets generelle egenskaber, kapacitet og begrænsninger samt algoritmer, data, trænings-, prøvnings- og valideringsprocesser foruden dokumentation for det relevante risikostyringssystem. Den tekniske dokumentation skal holdes ajour».

²⁴⁹ Veale et. al. (2021) s. 102.

²⁵⁰ I den danske forordningsteksten: «systemets konstruktionsspecifikationer, dvs. AI-systemets og algoritmenes generelle logik, navnlig de vigtigste designvalg, herunder begründelser og antagelser».

²⁵¹ Sovrano (2021) s. 132.

Forordningsforslagets generelle tilnærming er som nevnt risikobasert, jf. art. 5. De overordnede kravene om etterprøvbarehet synes til syvende og sist å skulle bidra til å minimere systemenes skadepotensiale.

4.1.2.2 Hvorfor er etterprøvbarehet viktig?

For at forvaltningens virksomhet skal være åpen og offentlig, og slik bidra til en opplyst samtale, må informasjonen lagres. Dokumentasjon gir også mulighet for nødvendig kontroll av forvaltningen. Videre er tilgjengelig og sporbar dokumentasjon en forutsetning for å kunne gi innsyn og begrunnelser, som igjen er viktig for borgernes rett til rettsikkerhetsgarantier som klage og omgjøring.

Det følger av arkivlova § 6 at det offentlige «pliktar å ha arkiv, og desse skal vera ordna og innretta slik at dokumenta er tryggja som informasjonsskjelder for samtid og ettertid». Forarbeidene presiserer at både selve dokumentet og den sammenheng det inngår i er av betydning.²⁵² For automatiserte saksbehandlingssystemer kan dette omfatte tolkningsvalg tatt av systemets utviklere.

Schartum har påpekt at tilgjengeliggjøring av rettslige beslutninger i forbindelse med systemutviklingen kan ses som en forutsetning for forsvarlig saksbehandling ved bruk av automatiserte systemer. Offentliggjøring av slik informasjon står etter hans oppfatning sentralt i ivaretagelsen av borgernes rettsikkerhet.²⁵³ Dette kommer vi nærmere tilbake til i kommentarene til regulering av systemenes bruk.

I NOU 2019: 5 foreslår forvaltningslovutvalget i en ny § 12 at forvaltningsorganet «skal dokumentere det rettslige innholdet i automatiserte saksbehandlingssystemer», og at denne dokumentasjonen som utgangspunkt «skal offentliggjøres», jf. s. 20. Arkivlovutvalget forslag til ny § 10 gir en mer presis angivelse av hva som skal dokumenteres. Utkastet lyder som følger:

- (1) Virksomheter som helt eller delvis automatiserer rettsanvendelse, skal dokumentere
 - (a) hvilke datatyper som anvendes
 - (b) hvilke kilder som benyttes for disse datatypene
 - (c) hvilke behandlingsregler som er utledet av rettsreglene og som er styrende for vedtakene, og
 - (d) hvilke endringer som er gjort av systemet på bakgrunn av vedtak av lover og forskrifter og andre politiske vedtak.
- (2) Kongen kan pålegge at det utarbeides dokumentasjon som gir grunnlag for rekonstruksjon av kjørbare programkode, etter at systemet er tatt ut av drift.

Hvordan tilstrekkelig etterprøvbarehet bør sikres på et operativt nivå kommer vi nærmere tilbake til.

4.1.3 Åpenhet

I forslaget til prinsippbestemmelse er kravet om at rettslige systemer skal anvendes med åpenhet en side av prinsippet om transparens. Åpenhetsprinsippet innebærer at brukeren skal gjøres klar over at hun er i kontakt med en algoritme. Det er med andre ord et krav om at brukeren gis informasjon om det, dersom hun interagerer med en algoritme.

²⁵² Ot.prp. nr. 77 (1991–92) s. 6.

²⁵³ Schartum (2018) s. 278.

KI-forordningen art. 52 innebærer at «natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use». Dersom mennesker interagerer med et KI-system må det etter dette informeres om det, såfremt det ikke fremgår som åpenbart. Åpenhetskravet gjelder ikke bare for høyrisiko KI-systemer, men også systemer med lavere risiko, slik som chat-bots, enkelte gjenkjennelsessystemer, eller bruk av falske bilder, lyd eller video (for eksempel «deepfake»²⁵⁴).

4.1.4 Målestokk og terskel for oppfyllelse av kravene om transparens

For å avgjøre om et system er tilstrekkelig transparent, må det oppstilles en terskel for oppfyllelse. En terskel for oppfyllelse krever at det er mulig å *måle* transparens, og det er med det et behov for å finne frem til en målestokk («metrics»)²⁵⁵. Dette vil for det første være nyttig før systemene tas i bruk, *ex ante*, ved vurderingen av om systemet oppfyller de standardiserte kravene. I tillegg vil en målestokk måtte stå sentralt *ex post*, når man skal vurdere om de standardiserte kravene er overholdt. KI-systemet som er under lupen til et overvåkningsorgan, vil måtte sammenlignes med den aktuelle standarden for etterlevelse.

Sovrano et. al. har tatt til orde for at en målestokk for transparens må inneholde flere karakteristikk.²⁵⁶ For det første vil hvilken målestokk som legges til grunn bero på den aktuelle risikoen systemet representerer for de aktuelle individenes grunnleggende rettigheter og friheter («risk-focused»). Dette samsvarer godt med forordningsforslagets generelle risikotilnærming, jf. art. 5. For det andre mener Sovrano et. al. at målestokken må være anvendelig på ulike typer algoritmer («model-agnostic»). Dette innebærer at målestokken må kunne benyttes på alt fra systemer som anvender maskinlæring til systemer kun basert i statistikk.

Sovrano et. al. påpeker for det tredje at målestokken må kunne justeres utfra hvem som er mottaker for forklaringen, og deres behov («goal-aware»). Mottakerne vil typisk være brukere, tilbydere eller tilsynsorganer. Målestokken bør etter dette være utformet på en slik måte at den kan tilpasses så mange mottakere som mulig. Som EU-kommisjonen har påpekt er det et behov for at informasjonen «should be tailored to the particular context».²⁵⁷

For det siste fremhever Sovrano et. al. at målestokken i seg selv må være forståelig og tilgjengelig. Hvis ikke vil forklaringens mottakere havne i en «ignotum per ignotius»-situasjon, der forklaringen er vanskeligere å forstå enn det som bli forklart. Det har lite for seg å motta et mål på graden av transparens, dersom det ikke er mulig å forstå bakgrunnen for resultatet.

Selv om utviklingen av en målestokk for transparens vil være en utfordring, er det en forutsetning for etterlevelse av KI-forordningen ved implementering av KI-systemer. I forvaltningsrettslig sammenheng vil en målestokk stå sentralt for forvaltningsrettslige prinsipper om rettssikkerhet, tillit og legitimitet, og regler om begrunnelse, innsyn og klage.

²⁵⁴ «Deepfakes» er når man benytter dyplæringsmetoder i kunstig intelligens til å lage bilder av falske hendelser. Se Sample (2020).

²⁵⁵ Sovrano (2021) s. 132.

²⁵⁶ Sovrano (2021) s. 133.

²⁵⁷ European Commission (2020) s. 20.

4.2 Forsvarlig forvaltning av data

4.2.1 Bakgrunn

Prinsippet om forsvarlig forvaltning av data er formulert med grunnlag i de etiske retningslinjene for kunstig intelligens, utviklet av EUs ekspertgruppe for kunstig intelligens. Prinsippet er utviklet med utgangspunkt i kravet om *personvern og datastyring* («privacy and data governance»²⁵⁸).

Vernet om privatlivets fred er en grunnleggende menneskerettighet, slik det følger av blant annet den europeiske menneskerettskonvensjonen (EMK) artikkel 8. Et element i denne rettigheten er at enhver skal ha kontroll over og kunne bestemme over egne personopplysninger, som også er bakgrunnen for EUs personvernforordning (GDPR). En økt bruk av kunstig intelligente systemer medfører en risiko for retten til personvern og privatliv. Dette gjelder spesielt i velferdsforvaltningen, hvor personopplysninger vil utgjøre en stor del av datagrunnlaget for slike systemer.

En forutsetning for å ivareta denne rettigheten er at data som anvendes i forbindelse med utvikling og bruk av kunstig intelligente systemer, blir *forvaltet* på en måte som respekterer personvernet. Dette innebærer å sikre kvaliteten og integriteten til slik data, hvem som har tilgang til den og at den blir behandlet forsvarlig.

Forholdet til personvernet vil bli nærmere behandlet i et eget punkt senere. I dette punktet vil utvalget fokusere på forvaltning av data og relaterte krav i forordningsforslaget, og hvilke implikasjoner disse vil ha for norsk velferdsforvaltning.

4.2.2 Hva er dataforvaltning?

Dataforvaltning kan forstås som prosessen med å opprette, innhente, endre, dele, beskytte, dokumentere og lagre *data*. Data er informasjon i form av et bestemt format – én bestemt representasjon av opplysninger. Begrepet brukes gjerne om opplysninger i elektronisk form, som kan lagres og brukes av datamaskiner.²⁵⁹

Det kan skilles mellom to hovedtyper kilder til data:²⁶⁰

1. Data fra mennesker som sanser og fortolker sine omgivelser og uttrykker sine forståelser på den formaliserte måten som det digitale systemet tillater
2. Data fra sensorer som er programmert til direkte å registrere ulike aspekter ved den fysiske virkeligheten og inngi disse dataene til det digitale systemet – det vil si uten å gå «omveien» om et menneskes fortolkning

For at data skal ha noen verdi for oss, må de tillegges en mening; vi må forstå hvordan de skal tolkes og hva de representerer. Data er både en fortolkning av opplysningene som de representerer, og en opplysning som selv må fortolkes. Det er først når vi kan tolke dem at data blir til informasjon. Data kan representere et bestemt forhold eller en egenskap, eller de kan være kontekstavhengig. Tallet 67 vil for eksempel bare være et tall, med mindre vi bestemmer at det representerer alderen til en person. Hvis vi i tillegg bestemmer at konteksten

²⁵⁸ European Commission (2019).

²⁵⁹ Cambridge Dictionary, s.v. «data», <https://dictionary.cambridge.org/dictionary/english/data>.

²⁶⁰ Oppsummeringen er skrevet med inspirasjon fra Schartum (2021).

er alderen for pensjonsuttak, vil dataen gi oss informasjon om at vedkommende kan ta ut alderspensjon. Informasjon kan følgelig sies å være *meningsinnholdet* i data.²⁶¹

Hvis denne informasjonen skal anvendes til et bestemt formål, må den *forvaltes*. Det betyr at informasjonen må innhentes, lagres og gjøres tilgjengelig for bruk og eventuell deling. Hvis informasjonen er mangelfull eller uriktig, må den endres. Informasjonen må også beskyttes for å sikre at uvedkommende ikke får tilgang til å redigere eller slette den.

Begrepet «dataforvaltning» er ikke uttrykkelig definert i noen norske kilder. Et begrep som anvendes i et langt større omfang, og som kan sies å være nært beslektet med «dataforvaltning», er «informasjonsforvaltning». Med informasjonsforvaltning menes et helhetlig syn på aktiviteter, verktøy og andre tiltak for å sikre best mulig kvalitet, utnytting og sikring av informasjonen i en virksomhet. Organiseringen av informasjonen skal være systematisk og henge sammen med virksomhetens arbeidsprosesser.²⁶²

Informasjonsforvaltning kan sies å være forvaltning av alle former for informasjon, inkludert data. Dataforvaltning vil i så fall være en underkategori av informasjonsforvaltning, som kun dreier seg om forvaltning av informasjon i form av data. Informasjon som forvaltes digitalt i dag er gjerne informasjon i form av data, og prinsipper som gjelder for informasjonsforvaltning vil følgelig gjelde også for dataforvaltning.

Det er videre en forskjell mellom åpne og *lukkede* data. Åpne data defineres som «digitale data som er gjort tilgjengelige med de tekniske og juridiske egenskaper som er nødvendige for at de skal kunne brukes fritt, og bli distribuert av hvem som helst, når som helst og hvor som helst».²⁶³ Åpne data er informasjon som er tilgjengelig for alle og som alle kan få tilgang til, bruke og dele, uten begrensning av personelle regler.²⁶⁴ Lukkede data er informasjon som kun autoriserte og autentiserte brukere har tilgang til. Det gjelder for eksempel personopplysninger, som er skjermet på grunn av personvern hensyn.²⁶⁵

4.2.2.1 Begrepene «data management» og «data governance»

På engelsk skilles det mellom begrepene «data management» og «data governance».

Betydningen av «data management» samsvarer i stor grad med betydningen av *dataforvaltning*. Dataforvaltning som disiplin består av flere komponenter, herunder «data governance», som kan oversettes til *datastyring*.

Dataforvaltning er en *praksis*. Det innebærer at data forvaltes som en verdifull ressurs, slik at verdien av dataen kan realiseres. En forutsetning for en effektiv dataforvaltning er at man har en *strategi* for hvordan man skal få tilgang til, innhente, endre, bruke og lagre data. Det er dette som er datastyring.

Datastyring handler om å gi retningslinjer for hvilke beslutninger som må tas ved forvaltning av data, hvem som skal ta disse beslutningene, og deres roller og ansvar i relasjon til disse beslutningene. Gjennom dataforvaltning sikrer man *implementering* av disse retningslinjene.²⁶⁶

²⁶¹ Store Norske Leksikon (2021).

²⁶² Digitaliseringsdirektoratet (u.å.l).

²⁶³ Kultur- og likestillingsdepartementet (2017).

²⁶⁴ Felles datakatalog, s.v. «åpne data», <https://data.norge.no/search-all?q=åpne%20data>.

²⁶⁵ Meld. St. 22 (2020–2021).

²⁶⁶ Meld. St. 22 (2020–2021).

Både de etiske retningslinjene for pålitelig kunstig intelligens utviklet av EUs ekspertgruppe for kunstig intelligens og forordningsforslaget refererer til «data governance» i de punktene som ligger til grunn for dette prinsippet. Utvalget har likevel valgt å formulere et prinsipp om forsvarlig *dataforvaltning* og ikke *datastyring*. Det har vi valgt fordi *dataforvaltning* er et mer helhetlig begrep som omfatter *både* *datastyring* og andre relaterte elementer. For å sikre en trygg bruk av kunstig intelligens i velferdsforvaltningen er det viktig at hele praksisen er forsvarlig: det vil ikke være tilstrekkelig å *kun* ha på plass en forsvarlig strategi, dersom den ikke fungerer i praksis.

4.2.2.2 *Dataforvaltning i forordningsforslaget*

Forordningsforslagets artikkel 10 stiller omfattende krav til datakvalitet og *datastyring* for datasett som brukes ved utviklingen av høyrisiko kunstig intelligente systemer som blir trent opp ved hjelp av data, det vil si *maskinlæring*. Kunstig intelligente systemer, spesielt systemer som involverer maskinlæring, krever mye data for utvikling og testing. Bestemmelsen skal sikre at slik data holder høy kvalitet.

Dette forutsetter implementering av *egnet datastyrings- og -forvaltningspraksis* («appropriate data governance and management practices»)²⁶⁷.

Datasett som brukes ved utviklingen av systemet skal være *tilstrekkelig relevant, representativ, fri for feil og fullstendig* («sufficiently relevant, representative, free of errors and complete»), sett i lys av systemets tiltenkte formål. De må også ha *egnede statistiske egenskaper* («appropriate statistical properties») med hensyn til systemets målgruppe.²⁶⁸ Disse kravene har mottatt sterk kritikk for å være umulige å oppfylle i praksis. For det første finnes det ingen nærmere retningslinjer eller felles forståelse for hva som ligger i kravene. Det er også uklart hvordan et system skal designes for å overholde dem. For det andre er kravene urealistiske. Det er for eksempel ikke mulig å ha et feilfritt datasett i tilfeller hvor det er tale om store mengder data, hvor store deler av dataen er personopplysninger og datasettet oppdateres kontinuerlig. I tillegg vil kunstig intelligente systemer ofte trenes på ufullstendige eller uriktige datasett, nettopp for å sikre at systemet er tilstrekkelig robust.

I rapportutkastet («draft report») som ble publisert tidligere i år har bestemmelsen blitt endret, slik at det kreves at datasett er relevante, representative, *oppdaterte*,²⁶⁹ og *i best mulig grad, i betraktning av teknikkens stand, frie for feil*²⁷⁰ og *så fullstendige som mulig*.²⁷¹ Den endrede ordlyden åpner i større grad for at det kan være feil i datasettet, så lenge antall feil holdes til et minimum. Den innrømmer også at feilmarginen kan variere ut fra hva slags teknologi det er tale om i et konkret tilfelle.

En vurdering av hvorvidt data overholder disse kvalitetskravene må foretas i lys av systemets tiltenkte formål. Sett i sammenheng med den reviderte ordlyden, tilsier det at den tillatte «feilmarginen» for datakvaliteten vil avhenge av hva systemet skal brukes til. Et system som skal brukes til å treffe mer eller mindre endelige vedtak vil innrømmes en mindre feilmargin, enn et system som kun skal anvendes som beslutningsassistanse til menneskelige saksbehandlere. En slik forståelse samsvarer også med den risikobaserte tilnærmingen generelt i forordningsforslaget.

²⁶⁷ COM(2021) 206 art. 10 nr. 2.

²⁶⁸ Ibid. art. 10 nr. 3 og fortalepunkt 44.

²⁶⁹ «up-to-date».

²⁷⁰ «and to the best extent possible, taking into account the state of the art, free of errors».

²⁷¹ «as complete as possible»; 2021/0106(COD).

Datasett som brukes ved utviklingen av systemet må videre ta hensyn til *trekkene, egenskapene eller elementene som er særegne for den spesifikke geografiske, adferdsmessige eller funksjonelle omgivelsen eller konteksten*²⁷² innenfor området hvor systemet skal brukes.²⁷³ I rapportutkastet er bestemmelsen utvidet til å også omfatte den *kulturelle* omgivelsen eller konteksten. Dette kan leses som et krav om mangfold og representativitet i datasettene. Det innebærer at datasett må reflektere variasjonene i befolkningen og omgivelsene i den delen av samfunnet som systemet skal interagere med.

For å sikre at ingen blir utsatt for urettmessig diskriminering som følge av bias i systemet, vil det være i det offentliges interesse at leverandørene/tilbyderne av systemet kan behandle særlige kategorier av personopplysninger for å overvåke, oppdage og korrigere *bias*.²⁷⁴

Artikkel 10 nevner ikke hvem som er ansvarlig for å sikre den påkrevde datakvaliteten. Dette følger av Artikkel 16, som fastslår at det er tilbydere av høyrisiko kunstig intelligente systemer som må sikre at deres høyrisiko kunstig intelligente systemer overholder kravene som stilles i forordningsforslagets kapittel 2.²⁷⁵

4.2.3 Hvorfor er dataforvaltning viktig?

For at offentlig sektor skal kunne tilby digitale og automatiserte tjenester til borgerne, er det nødvendig med data om personer, deres behov, preferanser, relasjoner, omstendigheter, tidligere hendelser mv. Slik data gjør det mulig å utvikle brukervennlige tjenester. Det innebærer også at dataforvaltning blir et nytt ansvarsområde for offentlig sektor.

Dataforvaltning innebærer at data ses på som en *ressurs*, det vil si at informasjonen som ligger i data er verdifull. Verdien kan realiseres ved å bruke informasjonen til forbedring, fornyelse eller etablering av virksomheter og tjenester. Et sentralt aspekt ved digitalisering er at *analog informasjon* – det vil si informasjon som representerer virkeligheten – blir transformert til data som kan leses og brukes av maskiner. Når informasjonen blir «oversatt» til et språk som maskinen forstår, kan maskinen tolke og analysere informasjonen, og ta beslutninger og utføre handlinger på egenhånd. Dette effektiviserer prosesser som tidligere måtte gjøres fysisk av mennesker, og frigjør menneskelig kapasitet til andre oppgaver.

Informasjon er blant de viktigste ressursene i offentlig sektor, og kan brukes til å utvikle og forbedre velferdsordninger for borgerne. Norge står i en særstilling sammenlignet med mange andre land når det gjelder informasjon om borgerne og samfunnet; vi har én enhetlig arbeids- og velferdsforvaltning (NAV), bestående av både kommunale og statlige tjenester. Det innebærer at vi har mye informasjon samlet på ett sted, og ikke i like stor grad er avhengige av å samle informasjon fra flere ulike organer eller hente informasjon fra andre (for eksempel private) aktører.

Digitaliseringen av det norske samfunnet har ført til at vi nå har store datamengder fra flere kilder. I den forbindelse oppstår spørsmålet om hvordan forvaltningen kan utnytte denne informasjonen til å lage bedre tjenester og til å løse samfunnsoppdrag på helt nye måter.

²⁷² «the features, characteristics or elements that are particular to the specific geographical, behavioural or functional setting or context».

²⁷³ COM(2021) 206, art. 10 nr. 3 og fortalepunkt 44.

²⁷⁴ Ibid.

²⁷⁵ «ensure that their high-risk AI systems are compliant with the requirements set out in Chapter 2».

Innovasjon, brukerorientering og god utnyttelse av data henger nært sammen.²⁷⁶

Dataforvaltning handler nettopp om å utvinne potensialet som ligger i utnyttelsen av data.

Mye data kan være en fordel fordi (1) bearbeiding av store datamengder er mer effektivt enn å behandle datasett enkeltvis, og (2) mer informasjon fører til ny kunnskap som kan gi større nytte for samfunnet.²⁷⁷ Samtidig kan store mengder data være problematisk hvis dataen ikke blir forvaltet på en forsvarlig måte.

God forvaltning av informasjon, herunder data, er en forutsetning for å realisere prinsippet om «kun én gang», som blant annet handler om å redusere borgernes byrder av dobbelrapportering til det offentlige.²⁷⁸ For å nå denne målsettingen, er det iverksatt et arbeid med «orden i eget hus», som nettopp er en veileder for god informasjonsforvaltning. Den skal gjøre det lettere for offentlig sektor å gjenbruke data i saksbehandling og annen oppgaveløsning ved å vite at data finnes, hvor data befinner seg og hva den kan brukes til.²⁷⁹

Mye data er en nødvendig forutsetning for å utvikle kunstig intelligente systemer, særlig systemer hvor maskinlæring blir benyttet. Mye data er imidlertid ikke tilstrekkelig for at vi skal kunne nyte fordelene av kunstig intelligens. Hvilken verdi data har, avhenger av hva den brukes til. Data som kan brukes til å forutsi borgernes behov i ulike livsstadier og tilby passende veiledning og tjenester vil være verdifull. Det forutsetter imidlertid at slik data *faktisk er egnet til* å brukes av et kunstig intelligent system til nettopp dette formålet; data må være av tilstrekkelig *kvalitet*. Dette stiller på den ene siden krav til de opplysningene som blir omgjort til lesbar data for systemet, og på den andre siden krav til at dataen må ha bestemte egenskaper, for eksempel riktig format og fullstendighet. Et godt datagrunnlag er essensielt for at den informasjonen som vi får fra kunstig intelligente systemer skal være forståelig og riktig.

Dataforvaltning er også viktig ved anvendelse av kunstig intelligente systemer. Systemet må klare å innhente riktig data fra riktig sted, som forutsetter at data er lagret og organisert. Et dataavhengig system vil heller ikke fungere hvis det plutselig ikke har tilgang på data, eller hvis dataen blir endret til å være uforståelig for systemet. Data må derfor beskyttes, slik at ingen uvedkommende kan slette eller redigere data.

I velferdsforvaltningen vil data ofte bestå av opplysninger som kan knyttes til enkeltpersoner, altså personopplysninger. Personvernregelverket stiller ytterlige krav til behandling av slike opplysninger. En forsvarlig forvaltning av data skal også sikre ivaretagelse av personvernet.

4.2.4 Hva innebærer prinsippet om forsvarlig forvaltning av data?

Prinsippet om forsvarlig forvaltning av data innebærer at (1) data som brukes til å utvikle høyrisiko kunstig intelligente systemer ved hjelp av maskinlæring må oppfylle bestemte kvalitetskrav, og at (2) data blir håndtert på en hensiktsmessig måte ved utviklingen av høyrisiko kunstig intelligente systemer.

²⁷⁶ Digitaliseringsdirektoratet (u.å.m).

²⁷⁷ Ibid.

²⁷⁸ Holte (2019).

²⁷⁹ Kommunal- og distriktsdepartementet (2019).

4.2.4.1 Hva er datakvalitet?

En generell definisjon av *datakvalitet* er «fitness for use», det vil si hvorvidt data er egnet for å brukes til et bestemt formål.²⁸⁰ Det er imidlertid ikke mulig å måle datakvalitet ut fra hvor egnet dataen er for et bestemt formål eller en bestemt bruk. Datakvalitet kan imidlertid måles gjennom andre parametere, både kvantitativt og kvalitativt. Sammen kan disse parameterne gi uttrykk for dataens datakvalitet. Det er ingen universell enighet om hvilke parametere som skal legges til grunn, og ulike aktører har ulik tilnærming.

Digitaliseringsdirektoratet har utarbeidet en spesifisering for beskrivelse av kvalitet på datasett som del av rammeverket for informasjonsforvaltning.²⁸¹ Rammeverket skal gi føringer for alle offentlige virksomheter i arbeidet med «orden i eget hus». Det følger av denne spesifiseringen at kvaliteten på et datasett kan vurderes og beskrives ved hjelp av flere *kvalitetsdimensjoner*, disse er: nøyaktighet, komplett/dekning, aktualitet, samsvar, tilgjengelighet og relevans.²⁸²

Datakvalitet kan måles på både enhetsnivå og egenskapsnivå.²⁸³ En enhet er hvert registrerte datapunkt, mens egenskaper er opplysninger knyttet til hvert datapunkt. I praksis vil for eksempel en person være en slik enhet, og datakvaliteten vil kunne måles på grunnlag av de ulike opplysningene som er registrert om vedkommende. En bestemt opplysning, for eksempel personens alder, vil være en egenskap. Datakvaliteten vil også kunne måles for den aktuelle egenskapen hos alle personene. Hvis noen registrerte personer unnlater å oppgi alder, vil datagrunnlaget for den egenskapen være mangelfullt og derfor ufullstendig.

Det er viktig at data er *korrekt*, det vil si at den i størst mulig grad fungerer som en representasjon for virkeligheten. Data må også være *nøyaktig*, ved at betegnelsen på data må samsvare med hvilke verdier den representerer. Videre må data være *dekkende*, i den forstand at den inneholder de forventede opplysningene. Hvis man mangler data som burde ha vært med i datasettet, vil man ha *underdekning* av data. Tilsvarende gjelder hvis datasettet inneholder nullverdier. Data må oppdateres, slik at datasettet er tilstrekkelig *aktuelt* for dataens spesifikke brukskontekst. Til slutt må data være *konsistent*, i den forstand at det må være en forståelig sammenheng mellom ulike data.²⁸⁴

Algoritmer som brukes i maskinlæringssystemer og kunstig intelligens kan bare være så gode som den dataen som brukes til å utvikle dem.²⁸⁵ Bruk av irrelevant data kan føre til at systemet finner vilkårlige statistiske sammenhenger som ikke burde være brukt som grunnlag for å ta beslutninger. Feil i data kan påvirke beslutningene som treffes av et kunstig intelligent system.

En særlig aktuell kvalitetsutfordring er *bias*.²⁸⁶ Bias oppstår i tilfeller hvor datasett kun inneholder opplysninger om en *del* av det relevante datagrunnlaget, som ikke er representative for hele datagrunnlaget. Bias kan også oppstå av andre årsaker. For eksempel kan treningsdatasett for veiledet læring ha bias som følge av menneskelige feilvurderinger eller historiske skjevheter i datagrunnlaget. Historisk skjevhet kan gå ut på underrepresentasjon av

²⁸⁰ Mahanti (2021).

²⁸¹ Digitaliseringsdirektoratet (u.å.n).

²⁸² Spesifisering for beskrivelse av kvalitet på datasett. Digitaliseringsdirektoratet (2017).

²⁸³ Digitaliseringsdirektoratet (2020).

²⁸⁴ Oslo kommune (u.å.).

²⁸⁵ European Union Agency for Fundamental Rights (2019).

²⁸⁶ Se definisjon i kapittel 1.4.

en minoritet, eller at mennesker havner i kategorier basert på egenskaper som egentlig ikke burde være representative i den aktuelle konteksten. Det kan for eksempel være at menn er underrepresentert i et datasett som skal brukes av et system til å beregne om det er behov for oppfølging etter å ha fått barn. Det kan også være at et system som skal hjelpe med sysselsetting ikke anbefaler bestemte stillinger til kvinner fordi det er en statistisk underrepresentert gruppe i disse stillingene. Kunstig intelligens kan også bli påvirket av hvem som definerer problemstillingene.²⁸⁷

4.2.4.2 Krav til datakvalitet i forordningsforslaget

Forordningsforslaget stiller følgende krav til datakvaliteten for høyrisiko kunstig intelligente systemer:²⁸⁸

«Training, validation and testing data sets shall be relevant, representative, free of errors and complete. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons on which the high-risk AI system is intended to be used. These characteristics of the data sets may be met at the level of individual data sets or a combination thereof. »

Selv om kravene er formulert på en absolutt måte, gir fortalen uttrykk for et mer nyansert meningsinnhold. Ifølge fortalepunkt 44 skal datasett som brukes til trening, validering og testing være «sufficiently relevant, representative and free of errors and complete *in view of the intended purpose of the system*». Dette åpner i realiteten opp for en kontekstuell vurdering av datakvaliteten, hvor det nærmere innholdet i kravene må ses i lys av systemets tiltenkte formål.

En slik forståelse kan også leses ut fra artikkel 8 nr. 2, som fastslår at «[t]he intended purpose of the high-risk AI system ... shall be taken into account when ensuring compliance with those requirements», det vil si at systemets tiltenkte formål må tas i betraktning ved vurderingen av om kravene til høyrisiko kunstig intelligente systemer er overholdt.

Bestemmelsen er foreslått endret i et rapportutkast («draft report») til forordningsforslaget som ble publisert tidligere i år. Den endrede versjonen stiller krav om at datasett må være relevante, representative, *oppdaterte* («up-to-date»), *og i best mulig grad, i betraktning av teknikkens stand, frie for feil* («and to the best extent possible, taking into account the state of the art, free of errors») og *så fullstendige som mulig* («as complete as possible»)²⁸⁹ Den endrede ordlyden åpner i større grad for at det kan være feil i datasettet, så lenge antall feil holdes til et minimum. Den innrømmer også at feilmarginen kan variere ut fra hva slags teknologi det er tale om i et konkret tilfelle.

Det kan være verdt å merke seg at endringer i systemets tiltenkte formål («intended purpose») nødvendigvis gjør en ny samsvarsvurdering («conformity assessment»)²⁹⁰ Til grunn for dette kravet ligger en erkjennelse av at en vurdering av et systems samsvar («conformity») med forordningsforslagets krav ikke kan foretas uavhengig av systemets tiltenkte formål; formålet med systemet må følgelig ha betydning for hvordan kravene tolkes.

4.2.4.3 Hva er hensiktsmessig håndtering av data?

Hensiktsmessig dataforvaltning er en viktig forutsetning for fungerende kunstig intelligente systemer. De ulike elementene i dataforvaltningsprosessen – fra behandling og lagring, til

²⁸⁷ Nasjonal strategi for kunstig intelligens, pkt. 5.1.

²⁸⁸ COM(2020) 206 art. 10 nr. 3.

²⁸⁹ 2021/0106(COD).

²⁹⁰ COM(2020) 206 art. 43 (4) og fortalepunkt 66.

formatering og anvendelse av data til bestemte formål – bidrar samlet til å sikre at systemdata er nøyaktig og tilgjengelig.

Det første trinnet i prosessen vil ofte være utviklingen av en *dataarkitektur*. Enkelt forklart kan dette sies å være virksomhetens infrastruktur for data. Det er modeller, regler og standarder som styrer hvilke data som samles inn, hvordan de lagres, systematiseres og anvendes av datasystemer.²⁹¹

Data lagres normalt i *databaser*. En database er en samling med data som er lagret på et elektronisk medium. Datasamlingen er organisert og strukturert etter en bestemt strategi eller modell.²⁹² En database kan sies å representere en forenkling av virkeligheten. Hvor *virkelighetsnær* denne forenklingen er, vil bero på informasjonen som finnes i dataen. Når det gjelder kunstig intelligente systemer som anvender data som grunnlag for å utføre forvaltningsoppgaver, er det spesielt viktig at informasjonen i databasen samsvarer med virkeligheten. Databasen vil her fungere som en *faktabase* for systemet.

Et viktig begrep i denne sammenheng er *dataintegritet*, som refererer til dataens nøyaktighet, konsistens og pålitelighet. Dataintegritet har nær sammenheng med datakvalitet. Det er imidlertid en viktig forskjell mellom begrepene. Når vi snakker om *datakvalitet*, snakker vi om hvor riktig og pålitelig informasjonen er til å tjene et bestemt formål. Begrepet *dataintegritet* dreier seg om informasjonens *fysiske og logiske validitet*. Det går blant annet ut på å sikre at data ikke blir endret eller ødelagt på en uautorisert måte, og at informasjonen er i overensstemmelse med virkeligheten og konsistent.²⁹³ Dataintegritet er viktig for at informasjonen skal være *brukbar*. Dataintegriteten kan være ivaretatt, selv om *datakvaliteten* er dårlig. Det vil for eksempel være tilfelle hvis en bruker oppgir feil alder. Et system som er avhengig av data om brukerens alder vil fortsatt kunne generere et resultat, selv om det blir feil. Samtidig kan datakvaliteten være god, men uten dataintegritet vil ikke dataen være til noen nytte. Hvis alderen oppgis i feil format, eller ikke kan knyttes til brukeren, vil ikke systemet kunne anvende informasjonen i riktig kontekst.

En database med systematisert og tilgjengelig data vil gjøre det enkelt å hente ut og administrere dataen. Administrering av databaser er et viktig element i dataforvaltningen. Det innebærer blant annet å sikre god opptid og responstid, slik at systemene som henter informasjon fra databasene kan fungere. Andre oppgaver inkluderer blant annet oppdateringer av databasene, sikkerhetskopiering og utføring av programvare- og sikkerhetsoppdateringer.

Andre viktige deler av dataforvaltningen er *datamodellering*, hvor datastrukturer beskrives ved hjelp av modeller; *dataintegrasjon*, som handler om kombinasjon av data fra forskjellige datakilder; *datastyring*, som går ut på å fastsette retningslinjer og prosedyrer for å sikre enhetlig og effektiv bruk av data til et bestemt formål; styring av *datakvalitet*, hvor formålet er å rette eventuelle feil og unøyaktigheter i dataen; og «master data management» eller håndtering av masterdata.

Program for dataforvaltning bør inkludere:

- Etablerte datastyringskontroller som gir sikkerhet ved å begrense tilgangen til data til kun autoriserte brukere, og gjør det enkelt å identifisere data man trenger.
- Lett tilgjengelig data.

²⁹¹ Dataforeningen (2019).

²⁹² Store Norske Leksikon (2019).

²⁹³ Hagen (2005).

- Infrastruktur som kan utvikle seg når virksomheten endrer seg.
- Muligheten til å jobbe med eksisterende teknologi uten å måtte erstatte den med ny teknologi.
- Konsistent og kontrollert datadeling på tvers av virksomheten.
- I tillegg bør data ha tilstrekkelig datakvalitet.

En mangelfull dataforvaltning kan føre til at det oppstår såkalte «datasiloer», inkonsekvente datasett og datakvalitetsproblemer. Dette vil påvirke det kunstig intelligente systemet som avhenger av dataen, og kan i verste fall føre til at systemet gir uriktige resultater.

4.2.5 Prinsippets føringer for norsk velferdsforvaltning

Dataforvaltning bør være en kjerneoppgave for forvaltningen. Det innebærer at det må legges planer, strategier og budsjettmessige prioriteringer for å sikre effektiv datainnsamling, god datakvalitet, sikker datadeling og optimal utnyttelse av data. Det er i dag få statlige etater som eksplisitt anser data- eller informasjonsforvaltning som en del av sitt samfunnsoppdrag. I realiteten er dette likevel en viktig del av de fleste offentlige virksomheter.

Data vært et *uregulert rettslig område*. Med gjennomføringen av forordningsforslaget vil offentlige virksomheter plutselig være nødt til å forholde seg til et regelverk som setter konkrete krav til dataforvaltningen. Det kan skape en bølge av «compliance»-arbeid, slik situasjonen var for mange virksomheter ved ikrafttreddelsen av personvernforordningen. I denne sammenhengen er det viktig å minne om at Norge er i toppen når det kommer til digitalisering, og at vi allerede har utviklet mange gode strategier og retningslinjer for informasjonsforvaltning som de fleste virksomheter klarer å forholde seg til. Det er derfor grunn til å tro at forordningsforslagets datarelaterte krav ikke vil medføre store endringer eller omorganiseringer i offentlig sektor.

Prinsippet om forsvarlig forvaltning av data kan ses i sammenheng med det generelle forsvarlighetsprinsippet i forvaltningen. Som nevnt i utredningens punkt 3.2.1.3, stiller prinsippet krav til både forvaltningens fremgangsmåte i saksbehandlingen, samt til sakens innhold. Forsvarlighetsprinsippet skal sikre retten til kontradiksjon, at saken belyses fra alle sider og at det ikke blir tatt irrelevante eller utenforliggende hensyn.

Når det gjelder retten til kontradiksjon, er det en forutsetning for å utøve denne retten at man vet at man har en sak, og hva saken dreier seg om. Personvernforordningen stiller krav om at den registrerte skal informeres når det samles inn personopplysninger fra vedkommende.²⁹⁴ Den registrerte skal blant annet vite hvilke opplysninger det gjelder, hva de skal brukes til og hvor lenge de blir lagret. I tillegg skal vedkommende informeres om sin rett til innsyn og klage. Hvis personopplysningene blir brukt i forbindelse med automatiserte avgjørelser, følger det et krav i artikkel 15 nr. 1 bokstav h om at den registrerte skal få informasjon «om den underliggende logikken samt om betydningen og de forventede konsekvensene av en slik behandling». Hva som ligger i dette, redegjøres for nærmere under punkt 4.5.1.

Forsvarlig forvaltning av data vil bidra til at personvernreglene blir overholdt. En viktig side av forsvarlig dataforvaltning er jo nettopp at man skal ha oversikt og kontroll over hvilke data man sitter på og hva den brukes til. Kravet om loggføring («record keeping») vil også ha betydning i denne sammenhengen. Ved å loggføre systemets operasjoner kan man dokumentere når systemet ble brukt, og i hvilken sammenheng. Når den registrerte informeres

²⁹⁴ Personvernforordningen artikkel 13.

om at et kunstig intelligent system behandler personopplysninger om vedkommende, gis vedkommende også mulighet til å gjennomgå og eventuelt korrigere opplysningene. Dette bidrar til at data holdes «up-to-date» og representerer virkeligheten i størst mulig grad, i samsvar med kravene til datakvalitet i forordningsforslaget. På den måten sikrer man også retten til kontradiksjon.

Kravene som stilles til kvaliteten på data kan bidra til en bedre opplysning av saken. Et representativt datagrunnlag vil sikre at det er tilstrekkelig mangfold i de opplysningene som systemet vil bruke til å komme frem til et resultat. Et oppdatert datagrunnlag vil sikre at også endrede eller nye opplysninger vil bli tatt i betraktning. Et nøyaktig og fullstendig datagrunnlag vil sikre at ingen viktige opplysninger «blir borte» i prosessen, og at systemet kommer frem til resultatet på riktig grunnlag.

Når data blir forvaltet på en forsvarlig måte, blir det også enklere å kontrollere hvorvidt data som ligger til grunn for et resultat er relevant, og luke ut eventuelle irrelevante opplysninger eller utenforliggende hensyn fra det aktuelle datagrunnlaget. Spørsmålet om beslutningsgrunnlaget i enkeltsaker er fullstendig og relevant er et spørsmål om *rettsikkerhet* for den enkelte.²⁹⁵ Brukeren av et kunstig intelligent system vil ha en interesse i å kunne endre og tilføre opplysninger om seg selv og sin sak for å sikre at den endelige beslutningen treffes ut fra riktig informasjon. På den måten kan vedkommende også sikre en så adekvat og sann som mulig beskrivelse av seg selv og sin sak. Dette bidrar til å oppfylle forvaltningens utredningsplikt, samtidig som brukeren får utøve sin rett til innsyn i og retting av personopplysninger etter personvernforordningen artikkel 13, 15 og 16. Videre vil et fullstendig og relevant beslutningsgrunnlag også være et *saklig* beslutningsgrunnlag.

Det er verdt å nevne at EU-kommisjonen har lagt frem et forslag om en ny dataforordning, som inneholder regler om hvem som kan bruke og få tilgang til data skapt i EU på tvers av alle økonomiske sektorer. Data betraktes her som et gode som skal være tilgjengelig for alle, og en fornybar ressurs som ikke vil forringes ved at mange mennesker har tilgang samtidig. Forslaget vil blant annet gjøre det lettere for offentlig sektor å bruke data som finnes i privat sektor. Norge samarbeider tett med EU på digitaliseringsområdet, og den nye dataforordningen forventes å være EØS-relevant.²⁹⁶

4.3 Teknisk robusthet og sikkerhet

4.3.1 Bakgrunn

I Nasjonal strategi for digital sikkerhet fra 2021 viser regjeringen til en stadig økende sårbarhet for digitale trusler i samfunnet. I takt med at infrastrukturer og IKT-systemer blir mer komplekse, globale og integrerte, bruken av skyløsninger øker og flere IKT-funksjoner settes ut til tredjeparter, blir digitale trusler stadig mer sammensatte og krevende å forutse.²⁹⁷ Med en økende grad av digitalisering vil både volumet og variasjonen av data som skapes og hastigheten de samles inn på, øke. Funn fra en undersøkelse gjennomført av EY viser at denne utviklingen har ført til økt bekymring i offentligheten for personvern og mangel på kontroll over hvordan ens egne data tas i bruk.²⁹⁸

²⁹⁵ Schartum (1993) s. 64.

²⁹⁶ Osloregionens Europakontor (2022).

²⁹⁷ Departementene (2021) pkt. 1.1.

²⁹⁸ Bertrand (2021).

Som nevnt i utredningen punkt 3.2.1.5 er tillit en forutsetning for en velfungerende forvaltningsmyndighet. Selv om Norge er i verdenstoppen når det gjelder tillit til myndighetene,²⁹⁹ er det viktig at myndighetene ikke tar denne tilliten for gitt i forvaltningens digitaliseringsprosess. Opprettholdelse og økning av denne tilliten er en viktig forutsetning for at forvaltningen kan drives mer effektivt.³⁰⁰ På bakgrunn av dette, har regjeringen uttrykt en visjon om at det i Norge skal være trygt å bruke digitale tjenester. I et digitalisert samfunn skal både privatpersoner og virksomheter kunne ha tillit til at den nasjonale sikkerheten, individenes velferd og deres demokratiske rettigheter blir ivaretatt.³⁰¹

Å sikre tillit er også bakgrunnen for at EUs ekspertgruppe for kunstig intelligens fremhevet «robust KI» som en av de grunnleggende retningslinjene for troverdig KI.³⁰² Ekspertgruppen trekker frem at det ikke er tilstrekkelig at KI-systemer overholder og beskytter de individuelle rettighetene og de etiske normene de springer ut av. Samfunnet må også være trygge på at KI-systemene ikke vil forårsake utilsiktet skade både fra et teknisk og fra et sosialt perspektiv.³⁰³

Ved utformingen av kravene til troverdig KI utledet ekspertgruppen en rekke prinsipper fra menneskerettighetene, herunder prinsippet om forebygging av skade.³⁰⁴ Kunstig intelligente systemer må verken forårsake eller forverre skade, eller på annen måte påvirke mennesker negativt. Dette innebærer blant annet at systemene og miljøene de opererer i må være trygge og sikre, samt at ekstra forsiktighet må utøves i møte med sårbare personer og asymmetriske makt-/informasjonsforhold.³⁰⁵ Ekspertgruppen formulerte på denne bakgrunn et krav om *teknisk robusthet og sikkerhet*.³⁰⁶ Kravet forutsetter systemer som er motstandsdyktige og sikre mot angrep, samt en plan for det tilfellet at noe skulle gå galt. I tillegg stiller det krav til nøyaktighet, pålitelighet og reproduserbarhet.³⁰⁷

I den følgende vurderingen av kravets innhold vil vi se nærmere på den sentrale bestemmelsen i EUs forslag til KI-forordning.

Article 15

Accuracy, robustness and cybersecurity

1. High-risk AI systems shall be designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of accuracy, robustness and cybersecurity, and perform consistently in those respects throughout their lifecycle.
2. The levels of accuracy and the relevant accuracy metrics of high-risk AI systems shall be declared in the accompanying instructions of use.
3. High-risk AI systems shall be resilient as regards errors, faults or inconsistencies that may occur within the system or the environment in which the system operates, in particular due to their interaction with natural persons or other systems.

²⁹⁹ Direktoratet for forvaltning og økonomistyring (2022).

³⁰⁰ Kommunal- og distriktsdepartementet (2018).

³⁰¹ Departementene (2021) pkt. 1.3.

³⁰² High-Level Expert Group on Artificial Intelligence (2019) s. 7.

³⁰³ L.c.

³⁰⁴ High-Level Expert Group on Artificial Intelligence (2019) s.10-14.

³⁰⁵ Ibid s.12.

³⁰⁶ Ibid s. 14 og s. 16-17.

³⁰⁷ Se også forklaring i utredningens punkt 4.1.

The robustness of high-risk AI systems may be achieved through technical redundancy solutions, which may include backup or fail-safe plans.

High-risk AI systems that continue to learn after being placed on the market or put into service shall be developed in such a way to ensure that possibly biased outputs due to outputs used as an input for future operations ('feedback loops') are duly addressed with appropriate mitigation measures.

4. High-risk AI systems shall be resilient as regards attempts by unauthorised third parties to alter their use or performance by exploiting the system vulnerabilities.

The technical solutions aimed at ensuring the cybersecurity of high-risk AI systems shall be appropriate to the relevant circumstances and the risks.

The technical solutions to address AI specific vulnerabilities shall include, where appropriate, measures to prevent and control for attacks trying to manipulate the training dataset ('data poisoning'), inputs designed to cause the model to make a mistake ('adversarial examples'), or model flaws.

Som bestemmelsen gir uttrykk for dekker kravet både nøyaktighet, robusthet og cybersikkerhet. Kravene til pålitelighet og reproduserbarhet er ikke nevnt, men som vi vil vise betyr dette ikke at kravene ikke er tatt høyde for i forslaget. I det følgende vil vi se nærmere på disse fem begrepene.

4.3.2 Nøyaktighet

Nøyaktighet innebærer at noe er nøyaktig. Adjektivet kan defineres som noe «som stemmer helt» med for eksempel virkeligheten, et forbilde eller liknende.³⁰⁸ Kravet forutsetter med dette en høy grad av riktighet, korrekthet eller presisjon i forhold til noe annet. Kjernen i kravet til nøyaktighet er KI-systemets evne til å gjøre korrekte vurderinger og foreta korrekte prognoser, anbefalinger eller beslutninger basert på data og modeller.³⁰⁹ I utredningens punkt 4.4.4.1 utdypes mulige forklaringer hva som ligger i riktige avgjørelser. Disse vurderingene er også relevante for vurderingen av nøyaktighet.

KI-forordningens artikkel 15 nr. 1 og nr. 2 krever at høyrisiko KI-systemer skal utformes og utvikles med «an appropriate level of accuracy».³¹⁰ Hva som regnes som et passende nivå avhenger av formålet til systemet, men en nærmere beskrivelse av kravets innhold gis ikke i forordningsteksten.

Et av formålene med forslaget er å sikre EU borgernes rettigheter i møte med kunstig intelligente systemer. Det er derfor nærliggende å anta at nivået vil måtte variere basert på risikoen for at systemet vil utfordre en eller flere av disse rettighetene. Tilsvarende antakelse følger av at bestemmelsen gjelder for høyrisiko KI-systemer. Videre vil nivået variere ut fra typen rettighet som utfordres. For eksempel vil et høyt nivå av nøyaktighet være spesielt avgjørende i situasjoner der KI-systemet direkte påvirker menneskeliv. Fortalen viser i punkt 39 videre til viktigheten av blant annet høy nøyaktighet overfor særlig sårbare parter.

I tillegg til å sørge for en passende nøyaktighet krever forslaget at nivået skal holdes konsekvent gjennom hele systemets livssyklus. Ordlyden³¹¹ tilsier at det ikke er snakk om å holde nøyaktigheten på et konsekvent nivå, men snarere at en skal sørge for at nivået av

³⁰⁸ Det Norske Akademis Ordbok, s.v. «nøyaktig». <https://naob.no/ordbok/n%C3%B8yaktig>

³⁰⁹ High-Level Expert Group on Artificial Intelligence (2019) s. 17.

³¹⁰ I den danske forordningsteksten: «et passende niveau af nøjagtighed».

³¹¹ «...and perform consistently in those respects throughout their lifecycle» jf. forslagets artikkel 15 nr. 1.

nøyaktighet alltid er *passende*. Med andre ord må et høyrisiko KI-system stadig kontrolleres, vurderes og videreutvikles idet graden av nøyaktighet som kreves kan variere over tid.

Selv om forordningsteksten ikke eksplisitt sier noe om hvordan passende nøyaktighet skal sikres er det klart at en rekke av kravene som stilles til høyrisiko KI vil kunne bidra til dette. Regler som sikrer forsvarlig forvaltning av data (se punkt 4.2) sørger for at vurderinger, prognoser, anbefalinger, beslutninger m.m. tas på så korrekt grunnlag som mulig. Risikostyringssystemer som tas i bruk gjennom hele livssyklusen til systemet – f.o.m. utvikling t.o.m. avvikling og etterfølgende lagring – kan bidra til å redusere og korrigere utilsiktede risikoer av unøyaktighet. I forslaget er det henholdsvis artikkel 10 og artikkel 9 som er ment å ivareta dette. Videre gjør regler som sikrer transparens (se punkt 4.1) det mulig for brukere, utbytere og andre å kontrollere graden av nøyaktighet og eventuelt avdekke hvor sannsynlig det er at unøyaktighet inntreffer. I tillegg muliggjør regler om tilsyn (jf. blant annet forslagets artikkel 14) og utpekelsen av tilsynsmyndighet (jf. forslagets artikkel 59) korreksjon av unøyaktighet både før og etter inntreffelse.

4.3.3 Robusthet

Begrepet «robusthet», i denne sammenhengen knyttet til et system, sikter til at noe er driftssikkert og effektivt og følgelig stabilt.³¹² Artikkel 15 nr. 3 i forslaget krever at høyrisiko KI-systemer skal være motstandsdyktige overfor «errors, faults or inconsistencies»³¹³ som kan oppstå i systemet eller miljøet som systemet opererer i. Dette gjelder særlig der årsaken er systemets interaksjon med fysiske personer eller andre systemer.

Fortalens punkt 50 utdyper at denne typen risiko særlig er knyttet til systemets begrensninger. For eksempel kan en bruker kommunisere informasjon til systemet på en uventet eller motsigende måte. Ord kan ha ulik betydning i ulike regioner og måten vi interagerer med digitale systemer vil variere avhengig av faktorer som alder, kulturell bakgrunn samt kunnskap om og tillit til systemet. Alternativt kan systemet operere i et miljø av andre systemer som sammen understøtter hverandre, deler informasjon og liknende. En hendelse som oppstår i ett av disse systemene, kan ha store konsekvenser for andre³¹⁴ blant annet for deres tilgang til relevant og korrekt data og liknende.

I EUs etiske retningslinjer for pålitelig KI uttrykkes det at teknisk robusthet krever at KI-systemer utvikles med en forebyggende tilnærming til disse risikoene. Systemene skal kunne operere som tiltenkt selv om risikoene skulle inntreffe og slik at uventede negative følger minimeres og uakseptable negative følger forhindres.³¹⁵

Forslaget til forordning om KI nevner kun et eksempel på sikkerhetstiltak som kan sikre passende robusthet – «technical redundancy solutions».³¹⁶ Reserveplaner og feilsikre planer nevnes eksplisitt som slike mulige teknisk overflødige løsninger. Dette kan blant annet innebære at KI-systemet teknisk bytter fra en statistisk til en regelbasert prosedyre, eller ber om og/eller kontakter en menneskelig operatør før behandlingen kan fortsette.³¹⁷

³¹² Det Norske Akademis Ordbok, s.v. «robust». <https://naob.no/ordbok/robust>

³¹³ I den danske forordningsteksten: «fejl, svigt og uoverensstemmelser».

³¹⁴ Kommunal- og moderniseringsdepartementet (2020) pkt. 3.2.

³¹⁵ High-Level Expert Group on Artificial Intelligence (2019) s. 16.

³¹⁶ Jf. artikkel 15 nr. 3. I den danske forordningsteksten: «tekniske redundansløsninger».

³¹⁷ High-Level Expert Group on Artificial Intelligence (2019) s. 16-17.

Andre tiltak enn dem som er nevnt i artikkel 15 kan også bidra til å sikre passende robusthet. Også her vil risikostyringssystemer som nevnt i artikkel 9 kunne bidra til å redusere og korrigere utilsiktede risikoer. En proaktiv utvikling og testing av systemet og sikkerhetstiltak vil potensielt kunne avdekke problemer før de oppstår. Sistnevnte kan ytterligere styrkes ved regler om menneskelig tilsyn.

4.3.4 Cybersikkerhet

I tillegg til motstandsdyktighet i relasjon til robusthet, er det også viktig å sikre KI-systemer mot sårbarheter som kan utnyttes av tredjeparter. Dette er forsøkt regulert i forslaget artikkel 15 nr. 4. Fortalens punkt 51 understreker at cybersikkerhet spiller en avgjørende rolle i dette. Målet er å sikre at KI-systemene er motstandsdyktige mot forsøk på å endre systemets bruk, oppførsel, ytelse eller svekke dets sikkerhetsegenskaper.

Cybersikkerhet omfatter alt som er relatert til beskyttelsen av data, herunder personopplysninger, myndighets- og bransjeinformasjon og intellektuell eiendom, mot cyberangrep.³¹⁸ For at samfunnet skal kunne fungere er vi avhengige av kritisk infrastruktur som sykehus, finansielle tjenesteprogrammer og kraftverk. På individnivå kan cyberangrep føre til skade på individets liv og helse gjennom identitetstyveri og utpressingsforsøk. Eksponering av informasjon som lagres om og hos organisasjoner og bedrifter, myndigheter, militæret og andre samfunnskritiske organisasjoner kan skade både innbyggernes tillit til institusjonene og bedriftene, bedrifters konkurranseevne og personlige omdømme.

Cyberangrep på KI-systemer kan i tillegg utnytte KI-spesifikke eiendeler.³¹⁹ Datasett (en organisert samling av data³²⁰) som tas i bruk av systemet og treningsdatasett som brukes til treningen av systemet kan bli offer for dataforgiftning ved at informasjon i datasettene endres, fjernes eller liknende.³²¹ KI-systemets modell kan bli endret eller på annen måte kompromittert. I tillegg er underliggende infrastruktur, både programvare («software») og maskinvare («hardware») potensielt sårbar for angrep. Angrep som dette kan åpne for endret systemadferd og føre til at systemet tar andre beslutninger enn tiltenkt, begår flere feil, utvikler seg til å ta diskriminerende beslutninger eller i verste fall slutter å virke helt.³²²

4.3.5 Pålitelighet og reproduserbarhet

For å sikre tilliten til KI-systemene er det essensielt at systemene er både reproduerbare og pålitelige. Ifølge EUs etiske retningslinjer for KI er et KI-system pålitelig når det oppfører seg som forventet og tiltenkt på et varierende utvalg av situasjoner.³²³ Dette kravet gir som vi har vist føringer for tolkningen av de øvrige kravene. Et nøyaktig, robust og sikkert KI-system vil oppføre seg som det skal uansett hvilke utfordringer det stilles overfor.

Videre er reproduserbarhet viktig for å kunne vurdere påliteligheten til et KI-system. Dersom oppførselen til et KI-system og systemets resultat kan gjenskapes dersom det gjentas under de

³¹⁸ Visma (u.å).

³¹⁹ Jf. fortalens punkt 51 og High-Level Expert Group on Artificial Intelligence (2019) s. 16.

³²⁰ Digitaliseringsdirektoratet (u.å).

³²¹ Bull (2021).

³²² High-Level Expert Group on Artificial Intelligence (2019) s. 16.

³²³ High-Level Expert Group on Artificial Intelligence (2019) s. 17.

samme forholdene blir det mulig å beskrive hva KI-systemet gjør.³²⁴ Dette kan videre bidra til brukernes rett til informasjon og begrunnelse og sikre økt tillit til systemet.

4.4 Rettferdige, riktige og ikke-diskriminerende avgjørelser

4.4.1 Bakgrunn

Bruken av kunstig intelligente systemer i (offentlig) beslutningstaking drives gjerne frem av idéen om at «objektive» maskiner utelukkende treffer beslutninger med grunnlag i faktum, og at de ikke kan diskriminere eller la seg overstyre av menneskelige følelser. Det er imidlertid bevist at kunstig intelligente systemer kan «ervert» og opprettholde menneskelig bias i beslutningsprosessen, dersom datagrunnlaget som systemet treffer beslutning ut fra inneholder menneskelig *bias*.³²⁵

Det finnes ulike former for bias. Fra et statistisk perspektiv vil en estimator være «biased» når den inneholder systematiske feil som gjør at den ikke konvergerer til den sanne verdien som den prøver å estimere. Hos mennesker kan bias bestå i avvikende oppfatninger eller dømmekraft, som kan føre til at vi forholder oss ulikt til mennesker på bakgrunn av deres tilhørighet til bestemte grupper. Dersom kunstig intelligente systemer ikke kontrolleres for bias, kan de komme til urettferdige og diskriminerende resultater.³²⁶

Det følger av Regjeringens nasjonale strategi for kunstig intelligens at kunstig intelligente systemer skal legge til rette for *inkludering, mangfold og likebehandling*. Det fremheves her at det er viktig å være oppmerksom på at kunstig intelligens kan bidra til *inkludering og likestilling*, og at diskriminering unngås, ved utvikling og bruk av kunstig intelligente systemer. Det vises videre til at identifiserbar og diskriminerende bias bør fjernes i innsamlingsfasen, og at slik bias kan motvirkes ved å innføre kontrollprosesser som analyserer og korrigerer systemets beslutninger i lys av systemets formål.³²⁷

Rettferdighet og ikke-diskriminering inngår videre i ett av de syv nøkkelvilkårene i de etiske retningslinjene for pålitelig kunstig intelligens, utarbeidet av EUs ekspertgruppe for kunstig intelligens.³²⁸ Vilrårene er anerkjent av EU-kommisjonen.³²⁹ Retningslinjene er ikke bindende, men flere av vilrårene følger allerede av annen bindende EU-lovgivning.

Det er en viktig forutsetning for rettssikkerhet at avgjørelser som blir truffet fra det offentlige side, er rettferdige og riktige. I det ligger blant annet at grunnlaget for avgjørelsene må følge av rettsregler, og ikke av vilkårlige vurderinger. Behandlingen av mennesker må heller ikke krenke deres menneskerettigheter. Kravet om rettssikkerhet har nær sammenheng med grunnleggende menneskerettigheter, som forbudet mot usaklig og uforholdsmessig forskjellsbehandling, legalitetsprinsippet og myndighetenes sikringsplikt.

Forslaget til forordning om kunstig intelligens inneholder ingen bestemmelse som eksplisitt stiller krav til at høyrisiko systemer må anvendes på en rettferdig og ikke-diskriminerende

³²⁴ High-Level Expert Group on Artificial Intelligence (2019) s. 17.

³²⁵ Songül (2018).

³²⁶ Ibid.

³²⁷ Nasjonal strategi for kunstig intelligens, pkt. 5.2.

³²⁸ High-Level Expert Group on Artificial Intelligence (2019).

³²⁹ COM (2019) 168.

måte. Den mangelfulle reguleringen med hensyn til rettferdighet og ikke-diskriminering har blitt påpekt og problematisert av mange ulike aktører.

Utvalget ønsker å bygge på de europeiske retningslinjene for pålitelig kunstig intelligens og den nasjonale strategien for kunstig intelligens. Det må gjelde et grunnleggende prinsipp om at kunstig intelligente systemer skal treffe rettferdige, riktige og ikke-diskriminerende avgjørelser.

4.4.2 Nærmere om problemstillingen

Kunstig intelligente systemer som bruker maskinlæring er ikke avhengige av *direkte uttrykk for informasjon*, det vil si at et menneske identifiserer de spesifikke trinnene i en algoritme, slik at algoritmen kan finne frem til riktig løsning. Maskinlæring kjennetegnes nettopp av at systemet får *implisitt informasjon* fra store mengder data, og på grunnlag av dette selv finner frem til løsningen. Hvis systemet baserer sine løsninger på et tilstrekkelig bredt og balansert datasett, vil det være større sjanser for at løsningene gir et riktig resultat. Et system som skal gjenkjenne katter og hunder på bilder må gå gjennom mange bilder av ulike katter og hunder, før det kan begynne å gi nøyaktige svar.

Det er likevel ingen automatikk i at store datasett omfatter alle variasjoner og nyanser i materialet som systemet anvender for å finne løsninger. Hvor *dekkende* datagrunnlaget er, vil bero på hva slags data man har tilgjengelig. Hvis hvite katter ikke finnes i systemets treningsdata, vil systemet heller ikke kunne identifisere hvite katter.

Løsningene som et kunstig intelligent system kommer til kan være «biased» som følge av uheldige påvirkninger av systemet, enten som følge av ubalansert eller snevert treningsdatagrunnlag, eller som følge av at *datakilden selv* er «biased» i relasjon til etnisitet, kjønn eller andre egenskaper. Et illustrerende eksempel er de unøyaktige resultatene til ansiktsgjenkjenningssystemer ved identifisering av mørkhudede kvinner, som følge av at det ikke ble brukt en tilstrekkelig mengde med bilder av denne befolkningsgruppen i systemets treningsprosess.³³⁰ Datagrunnlaget var ikke tilstrekkelig bredt og balansert, og denne mangelen resulterte i uriktige og diskriminerende avgjørelser. Det kan også være bias i datasett som reflekterer tidligere løsninger og standpunkter, som ikke nødvendigvis er rettferdige. Bias kan følgelig ikke bare føre til at et kunstig intelligent system gir *uriktige* løsninger, det kan også gi grunnlag for diskriminering av minoriteter og andre befolkningsgrupper. Systemet vil fungere som et *speil av samfunnet*, og replisere bias som resultat. Et system som lærer av mennesker, lærer også av menneskelige feil.

Et kunstig intelligent system som skal ta beslutninger *for enkeltmennesker* må trenes på relevant data *om enkeltmennesker*. I mange tilfeller vil det være tale om personopplysninger. Aktører som behandler personopplysninger, skal sikre at opplysningene er korrekte. Det kan være utfordrende å kvalitetssikre disse når data samles inn fra mange ulike kilder, og tilnærmet umulig når man tar i betraktning datamengden som er nødvendig for å trene et kunstig intelligent system. Samtidig løper det en risiko for uriktige resultater når det finnes feil i datagrunnlaget.

4.4.3 Hva er «rettferdige» avgjørelser?

Avgjørelser truffet av kunstig intelligente systemer (som har betydning for menneskers velferdsrettigheter) skal være *rettferdige*. Forskning innen sosialpsykologien viser at opplevelsen av en rettferdig saksbehandling øker sannsynligheten for at borgerne vil akseptere

³³⁰ Panel for the Future of Science and Technology (2021).

utfall som er negative for dem.³³¹ Det er derfor viktig å klarlegge hva som ligger i begrepet «rettferdighet» i denne konteksten.

Det finnes ingen universell definisjon av begrepet «rettferdighet». Rettferdighet kan defineres som «den type forhold der mennesker behandles på en rimelig måte og i overensstemmelse med moralske prinsipper», eller «den type forhold der alle mottar det de fortjener».³³² Fra et filosofisk standpunkt er rettferdighet gjerne knyttet til hva som er *moralsk riktig*. Fra et juridisk standpunkt har rettferdighet nær sammenheng med *ikke-diskriminering* og *likebehandling*, idet rettferdighet innebærer å sikre at ingen blir urettmessig forskjellsbehandlet. Videre kan det også stilles spørsmål ved hva som utgjør en rettferdig avgjørelse. Er det et rettferdig resultat eller en rettferdig beslutningsprosess?

Begrepets kompleksitet gjør det vanskelig å formulere og gjøre forståelig for en *maskin*. Et kunstig intelligent system trenger ulike parametere for å definere en slik størrelse. En statistisk formulering av rettferdighet fordrer at det finnes en rekke kriterier som kan anvendes til å beregne graden av rettferdighet i ulike tilfeller. Når begrepet kan forstås på flere ulike måter, vil det være vanskelig å finne slike kriterier som er dekkende for alle de ulike forståelsene.

I sandkasse-prosjektet til Datatilsynet/NAV tar Datatilsynet utgangspunkt i hovedprinsippene om lovlig, etisk og sikker kunstig intelligens ved tilnærmingen til begrepet «rettferdighet». Det ses hen til rettferdighetsprinsippet i personvernforordningen. Videre fremheves det at rettferdighetsprinsippet inneholder flere elementer enn ikke-diskriminering; det står sentralt i flere andre rettighetslovverk, blant annet ulike menneskerettighetsbestemmelser og likestillings- og diskrimineringsloven. Disse lovverkene vil også kunne få betydning for forståelsen av begrepet «rettferdighet».

«En annen side er kravet om rettferdig («fair») (...) saksbehandling. Dette går blant annet på at den som blir berørt av en avgjørelse, må få reelle muligheter til å fremme sine synspunkter. Om nødvendig må den som skal treffe avgjørelsen hjelpe klienten med dette. Her kommer praktisk viktige spørsmål om forskjellige former for bistand til klienten – også i form av rettshjelp – inn i bildet. Institusjonen, eller det forvaltningsorganet som skal treffe en avgjørelse, må i det hele opptre slik at saken får en tilfredsstillende behandling, og slik at klienten også selv opplever det slik. Dette siste – at ikke bare må saksbehandlingen være betryggende, men klienten må få nok forklaring til selv å kunne forstå dette – kan være særlig viktig i sosial- og helseinstitusjoner. (...) Reglene må også utformes slik at feil blir rettet så raskt og effektivt som mulig i de tilfelle der feil tross alt forekommer.»³³³

Det virker hensiktsmessig å legge den forståelsen til grunn at en avgjørelse tatt av et kunstig intelligent system vil være rettferdig dersom den medvirker til en fordeling av ytelser slik at de som har rett på ytelsen får den, og de som ikke har rett på den ikke får den. Det er følgelig nær sammenheng mellom kravet om rettferdige avgjørelser og systemets «treffsikkerhet», som vil bli redegjort for nærmere i forbindelse med kravet til riktige avgjørelser.

³³¹ NOU 2019: 5 pkt. 10.8.1.

³³² Store Norske Leksikon (2021).

³³³ St. meld. nr. 32 (1976–77) s. 8.

4.4.4 Hva er «riktige» avgjørelser?

Avgjørelser truffet av et kunstig intelligent system skal være *riktige*. En intuitiv tolkning tilsier at det betyr at avgjørelsene skal være korrekte. Hvorvidt noe kan sies å være korrekt, beror på om det er i overensstemmelse med faktum. Det oppstår følgelig et spørsmål om hvilket «faktum» som kan tjene som sammenligningsgrunnlag for slike avgjørelser.

For det første må avgjørelsen knytte seg til riktig person, og ta i betraktning de relevante forholdene knyttet til vedkommende person. Avgjørelsen må være *treffsikker*. Denne forståelsen har sammenheng med at avgjørelsene må være *individuellt begrunnede*, det vil si at de må gjelde riktig person og riktig saksforhold. Hvorvidt systemet *kan* ta i betraktning relevante forhold, beror på kvaliteten og fullstendigheten av datagrunnlaget.

For det andre må avgjørelsen være *materielt riktig*. Flere saksbehandlingsregler i forvaltningsloven har til formål å legge til rette for materielt riktige resultater. Dette er regler som tar sikte på å opplyse saken (krav til utredning, forhåndsvarsling, aktiv informasjonsplikt og eventuelle andre regler om innhenting og foreleggelse av opplysninger), regler som setter parten i stand til å forstå bedre hvilke faktiske opplysninger som kan ha betydning i saken, og til å legge dem frem for forvaltningsorganet (bidrar til å opplyse saken). Forvaltningens veiledningsplikt, partens rett til å benytte advokat og tolk, og retten til å forklare seg muntlig i møte med en representant for forvaltningsorganet – bidrar til riktige avgjørelser.³³⁴

Andre regler søker å forhindre misbruk av posisjon eller vektlegging av usaklige hensyn under saksbehandlingen. Til denne gruppen hører inhabilitetsreglene og den tradisjonelle læren om myndighetsmisbruk; forvaltningen skal ikke ta utenforliggende hensyn, treffe vilkårlige eller sterkt urimelige avgjørelser eller bedrive usaklig forskjellsbehandling. Disse reglene tar sikte på å begrense et menneskes muligheter til å treffe uriktige avgjørelser ut fra egeninteresse, og er følgelig ikke helt treffende når det kommer til algoritmisk beslutningstaking. Det kan likevel tenkes at et kunstig intelligent system kan ha tendenser til å *favorisere* enkelte individer, ved å for eksempel legge tidligere resultater tilknyttet vedkommende til grunn i senere saker. Det vil neppe være realistisk eller praktisk gjennomførbart å kreve at et kunstig intelligent system ikke kan ta beslutninger om samme person flere ganger. Kontrollen av kunstig intelligente systemer må følgelig skje på andre måter, for eksempel ved korrigerende av bias i datagrunnlaget. Felles for de eksisterende forvaltningsrettslige reglene og slike kontrollregler for kunstig intelligens er målet om *tillit* mellom borgerne og beslutningstakerne.

Endelig finnes det regler i forvaltningsloven som legger til rette for at gale eller uheldige avgjørelser kan bli rettet opp. Både reglene om klage og omgjøring og reglene om ugyldighet hører til denne kategorien.

4.4.4.1 Hensynet til riktige avgjørelser i forordningsforslaget

Forordningsforslaget bruker begrepet «accuracy» i forbindelse med kravene til «robustness» og «cybersecurity». Begrepet oversettes til *nøyaktighet* på norsk, og er ikke direkte treffende. EUs ekspertgruppe for kunstig intelligens har uttalt at «accuracy» refererer til et kunstig intelligent systems evne til å foreta riktige vurderinger, for eksempel å klassifisere informasjon i riktige kategorier. Det omfatter også et systems evne til å gi korrekte forutsigelser, anbefalinger eller beslutninger basert på aktuell data.³³⁵ Samtidig fremheves det at en god utviklings- og evalueringsprosess kan bidra til å redusere og korrigere *utilsiktede*

³³⁴ NOU 2019: 5 s. 144-145.

³³⁵ High-Level Expert Group on Artificial Intelligence (2019).

*risikoer for unøyaktige forutsigelser.*³³⁶ En unøyaktig beslutning vil fort kunne sies å være *uriktig*, både i lys av legalitetsprinsippet og med hensyn til et materielt riktig resultat. EUs ekspertgruppe for kunstig intelligens fremhever videre at i de tilfeller unøyaktige forutsigelser ikke kan unngås, er det viktig at systemet kan indikere sannsynligheten for slike feil. Et høyt nivå av nøyaktighet er spesielt avgjørende i tilfeller hvor det kunstig intelligente systemet direkte påvirker menneskeliv.³³⁷

4.4.5 Hva er «ikke-diskriminerende» avgjørelser?

Diskriminering er en av, om ikke *den* største risikoen ved bruk av kunstig intelligente systemer i velferdsforvaltningen. Det spesielle med kunstig intelligens er at den «overtar» og «integrerer» menneskelige fordommer. Algoritmer kan gjenspeile og bidra til å opprettholde bias som eksisterer i samfunnet, ved at slik bias reflekteres i data som bearbeides av algoritmen. Et spørsmål som oppstår i denne forbindelse er: hvis slik bias allerede eksisterer blant mennesker i samfunnet, hvorfor blir det særlig problematisk ved bruk av kunstig intelligens? Det er to årsaker til dette.

For det første kan det være vanskelig å forstå hva som ligger til grunn for algoritmens endelige resultat. Når man anvender maskinlæring, gir man systemet *input* i form av store datamengder. Ved veiledet læring lar man systemet finne frem til riktig resultat ut fra eksempler. Ved forsterket læring interagerer systemet direkte med et miljø som gir «straff» eller «belønning».³³⁸ I begge tilfeller må systemet selv finne en fremgangsmåte for å komme frem til det ønskede resultatet. Menneskers kontroll over denne prosessen er begrenset til å bestemme hvilken data systemet skal basere sin beslutning på. I tillegg kan systemets fremgangsmåte for å finne frem til resultatet endres hver gang datagrunnlaget suppleres med ny data. Vi kan si at beslutningsprosessen blir en *svart boks*, og det blir umulig å finne ut av *hvorfor* systemet diskriminerer.

For det andre kan det ha større konsekvenser når menneskelig bias ytterligere forsterkes av algoritmer. Siden kunstig intelligente systemer er langt mer effektive enn mennesker, kan de anvendes i større skala. Beslutninger tatt av et kunstig intelligent system er også mer konsistente enn om samme beslutninger blir tatt av ett menneske. Det gjør at systemet har potensial til å spre bias *konsekvent*. I tillegg har kunstig intelligente systemer en tendens til å forsterke det de oppfatter som den viktigste informasjonen. Det kan føre til at systemet blir mer diskriminerende enn mennesker.

4.4.5.1 Hva er algoritmisk diskriminering?

Algoritmisk diskriminering brukes her som en betegnelse på en beslutning truffet av et kunstig intelligent system som resulterer i diskriminering av enkeltpersoner.

Diskriminering kan forstås som urettmessig forskjellsbehandling; at noen blir mindre gunstig behandlet enn andre, uten at det foreligger noen legitim grunn for det. Diskriminering brukes følgelig normalt som betegnelse for usaklig eller urimelig forskjellsbehandling på grunnlag av bestemte egenskaper (diskrimineringsgrunnlag), som kjønn, etnisitet, religion, funksjonsnedsettelse, seksuell orientering mv.³³⁹ En «motpol» til urettmessig

³³⁶ Ibid.

³³⁷ Ibid.

³³⁸ Store Norske Leksikon (2022).

³³⁹ Se likestillings- og diskrimineringsloven § 6 første ledd.

forskjellsbehandling er *rettmessig* forskjellsbehandling eller lovlig forskjellsbehandling. Dette er forskjellsbehandling med formål om å utjevne eksisterende forskjeller.

Det er flere «former» for diskriminering. Når vi snakker om *menneskelig* diskriminering, kan diskriminering oppstå ved at i utgangspunktet like tilfeller blir behandlet ulikt, eller ved at handlinger baseres på bestemte egenskaper og får negative konsekvenser. Menneskelig diskriminering kan i tillegg ha form av bevisste handlinger med tiltenkte uheldige konsekvenser for en person basert på vedkommendes egenskaper. Det omfatter også krenkende, sårende eller trakasserende ytringer. Kunstig intelligente systemer mangler evnen til å diskriminere bevisst, da de ikke har noen bevissthet. Resultatene truffet av slike systemer kan likevel ha form av diskriminerende ytringer, for eksempel hvis en chatbot kommer med rasistiske svar.

Menneskelig diskriminering kan være motivert av fordommer overfor bestemte grupper mennesker, ved at disse uriktig tillegges negative egenskaper. Menneskelig diskriminering kan også være motivert av stereotypier, som innebærer at enkeltpersoner tillegges egenskaper ut fra *antagelser* basert på tilhørighet til en gruppe, og ikke deres faktiske egenskaper. Slik diskriminering er normalt ikke rasjonelt begrunnet.³⁴⁰

Selv om kunstig intelligente systemer i stor grad er autonome, er det en forutsetning at mennesker er involvert på et eller annet stadium i systemets livssyklus. Kunstig intelligente systemer kan per i dag ikke *skape seg selv*. Når det gjelder systemer som fungerer ved hjelp av maskinlæring, vil det være mennesker som forbereder og merker relevant treningsdata og gir algoritmen tilbakemeldinger underveis i læringsprosessen. I tillegg er det mennesker som bestemmer hvilke data som skal tilføres datagrunnlaget når systemet er i bruk, hva systemets resultater skal brukes til, og om systemet overhodet fungerer som tiltenkt. Mennesker spiller følgelig en betydelig rolle også i algoritmisk beslutningstaking.

Mange mennesker stoler på kunstig intelligente systemer nettopp fordi det er et menneske «i loopen». Samtidig er det viktig å være bevisst på at menneskelig involvering medfører en risiko for diskriminerende kunstig intelligente systemer. Det er et ubestridt faktum at mennesker kan ta irrasjonelle og ulogiske beslutninger, eller la seg påvirke av irrelevante forhold. Menneskelige utilstrekkeligheter, inkludert diskriminerende holdninger, kan overdras til kunstig intelligente systemer, for eksempel ved å programmere systemet til å vektlegge en bestemt egenskap mer enn andre, eller ved å unnlate å trene systemet på et tilstrekkelig representativt og balansert datagrunnlag. På tilsvarende måte kan menneskeavledet data som systemet mates med under opplærings- og anvendelsesstadiene også være biased, for eksempel ved å gjenspeile diskrimineringsmønstre. Dette kan føre til algoritmisk diskriminering, ved at det kunstig intelligente systemet *reflekterer sine skaperes verdier*.³⁴¹

Et annet aspekt ved algoritmisk diskriminering er at noen mennesker kan ha en tendens til å stole på beslutninger truffet av kunstig intelligente systemer basert på en overbevisning om at systemet «vet best». I situasjoner hvor kunstig intelligente systemer skal assistere mennesker i beslutningstakingsprosessen, kan dette føre til at diskriminerende resultater som systemet genererer ikke blir overprøvd. Det kan også føre til såkalt «ankring», hvor mennesket som skal treffe den endelige beslutningen ufrivillig forankrer den i resultatet som systemet kom til – selv om den endelige beslutningen kunne ha blitt annerledes hvis vedkommende ikke var kjent med systemets resultat.

³⁴⁰ NOU 2009: 14.

³⁴¹ På engelsk «reflects the values of their creators». European Commission, Directorate-General for Justice and Consumers (2021).

4.4.5.2 Hensynet til ikke-diskriminering i forordningsforslaget

Forslaget til forordning om kunstig intelligens erkjenner at kunstig intelligente systemer kan brukes på en måte som strider med retten til ikke-diskriminering.³⁴² Klassifikasjonen av ulike typer systemer avhenger av hvilken risiko systemet innebærer for grunnleggende rettigheter, herunder retten til ikke-diskriminering.

Når det gjelder kunstig intelligente systemer til bruk i velferdsforvaltningen, understrekes det at personer som søker om eller mottar offentlige ytelser typisk er avhengige av disse tjenestene og befinner seg i en sårbar posisjon i møte med myndighetene. En beslutning om hvorvidt slike ytelser skal innvilges eller ikke, reduseres eller opphøre vil ha en betydelig innvirkning på vedkommendes livssituasjon. Bruk av kunstig intelligente systemer i denne sammenhengen medfører derfor en vesentlig risiko for diskriminering – både fordi disse menneskene ofte tilhører minoriteter i samfunnet, og fordi et eventuelt diskriminerende resultat vil ramme dem spesielt hardt. Slike systemer klassifiseres derfor som høyrisiko i forordningsforslaget.³⁴³

Forordningsforslaget inneholder ingen bestemmelser som definerer eller handler om algoritmisk diskriminering. Selv om forordningsteksten inneholder få eksplisitte referanser til begrepet, er det tatt hensyn til å forhindre diskriminering i de ulike bestemmelsene som stiller krav til høyrisiko systemer. Dette er krav til design og kvalitet på datasett som brukes ved utviklingen av kunstig intelligente systemer, om testing, risikostyring, dokumentasjon og menneskelig tilsyn gjennom systemets livssyklus.³⁴⁴

Høy datakvalitet er avgjørende for ytelsen til kunstig intelligente systemer som bruker maskinlæring.³⁴⁵ Det stilles krav til at datasett som brukes til trening, validering og testing skal være *representative*. Slike datasett må også hensynta egenskaper og forhold som er spesifikke for de *geografiske, adferdsmessige eller funksjonelle omgivelsene* hvor det kunstig intelligente systemet skal anvendes.³⁴⁶

Den tekniske dokumentasjonen for et høyrisiko kunstig intelligent system skal blant annet inneholde en detaljert beskrivelse av systemets potensielt diskriminerende virkninger.³⁴⁷ Den skal også inneholde detaljert informasjon om hvordan systemet overvåkes og kontrolleres, med hensyn til blant annet de forutsigbare utilsiktede utfallene og kildene til risiko for diskriminering.³⁴⁸

Det skal i tillegg være mulig for *brukere av systemet* å forstå om systemet potensielt kan ha diskriminerende virkninger.³⁴⁹ Bruksanvisningen til systemet skal blant annet inneholde informasjon om enhver kjent eller forutsigbar omstendighet ved bruk av systemet som kan føre til risiko for diskriminering.³⁵⁰

³⁴² COM(2021) 206 fortalepunkt 15.

³⁴³ Ibid. fortalepunkt 37.

³⁴⁴ Ibid. pkt. 1.2.

³⁴⁵ Ibid. fortalepunkt 44.

³⁴⁶ Ibid. art. 10.

³⁴⁷ Ibid. art. 11 og vedlegg IV nr. 2 (g).

³⁴⁸ Ibid. vedlegg IV nr. 3.

³⁴⁹ Ibid. art. 13.

³⁵⁰ Ibid. art. 13 nr. 3 (b) (iii). Forordningsteksten nevner ikke eksplisitt risiko for «diskriminering», men bruker isteden begrepet «fundamental rights». Retten til ikke-diskriminering er en slik «fundamental right».

Kravet om menneskelig tilsyn skal gjøre det mulig for vedkommende som fører tilsyn med systemet å være bevisst på tendensen til å automatisk stole på resultatene generert av et kunstig intelligent system.³⁵¹ Dette gjelder særlig for systemer som brukes til å gi informasjon eller anbefalinger vedrørende beslutninger som skal tas av fysiske personer. Her siktes det trolig til situasjonen hvor mennesker i større grad kan ha ukritisk tillit til kunstig intelligente systemer begrunnet i at systemet «vet best». Det er imidlertid uklart hvordan kravet om menneskelig tilsyn skal bidra til å redusere denne risikoen.

4.4.5.3 Er eksisterende ikke-diskrimineringslovgivning egnet til å regulere algoritmisk diskriminering?

Forordningsforslaget er ment å være i overensstemmelse med og komplimentere EUs Charter om grunnleggende rettigheter og sekundær lovgivning om ikke-diskriminering og likestilling.³⁵² Charteret samsvarer med den europeiske menneskerettighetskonvensjonen når det gjelder rettigheter som har sitt opphav i konvensjonen.³⁵³ Det forutsettes derfor at forordningsforslaget er i overensstemmelse med Norges EU-rettslige menneskerettsforpliktelser. Annen relevant EU-lovgivning på området er EUs likebehandlingsdirektiver, som enten er bindende i kraft av EØS-avtalen eller frivillig implementert i norsk lovgivning.

Unionsrådet foretok i 2020 en vurdering av Charteret i lys av digitaliseringen og utviklingen av kunstig intelligens.³⁵⁴ Det ble her uttalt at kunstig intelligens kunne utfordre de grunnleggende rettighetene på grunn av enkelte systemers kompleksitet og opasitet, ettersom forståelse og kontroll av slike systemer krever spesialisert ekspertise.³⁵⁵ Det er med andre ord vanskelig å håndheve en rettighet hvis man ikke kan forstå om den overhodet har blitt krenket. Rådet støttet EU-kommisjonens forslag om å utarbeide krav til kunstig intelligente systemer. Rådet oppfordret videre medlemsland i EU om å vurdere hvorvidt eksisterende nasjonale rettsregler var tilstrekkelige til å utnytte mulighetene og håndtere potensielle risikoer knyttet til bruken av digital teknologi og kunstig intelligens.³⁵⁶

I 2021 publiserte EU-kommisjonen en rapport om algoritmer og diskriminering, hvor hovedtema er hvordan økende bruk av algoritmer på alle samfunnsområder skaper utfordringer relatert til diskriminering. Rapporten peker på at algoritmisk diskriminering risikerer å ikke fanges opp av EUs likestillings- og ikke-diskrimineringslovgivning på grunn av blant annet eksisterende hull i lovgivningens materielle omfang, mangel på fleksibilitet i personlig omfang og håndhevingsutfordringer.³⁵⁷

Norsk likestillings- og ikke-diskrimineringslovgivning gir et bedre vern enn EU-lovgivningen, men de samme utfordringene gjør seg likevel gjeldende. Selv om algoritmen ikke diskriminerer direkte på bakgrunn av et diskrimineringsgrunnlag, kan den opprette såkalte *proxy-variabler* som har veldig nær sammenheng med diskrimineringsgrunnlagene, og følgelig basere sine resultater på et uklart eller ubeskyttet grunnlag. Et spørsmål som

³⁵¹ COM(2021) 206 art. 14 nr. 4 (b).

³⁵² Ibid. pkt. 1.2.

³⁵³ European Commission (u.å.).

³⁵⁴ Council of the European Union (2020).

³⁵⁵ Ibid. pkt. 11.

³⁵⁶ Ibid., se pkt.12.

³⁵⁷ European Commission, Directorate-General for Justice and Consumers (2021) s. 152.

aktualiserer seg i denne sammenhengen er om det kan være behov for vern av andre grupper enn de som omfattes av dagens lovgivning.

I tillegg kan det være vanskelig å oppdage at diskriminering har funnet sted på grunn av kunstig intelligente systemers komplekse beslutningsprosess. Hvis slike systemer diskriminerer, vil det gjerne være på en indirekte, subtil og usynlig måte. Når man ikke vet eller forstår *hvordan* eller *hvorfor* systemet har kommet frem til et bestemt resultat, kan man heller ikke fastslå at det foreligger diskriminering. Et viktig virkemiddel for å motvirke diskriminering i algoritmisk beslutningstaking er kravet til transparens.

4.4.5.4 Kort om digitalt utenforskap

En annen side av diskriminering er *digitalt utenforskap*. Ca. 14 prosent av den norske befolkningen over 16 år er *ikke-digitale*.³⁵⁸ Det betyr at denne delen av befolkningen, av ulike årsaker, ikke klarer å holde følge med den digitale utviklingen, herunder omleggingen av viktige offentlige tjenester til digitalt format. De bruker for eksempel ikke internett eller digitale verktøy (som smarttelefon, datamaskin eller nettbrett), mangler grunnleggende digitale ferdigheter eller er rett og slett *ikke interesserte* i å bli digitale. Personer over 60 år og personer utenfor arbeid og utdanning er mer utsatt for digitalt utenforskap enn den øvrige befolkningen.³⁵⁹

Digital utenforskap knytter seg ikke til diskriminerende *avgjørelser* som sådan, men er fremdeles et viktig tema som bør tas i betraktning i forbindelse med innfasingen av kunstig intelligent teknologi i velferdsforvaltningen. På den ene siden kan kunstig intelligente systemer gjøre det *lettere* å inkludere den gruppen mennesker som i dag faller utenfor digitaliseringsprosessen. En automatisering av ulike prosesser, for eksempel vurdering av om man er kvalifisert for trygdeytelser, kan involvere disse menneskene i den digitale beslutningsprosessen uten at de trenger å foreta seg noe selv. Et system som fanger opp når en person blir arbeidsledig kan for eksempel automatisk registrere en søknad om dagpenger og utarbeide relevant informasjon, som kan skrives ut og sendes med brev til personen det gjelder. På den andre siden er det vanskelig å tenke seg at slike beslutninger kan automatiseres i så stor grad uten at den aktuelle personen er informert eller konsultert om det på forhånd – i hvert fall slik regelverket er i dag. Tilgangen til kunstig intelligent saksbehandling vil derfor trolig forutsette at man møter opp fysisk på et NAV-kontor, eller at man bruker digitale verktøy (hvor man også kan informeres om den automatiserte prosessen og samtykke til behandling av personopplysninger).

Et annet spørsmål som oppstår i forbindelse med digitalt utenforskap er hvorvidt det vil være mulig å bruke data knyttet til denne «utenforstående» delen av befolkningen i utviklingen av kunstig intelligente systemer i velferdsforvaltningen. Hvis slik data av ulike årsaker ikke blir inkludert i datagrunnlagene – for eksempel fordi det ikke foreligger samtykke til bruk, eller fordi saker relatert til denne gruppen blir behandlet «utenom den normale prosessen» og derfor ikke registreres på samme måte – risikerer man at over 1/10 av befolkningen ikke blir representert. Det kan diskuteres om velferdsforvaltningen bør utvikle kunstig intelligente systemer til å ta beslutninger som berører den norske befolkningen når systemet utvikles uten å involvere *hele* befolkningen.

4.4.5.5 Prinsippets føringer for norsk velferdsforvaltning

Det er viktig å sikre at avgjørelser som blir tatt ved bruk av kunstig intelligens ikke fører til at noen blir utsatt for diskriminering. Grunnlaget for behandlingen bør derfor testes på forhånd

³⁵⁸ Kompetanse Norge (2021).

³⁵⁹ Ibid.

for diskriminerende faktorer. Det gjelder særlig bruken av algoritmer og de data man bygger disse på°. Den enkelte bør også° ha mulighet til å klage på° avgjørelser i ettertid, eventuelt en mulighet til å be om at den automatiserte avgjørelsen blir gjenstand for menneskelig overprøving.³⁶⁰

Datagrunnlaget bør i enkelte tilfeller kunne korrigeres for å representere den relevante brukergruppen *slik den burde være*, og ikke *slik den er*. Selv om datagrunnlaget samsvarer med virkeligheten, er det ingen garanti for at det i virkeligheten ikke oppstår noen form for diskriminering eller ufordelaktig forskjellsbehandling. Slike holdninger forsterkes ytterligere av kunstig intelligente systemer, og kan medføre uheldige resultater. Det er derfor viktig å aktivt sørge for at datagrunnlaget er representativt og mangfoldig.

4.5 Personvern

Personvernet er blitt en av de best etablerte grunnleggende rettighetene i det digitale rom. Rettigheten har røtter i retten til respekt for privatliv og har i Europa fått et sterkt rettsvern gjennom vedtakelsen av GDPR (personvernforordningen). Forordningen har i stor grad styrket individers personvernrettigheter og skjerpet kravene til behandling av slike personlig data. Men samtidig som kravene stilles blir intensivert, øker etterspørselen etter data med fremveksten av KI-systemer som i hovedsak kun kan bli intelligente dersom de har nok data å lære av.³⁶¹

Nøyaktigheten av resultatene til KI-basert systemer avhenger med andre ord av mengden relevant data som systemet har tilgang til. Personvernforordningen stiller krav til behandlingen av disse dataene som opprinnelig oppgis av brukeren. En av problemstillingene knyttet til dette forholdet ble satt på spissen i Datatilsynets sandkasseprosjekt med NAV. Sluttrapporten konkluderer at selv om NAV kunne ha rettslig grunnlag for å behandle personopplysninger ved anvendelse av KI i den ønskede situasjonen var det tvilsomt om det rettslige grunnlaget også omfattet bruk av personopplysninger til opplæring av KI-systemet ved utvikling av en prediksjonsmodell.³⁶²

I tillegg genererer KI-systemer en stor mengde data i løpet av brukernes interaksjon med systemet.³⁶³ Dette kan være data som presenteres for brukeren, herunder avgjørelser, vedtak eller anbefalinger for veien videre, eller data om hvordan brukeren reagerer og møter KI-systemet som for eksempel underveisvurderinger om sannsynlig eller faktisk oppførsel knyttet til beslutninger eller liknende som tas at systemet.

Personvern som for kravet om teknisk robusthet og sikkerhet, nært knyttet til prinsippet om forebygging av skade.³⁶⁴ Dette betyr at det stilles krav ikke bare til behandlingen av personlig data, men også beskyttelsen. Blant annet stiller GDPR artikkel 32 krav til implementeringen av cybersikkerhet.

I tillegg krever forebygging av skade på personvernet som nevnt i punkt 4.2 også tilstrekkelig forsvarlig dataforvaltning.

³⁶⁰ Diskrimineringsnemnda v/ Kaland (2021).

³⁶¹ Datatilsynet (2018) s. 5.

³⁶² Datatilsynet (2022) pkt. 4.2.4.

³⁶³ High-Level Expert Group on Artificial Intelligence (2019) s. 17.

³⁶⁴ L.c.

Personvernforordningen alene og dets forhold til kunstig intelligente systemer er i stor grad blitt behandlet i litteraturen.³⁶⁵ Utvalget anerkjenner at kravet har en sentral rolle for velferdsforvaltningen, men henviser til tidligere og kommende litteratur på området for en nærmere drøftelse av kravet.

4.5.1 Nærmere om personvernforordningens krav ved automatiserte avgjørelser

Forordningsforslaget om regler for kunstig intelligens stiller først og fremst krav til utviklere og tilbydere av kunstig intelligente systemer. Forslaget gir ingen uttrykkelige rettigheter til enkeltindivider som slike systemer kan ha en innvirkning på. Slike rettigheter finnes imidlertid i personvernforordningen for automatiserte avgjørelser som bygger på personopplysninger. Når forvaltningen skal anvende kunstig intelligente systemer i saksbehandling og beslutningsprosesser, vil betydelige deler av datagrunnlaget bestå av personopplysninger. Det er derfor grunn til å redegjøre nærmere for de aktuelle reglene.

Det følger av Personvernforordningen artikkel 22 at «[d]en registrerte skal ha rett til ikke å være gjenstand for en avgjørelse som utelukkende er basert på automatisert behandling, herunder profilering, som har rettsvirkning for eller på tilsvarende måte i betydelig grad påvirker vedkommende». Avgjørelser som *utelukkende* er basert på automatisert behandling er avgjørelser som er tatt helt uten menneskelig involvering.

Det følger videre av artikkel 2 nr. 2 at dette ikke gjelder dersom avgjørelsen er nødvendig for inngåelse eller oppfyllelse av avtale, tillatt i henhold til nasjonal rett eller er basert på den registrertes uttrykkelige samtykke. Når det gjelder samtykke, følger det av artikkel 4 nr. 11 at det blant annet må være *frivillig avgitt* («freely given»). Kriteriet gir anvisning på en vurdering av den registrertes reelle muligheter og kontroll i den aktuelle situasjonen. Hvis den registrerte i realiteten ikke har noe valg enn å samtykke eller føler seg presset til det, vil samtykket ikke anses for å være frivillig avgitt. (Guidelines 05/2020) Dette vil typisk være tilfelle i situasjoner hvor det er et skjevt maktforhold mellom partene.

Personvernforordningens foralepunkt 43 gir klart uttrykk for at offentlige myndigheter trolig ikke vil kunne benytte samtykke som behandlingsgrunnlag, nettopp på grunn av ubalansen mellom partene.

På bakgrunn av dette har det oppstått et behov for offentlige myndigheter til å etablere nasjonale lovhjemler som kan gi grunnlag for automatiserte avgjørelser etter artikkel 22 nr. 2 bokstav b. Et eksempel på en slik lovhjemmel er NAV-loven § 4 a andre ledd, som åpner for at Arbeids- og velferdsetaten kan treffe avgjørelser som utelukkende er basert på automatisert behandling av personopplysninger, herunder sensitive opplysninger som nevnt i personvernforordningen artikkel 9 og 10. Den automatiserte avgjørelsen kan være både et enkeltvedtak og en annen beslutning som «i betydelig grad påvirker» en person.³⁶⁶ Bestemmelsens forarbeider fremhever at det ikke er ønskelig å åpne for automatisering av *alle* typer avgjørelser, og at det derfor oppstilles et krav om at behandlingen må være «forsvarlig». Det innebærer at det må foretas en vurdering av hvorvidt saksbehandlingen (inkludert de aktuelle lovhjemplene) er *egnet for helautomatisering*. Her vil det blant annet være relevant hvor enkel lovteksten er å fortolke, herunder hvor objektive og tydelige vilkårene i

³⁶⁵ Se for eksempel: Jarbekk (2019) eller Kuner (2020) om GDPR; Bygrave (2019), Hallinan (2021) eller Kerrigan (2022) om forholdet mellom personvern og kunstig intelligens; Ohnstad (2021) om forholdet mellom personvern og velferd; Choroszewicz (2020) om digital velferdsstat.

³⁶⁶ Prop. 135 L (2019–2020) s. 20.

bestemmelsen er, og om mer skjønnspregede bestemmelser lar seg operasjonalisere.³⁶⁷ Videre må det være mulighet for *manuell korrigerings* der opplysningene som ligger til grunn for den automatiserte behandlingen er feil. Her er det imidlertid verdt å merke seg at manuell korrigerings kan medføre at behandlingen ikke vil omfattes av personvernforordningens artikkel 22. Hvis et menneske foretar ytterligere vurderinger i beslutningsprosessen, vil det ikke være en utelukkende automatisert avgjørelse. Kriteriet for å konstatere menneskelig involvering er at vedkommende må ha autoritet og kompetanse til å endre beslutningen. (WP251 s. 21) Det er derfor viktig å avklare omfanget av en eventuell korreksjon etter bestemmelsen i § 4 a andre ledd.

Kravet til forsvarlighet vil videre sette grenser for bruk av helautomatiserte avgjørelser hvor det ikke er mulig å operasjonalisere skjønnnet. Bestemmelsen nevner også eksplisitt at den helautomatiserte avgjørelsen ikke kan bygge på skjønnsmessige vilkår i lov eller forskrift, med mindre avgjørelsen er utvilsom. Samtidig påpeker Departementet at en skjønnsmessig bestemmelse ikke nødvendigvis vil medføre en skjønnsmessig vurdering i NAV. Som eksempel nevnes folketrygdlovens regler om sykepenger, som forutsetter en konkret vurdering av om vilkårene for sykepenger er oppfylt. I slike tilfeller vil det kunne være tilstrekkelig med attest fra lege med gyldig diagnose for å fastslå at vilkårene er oppfylt – uten at NAV trenger å foreta noen ytterligere vurderinger. Slike *utvilsomme eller kurante avgjørelser* bør, ifølge Departementet, kunne helautomatiseres. Kriteriet om at avgjørelsen må være utvilsom refererer ikke til hvorvidt *systemet* vil være i tvil, men om utfallet av den skjønnsmessige vurderingen av vilkåret ville vært utvilsomt dersom vurderingen ble foretatt av en person. Det kan likevel tenkes tilfeller hvor vilkårene for en ytelse prinsippet er oppfylt (eller ikke oppfylt), men en nærmere vurdering av grunnlaget tilsier at resultatet burde bli annerledes. Det reiser spørsmål om hvem som til syvende og sist vil ha ansvar for å bestemme hvorvidt en avgjørelse er «utvilsom» eller ikke.

Det følger av forarbeidene til NAV-loven at det på nåværende tidspunkt ikke er aktuelt «å underlegge et krav om en ytelse som inneholder kompliserte vilkår maskinell behandling med såkalt intelligent beslutningsevne», hvor det er større fare for urettmessig og utilsiktet diskriminering. Departementet viser til at «en slik eventuell omlegging av saksbehandlingen [må] underlegges en mer helhetlig gjennomgang og konsekvensanalyse».

Personvernforordningen artikkel 22 nr. 2 bokstav b stiller videre krav om at det må fastsettes egnede tiltak for å verne den registrertes rettigheter, friheter og berettigede interesser. Det følger av personvernforordningens foralepunkt 71 at slike tiltak i det minste bør inkludere en rett for den registrerte til å kreve menneskelig inngripen, til å få forklaring på avgjørelsen som er truffet, til å gi uttrykk for sitt syn og til å motsette seg avgjørelsen.³⁶⁸ Forvaltningslovens regler om veiledningsplikt, kontradiksjon, innsyn, begrunnelse og klage må i utgangspunktet anses for å være et vesentlig bidrag til å oppfylle kravet.³⁶⁹ Det er imidlertid viktig å merke seg at garantiene i forvaltningsloven er *minimumsgarantier*, og at tiltakene må vurderes konkret for den aktuelle behandlingen. Det innebærer at det i enkelte tilfeller kan tenkes at en forsvarlig saksbehandling ikke vil være tilstrekkelig for å redusere risikoen som en automatisert behandling innebærer, og at forvaltningen må stille ytterligere garantier ved slik

³⁶⁷ L.c.

³⁶⁸ Skullerud (2022).

³⁶⁹ Prop. 135 L (2019–2020) s. 22.

behandling. Departementet åpner i forarbeidene til NAV-loven for å forskriftsregulere dette.³⁷⁰

Forarbeidene til NAV-loven forutsetter at det foreligger tilstrekkelige kvalitetssikringsmekanismer for å sikre at opplysninger som ligger til grunn for avgjørelsen er *riktige og dekkende*.³⁷¹ Det er uklart hva som er ment å ligge i dette. Når det gjelder automatiserte avgjørelser som har vesentlig betydning for den enkelte, i henhold til kravet i personvernforordningen artikkel 22 nr. 1 om «rettsvirkninger» eller «tilsvarende påvirkning», kan det være grunn til å stille strengere krav til kvaliteten på opplysningene. Det kan være nødvendig å kryssjekke informasjon fra ulike kilder før den tilføres datagrunnlaget for avgjørelsen, eller verifisere informasjonen på en annen måte. Strengt krav til datakvaliteten kan føre til at forvaltningen må bruke mer tid og ressurser på å etablere et tilstrekkelig godt datagrunnlag for maskinlæringsystemer. Et annet spørsmål som oppstår i den forbindelse er hvor langt forvaltningen skal strekke seg for å forsikre seg om at informasjonen er riktig og dekkende. Vil effektivitetsfordelen som følge av automatisering av beslutningsprosesser spises opp av ressursene som må brukes på å oppfylle de rettslige vilkårene for slike beslutningsprosesser?

Departementet åpner for at det kan være nødvendig å regulere krav til jevnlig kontroll og revisjon av de benyttede reglene og algoritmene, med formål om å begrense «faren for feil». Dette synspunktet har også støtte fra Teknologirådet, som mener at algoritmer bør revideres på samme måte som bedrifter må revidere sitt regnskap.³⁷² Synspunktet er fremmet i forbindelse med EU-forslaget til en forordning om digitale tjenester. Ifølge Teknologirådet bør dette gjelde alle systemer der maskinlærings og algoritmer tas i bruk på en måte som innebærer risiko for brukerne eller samfunnet. Samtidig presiseres det at forhåndsgodkjennelse av alle algoritmer «ville strupe utvikling og nødvendig effektivisering». Det fordrer igjen en avveining mellom hensynet til «riktige» avgjørelser og en effektiv forvaltning.

Ifølge personvernforordningens fortalepunkt 71 skal den registrerte som minimum ha rett til «menneskelig inngripen» ved helautomatiserte avgjørelser. Departementet tydeliggjør i NAV-loven § 4 a andre ledd fjerde punktum at den registrerte har rett til manuell overprøving av avgjørelsen i den grad det ikke foreligger formell klage- eller ankerett. Ved klage foretas manuell behandling.

For å sikre en rettferdig og åpen behandling, skal den registrerte ifølge personvernforordningen artikkel 14 nr. 2 bokstav g få «relevant informasjon om den underliggende logikken samt om betydningen og de forventede konsekvensene» av den automatiserte behandlingen. En tilsvarende rettighet for den registrerte følger av personvernforordningen artikkel 15 nr. 1 bokstav h. Plikten til å forklare «logikken» for den registrerte på en måte som er forståelig for vedkommende vil være utfordrende i tilfeller hvor den automatiserte avgjørelsesprosessen bygger på maskinlærings. Plikten kan imidlertid ikke innebære at den registrerte kan kreve å få fremlagt algoritmene i sin helhet. Informasjonsplikten er begrunnet i at informasjon er en forutsetning for at den registrerte skal kunne utøve sine rettigheter etter personvernforordningen kapittel III. Å få innsyn i en algoritme vil ikke gjøre et kunstig intelligent system noe mer *forståelig* for den registrerte. Det bør være tilstrekkelig å forklare hvilke kriterier som blir vektlagt i forbindelse med den

³⁷⁰ L.c.

³⁷¹ L.c.

³⁷² Bekkelund (2020).

automatiserte avgjørelsesprosessen, og hvorfor disse blir vektlagt. I tillegg bør den registrerte opplyses om hovedtrekkene i vurderingen, kildene til opplysningene som benyttes og hvorfor de er relevante.³⁷³

4.6 Individuell begrunnelse

Avgjørelser som treffes av et kunstig intelligent system skal være *individuell begrunnet*. Dette prinsippet består av flere ulike elementer, og er uløselig knyttet til de øvrige prinsippene som er behandlet i denne utredningen.

Først og fremst må vedtaket rette seg mot *riktig person*, det vil si vedkommende som har interesse i å få sin sak avklart. Dette vil typisk være *parten* i forvaltningslovens forstand, som er den personen en avgjørelse retter seg mot eller som saken ellers direkte gjelder.³⁷⁴

Videre må systemvedtaket bygge på *riktig* og *relevant* informasjon i relasjon til denne personen. Forvaltningsorganet må sikre at informasjonen som ligger til grunn for beslutningen er korrekt, oppdatert og relevant for den aktuelle saken. Det forutsetter at forvaltningsorganet har implementert gode rutiner for dataforvaltning og -styring. Forvaltningsorganet må ha innsikt i datagrunnlaget for å kunne vurdere hvorvidt data overholder de nødvendige kvalitetskravene, og for å kunne legge til eller fjerne bestemte datapunkter.

Forvaltningsorganet må også ha innsikt i *hvilke faktorer* som blir vektlagt av det kunstig intelligente systemet i beslutningsprosessen, for å sikre at det ikke blir lagt vekt på irrelevante forhold eller at systemet diskriminerer på grunnlag av bestemte opplysninger. Dette fordrer grundig loggføring av systemets aktiviteter.

For at forvaltningsorganet skal kunne sikre at beslutningsgrunnlaget er riktig og relevant, må den enkelte personen som beslutningen retter seg mot *involveres i prosessen*. Vedkommende må ha mulighet til å rette eller slette opplysninger om seg selv, i henhold til personvernforordningen artikkel 16 og 17. Dette forutsetter at den enkelte må være informert om hvilke opplysninger om vedkommende forvaltningsorganet behandler, i henhold til den enkeltes rett til innsyn etter personvernforordningen artikkel 15. Dette samsvarer også med innsynsretten som en part har i sakens dokumenter etter forvaltningsloven § 18.

Begrensningene i innsynsretten i forvaltningsloven § 18 a—19 må gjelde tilsvarende.

Her må det klart avgrenses mot opplysninger som *ikke* gjelder den enkelte. Rettighetene i personvernforordningen knytter seg til egne personopplysninger, mens innsynsretten i forvaltningsloven favner videre ved å også inkludere sakens dokumenter og andre faktiske opplysninger.³⁷⁵ Forholdet til *andres personopplysninger* er ikke eksplisitt omtalt, bortsett fra at innsyn i opplysninger som gjelder en annen persons helseforhold er betinget av at det må være *av vesentlig betydning* for parten.³⁷⁶ Det tilsier at terskelen for tilgang til andres sensitive opplysninger i henhold til personvernforordningen artikkel 9 er hvis partens interesse i å få innsyn i slike opplysninger for å ivareta *egne interesser* veier tyngre enn hensynet til beskyttelse av sensitive personopplysninger. Når det gjelder *øvrige personopplysninger*, bør innsynsretten modereres i samsvar med begrensningene som følger av forvaltningsloven § 18—19. Det må derfor legges til grunn at en part kun vil ha innsynsrett i andres personopplysninger dersom disse inngår i «sakens dokumenter» etter forvaltningsloven § 18,

³⁷³ Skullerud (2022).

³⁷⁴ Forvaltningsloven § 2 bokstav e.

³⁷⁵ Jf. forvaltningsloven § 18 og 18 c.

³⁷⁶ Forvaltningsloven § 19 andre ledd bokstav a.

og som har betydning for avgjørelsen. Vedkommende vil følgelig ikke ha noen rett til å få utlevert *hele* informasjonsgrunnlaget for det kunstig intelligente systemet, men kun informasjon om opplysningene som gjelder vedkommende personlig eller er relevante for den aktuelle saken.

Et sentralt hensyn bak forvaltningens informasjonsplikt og plikt til å gi innsyn i sakens dokumenter ved forespørsel er at personen som saken gjelder skal kunne ha et tilstrekkelig grunnlag for å kunne utøve sine rettigheter til klage og omgjøring etter forvaltningsloven kapittel VI, og øvrige rettigheter etter personvernforordningen kapittel III, blant annet retten til å ikke være gjenstand for avgjørelser som utelukkende er basert på automatisert behandling.³⁷⁷ Det er derfor viktig at vedkommende får informasjon om hvilke opplysninger som har betydning for saken på en enkel og forståelig måte. Det innebærer blant annet at forvaltningsorganet må kunne *omgjøre data til informasjon*. Også dette forutsetter en god dataforvaltningsstrategi.

Den enkelte må også få informasjon om hvordan forholdene som har betydning for den endelige beslutningen *vektlegges* av det kunstig intelligente systemet. Dette sikrer vedkommendes mulighet til å forutse egen rettsstilling. Forvaltningen bør for øvrig følge personvernforordningens prinsipper for kommunikasjon med den enkelte, nærmere bestemt at informasjonen skal gis på en kortfattet, åpen, forståelig og lett tilgjengelig måte og på et klart og enkelt språk.³⁷⁸

En annen forutsetning for at beslutningen skal kunne bygge på riktig og relevant informasjon er at systemet ikke må hensynta eller vektlegge forhold som *ikke* har betydning for den aktuelle saken. Det innebærer at både datagrunnlaget og algoritmene må gjennomgås for å sikre at det kunstig intelligente systemet ikke diskriminerer eller tar i betraktning irrelevante forhold. Samtidig ligger det i begrepet «individuell begrunnelse» at systemet må kunne ta i betraktning en rekke *individuelle forhold* som kan få betydning i den enkelte saken (men ikke nødvendigvis i en annen sak). Prinsippet om individuell begrunnelse har følgelig nær sammenheng med *mangfoldet* og *representativiteten* i datagrunnlaget. Beslutningen må ikke fremstå som generert for en større gruppe av befolkningen, og begrunnelsen for beslutningen må reflektere at partens individuelle forhold har blitt vurdert.

4.7 Borgernes livskvalitet og velferd

4.7.1 Bakgrunn

Ordet velferd stammer fra norrønt og er en sammensetning av ordene *vel* og *ferð*.³⁷⁹ Forenklet kan ordet bety noe som «god reise» og det viser gjerne til reisen gjennom livet. Velferd handler med andre ord om å ha det godt,³⁸⁰ men hva som regnes for «godt» er ikke entydig.³⁸¹ Velferdbegrepet blir i Helsedirektoratets rapport «Well-being på norsk» antatt å være nært sammenfallende med det engelske begrepet *well-being* der disse brukes som overordnede

³⁷⁷ Hva som ligger i dette, behandles nærmere i utredningens punkt 4.5.

³⁷⁸ Personvernforordningen art. 12.

³⁷⁹ Det Norske Akademis Ordbok, s.v. «velferd». <https://naob.no/ordbok/velferd>

³⁸⁰ Store Norske Leksikon (2021).

³⁸¹ Dette illustreres blant annet ved de 10 ulike definisjonene av *god* jf. Det Norske Akademis Ordbok, s.v. «god». <https://naob.no/ordbok/god>

generelle begreper.³⁸² Ingen av begrepene har en allmenn definisjon, men litteraturen knyttet til dem begge kan bidra til å belyse deres betydning.

Helsedirektoratets rapport søker å presentere og drøfte noen av de sentrale tilnærmingene til en definisjon av begrepet *well-being*, særlig i lys av norsk folkehelse. De ulike tilnærmingene kan sies å ligge på en skala fra rent subjektive til objektive.³⁸³ De subjektive teoriene tar utgangspunkt i individers egne opplevelser eller oppfatninger som målestokk av blant annet tilfredshet. Motsetningsvis fokuserer de objektive teoriene på at visse elementer og betingelser antas å utgjøre *well-being*, uavhengig av subjektive opplevelser.³⁸⁴ Den norske sosiologen Barstad omtaler disse ytterpunktene velferdsbegrepet som livskvalitet og levekår.³⁸⁵ Mellom disse finnes øvrige teorier om velferd på en mer eller mindre glidende skala.

Velferd er med andre ord ikke et noe enhetlig fenomen, men kan ha ulik betydning avhengig av kontekst og premisser.³⁸⁶ I denne utredningen er hovedfokuset digitalisering av forvaltningen, nærmere bestemt bruk av KI i forvaltningen. Forvaltningens formål og oppgaver vil derfor fungere som kontekst for velferdsbegrepet.

Forvaltningen er statens utøvende myndighet og dets overordnede oppgave er å bidra til å gi borgerne trygghet og forutsigbarhet. Dette gjøres gjennom regelhåndheving, sikring av grunnleggende velferd og ved annen allmenntilgitt virksomhet.³⁸⁷ På den ene siden har forvaltningen et ansvar for borgernes individuelle velferd. Den skal fungere adferdsregulerende med sikte på å hindre konflikt mellom enkeltborgere eller grupper³⁸⁸ i tillegg til å sikre at infrastruktur og tjenester finnes og fremme likhet for borgernes tilgang til dem.³⁸⁹ På den andre siden skal ikke forvaltningen bare tjene individene, men også samfunnet som helhet. Adferdsregulering kan også ha samfunnsmessige mål³⁹⁰ samtidig som at forvaltningen skal sikre samfunnet som helhet på områder som for eksempel forsvar og beredskap, diplomati og miljø.³⁹¹

Teknologirådets rapport «Saken forklart: Kunstig intelligens» fra 2017 presenterer kort en rekke muligheter som KI-systemer kan bidra til å skape, også for offentlig sektor. Blant annet vises det til «smart forvaltning» der KI kan brukes til å forutse behov, tilpasse tjenester, avdekke svindel og feil i tillegg til å gi forvaltningen og andre styrket beslutningsstøtte.³⁹² Men rapporten fremhever også en rekke utfordringer som ikke kan ignoreres. Blant disse trekkes det frem et åpent brev fra 2015 der blant annet fysikeren Steven Hawking, entreprenøren Elon Musk og forskningsdirektøren i Google Peter Norvig etterlyste mer forskning på samfunnsmessige konsekvenser som KI kunne medføre, særlig relatert til KI-

³⁸² Helsedirektoratet (2015) s. 12 og s.81.

³⁸³ Ibid s. 9.

³⁸⁴ L.c.

³⁸⁵ Ibid s. 12-13 med videre henvisninger.

³⁸⁶ Ibid s. 8-9.

³⁸⁷ Jf. NOU 2019: 5 pkt. 5.2.1.

³⁸⁸ Ibid pkt. 5.2.2.

³⁸⁹ Ibid pkt. 5.2.3.

³⁹⁰ Ibid pkt. 5.2.2.

³⁹¹ Ibid pkt. 5.2.4.

³⁹² Teknologirådet (2017) s. 2.

systemer som utvikler seg til å ha verdier og mål som avviker fra våre egne.³⁹³ Videre har World Economic Forum trukket frem KI som teknologiutviklingen med størst potensiell nytteverdi og samtidig størst skadepotensial.³⁹⁴

Med henvisning til prinsippene om rettferdighet, forebygging av skade og autonomi, er velferdshensynet søkt ivaretatt også i EUs etiske retningslinjer for KI. Retningslinjene krever at KI-systemer støtter menneskers personlige autonomi og beslutningsevne.³⁹⁵ I tillegg stilles det krav til sikring av samfunns- og miljømessig velferd³⁹⁶ som innebærer at også det bredere samfunnet, andre sansende vesener og miljøet må tas i betraktning som interessenter ved utviklingen og anvendelsen av KI-systemer.³⁹⁷ Retningslinjene tar for seg særlig tre områder av betydning; Bærekraftighet og miljøvennlighet, sosial påvirkning samt samfunn og demokrati.³⁹⁸

4.7.2 Bærekraftig og miljøvennlig

KI-systemer har stort potensiale for å bidra i kampen med noen av dagens mest påtrengende sosiale utfordringer. Samtidig kan hele livssyklusen til systemene bidra til å gjøre noen av disse utfordringene større. Dette gjelder særlig i forhold til den potensielle miljøbelastning som KI-systemer kan ha.

I en studie utført i 2021 trakk EUs *Special Committee on Artificial Intelligence in a Digital Age (AIDA)*³⁹⁹ frem at anvendelsen av KI krever bruk, forbedring og utvidelse av digital infrastruktur, herunder datasentre og nettverk, som direkte fører til økt forbruk av elektrisitet, materialer og ressurser.⁴⁰⁰ Studien viser blant annet til at rundt 7 % av verdens elektrisitet i dag brukes av datasentre, digitale nettverk og annen IKT, og innen 2030 anslås det å øke til 13 %.⁴⁰¹ En undersøkelse fra 2019 analyserte karbonavtrykket som følger av trening av KI-modeller blant annet til språkbehandling,⁴⁰² og viste at dette kunne tilsvare omtrent fem ganger livsløpsutslippet til en bil for én treningssesjon.⁴⁰³

I tillegg krever KI-systemer lagring av en stadig økende mengde data, herunder data til trening av systemet og data produsert av systemet ved interaksjon med brukere, prediksjoner, beslutninger m.m. Både databehandling og kjøling av datasentre krever store mengder strøm. Statkraft meldte i 2018 at verdens datasentre samlet kunne anslås å bruke mer energi enn hele Tyskland.⁴⁰⁴ Sammenholdt med faktumet at ingen KI-systemer er statiske og at de

³⁹³ L.c.

³⁹⁴ World Economic Forum (2017).

³⁹⁵ High-Level Expert Group on Artificial Intelligence (2019) s. 15.

³⁹⁶ På engelsk: «societal and environmental well-being». High-Level Expert Group on Artificial Intelligence (2019) s. 19.

³⁹⁷ High-Level Expert Group on Artificial Intelligence (2019) s. 19.

³⁹⁸ L.c.

³⁹⁹ På dansk: Det Særlige Udvalg om Kunstig Intelligens i en Digital Tidsalder.

⁴⁰⁰ European Parliament, Directorate-General for Internal Policies of the Union (2021) s. 30.

⁴⁰¹ Ibid s. 31.

⁴⁰² På engelsk: Natural Language Processing (NLP).

⁴⁰³ Sæbø (2021a) og Sæbø (2021b) med videre henvisninger.

⁴⁰⁴ Haga (u.å).

kontinuerlig vil behøve å bli tilpasset, oppdatert m.m. tegnes det et dystert bilde for verdens klima og miljø som allerede er under press.

Produksjon av digital infrastruktur krever også en rekke naturressurser. Råvarer som for eksempel visse typer metaller, herunder kobolt, palladium, kobber og litium, regnes gjerne enten som samfunnskritiske eller utvinnes bare i små mengder. Komponentene som tar dem i bruk har i tillegg ofte kort levetid slik at de må byttes ut jevnlig. Økt etterspørsel etter disse ressursene kan få store miljømessige og sosiale konsekvenser.⁴⁰⁵

I forordningsforslaget er hensynet til bærekraft og miljø primært søkt ivaretatt gjennom respekt for de fundamentale rettighetene slik de er nedfelt i EUs charter om grunnleggende rettigheter.⁴⁰⁶ Dette gjelder særlig artikkel 37 som omhandler EUs plikt til å sikre et høyt miljøvernivå og forbedring av miljøkvalitet, men også gjennom bestemmelser knyttet til borgernes helse og sikkerhet. I fortalen til forslaget omtales de fundamentale rettighetene gjennomgående og det vektlegges som tungtveiende allmenne hensyn som forslaget søker å forfølge og sikre.⁴⁰⁷

Hensynet kan også sies å være forsøkt regulert gjennom frivillige standarder.⁴⁰⁸ Forordningsforslagets artikkel 69 nr. 2. åpner for etableringen av frivillige retningslinjer og krav blant annet for miljømessig bærekraft.⁴⁰⁹

Article 69
Codes of conduct

...

2. The Commission and the Board shall encourage and facilitate the drawing up of codes of conduct intended to foster the voluntary application to AI systems of requirements related for example to environmental sustainability, accessibility for persons with a disability, stakeholders participation in the design and development of the AI systems and diversity of development teams on the basis of clear objectives and key performance indicators to measure the achievement of those objectives.

...

Denne reguleringen tyder på at miljørisiko ikke spiller en vesentlig selvstendig rolle i forslaget. I stedet reguleres miljørisiko primært i tilknytning til menneskerettigheter eller andre klart definerte interesser.⁴¹⁰

Mangelen på eksplisitt bindende regulering i forordningsforslaget betyr likevel ikke at forvaltningen ikke bør ha et ansvar for å sikre at KI-systemene den tar i bruk ivaretar hensynene til bærekraftighet og miljøvennlighet. Tvert imot følger det av forvaltningens arbeidsoppgaver å ivareta også allmenne interesser som dette. Blant annet følger det av Grunnloven § 112, også kjent som *miljøparagrafen*, at alle har «rett til et miljø som sikrer helsen, og til en natur der produksjonsevne og mangfold bevares». Denne retten gjelder også for fremtidige generasjoner gjennom et krav om langsiktig og allsidig forvaltning av naturressursene. Videre krever naturmangfoldloven § 7 at prinsippene i de etterfølgende

⁴⁰⁵ European Parliament, Directorate-General for Internal Policies of the Union (2021) s. 33.

⁴⁰⁶ Charter of Fundamental Rights of the European Union.

⁴⁰⁷ Se blant annet punkt (1) der det står «This Regulation pursues a number of overriding reasons of public interest, such as a high level of protection of health, safety and fundamental rights...».

⁴⁰⁸ Omtalt som «codes of conduct» (engelsk) og «adfærdskodekser» (dansk).

⁴⁰⁹ Omtalt som «environmental sustainability» (engelsk) og «miljømæssig bæredygtighed» (dansk).

⁴¹⁰ European Parliament, Directorate-General for Internal Policies of the Union (2021) s. 37.

paragrafene, herunder om krav til kunnskapsgrunnlag og miljøforsvarlige teknikker, «legges til grunn som retningslinjer ved utøving av offentlig myndighet».

Ut over lovregulerte tilfeller har regjeringen i sin nasjonale strategi for kunstig intelligens lagt vekt på at etiske vurderinger, herunder «hensyn til konsekvenser for miljø og hvorvidt løsningen bidrar til oppfyllelse av FNs bærekraftsmål», bør bygges inn i utviklingen av algoritmer til KI-systemer.⁴¹¹

Det er utvalgets vurdering at tiltak bør settes inn for alle stegene i livssyklusen til KI-systemene, herunder systemets utvikling, distribusjon og bruksprosess, samt forsyningskjeden. Konsekvensutredninger tilsvarende dem som benyttes for arealplanlegging⁴¹² kan benyttes for å «sikre at hensynet til miljø og samfunn blir tatt i betraktning»⁴¹³ ved både vurderingen av *om* og *hvordan* KI-systemer skal tas i bruk av forvaltningen. I tillegg bør ressursbruk og energiforbruk under treningen av KI-systemet underlegges kritisk undersøkelse og kontroll, og prioriteringen av mindre miljøskadelige valg ved utviklingen av systemene bør sikres på et tidlig stadium.⁴¹⁴

Ut over slike konkrete tiltak vil utvalget oppfordre forvaltningen til å legge til rette for ytterligere forskning på dette området som kan belyse særlig miljøbelastende områder og potensielle løsninger. For eksempel har AI@SINTEF⁴¹⁵ forsket på områder der KI-systemer har bidratt til en bedre prosess knyttet til vareproduksjon eller mer effektiv drift og rehabilitering av bygg.⁴¹⁶

4.7.3 Sosial påvirkning

EUs etiske retningslinjer anerkjenner at utbredt eksponering for KI-systemer som kommuniserer og interagerer med mennesker ved sosial simulering, kan endre måten vi oppfatter vår sosiale handlekraft og til og med påvirke sosiale relasjoner og tilknytning.⁴¹⁷ På den måten kan KI-systemer bidra til forringelse av sosiale ferdigheter og dermed påvirke vår fysiske og mentale velferd.

Det finnes flere måter KI kan overvåke og forstå følelsene til individer i samfunnet. Gjennom metoder som ansiktsgjenkjenning og taleforståelse kan KI-systemer fange opp ansiktsuttrykk og analysere stemmeleie og volum og bruke informasjonen til å tolke følelser og humør.⁴¹⁸ Våre handlinger på nettet blir i stadig større grad registrert og analysert til blant annet å tolke personlighet. Samtidig har antallet sensorer på kroppen økt gjennom bruk av treningsarmbånd og smartklokker. Disse samler data om puls, bevegelse, temperatur m.m. som likner på informasjonen som registreres av løgndetektorer.⁴¹⁹

Overvåkingen av individer gjøres sjelden uten et mål om å kunne påvirke handlingsmønstre og liknende. Et eksempel er forskningsprosjektet Intromat ved Haukeland universitetssykehus

⁴¹¹ Departementene (2019) pkt. 5.2.

⁴¹² Se særlig forskrift 21. juni 2017 nr. 859 om konsekvensutredninger.

⁴¹³ Jf. forskrift om konsekvensutredninger § 1.

⁴¹⁴ High-Level Expert Group on Artificial Intelligence (2019) s. 19.

⁴¹⁵ SINTEFs laboratorium for forskning på kunstig intelligens.

⁴¹⁶ SINTEF (2021) med videre henvisninger.

⁴¹⁷ High-Level Expert Group on Artificial Intelligence (2019) s. 19.

⁴¹⁸ Valevatn (2018).

⁴¹⁹ L.c.

som utvikler en applikasjon som skal overvåke humørsvingninger hos personer med bipolar lidelse med det formålet å fange opp symptomer tidlig og forebygge tilbakefall.⁴²⁰ Et annet eksempel er rollen Cambridge Analytica hadde under både Brexit-valgkampen i Storbritannia og presidentvalget i USA i 2016. Firmaet hevdet å ha analysert personligheten til en rekke velgere og brukt profilene til å skreddersy politiske budskap til individer.⁴²¹

Teknologi som dette kan som regel verken regnes for i seg selv å være god eller ond, men den er heller ikke nøytral. Emosjonell intelligens, hva enn den finnes naturlig i mennesker eller som et program i maskiner, kan brukes for å hjelpe oss og har samtidig evne til å manipulere oss. En økt grad av manipulasjon kan ha både negative og positive ringvirkninger for vår opplevelse av livskvalitet og -tilfredshet.

Utvalget mener at skadepotensialet krever at denne typen systemer overvåkes nøye og virkningene av dem granskes fortløpende. Blant annet bør det gjennomføres grundige vurderinger av *når, hvordan* og ikke minst *om* forvaltningen skal ta i bruk KI-systemer som kan manipulere borgerne. EUs etiske retningslinjer viser blant annet til viktigheten av at kunstig intelligente systemer støtter menneskelig handlekraft.⁴²² For forvaltningen betyr dette at behovet for adferdsregulering må veies opp mot hensynet til borgernes individuelle velferd, og at prinsippet om brukerautonomi må være sentralt for systemenes funksjonalitet. Brukerne bør settes si stand til å ta informerte selvstendige valg. Dette kan omfatte opplæring som sørger for å gi kunnskap og verktøy som kan bidra til en tilfredsstillende grad av forståelse for og evne til å samhandle med KI-systemer, inkludert evnen til å vurdere og utfordre dem.⁴²³

Bruken av manipulerende KI skal ikke komme på bekostning av tilliten til forvaltningen. Utvalget mener videre at transparens i hele KI-systemets livssyklus, retningslinjer for etisk bruk av KI, sertifiseringer, tilsyn med systemene og liknende, kan bidra til å minske de potensielle skadevirkningene som systemene kan ha på borgernes livskvalitet.

Kunstig intelligente systemer har også evnen til å påvirke individers grunnleggende rettigheter. I rapporten *Ny teknologi og menneskerettigheter* peker Norges institusjon for menneskerettigheter (NIM) blant annet på at teknologi ikke bare kan forsterke eksisterende diskriminerende praksis, men også skape nye former for diskriminering.⁴²⁴ Rapporten viser videre til covid-19-pandemien der smittesprende teknologi basert på kunstig intelligens på den ene siden bidro til å sikre retten til liv og helse, og samtidig utfordret retten til privatliv, ytringsfrihet, forsamlingsfrihet mv.⁴²⁵

For å bøte på denne utfordringen anbefaler både AI HLEG og NIM at det gjennomføres konsekvensanalyser som vektlegger menneskerettighetene.⁴²⁶ Utvalget er enig i denne vurderingen. Analysen bør kartlegge både tilsiktede, utilsiktede og tverrgående effekter⁴²⁷ som anvendelsen av nye KI-systemer kan ha på menneskerettighetene og borgernes livskvalitet. Den bør også vurdere hvorvidt disse risikoene kan reduseres eller regnes som nødvendige i et

⁴²⁰ Intromat (u.å).

⁴²¹ Se blant annet Grassegger (2017).

⁴²² High-Level Expert Group on Artificial Intelligence (2019) s. 15-16.

⁴²³ Ibid s. 16.

⁴²⁴ NIM (2022) s. 7.

⁴²⁵ Ibid s. 8.

⁴²⁶ High-Level Expert Group on Artificial Intelligence (2019) s. 15 og NIM (2022) s. 32-33.

⁴²⁷ NIM (2022) s. 32-33.

demokratisk samfunn.⁴²⁸ Konsekvensanalysen bør gjennomføres før systemet utvikles samtidig som at det bør innføres mekanismer som sikrer periodevis vurdering og tilbakemelding fra brukerne mens systemet er i bruk.

Forordningsforslaget forsøker å imøtegå disse problemstillingene ved å stille en rekke krav til høyrisiko KI-systemer i del III kapittel 2. I tillegg oppmuntres det til frivillig anvendelse av kravene på KI-systemer av lavere risikograd jf. artikkel 69 nr. 1.

Article 69

Codes of conduct

1. The Commission and the Member States shall encourage and facilitate the drawing up of codes of conduct intended to foster the voluntary application to AI systems other than high-risk AI systems of the requirements set out in Title III, Chapter 2 on the basis of technical specifications and solutions that are appropriate means of ensuring compliance with such requirements in light of the intended purpose of the systems.

...

Bestemmelsen oppfordrer videre i nr. 3 til at slike standarder utformes på en måte som sikrer involveringen av brukerne, andre interessenter og organisasjoner som måtte representere dem.

...

3. Codes of conduct may be drawn up by individual providers of AI systems or by organisations representing them or by both, including with the involvement of users and any interested stakeholders and their representative organisations. Codes of conduct may cover one or more AI systems taking into account the similarity of the intended purpose of the relevant systems.

Økt brukerinvolvering vil også være i tråd med regjeringens digitale agenda for 2017 som fastslår «Brukeren i sentrum» som en av fem hovedprioriteringer i IKT-politikken.⁴²⁹ Ifølge agendaen skal brukerne og deres behov være det sentrale utgangspunktet ved digitaliseringen av offentlig sektor. For å få til dette oppfordrer regjeringen blant annet til involvering av brukerne i den digitale tjenesteutviklingen samt til å trekke læring av andres erfaringer.⁴³⁰ Dette krever økt digital kompetanse samt reell brukermedvirkning som kan sikre god brukskvalitet og at brukernes reelle behov dekkes.⁴³¹ Det er utvalgets vurdering at økt deltakelse også kan medføre økt tillit til forvaltningen og bidra til at eventuelle KI-systemer tas i bruk i henhold til lovgivningen og retningslinjer for etisk KI.

4.7.4 Samfunn og demokrati

I tillegg til å vurdere virkningen på individer skal det bredere samfunnsperspektivet tas i betraktning. EUs retningslinjer legger vekt på at effekten på institusjoner, demokratiet og samfunnet for øvrig må følges nøye.⁴³²

Manipulerende teknologi som dem nevnt i utredningens punkt 4.7.3 har også potensiale for å kunne påvirke samfunnet som helhet. For eksempel vil individuelt tilpasset politisk markedsføring kunne ha en påvirkning på demokratiske valg. Hvorvidt og i hvor stor grad slik manipulering gav utslag i Brexit-valget eller det amerikanske presidentvalget er omstridt, men

⁴²⁸ High-Level Expert Group on Artificial Intelligence (2019) s. 15.

⁴²⁹ Meld. St. 27 (2015–2016) pkt. 1.

⁴³⁰ Ibid pkt. 6.1.

⁴³¹ L.c.

⁴³² High-Level Expert Group on Artificial Intelligence (2019) s. 19.

eksemplet illustrerer likevel effekten KI-systemer potensielt kan ha på demokratiet og samfunnet for øvrig.

Historikeren og forfatteren Yuval Harari advarer om utviklingen mot «digitalt diktatur» som følge av bruk av KI-systemer som kan forutse folks begjær, manipulere deres følelser eller til og med ta beslutninger på deres vegne.⁴³³ Frem til i dag har demokratier generelt utkonkurrert diktaturer fordi de gjennom å distribuere makt på en rekke ulike institusjoner og individer var bedre på å prosessere data og ta beslutninger. I lys av ny teknologi påpeker Harari at sentralisert prosessering av data gjennom bruk av KI vil utkonkurrere distribuert prosessering i løpet av dette årtusen⁴³⁴ og vi risikerer dermed at dagens institusjoner og demokratiet selv kollapser.

De samme vurderingene som er tatt av utvalget i øvrige deler av dette kapitlet gjelder i forhold til påvirkningen bruken av KI-systemer kan ha på samfunnet og demokratiet. Det bør blant annet gjennomføres konsekvensanalyser knyttet til KI-systemers effekt på samfunnets verdier og garantier. Ikke minst må det vurderes strengt hvorvidt KI-systemer skal tas i bruk på situasjoner knyttet til for eksempel demokratiske prosesser som valg.

4.7.5 Ivaretagelse av borgernes livskvalitet og velferd

De ovenstående betraktninger danner grunnlaget for utvalgets forslag om å innta et prinsipp om «ivaretagelse av borgernes livskvalitet og velferd». Blant forvaltningens hovedoppgaver ligger det å ivareta borgernes individuelle og kollektive velferd samt ivaretagelse av allmenne velferdsinteresser knyttet til miljø, demokrati og samfunnet for øvrig. Disse oppgavene er ikke uttrykkelig nevnt i forvaltningsloven, men følger indirekte av oppgavene som er formulert i loven og annen spesiallovgivning.

Utvalget har likevel valgt å fremheve dette som et prinsipp i endringslovforslaget fordi disse oppgavene med stor sannsynlighet vil bli utfordret av den økte digitaliseringen av forvaltningen, særlig i møte med kunstig intelligens. KI-systemer og deres enorme kapasitet til å behandle og prosessere data overgår mennesker både på effektivitet og presisjon. Samtidig er systemene rigide og har liten forståelse for individuelle forhold som kan kreve en skjønnsmessig vurdering. Forvaltningens arbeidsområder har som regel stor påvirkning på borgerne både individuelt og kollektivt. I tillegg møter forvaltningen ofte personer som er i en sårbar livssituasjon.

Det er etter dette utvalgets vurdering at forvaltningen må være svært forsiktige når det gjelder typen KI-systemer som tas i bruk og tenke nøye gjennom i hvilke tilfeller utviklingen av KI-systemer kan være gunstig for borgerne. Utvalget ønsker med dette forslaget å fremheve denne plikten som en påminnelse til forvaltningen om å ta i betraktning «borgernes livskvalitet», herunder individuell og kollektiv velferd, og generell allmenn «velferd», herunder hensyn til miljø og demokrati.

⁴³³ World Economic Forum (2018).

⁴³⁴ L.c.

5 Forvaltningens utvikling og bruk av kunstig intelligente systemer i lys av grunnleggende prinsipper

5.1 Hvilke krav til utvikling og bruk bør stilles i lov og forskrift?

Etter utvalgets oppfatning er det en forutsetning for bruk av kunstig intelligente systemer i forvaltningen at det gjøres endringer i forvaltningsloven. Lovendringene skal ivareta grunnleggende forvaltningsrettslige prinsipper, samt grunnprinsippene for bruk av kunstig intelligente systemer slik disse kommer til uttrykk i EUs forslag til forordning om kunstig intelligens. I det følgende vil vi utdype nærmere hva slags regulering utvalget ser det som nødvendig og hensiktsmessig å implementere.

5.1.1 Lovfesting av grunnleggende prinsipper og operasjonaliseringen av disse

De grunnleggende prinsippene for kunstig intelligente systemer, som gjennomgått i punkt 4, har stått sentralt for utvalget i utarbeidelsen av forslaget til endringer i forvaltningsloven. Utvalget ønsker å tydeliggjøre prinsippenes overordnede betydning ved å inn ta en prinsippbestemmelse i et nytt kapittel IX, som skal være førende for forvaltningens utvikling og bruk av kunstig intelligente systemer.

Disse prinsippene operasjonaliseres videre i flere av de andre bestemmelsene i utvalgets endringsforslag. Dette gjelder for eksempel for § 47 om loggføring av tekniske hendelser, som skal sikre kravet til transparens, herunder etterprøvbarehet. Prinsippet om forsvarlig forvaltning gis blant annet særlig klart utslag i kravet til tilstrekkelig datakvalitet, jf. forslagens § 48. Datakvalitet er en forutsetning for at avgjørelsene som treffes skal være treffsikre og rettsriktige, og med det ikke-diskriminerende.

De grunnleggende forvaltningsrettslige prinsipper gjennomgått i punkt 3, om blant annet rettssikkerhet, legalitet, forsvarlighet, effektivitet og tillit, må videre stå like sentralt i en algoritmisk forvaltning som ellers, om ikke sterkere. Etter utvalgets oppfatning tilsier særlige egenskaper ved algoritmisk forvaltning at individet gis utstrakte rettssikkerhetsgarantier, blant annet gjennom økt offentlighet og en noe utvidet klageadgang. Dette er bakgrunnen for utvalgets forslag i § 49 om offentliggjøring, samt §§ 53 flg. om klage, omgjøring og retting.

5.1.2 Utvikling av kunstig intelligente systemer

Utvalget mener det er et behov for å regulere utviklingen av kunstig intelligente systemer. En regulering av utviklingsprosessen vil kunne sikre at systemenes beslutninger er i tråd med de prinsippene vi ser det som avgjørende å ivareta. Mangler ved et systemvedtak vil i de aller fleste tilfeller kunne ledes tilbake til utviklingsprosessen av systemet, såfremt det ikke er snakk om en feil som følge av at det faktum systemet har behandlet i den konkrete saken er uriktig. Et viktig middel for å sikre treffsikre og rettsriktige avgjørelser vil etter utvalgets oppfatning være å gi formelle krav til organisering, kompetanse, rettslig deltakelse og menneskelig tilsyn i utviklingsprosessen.

Bakgrunnen for at utvalget ser det som hensiktsmessig å implementere «rettslige systemavgjørelser» som en egen beslutningstype i forvaltningsloven, er at vi mener det er nødvendig å stille særskilte krav om rettslig deltakelse ved denne typen avgjørelser. Med «rettslig systemavgjørelser» siktes det til avgjørelser med rettslig innhold som treffes som ledd i utviklingen av kunstig intelligente systemer.

Utvikling av kunstig intelligente systemer vil som den store hovedregel innebære at det fattes en rekke rettslige beslutninger for å gjøre rettsreglene mulige å programmere. En rettslig systemavgjørelse betegner blant annet en avgjørelse om et tolkningsvalg, for eksempel der det er nødvendig å definere et begrep som etter gjeldende rett har et uklart innhold. Begrepet betegner videre en situasjon der man er nødt til å supplere lov og forskrift med regler som skal fylle «hvite flekker» for å sikre en sammenhengende rettslig avklaring av et spørsmål.

Automatisert rettsanvendelse fordrer en detaljert og sammenhengende avklaring av rettsspørsmål, som kan gjøre behovet for utfyllende rettslige systemavgjørelser stort. For at tolkning og utfylling i programmeringsprosessen skal skje i tråd med gjeldende rett og god juridisk metode, er det avgjørende at jurister er med i prosessen. Det burde etter dette stilles krav om rettslig deltakelse i prosessen, herunder hvordan den rettslige delen av systemutviklingen skal organiseres. Regulering av organisering vil synliggjøre de juridiske oppgavene, og legge til rette for aktiv deltakelse fra jurister i prosessen.⁴³⁵

Krav til systemdokumentasjon og logging er nødvendig for å sikre at rettssikkerheten og grunnleggende prinsipper ivaretas. Derfor har vi § 47 om logging presisert at man i utviklingen av kunstig intelligente systemer er nødt til å ha funksjoner som muliggjør en automatisk registrering av tekniske hendelser, slik at det kan føres en logg gjennom hele systemets livssyklus. Dette dokumentasjonskravet må ses i sammenheng med dokumentasjonskravet i arkivloven, som presisert i utvalgets forslag.

Utvalget mener at en følge av at rettslige systemavgjørelser som ledd i «utfylling» av uregulerte rom bør sendes ut på høring, dersom det er snakk om avgjørelse av en viss betydning. Dette må skje før avgjørelsen implementeres i programkoden til et kunstig intelligent system. I denne sammenheng må det imidlertid trekkes en skjønnsmessig nedre grense, da ikke enhver rettslig systemavgjørelse realistisk sett kan sendes ut på høring.⁴³⁶ Her er det sentralt å peke på at bestemmelsen om offentlighet, samt kravene til tilsyn og kontroll, skal sikre at et system også kan angripes rettslig i ettertid.

Bakgrunnen for at utvalget mener at en slik høringsadgang er nødvendig er at de rettslige systemavgjørelsene vil ha egenskaper som minner om forskrifter, og som det er behov for at underlegges en demokratisk og åpen diskusjon. Som Schartum fremholder i sin høringsuttalelse til ny forvaltningslov kan trolig en slik høring gjøres enklere enn ordinær høring etter dagens forvaltningslov.⁴³⁷

Gjøres det endringer i systemet som følge av endringer i gjeldende rett, skal dette offentliggjøres. Det må være mulig for brukerne å forstå og vurdere endringens konsekvenser for dem selv i tråd med kravet om transparente systemer, herunder kravet om forklarbarhet.

Utvalget ønsker ikke å detaljere disse reglene nærmere i lov, men i forskrift. Utvalget har etter dette gitt en forskriftshjemmel for regler om utvikling i § 51.

5.1.3 Krav til systemvedtaks innhold

Utvalget presiserer i lovforslaget at ethvert system skal begrunnes individuelt. Dette skal skje automatisk, og er et absolutt krav. Systemvedtaket skal gi informasjon om hvilke individuelle

⁴³⁵ Schartum (2019) s. 5.

⁴³⁶ Schartum (2019) s. 6 og mailkorrespondanse 3.mai 2022.

⁴³⁷ Schartum (2019) s. 6.

opplysninger om parten som danner grunnlaget for systemvedtaket og hvordan disse har blitt vektlagt, med de begrensninger som følger av forvaltningsloven §§ 18 og 19.

I de fleste tilfeller vil hovedpoenget med begrunnelsen være at parten gis en mulighet til å vurdere om systemet har anvendt rett faktum. Kravet til offentliggjøring av systemene – som skal være relevant, forklarbar og tilgjengelig – gir videre brukeren en mulighet til å vurdere systemvedtaket i et bredere lys. Offentlighetsbestemmelsen i § 49 er en bestemmelse som i hovedsak er inntatt for å sikre forvaltningens aktive informasjonsplikt etter § 100, men som også kan få stor betydning i hver enkeltsak. Dersom brukeren mener at selve systemet ikke er i tråd med loven og dens grunnleggende prinsipper, kan brukeren påklage systemvedtaket. Klageprosessen i en slik situasjon kommer vi nærmere tilbake til om litt.

En offentliggjøring av systemet skal videre sikre forvaltningens veiledningsplikt. Forvaltningen må etter dette utvikle brukervennlige systemløsninger, og informasjonen som er offentliggjort må som allerede understreket være mulig for brukerne å forstå og anvende. For å oppfylle veiledningsplikten på en måte som ivaretar alle brukerne, uavhengig av teknisk kyndighet og individuelle ressurser, kan imidlertid ikke forvaltningen kun være tilgjengelig gjennom automatiserte prosesser på internett. Velutviklede og brukervennlige systemløsninger på forvaltningsorganets nettside kan imidlertid være et svært effektivt og rettssikkerhetsfremmende middel i ivaretagelsen av forvaltningens veiledningsplikt i den automatiserte tidsalder.

5.2 Hva bør være konsekvensene av feil i systemet og uriktige systemvedtak?

I dette kapitlet vil utvalget diskutere hvilke konsekvenser vi mener uriktige systemvedtak og feil i kunstig intelligente systemer til bruk i forvaltningen bør ha. Utvalget mener at en part eller annen med rettslig klageinteresse bør tilkjennes en klagerett, som etter vår oppfatning bør være noe utvidet sammenlignet med klageretten ved alminnelig saksbehandling.

Klageretten skal omfatte at systemet har basert seg på uriktige individuelle opplysninger om parten, samt andre feil i systemet. Utvalget mener videre at forvaltningen skal ha en plikt til å rette feil i systemet, og en plikt til å omgjøre systemvedtak fattet av et system med den feilen som oppdages. En forutsetning for omgjøring er imidlertid at det ikke er til skade for noen systemvedtaket retter seg mot eller direkte tilgodeser.

Et grunnleggende premiss i EUs forslag til forordning om kunstig intelligens er etter vår oppfatning at det inntas mekanismer for klage, omgjøring og retting ved bruk av høyrisiko-systemer. I det følgende vil vi se nærmere på forslagens krav til et effektivt rettsmiddel («effective redress») og ansvarlighet («accountability»).

5.2.1 «Effective redress». Et effektivt rettsmiddel

I EUs forordningsforslag fremheves det blant annet at forordningen skal sikre beskyttelse av retten til et effektivt rettsmiddel («effective remedy»), rettfærdig rettergang («fair trial») og prinsippet om god forvaltningsskikk («principle of good administration», «principle of good administration»).⁴³⁸ Kravet til effektive rettsmidler er også nedfelt i den europeiske menneskerettighetskonvensjonen artikkel 13.

Brudd på grunnleggende rettigheter skal etter forordningsforslaget til kunstig intelligens unngås gjennom en omfattende kontroll av systemet ex ante. I erkjennelsen av at brudd på

⁴³⁸ COM(2021) 206 s. 11.

tross av en slik kontroll fremdeles vil kunne forekomme, understrekes det at det er sentralt med en «effective redress for affected persons». En effektiv klageadgang skal sikres ved at systemene er transparente og ivaretar notoritet, slik at det er mulig med en effektiv kontroll ex post («strong ex post controls»).

Forordningsforslaget har imidlertid blitt kritisert for å ikke inneholde konkrete mekanismer for klage, omgjøring og retting («complaint and judicial redress»)⁴³⁹ Forslaget legger stor vekt på at systemet må kontrolleres ex post, slik at eventuelle feil kan oppdages og rettes opp, men det gis ingen uttrykkelige regler om klage, omgjøring og retting. Etter vår oppfatning er imidlertid eksistensen av en slik plikt uansett en grunnleggende forutsetning for forslaget.

5.2.2 «Accountability». Ansvarlighet

EUs forordningsforslag taler flere steder om «accountability», se blant annet fortalens punkt 38 om viktigheten av å «retain public trust and ensure accountability and an effective redress». I den danske versjonen av forordningsforslaget heter det at man må «bevare offentlighedens tillid og sikre ansvarlighed og effektive rettsmidler». Det fremholdes gjerne at maktbruk fordrer «accountability».⁴⁴⁰ Hva ligger egentlig i dette begrepet?

Begrepet «accountability» på engelsk kan oversettes til «ansvarlighet» på norsk, men dette er ikke en fullgod oversettelse. På engelsk skilles det mellom «accountability» og «responsibility», der sistnevnte refererer til noens plikt til å gjennomføre en oppgave de har for hånden, mens «accountability» handler om det som skjer *etterpå*, med andre ord sentreres begrepet rundt *konsekvensene* av en handling. «Accountability» er en aksept av at man er ansvarlig for sine egne handlinger.⁴⁴¹ Olsen oppgir at begrepet «refers to being answerable to somebody else, being obligated to explain and justify action and inaction».⁴⁴²

I en forvaltningsrettslig sammenheng kan vi si at «accountability» handler om at forvaltningen må stå til ansvar for de feil de har begått. En viktig side ved det å *svare for seg* er at forvaltningen er transparent, og begrunner sine avgjørelser. På denne måten kan offentligheten bedømme om avgjørelsene er riktige, og la det få konsekvenser dersom dette ikke er tilfellet. En effektiv ansvarliggjøring av forvaltningen innebærer en rett til at uriktige avgjørelser gjøres om.⁴⁴³ En effektiv ansvarliggjøring forutsetter etter dette implementeringen av mekanismer som klageadgang, omgjøring og retting.

En ansvarlig forvaltning er forsvarlig, noe annet ville vært uforsvarlig. Forvaltningen skal videre opptre effektivt, tillitsfullt og rettssikkert. Dette er en forutsetning for byråkratisk legitimitet. De grunnleggende forvaltningsrettslige prinsipper taler med tyngde for å innta mekanismer for ansvarliggjøring av forvaltningen ved bruk av kunstig intelligente systemer, tilpasset de særlige egenskaper ved bruken av slike.

Et særskilt poeng er at det kan være vanskelig for de klageberettigede å oppdage feil i kunstig intelligente systemer, fordi de gjerne er basert på kompliserte algoritmer. Uten å forstå hvordan systemene fungerer, og hvordan de ulike algoritmene har ledet frem til det aktuelle systemvedtaket, kan man vanskelig vurdere om det foreligger en klagegrunn. En helt grunnleggende forutsetning for en effektiv ansvarliggjøring er etter dette at forvaltningen

⁴³⁹ Veale mfl. (2021) s. 112.

⁴⁴⁰ Busuioc (2021) s. 827.

⁴⁴¹ Investopedia, s.v. «accountability». <https://www.investopedia.com/terms/a/accountability.asp>

⁴⁴² Olsen (2014) s. 107.

⁴⁴³ Busuioc (2021) s. 827.

opptrer i tråd med prinsippene gjennomgått over. Her står prinsippet om forklarbarhet og tilgjengelighet særlig sentralt.

5.2.3 Behovet for særskilte reguleringer om klage, omgjøring og retting av systemvedtak

Vår forutsetning er at forvaltningens systemer er tilstrekkelig transparente til at parter og andre med klageinteresse kan oppdage feil. Som allerede påpekt mener utvalget at klageadgangen for systemvedtak bør være noe utvidet sammenlignet med klageretten ved alminnelig saksbehandling. Dette har sammenheng med at rettslige systemavgjørelser og systemet som sådan har en forskriftsaktig karakter.

Systemvedtak vil kunne være avgjørende og inngripende for enkeltpersoner, og behovet for å ivareta sentrale prinsipper som rettssikkerhet, tillit og forsvarlighet gjør seg like sterkt gjeldende for systemvedtak som for alminnelige enkeltvedtak. Systemvedtaks enorme potensiale til å behandle mange saker på kort tid, og potensielt fatte en stor mengde uriktige systemvedtak, kan videre tilsi et større behov for en mulighet til å gripe inn dersom man mener det foreligger en feil. I tillegg taler det at kunstig intelligent systemer ofte vil basere seg på kompliserte algoritmer som kan være utfordrende for den alminnelige borger å forstå seg på, og dermed stole på, at man tilkjenner en adgang til å klage på et systemvedtak. En slik ordning må antas å inngi tillit.

For det første mener vi at en part eller annen med klageinteresse bør ha klagerett der hun oppdager at det kunstig intelligente systemet har lagt feil faktum til grunn for systemvedtaket. For eksempel dersom parten ser av begrunnelsen at systemet har basert seg på at hun ikke har vært i jobb det siste året, og dette rett og slett ikke stemmer. En klagerett i et slikt tilfelle vil ikke være kontroversielt, og det vil få den konsekvens at informasjonsgrunnlaget endres og saken undergis en ny behandling i systemet.

For det andre mener vi at en klageberettiget kan klage dersom hun mener at den rettslige systemavgjørelsen som ligger til grunn for det anvendte kunstig intelligente systemet, er uriktig. Dette vil typisk være tilfellet der man mener at den rettslige systemavgjørelsen ikke er i tråd med gjeldende rett, for eksempel ved at tolkningen som er lagt til grunn er i strid med trinnhøyere rettsnormer eller internasjonale forpliktelser. Eller dersom utfyllingen av «hvite flekker» i lovverket er i strid med gjeldende rett etter rettskildelæren.

For det tredje mener vi det vil være en klagegrunn om systemet er i strid med de krav og prinsipper som oppstilles i kapittelet ellers. Systemet kan for eksempel være diskriminerende, i strid med personvernet, eller basere seg på et uforsvarlig datagrunnlag. Det kan også være forhold ved utviklingen av systemet som ikke er i tråd med krav etter lov og forskrift, for eksempel at det viser seg at det ikke har vært noen rettslig deltakelse i utviklingsprosessen.

I vårt lovforslag har vi satt det som en forutsetning for en ny behandling av saken etter klage at feilen har fått innvirkning på resultatet. Det klare utgangspunktet er at dette vil være tilfellet dersom det først kan konstateres at det foreligger en feil. Poenget er å avgrense mot de tilfellene der den aktuelle feilen åpenbart ikke har fått innvirkning. Terskelen for at vilkåret ikke er oppfylt, er med det svært høy.

5.2.4 Rettslig systeminteresse?

Utvalget har diskutert om vi bør innta en bestemmelse om rettslig systeminteresse, men vi har landet på at dette ikke er noe vi ønsker å implementere. Tanken var at loven kunne gi en adgang til å klage på en rettslig systemavgjørelse eller et kunstig intelligent system i

forvaltningen, selv om det ikke foreligger et systemvedtak, og selv om den som ønsker å klage ikke faller innunder gruppen av klageberettigede etter bestemmelsen om klage på systemvedtak.

Bakgrunnen for at vi tenker at det ikke er hensiktsmessig med en slik klageadgang, er at det vil være vanskelig å oppstille en begrensning for når en slik klage kan fremsettes. Kan alle og enhver sies å ha rettslig systeminteresse, og skal man kunne klage på alt mulig? Dette kan potensielt bli svært ressurskrevende for det aktuelle forvaltningsorganet. Vi mener organisasjoner og andre borgeres klageadgang tilfredsstilles ved adgangen til å klage for dem som har «rettslig klageinteresse» etter regelen om klage på systemvedtak.

I denne sammenheng er det sentralt å peke på at forvaltningen har en plikt til å rette systemet dersom det oppdages feil. Systemet vil i tillegg underlegges tilsyn *ex ante*, og en kontinuerlig kontroll *ex post*, som vi straks kommer nærmere tilbake til. Dersom andre skulle mene at det foreligger en feil i systemet, vil disse kunne underrette kontrollorganet om det, slik at kontrollorganet eventuelt kan ta det videre. Vi ønsker imidlertid ikke å implementere en egen bestemmelse om rettslig systeminteresse.

5.3 Hva vil skille saksbehandlingen ved bruk av kunstig intelligente systemer fra tradisjonell saksbehandling?

Spørsmålet er hvordan forvaltningens saksbehandling ved bruk av kunstig intelligente systemer vil bli etter dette lovforslaget. Etter vår oppfatning er det ikke grunn til å la saksgangen ved bruk av kunstig intelligente systemer skille seg markant fra den tradisjonelle saksbehandlingen. Hovedpoenget med å ta i bruk kunstig intelligente systemer er å forbedre og effektivisere, og gjøre forvaltningens avgjørelser med effektive, treffsikre og forutberegnelige. Brukernes *behov* vil være det samme, og *resultatet av prosessen* vil være det samme, nemlig at saksbehandlingen munner ut i et vedtak som er bestemmende for partens rettigheter og plikter. Det er imidlertid enkelte forskjeller det er verdt å trekke frem.

5.3.1 Effektiv saksgang

En stor forskjell ved bruk av kunstig intelligente systemer vil være muligheten for en mer effektiv saksgang. Bruken vil merke at søknadsprosessen går kjappere enn det som er utgangspunktet ved ordinær behandling. Hva gjelder selve saksgangen, altså rekkefølgen på de ulike bestanddelene av saksbehandlingen, vil imidlertid ikke forskjellene være særlig store.

Kravet om forsvarlig menneskelig tilsyn med helautomatiserte avgjørelser tilsier at svaret ikke kan gis umiddelbart etter at søknaden er sendt inn. Det er sentralt å merke seg at det her kun vil være snakk om *tilsyn*, herunder en vurdering av om resultatet er i samsvar med tidligere avgjørelser og overordnede rettsnormer, som sikrer rettferdige og rettsriktige resultater. Tilsynet innebærer ikke at saken skal underlegges en full menneskelig behandling. Dette innebærer at vi etter alle solemerker vil få en mer effektiv saksgang ved bruk av kunstig intelligente systemer.

Vi har tatt utgangspunkt i at søkeren er nødt til å legge inn en søknad med sine individuelle opplysninger for å få tilkjent et systemvedtak. Det betyr at vi ikke har sett nærmere på spørsmålet om automatisering av selve søknadsprosessen, med andre ord for eksempel det scenario at pengene kommer inn på konto idet du går ut i pappaperm, uten at du trenger å søke om det. En interessant problemstilling i dette henseende kan være om en slik fremgangsmåte ivaretar borgernes autonomi.

5.3.2 Systemoffentlighet og veiledningsplikt

En forskjell mellom tradisjonell og systembasert saksbehandling vil være at resultatet av systemvedtaket i stor grad vil være forhåndsbestemt. Dette muliggjør systemoffentlighet, altså at selve systemet gjøres offentlig tilgjengelig. Systemet kan for eksempel ligge ute på forvaltningsorganets nettsider til enhver tid, slik at søkeren kan teste ut hva som med stor sannsynlighet vil bli resultatet i et systemvedtak, før hun sender inn selve søknaden. Dette kan spare forvaltningen for ressurser, da det er nærliggende å tenke seg at en slik løsning vil innebære at færre har et behov for veiledning. En slik systemoffentlighet vil i tillegg gi økt forutberegnelighet for brukeren.

Vi mener allikevel at tilbudet om menneskelig veiledningsplikt i forvaltningen bør opprettholdes for dem som måtte ønske å benytte seg av det. Bakgrunnen for det er å ivareta hensynet til dem som er mindre teknisk kyndige, og hindre digitalt utenforskap. I tillegg ønsker vi at et slikt tilbud skal eksistere utfra den erkjennelse av at man av og til vil føle et behov for å snakke med et menneske – for eksempel dersom man mener det er særlige forhold ved sin sak man lurer på om kan gi utslag i behandlingen.

5.4 Tilsyn og kontroll

5.4.1 Bakgrunn

Det finnes en rekke statlige tilsynsorganer⁴⁴⁴ som fører kontroll med at private og offentlige aktører (tilsynsobjektene) i Norge overholder landets lover, forskrifter og enkeltvedtak samt står for reaksjonsfastsettelse ved avvik.⁴⁴⁵ Hovedoppgaven til tilsynsorganene er gjerne å ivareta en rekke kunde- og brukerhensyn for å sikre trygghet og forutsigbarhet for borgerne. Dette kan bety å sikre borgerne mot maktovergrep, sørge for at bedrifter og produkter holder den kvaliteten de lover og i noen tilfeller å beskytte borgerne mot seg selv. Tilsynene har med andre ord en rolle både som korrektiv og kvalitetsgarantist.⁴⁴⁶ Oppgaven må samtidig ikke blandes med den som ligger til rettsvesenet gjennom politiet og domstolene.

Mer konkret har tilsynsorganer i Norge særlig i oppgave å innhente kunnskap om tilsynsobjektene med formål om å finne ut om det foreligger avvik i virksomheten.⁴⁴⁷ Dette gjøres blant annet ved å oppsøke tilsynsobjektene enten uanmeldt eller forhåndsvarslet. Eventuelle avvik skal deretter følges opp med nødvendige reaksjoner som blant annet kan innebære veiledning, pålegg om oppretting, stansing av virksomheten, bøter eller andre egnede tiltak som kan sikre overholdelse av kravene.⁴⁴⁸ I tillegg har tilsynsorganene andre beslektede oppgaver knyttet til generell informasjonsyting i henhold til forvaltningens veiledningsplikt og relevant særlovgivning, områdeovervåkning ved utarbeidelse av rapporter, statistikker og lignende, samt fastsettelse av normerende vedtak for eksempel i form av konsesjoner.⁴⁴⁹

⁴⁴⁴ I St.meld. nr. 17 (2002–2003) var tallet estimert til nær 40 og det er nærliggende å anta at dette har økt.

⁴⁴⁵ St.meld. nr. 17 (2002–2003) pkt. 1 og pkt. 3.2.

⁴⁴⁶ Ibid pkt. 1.

⁴⁴⁷ Ibid pkt. 3.2.

⁴⁴⁸ L.c.

⁴⁴⁹ L.c.

Som eksempler på tilsynsorganer i Norge kan vi nevne blant annet Konkurransetilsynet som fører tilsyn med konkurranselovgevingen,⁴⁵⁰ Statens legemiddelverk fører tilsyn med produksjon, utprøving og omsetning av legemidler i henhold til legemiddeloven⁴⁵¹ og Datatilsynet som har tilsynsoppgaver knyttet til overholdelsen av personopplysningslova.⁴⁵²

Tilsynsobjektene varierer ut fra tilsynsorganets arbeidsoppgaver, og forvaltningsorganene selv kan også være blant disse. For eksempel er det Arkivverkets oppgave å sørge for forsvarlig arkivhold og dokumentasjonsforvaltning i offentlig sektor.⁴⁵³ I henhold til formålsparagrafen⁴⁵⁴ omfatter dette blant annet arkiver som inneholder «viktig forvaltningsmessig dokumentasjon».

Kontroll med forvaltningen kan utføres enten på systemnivå eller i enkeltsaker.⁴⁵⁵ I enkeltsaker innebærer dette som regel enten intern kontroll gjennom klagebehandling og omgjøring på eget initiativ eller ekstern kontroll via domstolene. Etter en kort gjennomgang av kompetansen til å gjennomføre kontroll i enkeltsaker, vil utvalgets vurdering fokusere på behovet for kontroll av forvaltningen på systemnivå og gjennomføringen av dette der KI-systemer tas i bruk av forvaltningen.

5.4.2 Kontroll i enkeltsaker

5.4.2.1 Om kompetanse til å behandle klager

Klageordninger gir parter rett til å kreve en ny behandling av sakene deres.⁴⁵⁶ I utredningens punkt 5.2 har utvalget gjennomgått de situasjonene som fordrer regler om klageadgang for brukerne i forvaltningen ved anvendelsen av kunstig intelligente systemer. Spørsmålet som gjenstår, angår plasseringen av kompetanse til å behandle klagesakene.

Etter dagens klageordning i fvl. §§ 28-34 avgjøres klager ikke av forvaltningsorganet som opprinnelig traff vedtaket, men av en klageinstans som kan overprøve alle sider av saken. Dersom klageinstansen tar klagen til følge, kan den treffe et nytt realitetsvedtak i saken⁴⁵⁷ eller oppheve vedtaket og sende saken tilbake til underinstansen for helt eller delvis ny behandling.⁴⁵⁸

Utvalget ønsker som utgangspunkt ikke å endre dagens klageordning, men anerkjenner samtidig at behandlingen av klager over systemvedtak vil kreve en særlig kompetanse om kunstig intelligente systemer som store deler av forvaltningen besitter i dag. For å bøte på denne utfordringen finnes det flere mulige løsninger.

For det første kan det kreves et kompetanseløft i forvaltningen om kunstig intelligente systemer som gjør eksisterende organer i stand til å vurdere de KI-systemene som tas i bruk

⁴⁵⁰ For mer info om Konkurransetilsynet se blant annet regjeringen.no eller konkurransetilsynet.no.

⁴⁵¹ For mer info om Statens legemiddelverk se blant annet regjeringen.no eller legemiddelverket.no.

⁴⁵² For mer info om Datatilsynet se blant annet regjeringen.no eller datatilsynet.no.

⁴⁵³ Se arkivlova, blant annet kapittel II jf. §4. Arkivverket har også enkelte oppgaver knyttet til private arkiv jf. lovens kapittel III, men dette gjelder primært dersom de er verneverdige. For mer info om Arkivverket se blant annet arkivverket.no.

⁴⁵⁴ Arkivlova § 1.

⁴⁵⁵ Bragdø-Ellenes (2017) s. 207.

⁴⁵⁶ NOU 2019: 5 pkt. 24.1.

⁴⁵⁷ I.c.

⁴⁵⁸ Ibid pkt. 24.2.1

av organet samt de rettslige systemavgjørelser som er foretatt ved utviklingen av dem. Det er nærliggende å anta at et slikt løft vil være både kostbart og tidkrevende idet det vil kreve etterutdanning av nåværende ansatte og potensielt at det ansettes nye med den påkrevde kompetansen. Denne løsningen bidrar heller ikke i seg selv til deling av kompetanse og erfaring og kan føre til at behandlingen av systemvedtak vil variere avhengig av fagområde.

For det andre kan det etableres en egen klagenemnd for kunstig intelligente systemer. Klagenemnda vil måtte bestå av kompetente fagfolk innenfor kunstig intelligens, digital velferdsstat, velferdsrett og menneskerett for å sikre best mulig løsning innenfor lovens rammer. Denne løsningen vil kunne bidra til en mer helhetlig saksbehandling og støtte opp om visjonen om en sammenhengende og helhetlig offentlig forvaltning.⁴⁵⁹ Løsningen kan også virke ressursbesparende idet det senker behovet for dyptgående kompetanse om kunstig intelligens i de ulike forvaltningsorganene. På den andre siden risikerer denne løsningen en forringelse av forvaltningsskjønnet ved at kompetansen om fagområdets nyanser går tapt i forsøket på å oppnå likebehandling. I tillegg vil økt bruk av kunstig intelligente systemer føre til at klagenemnda må behandle en stadig større andel klagesaker mens eldre klageinstanser mister hovedvekten av sine arbeidsoppgaver. I verste fall kan arbeidsmengden bli så stor for klagenemnda at formålet med klageretten bortfaller helt og dermed føre til at borgernes rettsikkerhet utfordres.

Utvalget mener likevel at sistnevnte løsning bør gjennomføres, med forbehold om at visse tilpasninger rettet mot gjennomføringen av kontrollen vurderes. Klagesaker som bygger på at individuelle opplysninger som systemvedtaket er basert på, er uriktige kan etter vår mening behandles tilsvarende som enkeltvedtak. Kunnskapen som kreves for disse vedtakene er ikke relatert til kunstig intelligens som sådan, men omhandler faktisk informasjon. Vurderingen av denne typen informasjon gjøres daglig av forvaltningsorganene allerede og hensynet til effektivitet tilsier at oppgaven kan forbli hos de aktuelle organene. Klagesaker som derimot omhandler en rettslig systemavgjørelse eller overenstemmelsen med de krav som stilles til KI-systemet bør tilfalle KI-klagenemnda.

Utvalget lar det være opp til Kongen å fastsette nærmere regler for klagenemndas organisering og saksbehandling. For å sikre at klagebehandlingen oppfyller kravene som følger av utredningsplikten anbefaler vi at det opprettes mekanismer som sørger for inkluderingen av fagkompetanse på det aktuelle forvaltningsområdet. Dette kan blant annet gjøres ved å oppstille krav om at representanter fra den aktuelle klageinstansen sikres deltakelse i klagebehandlingen, gjerne som et midlertidig medlem av KI-klagenemnda.

5.4.2.2 Om kompetanse til å omgjøre systemvedtak

I noen tilfeller har vedtaksorganet og overordnede organer adgang til å rette eller omgjøre et vedtak uten at det er påklaget. *Retting* handler om å rette opp klare feil som gjør at vedtakets innhold tilsynelatende er annerledes enn det som er tilsiktet fra forvaltningens side.⁴⁶⁰ *Omgjøring* innebærer derimot at vedtaket får et annet innhold enn opprinnelig ved mer eller mindre omfattende endringer.⁴⁶¹ Reglene om retting og omgjøring følger henholdsvis av ulovfestet rett og fvl. § 35.

Utvalget mener at reglene om retting og omgjøring bør få anvendelse også på systemvedtak så langt det er mulig. Samtidig fordrer saksbehandling med kunstig intelligente systemer noe særregulering. På samme måte som at saksbehandlere kan begå feil, kan det også forekomme

⁴⁵⁹ Se utredningens punkt 2.3 med videre henvisninger.

⁴⁶⁰ NOU 2019: 5 pkt. 25.1 og 25.5.

⁴⁶¹ Ibid pkt. 25.1.

feil i det kunstig intelligente systemet for eksempel i programkoden, datagrunnlaget osv. Feil i KI-systemet kan videre føre til feil i systemvedtakene. Fordi KI-systemer på kort tid kan behandle et vesentlig større antall saker enn mennesker, kan en slik feil skape usikkerhet om hvorvidt brukernes rettigheter er overholdt. I verste fall vil forvaltningsorganet skulle behandle et uoversiktlig antall klagesaker eller et vesentlig antall brukerne vil måtte leve med en potensielt urettmessig innskrenkning av deres rettigheter.

Følgelig foreslår vi en lovfestet plikt for forvaltningen til å rette kunstig intelligente systemer når det oppdages feil i systemet. Dette gjelder først og fremst dersom det avdekkes at noen av de rettslige systemavgjørelsene som ligger til grunn for det anvendte kunstig intelligente systemet er uriktige eller at det kunstig intelligente systemet ikke er i overenstemmelse med de kravene og prinsippene som følger av lovforslaget. Inntil feilen er rettet skal det kunstig intelligente systemet ikke tas i bruk.

Plikten til å rette kan ligge hos vedtaksorganet selv eller klageorganet dersom feilen er oppdaget av dette. For det tilfellet at feilen ikke enkelt kan rettes av disse organene, skal tilsynsorganet for kunstig intelligente systemer gis adgang til å rette feilen. Tilsynsorganet skal uansett rette feilen dersom særlige grunner taler for dette. Ved vurderingen av om særlige grunner foreligger kan det blant annet tas i betraktning feilens art, hvor lenge feilen har vært til stede i systemets anvendelsesperiode og om rettingen vil kreve omfattende testing før systemet igjen kan tas i bruk.

I tillegg foreslås et krav om omgjøring av alle systemvedtak som er truffet av et kunstig intelligent system med slike feil. For å sikre hensynet til forutberegnelighet kan dette likevel ikke gjelde dersom omgjøring vil være til skade for noen som systemvedtaket retter seg mot eller direkte tilgodeser.

5.4.3 Kontroll på systemnivå

5.4.3.1 Særlig om behovet for menneskelig tilsyn

I utredningens punkt 4.7 om borgernes livskvalitet og velferd har utvalget vist til forvaltningens overordnede oppgaver knyttet til blant annet å sikre borgernes grunnleggende velferd. Gjennom forskrifter og enkeltvedtak har forvaltningen stor makt over borgerne både kollektivt og individuelt og når forvaltningen begår feil kan dette ha store innvirkninger på dem. Kontroll med forvaltningen er derfor svært viktig både for å unngå fremtidige feil og for å oppdage og rette opp feil som allerede er begått.

Som vi har vist til tidligere i rapporten har innføringen av KI i forvaltningen stort potensiale for å utfordre en stor del av menneskerettighetene og andre grunnleggende forvaltningsrettslige rettigheter som borgerne har i dag. Innføringen kommer gjerne med løfter om bedre kvalitet på tjenester, økt kapasitet og fleksibilitet i forvaltningen samt økt effektivitet både i forhold til tid og ressurser. På den andre siden betyr den økte kapasitet for eksempel i form av automatisk massesaksbehandling⁴⁶² også økt skadepotensiale dersom KI-systemene begår feil. En feil i algoritmene til systemet kan føre til at et vesentlig større antall saker vurderes feil over en kortere tidsperiode. Utfallet vil kunne være en skandale tilsvarende trygdeskandalen⁴⁶³ som ble avdekket i 2019, men med et større antall berørte eller samme antall berørte over en kortere tidsperiode.

⁴⁶² Se blant annet JOU 2020: 5 kap. 2.

⁴⁶³ For mer informasjon om trygdeskandalen: Institutt for offentlig rett UiO (u.å) med videre henvisninger.

På bakgrunn av dette vurderer utvalget det som nødvendig å fastsette særlige kontrollmekanismer som kan imøtegå noen av de utfordringene som bruk av KI i forvaltningen kan medføre. Også EUs etiske retningslinjer for KI oppstiller et krav om menneskelig tilsyn.⁴⁶⁴ Kravet har opphav i prinsippet om respekt for personlig autonomi og anbefaler at KI-systemer utformes på en måte som støtter dette.

I den videre behandlingen vil utvalget gjøre rede for hvordan dette tilsynet bør føres. Siden feil kan oppstå både før og etter at KI-systemene tas i bruk, vil dette være utgangspunktet for skillet mellom forhånds- og etterhåndskontroll. Ved vurderingen av hvilke kontrollmekanismer som skal tas i bruk har utvalget tatt utgangspunkt i forordningsforslagets tilsynsmekanismer tilpasset en forvaltningsrettslig kontekst.

5.4.3.2 Kort om kompetanse til å føre tilsyn

I henhold til forordningsforslaget vil kompetansen til å føre eksternt tilsyn⁴⁶⁵ ligge hos en eller flere nasjonale tilsynsmyndigheter.⁴⁶⁶ I henhold til artikkel 3 nr. 42 vil dette organet blant annet ha ansvaret for gjennomføringen og anvendelsen av forordningen og etter artikkel 59 nr. 1 skal det organiseres på en måte som sikrer objektiviteten og uavhengigheten til organet. Hvilket organ som vil bli tildelt dette ansvaret, herunder om det skal etableres et nytt tildelingsorgan eller om ansvaret kan tildeles et allerede eksisterende organ, ligger utenfor utvalgets mandat å vurdere. Dette spørsmålet vil måtte besvares dersom forslaget skulle bli vedtatt av EU og innlemmet i EØS-avtalen.

Utvalget vil likevel bemerke at Datatilsynet har argumentert for at tilsynsorganene etter GDPR art. 51 også er egnet som tilsynsorgan etter KI-forordningen.⁴⁶⁷ Datatilsynet peker blant annet på at datatilsynsmyndighetene allerede har forståelsen og kunnskapen om KI-teknologier, data- og databehandling samt grunnleggende rettigheter, som kreves i forordningsforslaget. I tillegg vil en stor andel av KI-systemene som utvikles enten være basert på behandlede personopplysninger eller behandle personopplysninger selv. Derfor vil arbeidsoppgavene til de allerede etablerte datatilsynsmyndighetene i stor grad overlappe med oppgavene til tilsynsmyndighetene etter forordningsforslaget. Utvalget støtter denne vurderingen, men vil også påpeke at dersom dette skulle bli løsningen i Norge og andre Europeiske land må kompetansen hos datatilsynsmyndighetene løftes opp på et nivå som tillater dem å ikke bare sikre personvern hensynene, men også de særlige hensynene som følger med bruken av KI-systemer.

I tillegg til å ha det generelle ansvaret for gjennomføringen av forordningen, vil tilsynsorganet også fungere som *autoriserende myndighet*.⁴⁶⁸ Dette innebærer å sørge for autorisering av *overensstemmelsesvurderingsorganer*⁴⁶⁹ som ønsker å få kompetanse til å foreta vurderinger

⁴⁶⁴ På engelsk: Human agency and oversight. På dansk: «Menneskelig utførelse og kontrol».

⁴⁶⁵ Med eksternt tilsyn siktes det til tilsyn som føres av andre enn tilbyderer av KI-systemet.

⁴⁶⁶ På engelsk: «national supervisory authority». På dansk: «national tilsynsmyndighed».

⁴⁶⁷ Datatilsynet (2021) s. 2.

⁴⁶⁸ Jf. artikkel 3 nr. 19. På engelsk: «notifying authority». På dansk: «bemyndigende myndighed». I henhold til art. 59 nr. 2 kan denne kompetansen tillegges et separat nasjonalt kompetent organ dersom organisatoriske og administrative grunner taler for dette.

⁴⁶⁹ Jf. artikkel 3 nr. 21. På engelsk: «conformity assessment body». På dansk: «overensstemmelsesvurderingsorganer».

av om KI-systemer oppfyller kravene som følger av forordningens del II kapittel 2 (*overenstemmelsesvurdering*⁴⁷⁰).

Øvrig kontroll med KI-systemer foretas i hovedsak internt i tilbyderens virksomhet. For forvaltningen betyr dette at tilsynet først og fremst gjennomføres av den instansen eller det organet som direkte kan regnes som tilbyder. I tillegg vil eventuelle overordnede organer ha organisasjons- og instruksjonsmyndighet etter en konkret vurdering.

5.4.3.3 Før systemet tas i bruk (*ex ante*)

Forhåndskontroll med KI-systemer sikres på flere måter i EUs forordningsforslag. Utvalget vil gå gjennom noen av de viktigste bestemmelsene og deres betydning for forvaltningen.

For det første fastslår artikkel 59 nr. 7 at tilsynsorganet⁴⁷¹ kan yte veiledning og rådgivning til tilbydere av KI-systemer om gjennomføringen av forordningen. Veiledningen kan blant annet foregå gjennom et sentralt kontaktpunkt for kommunikasjon med de relevante aktørene.⁴⁷² For forvaltningen vil dette i stor grad bli oppfylt av veiledningsplikten etter forvaltningsloven § 11 som utvalget har gjort rede for ovenfor i punkt 3.2.2.1 samt i henhold til relevant særlovgivning.

Samtidig krever forslaget at dersom veiledningen også berører kompetanseområder som er omfattet av andre EU-rettsakter, skal det kompetente organet på dette området konsulteres der det er relevant. Dette er en problemstilling som de færreste tilsynsorganer står overfor, men som er svært relevant for tilsynsmyndigheten for KI-systemer da slike systemer tas i bruk i en rekke ulike områder som forvaltningen har ansvar for. Det nærmeste vi kommer slik sektorovergripende EU-lovgivning i dag er personopplysningsloven og tilhørende forordning.

For det andre følger det av artikkel 19 før høyrisiko KI-systemer plasseres på markedet eller tas i bruk, skal de underlegges en overenstemmelsesvurdering i henhold til de fastsatte prosedyrene i vedlegg VI eller VII jf. artikkel 43.

Article 19

Conformity assessment

1. Providers of high-risk AI systems shall ensure that their systems undergo the relevant conformity assessment procedure in accordance with Article 43, prior to their placing on the market or putting into service. Where the compliance of the AI systems with the requirements set out in Chapter 2 of this Title has been demonstrated following that conformity assessment, the providers shall draw up an EU declaration of conformity in accordance with Article 48 and affix the CE marking of conformity in accordance with Article 49.

Tilbyderen av KI-systemet har plikt til å sørge for at dette gjennomføres, men forordningen krever også at andre aktører kontrollerer at vurderingen er foretatt. Dette gjelder blant annet for importører jf. artikkel 26 nr. 1 bokstav a og distributører jf. artikkel 27 nr. 1. I tillegg har begge aktørene ansvar for at KI-systemet de ønsker å henholdsvis importere og distribuere ikke blir plassert på markedet dersom de anser eller har grunn til å anse at systemet ikke oppfyller kravene etter forordningen jf. nr. 2.

Overenstemmelsesvurderingen etter vedlegg VI innebærer en intern kontroll av kvalitetsstyringssystemet og den tekniske dokumentasjonen som skal være etablert og utarbeidet jf. henholdsvis artikkel 17 og 18. Tilbyderen skal vurdere hvorvidt KI-systemet

⁴⁷⁰ Jf. artikkel 3 nr. 20. På engelsk: «conformity assessment». På dansk: «overensstemmelsesvurdering».

⁴⁷¹ Samt andre nasjonale kompetente myndigheter etter artikkel 3 nr. 43.

⁴⁷² Bestemmelsen bruker ordet «operator» som er definert i art. 3 nr. 8 og omfatter blant annet brukere, tilbydere, importører og distributører.

er i overensstemmelse med kravene som stilles til høyrisiko KI-systemer. I tillegg skal tilbydereren kontrollere at design- og utviklingsprosessen av KI-systemet samt overvåkningssystemet etablert i henhold til artikkel 61, er i overensstemmelse med den tekniske dokumentasjonen.

ANNEX VI

CONFORMITY ASSESSMENT PROCEDURE BASED ON INTERNAL CONTROL

1. The conformity assessment procedure based on internal control is the conformity assessment procedure based on points 2 to 4.
2. The provider verifies that the established quality management system is in compliance with the requirements of Article 17.
3. The provider examines the information contained in the technical documentation in order to assess the compliance of the AI system with the relevant essential requirements set out in Title III, Chapter 2.
4. The provider also verifies that the design and development process of the AI system and its post-market monitoring as referred to in Article 61 is consistent with the technical documentation.

Vedlegg VII oppstiller en overensstemmelsesvurdering som må foretas av et autorisert overensstemmelsesvurderingsorgan, også omtalt i forslaget som et *autorisert organ* jf. art. 3 nr. 22. Prosedyren er mer omfattende enn den som følger av vedlegg VI, men forskjellen er i hovedsak knyttet til oppstillingen av rettighetene og pliktene til tilbyderne og det autoriserte organet. Kontrollen og vurderingene som foretas er dermed overordnet de samme.

Forutsatt at forvaltningens KI-system ikke er «beregnet for biometrisk fjernidentifikasjon i sanntid og påfølgende biometrisk fjernidentifikasjon av fysiske personer»⁴⁷³ jf. punkt 5 (a) i vedlegg III om høyrisiko KI-systemer, vil dets høyrisiko KI-systemer kun kreve intern kontroll etter vedlegg VI jf. art. 43 nr. 2 første punktum.

I vårt lovforslag er disse kravene til forhåndskontroll ivaretatt gjennom flere bestemmelser. Tilsynsmyndighetens kompetanse fremkommer av lovforslagets § 56 femte ledd. Innholdet i kompetansen vil måtte følge av gjennomføringsloven med henvisning til de relevante bestemmelsene i forordningen. Kravene til gjennomføring av overensstemmelsesvurdering vil følge primært av § 45 i lovforslaget som henviser videre til gjennomføringsloven jf. ordlyden «formell lov» og KI-forordningen jf. ordlyden «internasjonale forpliktelser».

5.4.3.4 Etter systemet er tatt i bruk (ex post)

Også etterhåndskontroll sikres på flere måter i forslaget til forordning om KI, først og fremst gjennom artikkel 14 om menneskelig tilsyn.

Article 14

Human oversight

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use.
2. Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter.
3. Human oversight shall be ensured through either one or all of the following measures:

⁴⁷³ Oversettelse av den danske teksten i forslaget: «beregnet til biometrisk fjernidentifikasjon i realtid og efterfølgende biometrisk fjernidentifikasjon af fysiske personer».

- a) identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;
 - b) identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the user.
4. The measures referred to in paragraph 3 shall enable the individuals to whom human oversight is assigned to do the following, as appropriate to the circumstances:
- a) fully understand the capacities and limitations of the high-risk AI system and be able to duly monitor its operation, so that signs of anomalies, dysfunctions and unexpected performance can be detected and addressed as soon as possible;
 - b) remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system ('automation bias'), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;
 - c) be able to correctly interpret the high-risk AI system's output, taking into account in particular the characteristics of the system and the interpretation tools and methods available;
 - d) be able to decide, in any particular situation, not to use the high-risk AI system or otherwise disregard, override or reverse the output of the high-risk AI system;
 - e) be able to intervene on the operation of the high-risk AI system or interrupt the system through a «stop» button or a similar procedure.
5. For high-risk AI systems referred to in point 1(a) of Annex III, the measures referred to in paragraph 3 shall be such as to ensure that, in addition, no action or decision is taken by the user on the basis of the identification resulting from the system unless this has been verified and confirmed by at least two natural persons.

Bestemmelsen tildeler ikke kompetanse til å foreta tilsynet til spesifiserte personer eller organer, men krever at systemet tilrettelegges for menneskelig tilsyn mens det er i bruk. Formålet med tilsynet er å forebygge eller minimere risikoen som kan oppstå for helse, sikkerhet eller grunnleggende rettigheter ved bruk eller forutsigbart misbruk av KI-systemer. Menneskelig tilsyn sikres gjennom passende tiltak som må identifiseres av tilbyderen før det plasseres på markedet. Fortalens punkt 48 foreslår tiltak som for eksempel innebygde operasjonelle begrensninger i systemet som ikke kan overstyres av systemet selv samt å sikre at de fysiske personene som skal utføre tilsynet har den nødvendige kompetansen, opplæringen og myndigheten for utførelsen av det.

Kompetansen til å føre tilsyn er dels lagt til tilbyderen. Blant annet følger det av artikkel 43 nr. 4 at dersom et høyrisiko KI-system vesentlig endres,⁴⁷⁴ skal det underkastes en overenstemmelsesvurdering. Under visse vilkår gjelder dette ikke dersom endringene er forhåndsbestemt av tilbyderen ved tidspunktet for den første vurderingen. Forutsatt at klassifiseringen av systemet etter vedlegg II ikke også har endret seg betyr dette at den samme prosedyren følges på nytt. For forvaltningen vil dette stort sett innebære en intern kontroll i henhold til vedlegg VI.

Videre sikres etterhåndskontroll gjennom den nasjonale tilsynsmyndigheten som blant annet er tildelt en rett til å anmode en rekke dokumenter fra både tilbydere, importører og distributører av KI-systemer. Tilbydere har etter artikkel 23 plikt til å samarbeide med tilsynsorganet ved å oppgi all informasjon og dokumentasjon som er nødvendig for å vurdere overenstemmelse med kravene for høyrisiko KI-systemer. Artikkel 50 lister opp en rekke av dokumenter som typisk vil være relevante.

⁴⁷⁴ På engelsk: «are substantially modified». På dansk: «ændres væsentlig».

Article 50

Document retention

The provider shall, for a period ending 10 years after the AI system has been placed on the market or put into service, keep at the disposal of the national competent authorities:

- a) the technical documentation referred to in Article 11;
- b) the documentation concerning the quality management system referred to Article 17;
- c) the documentation concerning the changes approved by notified bodies where applicable;
- d) the decisions and other documents issued by the notified bodies where applicable;
- e) the EU declaration of conformity referred to in Article 48.

Tilbyderen skal også selv informere tilsynsorganet om manglende overenstemmelse med kravene for høyrisiko KI-systemer og eventuelle iverksatte tiltak som skal rette opp mangelen jf. art. 16 bokstav h.

Den siste kontrollmekanismen som vi vil nevne følger av artikkel 71 som åpner for fastsettelsen av ulike sanksjoner ved overtredelsen av forordningen.

Article 71

Penalties

1. In compliance with the terms and conditions laid down in this Regulation, Member States shall lay down the rules on penalties, including administrative fines, applicable to infringements of this Regulation and shall take all measures necessary to ensure that they are properly and effectively implemented. The penalties provided for shall be effective, proportionate, and dissuasive. They shall take into particular account the interests of small-scale providers and start-up and their economic viability.

...

Forslaget lar det være opp til nasjonal rett å bestemme om tilsynsorganet skal ha adgang til å ilegge overtredelsesgebyr mot offentlige myndigheter jf. art. 71 nr. 7. Denne løsningen tilsvarende i personvernforordningen art. 83 nr. 7 og Datatilsynet er gitt slik myndighet etter personopplysningsloven § 26 annet ledd.

I likhet med høringsinstansenes syn i Prop. 56 LS (2017–2018) punkt 20.4.3 vil utvalget bemerke at det som utgangspunkt ikke bør legges opp til forskjellsbehandling av privat og offentlig sektor når det gjelder fastsettelsen av overtredelsesgebyr. Formålet med reglene i forslaget støtter dette. Det er også nærliggende å anta at faren for omdømmetap og konsekvensene en eventuell overtredelse vil ha på tilliten brukerne har til tjenesteleverandøren vil fungere som en større avskrekkende effekt enn et gebyr. På den andre siden vil ikke gebyrene nødvendigvis ha den effekten i offentlig sektor idet de vil fungere som en omfordelingsmekanisme internt i staten eller fra kommune til stat. Gebyret vil derfor innebære et inngrep i politiske myndigheters prioritering av de ressursene som skal løse deres forvaltningsoppgaver. I verste fall kan det påvirke organets evne til å levere samfunnskritiske tjenester.

Både personvernforordningen og forslaget til forordning om kunstig intelligens har et vidt anvendelsesområde i den forstand at regelverket skal følges uavhengig av livs- eller rettsområde. Samtidig innebar personopplysningsloven § 26 annet ledd innebar i realiteten en videreføring av gjeldende rett, men for KI-systemer finnes det ingen slik rettshistorie.

Størrelsen på gebyrene i forslaget tilsvarer i stor grad dem som er fastsatt i GDPR⁴⁷⁵ med unntak av at visse overtredelser⁴⁷⁶ kan medføre gebyr på 10 000 000 euro/2 prosentpoeng mer enn den øvre grensen som er satt i GDPR. Beløpsgrensene angir riktignok kun maksimalgrenser og ingen minimumsgrenser, men listen over momenter som det skal legges vekt på ved utmålingen er vesentlig lenger i GDPR art. 83 nr. 2 sammenlignet med forslaget art. 71 nr. 6. Graden av skjønn som kan utøves er tilsynelatende noe mindre enn den som ble lagt til grunn i departementets vurdering i Prop. 56 LS (2017–2018) punkt 20.5.3.

Den endelige beslutningen om hvorvidt offentlig sektor skal kunne ilegges overtredelsesgebyr må tas ved den eventuelle gjennomføringen av forslaget til forordning om kunstig intelligens. Utvalget mener at det som utgangspunkt ikke er noe som taler for at dette ikke skal være tilfellet, men at det bør vurderes om gebyret beløpsmessig bør begrenses.

I vårt lovforslag er kravene til etterhåndskontroll i § 52 tredje ledd og § 56 femte ledd. Førstnevnte krever at det skal føres «forsvarlig menneskelig tilsyn med helautomatisert saksbehandling utført av kunstig intelligente systemer» og viser særlig til den interne kontrollen av systemet. Innholdet i tilsynet vil følge av gjennomføringsloven med henvisning til forordningen (nåværende artikkel 14 i forslaget). Sistnevnte fastslår tilsynsmyndighetens kompetanse til å gjennomføre tilsyn og kontroll.

6 Administrative og økonomiske konsekvenser

Utvalget har ikke hatt anledning til å gjennomføre en omfattende analyse av de administrative og økonomiske konsekvensene endringsforslaget vil ha. Dette vil være nødvendig å utrede nærmere, jf. utredningsinstruksen punkt 2.1 nr. 4. Vi viser til JOU 2020: 5 *Kunstig intelligens – Muligheter og risikoer i velferdsforvaltningen* (punkt 8) og JOU 2020: 2 *Automatisk kontroll i Arbeids- og velferdsetaten – Grunnleggende rettigheter ved automatisk kontroll med borgernes etterlevelse av velferdslovgivningen* (punkt 7).

I utredningen har vi pekt på behovet for muligheter og risikoer ved kunstig intelligente systemer, og behovet for lovendringer i forvaltningsloven. Endringsforslaget skal sikre at forvaltningens utvikling og bruk av kunstig intelligente systemer er i tråd med EUs forordning om harmonisering av regler om kunstig intelligens, samt overordnede prinsipper. Dagens regelverk er ikke tilpasset implementeringen av kunstig intelligente systemer, og lovendring er nødvendig for at forvaltningen skal kunne benytte seg av de muligheter slike systemer gir.

Reglene for utvikling og bruk av kunstig intelligente systemer skal påse at forvaltningen opptrer på en rettssikker, effektiv og tillitsfull måte. En forutsetning for å overholde disse grunnleggende forvaltningsrettslige prinsipper ved implementeringen av kunstig intelligente systemer i forvaltningen, er at forvaltningen opptrer i samsvar med forordningsforslagets formulerte grunnprinsipper om transparens, forsvarlig dataforvaltning, teknisk robusthet og sikkerhet med videre. Endringsforslaget skal videre påse at sentrale rettssikkerhetsgarantier som klage og omgjøring ivaretas på en forsvarlig måte, og at det føres tilstrekkelig tilsyn og kontroll.

⁴⁷⁵ Beløpsgrensene som er brukt i GDPR art. 83 nr. 4 samt 5 og 6 er på henholdsvis 10 000 000 euro/2% av samlet global årsomsetning og 20 000 000 euro/4% av samlet global årsomsetning. Dette tilsvarer forslagets art. 71 nr. 5 og 4.

⁴⁷⁶ Herunder manglende etterlevelse av forbudet i artikkel 5 og KI-systemets manglende oppfyllelse av kravene i artikkel 10 jf. artikkel 71 nr. 3.

7 Lovforslag

Forslag til lov om endringer i forvaltningsloven (kunstig intelligente systemer, rettslige systemavgjørelser, og systemvedtak mv.)

I

I lov 10. februar 1967 om behandlingsmåten i forvaltningssaker gjøres følgende endringer:

§ 2 første ledd bokstav c) og g) skal lyde:

- c) *systemvedtak, et enkeltvedtak generert av et kunstig intelligent system*
- g) *rettslig systemavgjørelse, en avgjørelse med rettslig innhold som treffes som ledd i utviklingen av kunstig intelligente systemer;*

Nåværende § 2 første ledd bokstav c), d), e), f), g) og h) blir bokstav d), e), f), h), i) og j).

Nytt kapittel IX skal lyde:

Kapittel IX: Om rettslige systemavgjørelser, kunstig intelligente systemer og systemvedtak

§ 43. *(formål)*

Kapitlet søker å ivareta borgernes rettigheter i møte med forvaltning utøvet ved bruk av kunstig intelligente systemer.

§ 44. *(prinsipper)*

Kunstig intelligente systemer til bruk i forvaltningen skal være transparente og være utformet på en måte som ivaretar personvernet, en forsvarlig forvaltning av data, samt borgernes livskvalitet og velferd. Kunstig intelligente systemer skal være teknisk robuste og sikre.

Systemvedtak skal være rettfærdige og saklige. Den individuelle begrunnelsen for systemvedtaket må være mulig for parten å forstå, vurdere og etterprøve.

§ 45. *(forholdet til annen lovgivning)*

Et kunstig intelligent system skal være i tråd med krav som ellers stilles i formell lov, grunnlov og andre internasjonale forpliktelser.

§ 46. *(overordnet krav til kunstig intelligente systemer til bruk i forvaltningen)*

Ethvert kunstig intelligent system til bruk i forvaltningen skal oppfylle kravene som stilles til systemer som er klassifisert som høyrisiko-systemer etter EUs forordning om kunstig intelligens.

§ 47. *(loggføring av tekniske hendelser i kunstig intelligente systemer)*

Et kunstig intelligent system skal utvikles med funksjoner som muliggjør en automatisk registrering av tekniske hendelser (loggføring). Slik registrering av tekniske hendelser skal føres gjennom hele systemets livssyklus.

Registrering av tekniske hendelser (loggføring) skal sikre et tilstrekkelig nivå av etterprøvnbarhet av systemets virkemåte i henhold til systemets tiltenkte formål.

Registreringen skal utformes på en måte som tar hensyn til behovet for en forsvarlig arkivering i henhold til arkivloven.

§ 48. *(krav til datakvalitet i kunstig intelligente systemer)*

Datagrunnlaget til et kunstig intelligent system skal være av tilstrekkelig kvalitet. Datagrunnlaget skal være representativt og sikre treffsikre og rettsriktige systemvedtak.

§ 49. *(offentliggjøring av kunstig intelligente systemer)*

Informasjon om det kunstig intelligente systemet skal gjøres offentlig tilgjengelig av organet som er ansvarlig for systemet etter § 50. Informasjonen skal være relevant, forklarbar og tilgjengelig.

§ 50. *(ansvarlig forvaltningsorgan)*

Ansvarlig forvaltningsorgan er det forvaltningsorganet som bruker et kunstig intelligent system i utøvelsen av sine forvaltningsoppgaver.

Dersom flere forvaltningsorganer utvikler eller bruker et kunstig intelligent system sammen, er forvaltningsorganene som utgangspunkt solidarisk ansvarlige for systemet, med mindre en annen ansvarsfordeling er avtalt.

§ 51. *(forskrift om utvikling og bruk av kunstig intelligente systemer)*

Kongen kan gi nærmere bestemmelser i forskrift om utvikling og bruk av kunstig intelligente systemer.

§ 52. *(om systemvedtak)*

Systemvedtaket skal begrunnes individuelt. Forvaltningsorganet skal gi parten begrunnelsen samtidig som systemvedtaket treffes.

Systemvedtaket skal inneholde informasjon om hvilke individuelle opplysninger om parten som danner grunnlag for systemvedtaket og hvordan disse har blitt vektlagt, med de begrensninger som følger av §§ 18 og 19.

Det skal føres forsvarlig menneskelig tilsyn med helautomatisert saksbehandling utført av kunstig intelligente systemer.

§ 53. *(klage på systemvedtak)*

Systemvedtak kan påklages av en part eller annen med rettslig klageinteresse i saken til klagenemnda for kunstig intelligente systemer (KI-klagenemnda). En klage på et systemvedtak kan fremsettes dersom det påstås at

- a) individuelle opplysninger som systemvedtaket er basert på, er uriktige.
- b) en rettslig systemavgjørelse som ligger til grunn for det anvendte kunstig intelligente systemet, er uriktig.
- c) det kunstig intelligente systemet er i strid med de krav og prinsipper som stilles til kunstig intelligente systemer etter dette kapitlet.

Klage basert på bokstav a skal i samsvar med §§ 32 og 33 fremmes for det organet som er ansvarlig for vedtaket. Dersom det er snakk om grunn som nevnt i første ledd bokstav a, skal systemet gis riktige individuelle opplysninger, og saken undergis ny systembehandling. Saken kan undergis menneskelig saksbehandling dersom særlige grunner tilsier det.

Klage basert på bokstav b og c kan fremmes direkte for nemnda. Dersom det er snakk om en grunn som nevnt i første ledd bokstav b eller c, må det kunstig intelligente systemet rettes, jf. § 54, før systemvedtaket undergis ny systembehandling.

Ny behandling av saken etter annet og tredje ledd forutsetter at feilen kan ha fått innvirkning på det påklagede systemvedtaket.

§ 54. *(retting av kunstig intelligente systemer)*

Dersom det foreligger en feil som nevnt i § 53 første ledd bokstav b eller c, skal det kunstig intelligente systemet rettes.

Tilsynsorganet for kunstig intelligente systemer skal rette feilen hvis forvaltningsorganet ikke kan rette feilen på egenhånd, eller det foreligger særlige grunner som tilsier at feilen bør rettes av tilsynsorganet.

§ 55. *(omgjøring av systemvedtak)*

Alle systemvedtak som er truffet av et kunstig intelligent system med feil som nevnt i § 53 første ledd bokstav b eller c skal omgjøres uten at det er påklaget dersom endringen ikke er til skade for noen som systemvedtaket retter seg mot eller direkte tilgodeser.

§ 56. (*tilsyn og kontroll*).

Klagenemnda for kunstig intelligente systemer i forvaltning (KI-klagenemnda) er et uavhengig forvaltningsorgan administrativt underordnet Kongen og departementet. Nemnda kan ikke instrueres om behandlingen av enkeltsaker eller om den faglige virksomheten for øvrig. Kongen og departementet kan ikke omgjøre nemndas vedtak.

Klagenemnda avgjør klager på systemvedtak fremsatt i henholdt til § 53 første ledd bokstav b og c. For klagesaker som hovedsakelig gjelder § 53 første ledd bokstav a, kan klagen behandles i henhold til særlovgivningen.

Kongen kan gi forskrift om klagenemndas organisering og saksbehandling.

Tilsynsorganet for kunstig intelligente systemer skal føre tilsyn og kontroll med kunstig intelligente systemer i forvaltningen.

Nåværende kapittel IX, X og XI blir kapittel X, XI og XII.

Nåværende §§ 43-52 blir § 57-65.

II

1. Loven gjelder fra den tid Kongen bestemmer.
2. Kongen kan gi overgangsbestemmelser.

8 Merknader til bestemmelsene

§ 2 første ledd bokstav c) og g) skal lyde:

- c) *systemvedtak, et enkeltvedtak generert av et kunstig intelligent system*
- g) *rettslig systemavgjørelse, en avgjørelse med rettslig innhold som treffes som ledd i utviklingen av kunstig intelligente systemer;*

Det skal inntas to nye definisjoner i forvaltningslovens § 2. Definisjonen av systemvedtak i bokstav c bygger på definisjonene av vedtak og enkeltvedtak i bokstav a og b, med den forskjellen at resultatet helt eller delvis er generert av et kunstig intelligent system. Loven legger til grunn den samme definisjonen av kunstig intelligente systemer som Lov om kunstig intelligens, som gjennomfører EUs forordning om kunstig intelligens.

En egen definisjon av rettslige systemavgjørelser er nødvendig for å kunne gi en mer treffsikker adgang til klage, omgjøring og retting av systemvedtak og systemer som sådan. Ved å la slike beslutninger defineres særskilt rettes det også oppmerksomhet mot det faktum at det underveis i utviklingen av systemer som skal avgi systemvedtak, er nødvendig å ta avgjørelser om hvordan gjeldende rett er å forstå, både gjennom tolkning og utfylling.

Til nytt Kapittel IX: Om rettslige systemavgjørelser, kunstig intelligente systemer og systemvedtak.

Kapittelet regulerer forvaltningens utvikling og bruk av kunstig intelligente systemer.

§ 43. (formål).

Kapittelet søker å ivareta borgernes rettigheter i møte med forvaltning utøvet ved bruk av kunstig intelligente systemer.

Paragrafen angir kapittelets formål. Hensikten er å understreke at kapittelet har borgernes grunnleggende rettigheter for øye, og at bestemmelsene skal tolkes i lys av dette.

§ 44. (prinsipper).

Kunstig intelligente systemer til bruk i forvaltningen skal være transparente og være utformet på en måte som ivaretar personvernet, en forsvarlig forvaltning av data, samt borgernes livskvalitet og velferd. Kunstig intelligente systemer skal være teknisk robuste og sikre.

Systemvedtak skal være rettferdige og saklige. Den individuelle begrunnelsen for systemvedtaket må være mulig for parten å forstå, vurdere og etterprøve.

Paragrafen er ny og angir kapittelets grunnleggende prinsipper. Prinsippene skal utgjøre utgangspunktet for tolkningen av bestemmelsene i resten av kapittelet, og ved enhver utvikling og bruk av kunstig intelligente systemer til bruk i forvaltningen.

Med «transparente» siktes det til at de kunstig intelligente systemene skal være forklarbare og etterprøvbare, og at systemene må anvendes med åpenhet.

Et forklarbart system er et system hvis virkemåte er mulig å gjøre forståelig gjennom en nærmere redegjørelse. Systemet skal være forklarbart overfor enhver som har befatning med systemet eller systemets resultat, herunder også brukere av systemet, som i forvaltningsmessig sammenheng som hovedregel vil være «part» i lovens forstand. Brukerne av systemene må kunne fortolke systemets output – altså systemet resultat – på en måte som gjør dem i stand til å anvende denne informasjonen, for eksempel i en vurdering av om de ønsker å benytte seg av adgangen til å klage. Loven stiller ikke krav om fullstendig transparens, men tilstrekkelig transparens. Hva som vil være tilstrekkelig transparent vil bero på den aktuelle risikoen systemet representerer for individenes grunnleggende rettigheter og friheter, og hvor komplekse algoritmer det er snakk om. Hvem som er mottaker for forklaringen, vil også stå sentralt.

At systemet er etterprøvbart innebærer at systemet er mulig å kontrollere i ettertid. Kravet fordrer at hele livssyklusen til systemet dokumenteres, slik at enhver forutsetning i systemet, enhver prediksjon som er foretatt også videre, er mulig å se tilbake på. Formålet med kravet om etterprøvbart er å minimere systemenes skadepotensiale. Prinsippet operasjonaliseres særlig i § 47 om krav til loggføring.

Kravet om at et kunstig intelligent system må anvendes med åpenhet innebærer at forvaltningen må gjøre det tydelig overfor brukeren at denne er i kontakt med et kunstig intelligent system.

Med «ivaretar personvernet» menes det respekt for og overholdelse av de til enhver tid gjeldene regler om personvern, herunder personopplysningsloven, personvernforordningen og annen regulering i særlovgivningen.

Data som anvendes ved utvikling og bruk av kunstig intelligente systemer skal forvaltes på en forsvarlig måte. Det innebærer for det første et kvalitetskrav, og for det andre et krav om hensiktsmessig håndtering. Kravet om datakvalitet innebærer at datagrunnlaget må være relevant, representativt, nøyaktig og fullstendig.

Med «relevant» menes at dataen må ha betydning for systemets funksjon. Inkludering av data som *ikke* er relevant kan bidra til at systemet tar i betraktning forhold som ikke bør virke inn på beslutningen. Det motvirker også kravet om transparens, idet det gjør det mulig for systemet å komplisere beslutningsprosessen ved å basere resultatet på flere forhold enn nødvendig. Kravet om relevans må i tillegg ses i sammenheng med personvernlovgivningens regler og prinsipper, spesielt om formålsbegrensning og dataminimering.

Med «representativt» menes at dataen må reflektere mangfoldet innenfor den kategorien som den tilhører. Med «nøyaktig» menes at dataen ikke må inneholde forstyrrende elementer eller feil. Med «fullstendig» menes at datagrunnlaget ikke må være mangelfullt.

Kravet om hensiktsmessig håndtering av data refererer til forvaltningens plikt til å etablere en strategi for forvaltning av data.

Kravet om ivaretagelse av «borgernes livskvalitet og velferd» understreker forvaltningens ansvar for borgernes individuelle og kollektive velferd samt samfunnet som helhet. Prinsippet skal tas hensyn til gjennom hele livsløpet til de kunstig intelligente systemene som forvaltningen ønsker å ta i bruk, inkludert ved vurderingen om et slikt system skal utvikles, anvendes og ekspanderes.

For det første skal det tas behørig hensyn til borgerne både som individer og som gruppe, herunder deres individuelle nåtidige og fremtidige behov og livskvalitet. Borgerne som gruppe kan både omfatte mindre grupper brukere som for eksempel blir direkte berørt av virksomheten som søkes gjennomført, men også større samfunnsgrupper eller befolkningen som helhet.

For det andre skal det tas behørig hensyn til samfunnet, herunder de verdiene og rettsstatsprinsippene som ligger til grunn for dette.

For det tredje er forvaltningen ansvarlig for å sikre bærekraftige og miljøvennlige løsninger. Dette kan blant annet gjøres ved innføring av miljø- og bærekraftstiltak under både utviklingen, distribusjonen og bruksprosessen, samt forsyningskjeden til de kunstig intelligente systemene som forvaltningen ønsker å ta i bruk.

Første ledd tredje punktum krever at kunstig intelligente systemer er «teknisk robuste og sikre». Dette sikres gjennom å sørge for passende nøyaktighet samt motstandsdyktighet overfor feil som kan oppstå i systemet og sårbarheter som kan utnyttes av tredjeparter.

Hva som regnes som passende nøyaktighet kan variere avhengig av systemets formål og kan også endre seg over tid. Nøyaktighet kan sikres blant annet gjennom forsvarlig forvaltning av data.

Motstandsdyktighet overfor feil som kan oppstå i systemet sikres ved å bygge teknisk robuste systemer. Systemene må for eksempel kunne takle at brukere interagerer med systemet på en uventet måte, herunder at språkkunnskap og kompetanse kan variere mellom individer. Det må også sikres god interaksjon med andre systemer som KI-systemet skal interagere med blant annet for å oppsøke informasjon.

Motstandsdyktighet overfor sårbarheter som kan utnyttes av tredjeparter sikres gjennom en passende grad av cybersikkerhet. Systemene må minimum være utformet på en måte som sikrer at data, herunder personopplysninger, myndighets- og bransjeinformasjon og intellektuell eiendom, sikres mot forutsigbare cyberangrep.

§ 45. (forholdet til annen lovgivning).

Et kunstig intelligent system skal være i tråd med krav som ellers stilles i formell lov, grunnlov og andre internasjonale forpliktelser.

Paragrafen understreker at forvaltningen ved utvikling og bruk av kunstig intelligente systemer også skal vurdere lovligheten opp mot annen formell lov, samt overordnede rettsnormer. Bestemmelsens hensikt er å klargjøre dette, og skal med det fungere som en påminnelse til forvaltningen.

§ 46. (overordnet krav til kunstig intelligente systemer til bruk i forvaltningen).

Ethvert kunstig intelligent system til bruk i forvaltningen skal oppfylle kravene som stilles til systemer som er klassifisert som høyrisiko-systemer etter EUs forordning om kunstig intelligens.

Bestemmelsen stiller krav om at alle kunstig intelligente systemer som brukes i forvaltningen skal oppfylle kravene til systemer som er klassifisert som høyrisiko-systemer i EUs forordning om kunstig intelligens. Begrunnelsen for denne bestemmelsen er at kunstig intelligente systemer som brukes av det offentlige i møte med borgerne vil innebære en betydelig risiko for borgernes rettigheter og plikter. Borgerne møter gjerne offentlige myndigheter i situasjoner hvor de er ekstra sårbare; enten når de henvender seg om offentlig støtte og hjelp, eller når de mottar vedtak som får innvirkning på deres liv og hverdag. Ved bruk av kunstig intelligens i slike sammenhenger er det spesielt viktig å sikre at borgernes rettigheter ivaretas, at de får riktig og adekvat hjelp og at avgjørelser blir truffet på en forsvarlig måte.

§ 47. (loggføring av tekniske hendelser i kunstig intelligente systemer).

Et kunstig intelligent system skal utvikles med funksjoner som muliggjør en automatisk registrering av tekniske hendelser (loggføring). Slik registrering av tekniske hendelser skal føres gjennom hele systemets livssyklus.

Registrering av tekniske hendelser (loggføring) skal sikre et tilstrekkelig nivå av etterprøvbarehet av systemets virkemåte i henhold til systemets tiltenkte formål.

Registreringen skal utformes på en måte som tar hensyn til behovet for en forsvarlig arkivering i henhold til arkivloven.

Paragrafen gjennomfører EUs forordning om kunstig intelligens artikkel 12 nr. 1 og 2. Bestemmelsen må ses i sammenheng med det grunnleggende kravet til transparens, som omfatter etterprøvbarehet.

Med «tekniske hendelser» siktes det til ethvert steg i systemets livssyklus, med andre ord enhver beslutning eller konklusjon algoritmen tar, samt alt datamateriale systemet baserer seg på. Bakgrunnen for at lovteksten presiserer at det er snakk om *tekniske* hendelser, er for å unngå forvirring med annen bruk av termen «hendelse», som blant annet i hendelsesorientert forvaltning.

Hva som vil være en «et tilstrekkelig nivå av etterprøvbarehet» må underlegges en konkret vurdering der man ser hen til hva systemet søker å oppnå, samt hvilken risiko systemet representerer overfor brukernes grunnleggende rettigheter. Poenget er at det skal være mulig å i ettertid prøve om systemet er i tråd med grunnleggende prinsipper og ellers er lovlig. En tilstrekkelig grad av etterprøvbarehet gjør en slik prøving mulig.

Bestemmelsens annet ledd benytter ordet «virkemåte» i stedet for forordningstekstens bruk av «funksjon» på dansk og «function» på engelsk. Bakgrunnen for det er at «virkemåte» er et mer brukt norsk ord, som etter utvalgets oppfatning i større grad beskriver det loven sikter til. Virkemåte er ment å ramme enhver teknisk hendelse systemet foretar seg, og enhver virkning

dette har og har hatt. I bestemmelsens tredje ledd presiseres det at bestemmelsens må ses i sammenheng med arkivlova.

§ 48. (krav til datakvalitet i kunstig intelligente systemer).

Datagrunnlaget til et kunstig intelligent system skal være av tilstrekkelig kvalitet. Datagrunnlaget skal være representativt og sikre treffsikre og rettsriktige systemvedtak.

Bestemmelsen stiller krav til datakvalitet, og må leses i sammenheng med prinsippbestemmelsen i lovens § 44. Datagrunnlaget skal være representativt, i den forstand at det skal reflektere mangfoldet i samfunnet og blant befolkningen som det kunstig intelligente systemet skal interagere med. Dette er viktig for å sikre at systemet ikke diskriminerer og kan anvendes av ulike befolkningsgrupper. Videre må datagrunnlaget bidra til treffsikre og rettsriktige systemvedtak. Med treffsikre siktes her til at vedtakene må være riktige for den personen de gjelder; de må ta i betraktning de relevante forholdene, legge vekt på de hensynene som gjør seg gjeldende og treffe avgjørelsen for riktig person. Med rettsriktige menes at vedtakene må treffes i overensstemmelse med gjeldende rett, både rettsregler og praksis. Det er også viktig at personvernet og grunnleggende menneskerettigheter respekteres og ivaretas.

§ 49. (offentliggjøring av kunstig intelligente systemer).

Informasjon om det kunstig intelligente systemet skal gjøres offentlig tilgjengelig av organet som er ansvarlig for systemet etter § 50. Informasjonen skal være relevant, forklarbar og tilgjengelig.

Paragrafen oppfyller forvaltningens aktive informasjonsplikt, som går ut på å «legge forholdene til rette for en åpen og opplyst offentlig samtale», jf. Grunnloven § 100 sjette ledd. Plikten muliggjør demokratisk kontroll av forvaltningen. Den aktive informasjonsplikten er ikke presisert i lovgivningen, og forvaltningslovutvalget av 2019 la til grunn at forvaltningens veiledningsplikt i stor grad ivaretar informasjonsbehovet, se NOU 2019:5 s. 245. Med denne bestemmelsens gis imidlertid forvaltningen en plikt til å offentliggjøre systemet.

For at forvaltningen skal oppfylle kravet om tilgjengelighet som presisert i bestemmelsens annet punktum, bør det kunstig intelligente systemet være offentlig tilgjengelig på forvaltningsorganets nettsider. Brukeren skal kunne teste sin sak i systemet i forkant av innsendt søknad. Som tidligere trukket frem, vil dette være ressursbesvarende for forvaltningen ved at behovet for veiledning vil bli mindre, i tillegg til at det vil gi økt forutberegnelighet for borgeren.

Det er grunn til å merke seg at offentlighetskravet i § 49 går lenger enn de krav som stilles i offentleglova. Offentleglova bidrar til å realisere Grunnloven § 100 femte ledd første punktum og forvaltningen passive informasjonsplikt, og gjelder innsyns krav. Offentlighetsplikten i § 49 gjelder offentlighet overfor allmennheten, for slik å kunne legge til rette for et deltakende og deliberativt demokrati.

Informasjonen skal videre være relevant og forklarbar. Dette må ses i sammenheng med at det offentlige «skal kommunisere på eit klart og korrekt språk som er tilpassa målgruppa», jf. språkløva § 9. For at det offentlige skal kunne benytte seg av informasjonen må det være mulig å kunne vurdere systemets virkemåte. Full tilgang til programkoden vil for de aller fleste være uforståelig, og kan videre komme i konflikt med grunnleggende hensyn til personvern hva gjelder data systemet har benyttet. Dette taler for at forvaltningen plikter å utarbeide forklaringer som er tilstrekkelige for å oppfylle bestemmelsens formål om å legge til rette for en åpen og opplyst samtale.

Med «informasjon» siktes det med andre ord til en offentliggjøring av selve systemet, supplert med forklaringer av systemets virkemåte der dette er nødvendig for at brukeren skal kunne benytte seg av informasjonen.

§ 50. (ansvarlig forvaltningsorgan).

Ansvarlig forvaltningsorgan er det forvaltningsorganet som bruker et kunstig intelligent system i utøvelsen av sine forvaltningsoppgaver.

Dersom flere forvaltningsorganer utvikler eller bruker et kunstig intelligent system sammen, er forvaltningsorganene som utgangspunkt solidarisk ansvarlige for systemet, med mindre en annen ansvarsfordeling er avtalt.

For å kunne føre effektiv kontroll med forvaltningen er det sentralt at det er mulig å ansvarliggjøre et forvaltningsorgan, og med det unngå ansvarspulverisering. Utvikling og bruk av kunstig intelligente systemer vil som utgangspunkt involvere mange parter. Bestemmelsens angir ansvarlig forvaltningsorgan.

Med «i utøvelsen av sine forvaltningsoppgaver» siktes det både til forvaltningens egen bruk av systemet som beslutningsstøtte, samt tilgjengeliggjøring av systemet for brukere, slik kravet til offentlighet i § 49 gir anvisning om.

§ 51. (forskrift om utvikling og bruk av kunstig intelligente systemer).

Kongen kan gi nærmere bestemmelser i forskrift om utvikling og bruk av kunstig intelligente systemer.

Etter denne paragrafen gis Kongen en hjemmel til å i forskrift gi nærmere regler for utvikling av kunstig intelligente systemer. Utvalget ser for seg at dette vil omfatte regler om organisering, kompetanse, høring og rettslig deltakelse.

§ 52. (om systemvedtak).

Systemvedtaket skal begrunnes individuelt. Forvaltningsorganet skal gi parten begrunnelsen samtidig som systemvedtaket treffes.

Systemvedtaket skal inneholde informasjon om hvilke individuelle opplysninger om parten som danner grunnlag for systemvedtaket og hvordan disse har blitt vektlagt, med de begrensninger som følger av §§ 18 og 19.

Det skal føres forsvarlig menneskelig tilsyn med helautomatisert saksbehandling utført av kunstig intelligente systemer.

Bestemmelsen stiller grunnleggende materielle krav til systemvedtak i første og andre ledd, og krav til menneskelig tilsyn med helautomatisert saksbehandling utført av kunstig intelligente systemer i tredje ledd.

Ifølge *første ledd* skal et systemvedtak være individuelt begrunnet. I dette ligger at systemvedtaket må rette seg mot riktig person, det vil si vedkommende som har interesse i å få sin rettsstilling avklart. Videre må systemvedtaket bygge på riktig og relevant informasjon. Det stiller for det første et krav om at den enkelte skal være informert om hvilket informasjonsgrunnlag beslutningen treffes på, slik at vedkommende kan endre eller supplere opplysningene ved behov. For det andre må systemet ikke vektlegge forhold som ikke er relevante for den aktuelle saken. Til slutt må systemvedtaket inneholde en tilstrekkelig begrunnelse i form av informasjon om hvilke forhold som har blitt vektlagt i beslutningsprosessen og hvorfor. Dette er også eksplisitt uttrykt i bestemmelsens *andre ledd*. Forvaltningsorganet som bruker det kunstig intelligente systemet, skal gi parten begrunnelsen samtidig som systemvedtaket treffes.

Bestemmelsens *andre ledd* refererer til forvaltningsorganets informasjonsplikt og begrunnelsesplikt. Forvaltningsorganet skal sikre at parten får informasjon om hvilke individuelle opplysninger om vedkommende som danner grunnlag for systemvedtaket, og hvordan disse har blitt vektlagt, med de begrensninger som følger av §§ 18 og 19. Det innebærer at forvaltningsorganet må ha innsikt i datagrunnlaget for det aktuelle systemet, samt hvilke faktorer systemet tar i betraktning ved beslutningstakingen. Forvaltningsorganet må i tillegg kunne identifisere hvilke opplysninger som knytter seg til den aktuelle parten. Denne informasjonen må gis til parten på en forståelig måte, slik at vedkommende får et grunnlag for å utøve sine rettigheter til klage og omgjøring.

Bestemmelsens *tredje ledd* pålegger forvaltningen en plikt til å sikre at det føres menneskelig tilsyn med helautomatisert saksbehandling som utføres av kunstig intelligente systemer. Plikten er i overensstemmelse med personvernforordningens artikkel 22, og forordningen om kunstig intelligens artikkel 14. Med «forsvarlig» menneskelig tilsyn menes at tilsynet må være effektivt, i den forstand at tilsynet må være egnet til å påse at systemvedtakene er i tråd med de overordnede prinsipper.

§ 53. (klage på systemvedtak).

Systemvedtak kan påklages av en part eller annen med rettslig klageinteresse i saken til klagenemnda for kunstig intelligente systemer (KI-klagenemnda). En klage på et systemvedtak kan fremsettes dersom det påstås at

- a. individuelle opplysninger som systemvedtaket er basert på, er uriktige.***
- b. en rettslig systemavgjørelse som ligger til grunn for det anvendte kunstig intelligente systemet, er uriktig.***

c. det kunstig intelligente systemet er i strid med de krav og prinsipper som stilles til kunstig intelligente systemer etter dette kapittelet.

Klage basert på bokstav a skal i samsvar med §§ 32 og 33 fremmes for det organet som er ansvarlig for vedtaket. Dersom det er snakk om grunn som nevnt i første ledd bokstav a, skal systemet gis riktige individuelle opplysninger, og saken undergis ny systembehandling. Saken kan undergis menneskelig saksbehandling dersom særlige grunner tilsier det.

Klage basert på bokstav b og c kan fremmes direkte for nemnda. Dersom det er snakk om en grunn som nevnt i første ledd bokstav b eller c, må det kunstig intelligente systemet rettes, jf. § 54, før systemvedtaket undergis ny systembehandling.

Ny behandling av saken etter annet og tredje ledd forutsetter at feilen kan ha fått innvirkning på det påklagede systemvedtaket.

Paragrafen angir hvem og på hvilket grunnlag en klage på et systemvedtak kan fremsettes. De klageberettigede, som vil være parten eller annen med rettslige klageinteresse, skal forstås likt som etter § 28 første ledd.

Med «individuelle opplysninger» i bokstav a siktes det til de opplysninger som systemet er gitt i behandlingen av den konkrete saken. Begrepet skal etter dette ikke forstås likt som personvernforordningen artikkel 4 nr. 1, selv om «individuelle opplysninger» ofte ville falle innenfor denne definisjonen.

§ 54. (retting av kunstig intelligente systemer).

Dersom det foreligger en feil som nevnt i § 53 første ledd bokstav b eller c, skal det kunstig intelligente systemet rettes.

Tilsynsorganet for kunstig intelligente systemer skal rette feilen hvis forvaltningsorganet ikke kan rette feilen på egenhånd, eller det foreligger særlige grunner som tilsier at feilen bør rettes av tilsynsorganet.

Bestemmelsen innebærer at forvaltningen ikke kan benytte seg av et system der det er oppdaget feil. Systemet skal være i tråd med de overordnede prinsippene og annen lov, og det vil stride mot denne lov dersom forvaltningen skulle fortsette å benytte seg av et system etter at det er påvist feil.

§ 55. (omgjøring av systemvedtak).

Alle systemvedtak som er truffet av et kunstig intelligent system med feil som nevnt i § 53 første ledd bokstav b eller c skal omgjøres uten at det er påklaget dersom endringen ikke er til skade for noen som systemvedtaket retter seg mot eller direkte tilgodeser.

Bestemmelsens oppbygning er i tråd med § 35 om omgjøring av enkeltvedtak. En stor fordel ved bruken av systemvedtak er at omgjøring kan skje svært raskt. Kravet om loggføring innebærer at forvaltningen vil ha kontroll på hvilke systemvedtak som er rammet av feilen,

slik at disse kan kjøres gjennom systemet etter at feil som nevnt i § 53 bokstav b og c er rettet i tråd med § 54. Forvaltningen vil etter dette kunne «masseomgjøre» alle rammede systemvedtak, og slik rette uriktige systemvedtak på en effektiv og rettssikker måte.

§ 56. (tilsyn og kontroll).

Klagenemnda for kunstig intelligente systemer i forvaltning (KI-klagenemnda) er et uavhengig forvaltningsorgan administrativt underordnet Kongen og departementet. Nemnda kan ikke instrueres om behandlingen av enkeltsaker eller om den faglige virksomheten for øvrig. Kongen og departementet kan ikke omgjøre nemndas vedtak.

Klagenemnda avgjør klager på systemvedtak fremsatt i henholdt til § 53 første ledd bokstav b og c. For klagesaker som hovedsakelig gjelder § 53 første ledd bokstav a, kan klagen behandles i henhold til særlovgivningen.

Kongen kan gi forskrift om klagenemndas organisering og saksbehandling.

Tilsynsorganet for kunstig intelligente systemer skal føre tilsyn og kontroll med kunstig intelligente systemer i forvaltningen.

Paragrafen gir regler om klagenemnda for kunstig intelligente systemer i forvaltningen (KI-klagenemnda).

I *første ledd* gis det regler om nemndas organisasjon og uavhengighet.

I *annet ledd* følger regler om nemndas oppgaver og kompetanse. Nemnda skal etter annet ledd første punktum avgjøre klager som fremsettes på grunnlag av § 53 første ledd bokstav b og c. Det følger av annet ledd annet punktum at klager som hovedsakelig gjelder § 53 første ledd bokstav a kan behandles enten av det klageorgan som etter den aktuelle særlovgivningen er tildelt myndighet som klageinstans eller av klagenemnda. Med «hovedsakelig» menes en vesentlig del av klagen.

Det følger av *tredje ledd* at klagenemndas organisering og saksbehandling fastsettes av Kongen.

Fjerde ledd fastslår at tilsynsorganet for kunstig intelligens har kompetanse til å føre tilsyn og kontroll med de kunstig intelligente systemene som forvaltningen tar i bruk. Hvilket organ dette gjelder og utformingen av tilsynet vil følge av lov om kunstig intelligens som gjennomfører EU-forordningen om kunstig intelligens.

9 Kilder

9.1 Norske rettskilder

Lover og forskrifter

1814	Lov 17. mai 1814 Kongeriket Norges Grunnlov (Grunnloven)
1967	Lov 10. februar 1967 om behandlingsmåten i forvaltningssaker (forvaltningsloven)
1992	Lov 27. desember 1992 nr. 109 om gjennomføring i norsk rett av hoveddelen i avtale om Det europeiske økonomiske samarbeidsområde (EØS) m.v. (EØS-loven)
1992	Lov 04. desember 1992 nr. 126 om arkiv (arkivlova)
1997	Lov 28. februar 1997 nr. 19 om folketrygd (folketrygdloven)
1999	Lov 21. mai 1999 nr. 30 om styrking av menneskerettighetenes stilling i norsk rett (menneskerettsloven)
2004	Forskrift 25. juni 2004 nr. 988 om elektronisk kommunikasjon med og i forvaltningen (eForvaltningsforskriften)
2006	Lov 19. mai 2006 nr. 16 om rett til innsyn i dokument i offentlig verksemd (offentleglova)
2006	Lov 16. juni 2006 nr. 20 om arbeids- og velferdsforvaltningen (arbeids- og velferdsforvaltningsloven) [NAV-loven]
2006	Forskrift 15. desember 2006 nr. 1456 til forvaltningsloven (forvaltningslovforskriften)
2009	Lov 19. juni 2009 nr. 100 om forvaltning av naturens mangfold (naturmangfoldloven)
2009	Lov 18. desember 2009 nr. 131 om sosiale tjenester i arbeids- og velferdsforvaltningen (sosialtjenesteloven)
2017	Lov 16. juni 2017 nr. 51 om likestilling og forbud mot diskriminering (likestillings- og diskrimineringsloven)
2017	Forskrift 21. juni 2017 nr. 859 om konsekvensutredninger
2018	Lov 15. juni 2018 nr. 38 om behandling av personopplysninger (personopplysningsloven)
2021	Lov 21. mai 2021 nr. 42 om språk (språklova)

Forarbeider

NUT 1958: 3	<i>Innstilling fra Komiteen til å utrede spørsmålet om mer betryggende former for offentlig forvaltning</i>
St. meld. nr. 32 (1976–1977)	<i>Rettsikkerhet i sosial- og helseinstitusjoner</i>
Ot.prp.nr. 3 (1976–1977)	<i>Om lov om endringer i lov 10. februar 1967 om behandlingen i forvaltningssaker (regler om taushetsplikt m.m.)</i>
Ot.prp. nr. 77 (1991–92)	<i>Om lov om arkiv</i>
NOU 1999: 27	<i>«Ytringsfrihet bør finde Sted» – Forslag til ny Grunnlov § 100</i>
St. meld. nr. 17 (2002–2003)	<i>Om statlige tilsyn</i>
NOU 2003: 30	<i>Ny offentlighetslov</i>
NOU 2009: 14	<i>Et helhetlig diskrimineringsvern – Diskrimineringslovutvalgets utredning om en samlet diskrimineringslov, grunnlovsværn og ratifikasjon av tilleggsprotokoll nr. 12 til EMK</i>
Dokument 16 (2011–2012)	<i>Rapport til Stortingets presidentskap fra Menneskerettighetsutvalget om menneskerettigheter i Grunnloven</i>
Prop. 1 S (2011–2012)	<i>For budsjettåret 2012 – Utgiftskapitler: 1–2, 1500–1592 og 2445 Inntektskapitler: 4510–4592 og 5445–5446</i>
Meld. St. 5 (2012–2013)	<i>EØS-avtalen og Norges øvrige avtaler med EU</i>
Meld. St. 23 (2012–2013)	<i>Digital agenda for Norge – IKT for vekst og verdiskaping</i>
Meld. St. 27 (2015–2016)	<i>Digital agenda for Norge – IKT for en enklere hverdag og økt produktivitet</i>
Prop. 56 LS (2017–2018)	<i>Lov om behandling av personopplysninger (personopplysningsloven) og samtykke til deltakelse i en beslutning i EØS-komiteen om innlemmelse av forordning (EU) nr. 2016/679 (generell personvernforordning) i EØS-avtalen</i>
NOU 2019: 5	<i>Ny forvaltningslov – Lov om saksbehandlingen i offentlig forvaltning (forvaltningsloven)</i>
NOU 2019: 9	<i>Fra kalveskinn til datasjø – Ny lov om samfunnsdokumentasjon og arkiver</i>
Prop. 108 L (2019–2020)	<i>Lov om språk (språklova)</i>
Prop. 135 L (2019–2020)	<i>Endringer i arbeids- og velferdsforvaltningsloven, sosialtjenesteloven, lov om Statens pensjonskasse og enkelte andre lover (behandling av personopplysninger)</i>
Meld. St. 22 (2020–2021)	<i>Data som ressurs – Datadrevet økonomi og innovasjon</i>

Styringsverktøy

Finansdepartementet. *Instruks om utredning av statlige tiltak (utredningsinstruks)*. 01.03.2016. Dokumentnr. FOR-2016-02-19-184 [Lest i Lovdata Pro].

Direktoratet for forvaltning og økonomistyring (DFØ). *Veileder til utredningsinstruks*. [u/å]. <https://dfo.no/fagomrader/utredning/veileder-til-utredningsinstruks>

Uttalelser m.m.

Kommunal- og moderniseringsdepartementet. *Én digital offentlig sektor. Digitaliseringsstrategi for offentlig sektor 2019-2025*. 11. juni 2019.

<https://www.regjeringen.no/no/tema/statlig-forvaltning/ikt-politikk/digitaliseringsstrategi-for-offentlig-sektor/id2612415/>

Departementene. *Nasjonal strategi for digital sikkerhet*. 2019.

<https://www.regjeringen.no/contentassets/c57a0733652f47688294934ffd93fc53/nasjonal-strategi-for-digital-sikkerhet.pdf>

Schartum, Dag Wiese. *Høringsuttalelse til NOU 2019:5 (2019)*.

<https://www.regjeringen.no/no/dokumenter/horing---nou-2019-5-ny-forvaltningslov/id2652945/Download/?vedleggId=60b4ef70-4584-4330-aec8-16e45dd400e3>

Kommunal- og moderniseringsdepartementet. *Nasjonal strategi for kunstig intelligens*. 2020.

<https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/>

UiO ved Senter for rettsinformatikk. *Kommentarer til EU-kommisjonens forslag til forordning om kunstig intelligens*. 09.06.2021.

<https://www.regjeringen.no/contentassets/02907dbb11ba45c6bfa2c7e0912793a9/uio-seri.pdf>

Datatilsynet. *Innspill fra Datatilsynet til EU-kommisjonens forslag til regulering av kunstig intelligens*. 10.06.2021.

<https://www.regjeringen.no/contentassets/02907dbb11ba45c6bfa2c7e0912793a9/datatilsynet.pdf>

Kommunal- og moderniseringsdepartementet. *Norwegian Position Paper on the European Commission's Proposal for a Regulation 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)*. 2021. Kan lastes ned her:

https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665314_en

Digitaliseringsdirektoratet (Digdir). *Innspill fra Digdir til norsk posisjonsnotat: EU-kommisjonens utkast til forordning for regulering av kunstig intelligens*. [2021?]

<https://www.regjeringen.no/contentassets/02907dbb11ba45c6bfa2c7e0912793a9/digdir.pdf>

9.2 Internasjonale rettskilder

Traktater

EMK	<i>Konvensjon om beskyttelse av menneskerettighetene og de grunnleggende friheter, Roma 4. november 1950. [Offisiell norsk oversettelse].</i>
SP	<i>Internasjonal konvensjon om sivile og politiske rettigheter, New York 16. desember 1966. [Offisiell norsk oversettelse].</i>
ØSK	<i>Internasjonal konvensjon om økonomiske, sosiale og kulturelle rettigheter, New York 16. desember 1966. [Offisiell norsk oversettelse].</i>
EU Charter	<i>Charter of Fundamental Rights of the European Union (2016/C 202/02)</i>

EU-direktiver og forordninger

Forordning (EF) 300/2008	<i>Europaparlaments- og rådsforordning (EF) nr. 300/2008 av 11. mars 2008 om felles bestemmelser om sikkerhet i sivil luftfart og om oppheving av forordning (EF) nr. 2320/2002.</i>
Direktiv 2014/90/EU	<i>Europaparlaments- og rådsdirektiv 2014/90/EU av 23. juli 2014 om skipsutstyr og om oppheving av rådsdirektiv 96/98/EF om skipsutstyr [Skipsutstyrsdirektivet 2014].</i>
Forordning (EU) 2016/679	<i>Europaparlaments- 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) [GDPR]</i>
Forordning (EU) 2018/1807	<i>Kommisjonsforordning (EU) nr. 2018/1807 av 14. november 2018 om et rammeverk for fri flyt av ikke-personlig data i den Europeiske Union.</i>
Forordning (EU) 2019/881	<i>Kommisjonsforordning (EU) nr. 2019/881 av 17. april 2019 om ENISA (Den Europeiske Unions Byrå for Cybersikkerhet), om cybersikkerhetssertifisering av informasjons- og kommunikasjonsteknologi og om oppheving av forordning (EU) nr. 526/2013 (forordningen om cybersikkerhet).</i>
Direktiv 2019/1024	<i>Europaparlaments- og rådsdirektiv (EU) nr. 2019/1024 av 20. juni 2019 om åpne data og viderebruk av informasjon fra offentlig sektor.</i>

Forarbeider

COM(2010) 2020	<i>Communication from the Commission, Europe 2020: A strategy for smart, sustainable and inclusive growth.</i>
COM(2018) 795	<i>Communication From The Commission To The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions: Coordinated Plan on Artificial Intelligence.</i>
COM(2020) 65	<i>WHITE PAPER On Artificial Intelligence – A European approach to excellence and trust [Offisiell engelsk tittel]</i> <i>HVIDBOG om kunstig intelligens - en europæisk tilgang til ekspertise og tillid [Offisiell dansk tittel]</i>
COM(2021) 205	<i>Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Fostering a European approach to Artificial Intelligence.</i>
COM(2021) 206	<i>Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts [Offisiell engelsk tittel]</i> <i>Forslag til Europa-parlamentets og Rådets forordning om harmoniserede regler for kunstig intelligens (retsakten om kunstig intelligens) og om ændring af visse af unionens lovgivningsmæssige retsakter [Offisiell dansk tittel]</i>
2021/0106(COD)	<i>Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – Presidency compromise text</i>

Rettspraksis

EMD	Marckx v. Belgium (No. 6833/74)
EMD	Söderman v. Sweden (No. 5786/08)
EFTA	E-2/03 Ásgeirsson

Uttalelser m.m.

High-Level Expert Group on Artificial Intelligence (AI HLEG). *A definition of AI: Main capabilities and scientific disciplines*. 18. desember 2018.

https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

High-Level Expert Group on Artificial Intelligence (AI HLEG). *Ethics Guidelines for Trustworthy AI*. 8. april 2019. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

OECD. *Recommendation of the Council on Artificial Intelligence*. 22. mai 2019. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

European Parliament. *Framework of ethical aspects of artificial intelligence, robotics and related technologies*. 20. Oktober 2020. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020IP0275>

European Parliament, Directorate-General for Internal Policies of the Union. *The role of artificial intelligence in the European Green Deal*. Mai 2021. Dokumentnr. PE 662.906 <https://data.europa.eu/doi/10.2861/882830>

Bundesrat Deutschland (Forbundsrådet Tyskland). *Beschluss des Bundesrates - Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union COM(2021) 206 final* [Norsk oversettelse: Vedtak fra Forbundsrådet - Forslag til en forordning fra Europaparlamentet og rådet om fastsettelse av harmoniserte regler for kunstig intelligens (forordning om kunstig intelligens) og om endring av visse EU-rettsakter COM(2021) 206 final]. 2021. Dokumentnr. Drucksache 488/21 (Beschluss).

[https://www.europarl.europa.eu/RegData/docs_autres_institutions/parlements_nationaux/com/2021/0206/DE_BUNDESRAT_CONT1-COM\(2021\)0206_DE.pdf](https://www.europarl.europa.eu/RegData/docs_autres_institutions/parlements_nationaux/com/2021/0206/DE_BUNDESRAT_CONT1-COM(2021)0206_DE.pdf).

European Parliamentary Research Service (EPRS). *Briefing EU Legislation in Progress – Artificial intelligence act*. Januar 2022. Dokumentnr. PE 698.792.

[https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf).

9.3 Litteratur

Bøker og artikler

Almestad, Knut. «The Notion of ‘Opting Out’». I *The Handbook of EEA Law*. Carl Baudenbacher red., 1.utg., Cham: Springer International Publishing, 2016, s. 85-94.

Baur, Georges. «Decision-Making Procedure and Implementation of New Law». I *The Handbook of EEA Law*. Carl Baudenbacher red., 1.utg., Cham: Springer International Publishing, 2016, s. 45-68.

Beaudouin, Valérie, Isabelle Bloch og David Bounie mfl. «Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach». *Telecom Paris, Institut Polytechnique de Paris*. (2020).

Bekkelund, Anne Siri Koksrud. «Algoritmer bør revideres.» (2020) <https://teknologiradet.no/algoritmer-bor-revideres/>

Bernt, Jan Fridtjof og Ørnulf Rasmussen. *Frihagens forvaltningsrett Bind I*. 1.utg., Bergen: Vigmostad & Bjørke AS, 2003.

Bing, Jon. «Automatiseringsvennlig lovgivning». *Tidsskrift for rettsvitenskap* (1977) s. 195-229.

- Boe, Erik Magnus. *Innføring i juss Bind 2*. 1.utg., Oslo: Tano Aschehoug, 1993.
- Bragdø-Ellenes, Sunniva Cristina og Terje Abusland. «Forvaltningsrett i et nøtteskall». Oslo: Gyldendal Norsk Forlag AS, 2017.
- Busuioc, Madalina. «Accountable Artificial Intelligence: Holding Algorithms to Account». *Public Administration Review Volume 81* (2021) s. 825-836.
- Bygrave, Lee A. «Minding the Machine v2.0: The EU General Data Protection Regulation and Automated Decision Making.» I *Algorithmic Regulation*. Karen Yeung og Marting Lodge red., 1. utg., Oxford: Oxford University Press, 2019.
- Choroszewicz, Marta og Beata Mäihäniemi. «Developing a Digital Welfare State: Data Protection and the Use of Automated Decision-Making in the Public Sector across Six EU Countries.» *Global Perspectives* 1 (1) (2020). <https://doi.org/10.1525/gp.2020.12910>
- Eckhoff, Torstein og Eivind Smith. *Forvaltningsrett*. 10.utg., Oslo: Universitetsforlaget, 2014.
- Eng, Svein. *Rettsfilosofi*. 1.utg., Oslo: Universitetsforlaget, 2007.
- Fertik, Michael. «Why Your Data Will Never Be Deleted.» (2015) <https://www.forbes.com/sites/michaelfertik/2015/06/09/why-your-data-will-never-be-deleted/?sh=181dec7d2371>
- Frihagen, Arvid. Lov og rett 1972/1, Årgang 11 s. 3.
- Graver, Hans Petter. *Alminnelig forvaltningsrett*. 5.utg., Oslo: Universitetsforlaget, 2019.
- Hallinan, Dara, Ronald Leenes og Paul de Hert (red.). *Data Protection and Privacy: Data Protection and Artificial Intelligence*. Oxford: Hart Publishing, 2021. <http://dx.doi.org/10.5040/9781509941780>
- Jarbekk, Eva I. E. og Simen Sommerfeldt. *Personvern og GDPR i praksis*. Oslo: Cappelen Damm akademisk, 2019.
- Kerrigan, Charles (red.). *Artificial Intelligence: Law and Regulation*. London: Edward Elgar Publishing, 2022.
- Korsgaard, Haakon og Jens Ludvigsen. «Rapport – Den digitale borger: En kvalitativ studie av den digitale hverdagen – for Kommunal- og moderniseringsdepartementet.» (2019) https://www.regjeringen.no/contentassets/9265ea2cbfde47e09d43c1b863e2c5bf/den_digitale_borger.pdf
- Kuner, Christopher, Lee A. Bygrave, Christopher Docksey mfl. *The EU General Data Protection Regulation (GDPR): A Commentary*. Oxford: Oxford University Press USA, 2020.
- Mahanti, Rupa. *Data Governance and Data Management: Contextualizing Data Governance Drivers, Technologies, and Tools*. 2021.
- Ohnstad, Bente. *Velferd, rettsikkerhet og personvern*. 2. utg, Bergen: Fagbokforlaget, 2021.
- Olsen, Johan P. «Accountability and Ambiguity». In *The Oxford Handbook of Public Accountability*, edited by Mark Bovens, Robert E. Goodin, and Thomas Schillemans. Oxford: Oxford University Press, 2014, s. 106-124.
- Schartum, Dag Wiese. «Om den offentlige forvaltnings edb-programmer, deres rettslige innhold og stilling som forskrift». *Tidsskrift for Rettsvitenskap* (1989), s. 650-678.
- Schartum, Dag Wiese. «Automatisk virkende rettsregler – om pling-juss og systemrettssikkerhet i elektronisk forvaltning» I Stub, Marius & Kraby, Ida Hjort

(Red.), Forsker og formidler. Festskrift til Erik Magnus Boe på 70-årsdagen 17. april 2013. Universitetsforlaget (2013), s. 253–266.

Schartum, Dag Wiese. *Digitalisering av offentlig forvaltning*. 1.utg., Oslo Fagbokforlaget, 2018.

Schartum, Dag Wiese. *Jus og digitalisering*, Lov og Rett 2021/2, Juridika.

Schartum, Dag Wiese. *Rettsikkerhet i systemutvikling og offentlig forvaltning*. 1993.

Skoghøy, Jens Edvin A. *Rett og rettsanvendelse*. 1.utg., Oslo: Universitetsforlaget, 2018.

Sovrano, Francesco, Salvatore Sapienza, Monica Palmirani mfl. «Metrics, Explainability and the European AI Act Proposal». *Multidisciplinary Scientific Journal* (2021).

Tennøe, Tore og Robindra Prabhu. «Kunstig intelligens: smart eller skremmende?» (2017) <https://www.aftenposten.no/viten/i/AnoWr/kunstig-intelligens-smart-eller-skremmende>

Thorgersen, Anders. «Digitale tjenester og brukernes tillit til NAV». *Arbeid og velferd nr. 3* (2017), s. 97-109.

Tolan, Songül. «Fair and Unibased Algorithmic Decision Making: Current State and Future Challenges». JRC Digital Economy Working Paper 2018-10 (2018).

Veale, Michael og Frederik Zuiderveen Borgesius. «Demystifying the Draft EU Artificial Intelligence Act. *Computer Law Review International* (2021).

Wachter, Sandra, Brent Mittelstadt og Luciano Floridi. «Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation». *International Data Privacy Law* (2017).

Rapporter

Helsedirektoratet. *Well-being på norsk*. Oslo: Juli 2015.

https://www.helsedirektoratet.no/rapporter/well-being-pa-norsk/Well-being%20p%C3%A5%20norsk.pdf/_attachment/inline/46a66c5f-e872-4e52-96b5-4ae1c95c5d23:488beb667da23e74e06e64a4e800417c2f205c90/Well-being%20p%C3%A5%20norsk.pdf

World Economic Forum. *The Global Risks Report 2017*. Genève: 11. januar 2017.

https://www3.weforum.org/docs/GRR17_Report_web.pdf

Teknologirådet. *Saken forklart: Kunstig intelligens*. Oslo: 09. desember 2017

https://teknologiradet.no/wp-content/uploads/sites/105/2019/01/SF_Kunstig-intelligens_web.pdf

Teknologirådet. *Kunstig Intelligens – muligheter, utfordringer og en plan for Norge*. Oslo: September 2018. <https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf>

Datatilsynet. *Kunstig intelligens og personvern*. Oslo: Januar 2018.

<https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>

Regjeringen. *Klima, digitalisering og likestilling: Dette er målene til den nye Europakommisjonen*. Oslo: 9. desember 2019.

<https://www.regjeringen.no/no/dokumenter/i/id2681600/>

World Economic Forum. *The Global Risks Report 2021: 16th Edition*. Genève: 2021.

https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf

Datatilsynet. *Sluttrapport fra sandkasseprosjektet med NAV*. Oslo: Januar 2022.
<https://www.datatilsynet.no/link/81ae9d7b5ebb47e688b66923afcef5ce.aspx/download>

European Commission. *Digital Economy and Society Index 2021: Norway*. 2021.
<https://ec.europa.eu/newsroom/dae/redirection/document/80543>

Norges institusjon for menneskerettigheter (NIM). *Ny teknologi og menneskerettigheter*. Oslo: 28. mars 2022. <https://www.nhri.no/wp-content/uploads/2022/03/Ny-teknologi-og-menneskerettigheter.pdf>

Oxford Insights. *Government AI Readiness Index 2021*. 21. januar 2022.
https://static1.squarespace.com/static/58b2e92c1e5b6c828058484e/t/61ead0752e7529590e98d35f/1642778757117/Government_AI_Readiness_21.pdf

Andre publikasjoner:

- | | |
|-------------|---|
| JOU 2020: 2 | <i>Automatisk kontroll i Arbeids- og velferdsetaten</i> |
| JOU 2020: 5 | <i>Kunstig intelligens – Muligheter og risikoer i velferdsforvaltningen</i> |

9.4 Annet

Grassegger, Hannes og Mikael Krogerus. «Dataene som snudde verden på hodet» (04.02.2017) <https://nrkbeta.no/2017/02/04/dataene-som-snudde-verden-pa-hodet/> lest 21.05.2022.

Valevatn, Joakim og Tore Tennøe. «Slik overvåker maskiner følelsene dine» (10.03.2018) <https://teknologiradet.no/overvaker-maskiner-folelsene-dine/> lest 21.05.2022.

Kommunal- og distriktsdepartementet. «Åpenhet og tillit» (09.10.2018) <https://www.regjeringen.no/no/tema/statlig-forvaltning/forvaltningsutvikling/apenhet-og-tillit/id2612413/> lest 18.05.2022.

World Economic Forum. *Will the Future Be Human?* [Videoklipp]. (2018) <https://webcasts.weforum.org/widget/1/davos2018?p=1&pi=1&hl=english&id=a0Wb000000ALGyjEAF&auto=1>

Sample, Ian. «What are deepfakes, and how can you spot them?» (2020). Hentet fra: <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>

Bertrand, Arnauld, Julie McQueen. «How can digital government connect citizens without leaving the disconnected behind?» (24.02.2021) https://www.ey.com/en_uk/government-public-sector/how-can-digital-government-connect-citizens-without-leaving-the-disconnected-behind lest 18.05.2022.

Digitaliseringsrådet. «Slik vil Digitaliseringsrådet få offentlige virksomheter til å jakte på fornyelse.» (2021) <https://www.digdir.no/digitaliseringsradet/slik-vil-digitaliseringsradet-fa-offentlige-virksomheter-til-jakte-pa-fornyelse/2936>

Regjeringen. «Forslag til forordning om kunstig intelligens (KI-forordningen).» (2021) <https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2021/juni/forslag-til-forordning-om-kunstig-intelligens-ki-forordningen/id2884935/>

Sæbø, Solve. «How to make AI more sustainable» (10.03.2021a) <https://blogg.nmbu.no/solvesabo/2021/03/how-to-make-ai-sustainable/> lest 20.05.2022.

Sæbø, Solve. «Innlegg: Dagens kunstige intelligens er ikke bærekraftig» (08.04.2021b) <https://www.dn.no/teknologi/teknologi/kunstig-intelligens/barekraft/innlegg-dagens-kunstige-intelligens-er-ikke-barekraftig/2-1-989508> lest 20.05.2022.

SINTEF. «Kunstig intelligens i klimakampen» (10.06.2021) <https://www.sintef.no/fagomrader/kunstig-intelligens/kunstig-intelligens-i-klimakampen/> lest 20.05.2022.

Bull, Mathew. «Exclusive: What is data poisoning and why should we be concerned?» (13.09.2021) <https://internationalsecurityjournal.com/what-is-data-poisoning/> lest 19.05.2022.

Store Norske Leksikon. «data» (13.09.2021) <https://snl.no/data> lest 12.03.2022.

Store Norske Leksikon. «velferd» (21.12.2021) <https://snl.no/velferd> lest 20.05.2022.

Store Norske Leksikon. «rettferdighet» (20.05.2021) <https://snl.no/rettferdighet> lest 12.03.2022.

Kiseleva, Anastasya. «Making AI's transparency transparent: notes on the EU proposal for the AI Act». (2021). Hentet fra: <https://europeanlawblog.eu/2021/07/29/making-ais-transparency-transparent-notes-on-the-eu-proposal-for-the-ai-act/>

Stub, Marius. «Karnov lovkommentar til forvaltningsloven § 11». I *Lovdata Pro* (2021).

Store Norske Leksikon. «Dyplæring». (2021) <https://snl.no/dyplæring>

Store Norske Leksikon. «Ekspertsystem.» (2021) <https://snl.no/ekspertsystem>

Direktoratet for forvaltning og økonomistyring (DFØ). «OECD-rapport om tillit: Bare én av to nordmenn mener myndighetene gjør nok for å møte fremtidige utfordringer (15.03.2022) <https://dfo.no/nyhetsarkiv/oecd-rapport-om-tillit-bare-en-av-to-nordmenn-mener-myndighetene-gjor-nok-for-a-mote-fremtidige-utfordringer> lest 18.05.2022.

Store Norske Leksikon. «Maskinlæring» (18.01.2022) <https://snl.no/maskinlæring> lest 15.03.2022.

Digitaliseringsdirektoratet (Digdir). «Hva er et datasett og hvilke datasett skal beskrives?» (u.å) <https://www.digdir.no/informasjonsforvaltning/hva-er-et-datasett-og-hvilke-datasett-skal-beskrives/2199>

Digitaliseringsstyrelsen. «Nu igangsættes yderligere tolv signaturprojekter med kunstig intelligens.» (2022) <https://digst.dk/nyheder/nyhedsarkiv/2022/januar/nu-igangsættes-yderligere-tolv-signaturprojekter-med-kunstig-intelligens/>

Digitaliseringsdirektoratet. «Innsatsområder i handlingsplanen.» (u.å.a) <https://www.digdir.no/handlingsplanen/innsatsomrader-i-handlingsplanen/2593>

Digitaliseringsdirektoratet. «Arbeidet med de sju livshendelsene.» (u.å.b) <https://www.digdir.no/sammenhengende-tjenester/arbeidet-med-de-sju-livshendelsene/1170>

Digitaliseringsdirektoratet. «Sammenhengende tjenester med brukeren i sentrum.» (u.å.c) <https://www.digdir.no/handlingsplanen/sammenhengende-tjenester-med-brukeren-i-sentrum/1255>

Digitaliseringsdirektoratet. «Økt deling av data og verdiskaping.» (u.å.d) <https://www.digdir.no/handlingsplanen/okt-deling-av-data-og-verdiskaping/1228>

Digitaliseringsdirektoratet. «Klart og digitaliseringsvennlig regelverk.» (u.å.e) <https://www.digdir.no/handlingsplanen/klart-og-digitaliseringsvennlig-regelverk/1227>

Digitaliseringsdirektoratet. «Felles økosystem for nasjonal digital samhandling og tjenesteutvikling.» (u.å.f) <https://www.digdir.no/handlingsplanen/felles-okosystem-nasjonal-digital-samhandling-og-tjenesteutvikling/1256>

Digitaliseringsdirektoratet. «Styring og samordning for en mer sammenhengende offentlig sektor.» (u.å.g) <https://www.digdir.no/handlingsplanen/styring-og-samordning-en-mer-sammenhengende-offentlig-sektor/1258>

Digitaliseringsdirektoratet. «Styrket samarbeid med privat sektor.» (u.å.h) <https://www.digdir.no/handlingsplanen/styrket-samarbeid-med-privat-sektor/1354>

Digitaliseringsdirektoratet. «Økt kompetanse i offentlig sektor.» (u.å.j) <https://www.digdir.no/handlingsplanen/okt-digital-kompetanse-i-offentlig-sektor/1296>

Digitaliseringsdirektoratet. «Digital sikkerhet.» (u.å.k) <https://www.digdir.no/handlingsplanen/digital-sikkerhet/1263>

Digitaliseringsdirektoratet. «Hva er DIGITAL?» (u.å.l) <https://www.digdir.no/internasjonalt-samarbeid/hva-er-digital/3198>

Europa-Parlamentet. «Europa-Parlamentets Forskningstjeneste (EPRS).» (u.å) <https://www.europarl.europa.eu/at-your-service/da/stay-informed/research-and-analysis> lest 09.05.2022.

Europalov. «KI-forordningen om europeisk regelverk for kunstig intelligens.» (u.å.) <https://www.europalov.no/rechtsakt/ki-forordningen-om-europeisk-regelverk-for-kunstig-intelligens/id-28218>

European Commission. «The Digital Economy and Society Index (DESI).» (u.å.) <https://digital-strategy.ec.europa.eu/en/policies/desi>

European Commission. «The Digital Economy and Society Index – Countries’ performance in digitisation.» (2022) <https://digital-strategy.ec.europa.eu/en/policies/countries-digitisation-performance>

The European Commission. «The Digital Europe Programme.» (u.å) <https://digital-strategy.ec.europa.eu/en/activities/digital-programme> lest 09.05.2022.

The European Parliamentary Research Service (EPRS). «About.» (u.å.a) <https://epthinktank.eu/about/> lest 09.05.2022.

The European Parliamentary Research Service (EPRS). «EU Legislation in Progress.» (u.å.b) <https://epthinktank.eu/eu-legislation-in-progress/> lest 09.05.2022.

Haga, Atle. «Datasentre: Den nye kraftkrevende industrien» (u.å) <https://www.statkraft.no/nyheter/nyheter-og-pressemeldinger/arkiv/2018/datasentre-den-nye-kraftkrevende-industrien/> lest 20.05.2022.

Intromat. «Relapse prevention for bipolar disorder» (u.å) <https://intromat.no/cases/bipolardisorder/> lest 21.05.2022.

Investopedia. «Accountability» (31.10.2021) <https://www.investopedia.com/terms/a/accountability.asp>

Institutt for offentlig rett UiO. «Trygdeskandalen» (u.å.) <https://www.jus.uio.no/ior/forskning/prosjekter/trygdeskandalen/>

Visma. «Why is cybersecurity important?» (u.å.) <https://www.visma.com/cyber-security/why-is-cyber-security-important/> lest 19.05.2022.

Magiega, Tambiama. «Artificial intelligence act.» (2022) [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

Clarke, Laurie. «MEPs are preparing to debate Europe's AI Act. These are the most contentious issues.» (2021) <https://techmonitor.ai/policy/meps-are-preparing-to-debate-europes-ai-act-these-are-the-most-contentious-issues> hentet 05.03.2022

Barcentewicz, Mikolaj og Benjamin Mueller. «More Than Meets The AI: The Hidden Costs of a European Software Law.» (2021) <https://www2.datainnovation.org/2021-more-than-meets-the-ai.pdf>

Kayser-Bril, Nicolas. «European Council and Commission in agreement to narrow the scope of the AI Act.» (2021) <https://algorithmwatch.org/en/eu-narrow-scope-of-ai-act/>

Bryson, Joanna. «Europe Is in Danger of Using the Wrong Definition of AI.» (2022) <https://www.wired.com/story/artificial-intelligence-regulation-european-union/>

Digitaliseringsdirektoratet. «Hva er digital transformasjon?» (u.å.) <https://www.digdir.no/innovasjon/hva-er-digital-transformasjon/1589>

Sansone, Dario og Anna Zhu. «Using Machine Learning to Create an Early Warning System for Welfare Recipients.» (2021) <https://docs.iza.org/dp14377.pdf>

Raghu, Maithra, Katy Blumer, Greg Corrado, Jon Kleinberg, Ziad Obermeyer og Sendhil Mullainathan. «The algorithmic automation problem: Prediction, triage, and human effort.» (2019) <https://arxiv.org/pdf/1903.12220.pdf>

Kuziemski, Maciej og Gianluca Misuraca. «AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings.» (2020) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7164913/>

Obermeyer, Ziad, Brian Powers Christine Vogelli og Sendhil Mullainathan. «Dissecting racial bias in an algorithm used to manage health of populations.» (2019) <https://doi.org/10.1126/science.aax2342>

Larson, Jeff, Surya Mattu, Lauren Kirchner og Julia Angwin. «How We Analyzed the COMPAS Recidivism Algorithm.» (2016) <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

Kayser-Bril, Nicolas. «Female historians and male nurses do not exist, Google Translate tell sits European users.» (2020) <https://algorithmwatch.org/en/google-translate-gender-bias/>

European Commission. «New Rules for Artificial Intelligence – Questions and Answers.» (2021) https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683

European Commission. «A Europe fit for the digital age.» (u.å.) https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age_en

Statista Research Department. «Amount of data created, consumed and stored 2010-2025.» (2022) <https://www.statista.com/statistics/871513/worldwide-data-created/>

Fremtind. «Nordmenn er skeptiske til kunstig intelligens.» (2021) <https://kommunikasjon.ntb.no/pressemedling/nordmenn-er-skeptiske-til-kunstig-intelligens?publisherId=17846913&releaseId=17920945>

European Commission. «Member States and Commission to work together to boost artificial intelligence “made in Europe”.» (2018a)

https://ec.europa.eu/commission/presscorner/detail/en/IP_18_6689

European Commission. «Commission appoints expert group on AI and launches the European AI Alliance.» (2018b) <https://digital-strategy.ec.europa.eu/en/news/commission-appoints-expert-group-ai-and-launches-european-ai-alliance>

von der Leyen, Ursula. «A Union that strives for more: My agenda for Europe.» (2018)

https://ec.europa.eu/info/sites/default/files/political-guidelines-next-commission_en_0.pdf

European Commission. «Shaping Europe’s digital future.» (2019)

https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/shaping-europe-digital-future_en

European Commission. «New rules for Artificial Intelligence – Questions and Answers.» (2021) https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683

(EØS-notat) Regjeringen. «Forslag til forordning om kunstig intelligens (KI-forordningen).» (2021) <https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2021/juni/forslag-til-forordning-om-kunstig-intelligens-ki-forordningen/id2884935/>

Kop, Mauritz. «EU Artificial Intelligence Act: The European Approach to AI.» (2021)

<https://www-cdn.law.stanford.edu/wp-content/uploads/2021/09/2021-09-28-EU-Artificial-Intelligence-Act-The-European-Approach-to-AI.pdf>

Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021.

<https://law.stanford.edu/publications/eu-artificial-intelligence-act-the-european-approach-to-ai/>

Ipek, Mikal. «EU Draft Artificial Intelligence Regulation: Extraterritorial Application and Effects.» (2022) <https://europeanlawblog.eu/2022/02/17/eu-draft-artificial-intelligence-regulation-extraterritorial-application-and-effects/>

European Data Protection Board. Guidelines 05/2020 on consent under Regulation 2016/679. 2020.

https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf

Åste Marie Bergseng Skullerud mfl., *Personvernforordningen. Lovkommentar*, [Artikkel 22. Automatiserte individuelle avgjørelser, herunder profilering](#), *Juridika* (kopierte 07. juni 2022)

Osloregionens Europakontor. «Ny måling: Europa gjør digitale fremskritt – Norge på en femteplass.» (2021) <https://osloregion.org/no/hva-skjer/siste-nytt/948-ny-maling-europa-gjor-digitale-fremskritt-norge-pa-en-femteplass>

Sanner, Jan Tore. «Digital kommunikasjon hovedregel – Viktig informasjon om endringer i forvaltningsloven og eForvaltningsforskriften.» (2014) Kommunal- og distriktsdepartementet

https://www.regjeringen.no/no/dokument/dep/kdd/andre-dokumenter/brev/utvalgte_brev/2014/digital-kommunikasjon-hovedregel/id751013/

Digitalisering i offentlig sektor (2014) Kommunal- og distriktsdepartementet

<https://www.regjeringen.no/no/tema/statlig-forvaltning/ikt-politikk/digitaliseringen-i-offentlig-sektor/id2340245/>

Én digital offentlig sektor: Digitaliseringsstrategi for offentlig sektor 2019-2025. (2019)

Kommunal- og distriktsdepartementet <https://www.regjeringen.no/no/tema/statlig-forvaltning/ikt-politikk/digitaliseringsstrategi-for-offentlig-sektor/id2612415/>

Bjornsson, Thorsteinn. «EU Programmes with EEA EFTA Participation.» (u.å.)
<https://www.efta.int/eea/eu-programmes>

Osloregionens Europakontor. «Nytt EU-regelverk for kunstig intelligens.» (2021)
<https://osloregion.org/no/hva-skjer/siste-nytt/843-nytt-eu-regelverk-for-kunstig-intelligens>

Digitaliseringsstyrelsen. «Ny FN-måling: Danmark er fortsatt verdensmestre i offentlig digitalisering.» (2020) <https://digst.dk/nyheder/nyhedsarkiv/2020/juli/ny-fn-maaling-danmark-er-fortsat-verdensmestre-i-offentlig-digitalisering/>

Digitaliseringsstyrelsen. «Kommuner og regioner skal afprøve kunstig intelligens for at løfte kvaliteten i den offentlige service.» (2019)
<https://digst.dk/nyheder/nyhedsarkiv/2019/oktober/kommuner-og-regioner-skal-afproeve-kunstig-intelligens-for-at-loefte-kvaliteten-i-den-offentlige-service/>

Signaturprosjekt (2021a) <https://videncenter.kl.dk/media/26094/soenderborg-kommune-hurtigere-sagsbehandling-ved-bevilling-af-kropsbaarne-hjaelpemidler.pdf>

Signaturprosjekt (2021b) <https://videncenter.kl.dk/media/26092/roskilde-kommune-digital-inklusion-og-support-ved-talegenkendelse.pdf>

Digitaliseringsstyrelsen. «Algoritmer på arbejde: 12 nye projekter med kunstig intelligens flytter ind på rådhus og hospitaler.» (2021)
<https://digst.dk/nyheder/nyhedsarkiv/2021/september/algoritmer-paa-arbejde-12-nye-projekter-med-kunstig-intelligens-flytter-ind-paa-raadhuse-og-hospitaler/>

Digitaliseringsstyrelsen. «Digitaliseringsstyrelsen, KL og Danske Regioner opsamler erfaringer og utfordringer med kunstig intelligens i det offentlige.» (2021)
<https://digst.dk/nyheder/nyhedsarkiv/2021/april/digitaliseringsstyrelsen-kl-og-danske-regioner-opsamler-erfaringer-og-udfordringer-med-kunstig-intelligens-i-det-offentlige/>

European Commission. «Ethics guidelines for trustworthy AI.» (2019) <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Digitaliseringsdirektoratet. «Kva er informasjonsforvaltning?» (u.å.)
<https://www.digdir.no/informasjonsforvaltning/kva-er-informasjonsforvaltning/2116>

Kultur- og likestillingsdepartementet. «Kulturdepartementets strategi for åpne data.» (2017)
<https://www.regjeringen.no/no/dokumenter/kulturdepartementets-strategi-for-apne-data/id2576038/>

Felles datakatalog. «Åpne data.» (u.å.) <https://data.norge.no/search-all?q=åpne%20data>

Digitaliseringsdirektoratet. «Hva er digital transformasjon?» (u.å.)
<https://www.digdir.no/innovasjon/hva-er-digital-transformasjon/1589>

Holte, Hans Christian. «Viktige avklaringer mangler før vi kan gjøre data til gull.» (2019)
<https://www.digi.no/artikler/kommentar-viktige-avklaringer-mangler-for-vi-kan-gjore-data-til-gull/464878>

Digitaliseringsdirektoratet. «Rammeverk for informasjonsforvaltning.» (u.å.)
<https://www.digdir.no/informasjonsforvaltning/rammeverk-informasjonsforvaltning/2118>
<https://data.norge.no/specification/spesifikasjon-for-beskrivelse-av-kvalitet-pa-datasett/#dimensjonene>
<https://data.norge.no/guide/veileder-kvantifiserbar-kvalitet/#kvalitet-kan-males-pa-ulike-nivaer>

Oslo kommune. «Datakvalitet.» (u.å.) <https://oslokommune.github.io/dataplattform/om-data/datakvalitet>

European Union Agency for Fundamental Rights. «Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights.» (2019) https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf

Dataforeningen. «2 grep for en fremtidssikker dataarkitektur.» (2019) <https://www.dataforeningen.no/2-grep-for-en-fremtidssikker-dataarkitektur/>

Store Norske Leksikon. «Database.» (2019) <https://snl.no/database>

Hagen, Jan-Erik. «Kvalitetsmåling av ansattdata.» Masteroppgave, Master i informasjonssikkerhet, Institutt for informatikk og medieteknikk, Høgskolen i Gjøvik, 2005

Osloregionens Europakontor. «EU ønsker en innovativ og rettferdig dataøkonomi.» (2022) <https://osloregion.org/no/hva-skjer/siste-nytt/1007-eu-onsker-en-innovativ-og-rettferdig-dataokonomi>

Person identification, human rights and ethical principles: Rethinking biometrics in the era of artificial intelligence. Study by the Panel for the Future of Science and Technology, EPRS (2021). [https://www.europarl.europa.eu/stoa/en/document/EPRS_STU\(2021\)697191](https://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2021)697191)

European Commission, Directorate-General for Justice and Consumers, Gerards, J., Xenidis, R. (2021) *Algorithmic discrimination in Europe : challenges and opportunities for gender equality and non-discrimination law*. Publications Office. <https://data.europa.eu/doi/10.2838/544956>

European Commission. «Why do we need the Charter?» (u.å.) https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights/why-do-we-need-charter_en

Council of the European Union <https://www.consilium.europa.eu/media/46496/st11481-en20.pdf>

Council of the European Union, Presidency conclusions - The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change, 11481/20, 2020. Se punkt 11.

Kompetanse Norge. «Befolkningens digitale kompetanse og deltakelse.» (2021) <https://www.kompetansenorge.no/statistikk-og-analyse/publikasjoner/befolkningens-digitale-kompetanse-og-deltakelse/>