VF-rapport nr. 7-2022

Bruk av kunstig intelligens i offentlig sektor og risiko for diskriminering

Kunnskapsgrunnlag for arbeidet med å forebygge diskriminerende effekter ved bruk av kunstig intelligens i offentlig virksomhet

Hilde G. Corneliussen, Aisha Iqbal, Gilda Seddighi og Rudolf Andersen

VESTLANDSFORSKING

RAMBØLL

VF-rapport	7-2022
Dato	16. desember 2022
Antall sider	80
Utgitt av **Adresse**	Vestlandsforsking Postboks 163, 6851 Sogndal
Prosjekt	Rambøll Management Consulting og Vestlandsforsking, på oppdrag fra Barne-, ungdoms- og familiedirektoratet (Bufdir)I
Prosjekttittel	Bruk av kunstig intelligens i offentlig sektor og risiko for diskriminering
Prosjektleder	Rudolf Andersen, Rambøll Management Consulting
Prosjektdeltakere	Hilde G. Corneliussen, Gilda Seddighi og Rajendra Akerkar, Vestlandsforsking Aisha Iqbal, Rambøll Management Consulting
Oppdragsgiver	Barne-, ungdoms- og familiedirektoratet (Bufdir)
På fremsiden	Illustrasjon under lisens Rambøll Management Consulting fra stock.adobe.com
ISBN	978-82-428-0453-2

www.vestforsk.no

Innholdsfortegnelse

Forord

Prosjektet *Bruk av kunstig intelligens i offentlig sektor og risiko for diskriminering* er gjennomført på oppdrag fra Barne-, ungdoms- og familiedirektoratet (Bufdir) av Rambøll Management Consulting i samarbeid med Vestlandsforsking i perioden november 2021 til desember 2022.

Vi ønsker å takke alle samarbeidspartnere til prosjektet, særlig referansegruppen bestående av Heather Broomfield fra Digitaliserings-direktoratet, professor Nicola Marsden fra Hochschule Heilbronn, Kari Laumann fra Datatilsynet, Kathinka Theodore Aakenes-Vik fra Likestillings- og diskrimineringsombudet og professor Ingunn Ikdahl fra juridisk fakultet, Universitetet i Oslo. I tillegg har vi fått god hjelp fra Alex Moltzau, Birte Hansen og Klas H. Pettersen i NORA (Norwegian Artificial Intelligence Research Consortium), professor Morten Goodwin fra Universitetet i Agder, IBM Fellow og leder for IBMs *AI Ethics Global* Francesca Rossi, Frank Vevle fra Bufdir og deltakere i Fagforum for Kunstig intelligens i offentlig sektor i regi av Digitaliseringsdirektoratet.

Vi retter også en stor takk til alle bidragsytere som har svart på spørreundersøkelsen og stilt opp til intervju i prosjektet!

Oslo/Sogndal

16. desember 2022

Rudolf Andersen

Hilde G. Corneliussen

Oppsummering

Kunstig intelligens (heretter KI) har gjort betydelige fremskritt de siste årene. Det forventes at KI vil få stor betydning for vår evne til å løse både små og store samfunnsutfordringer, og at det vil påvirke både privat og offentlig sektor i tillegg til enkeltindivid. For norsk offentlig sektor anses KI som et sentralt verktøy i reisen mot en enda bedre og mer effektiv offentlig sektor. Å ta i bruk KI innebærer imidlertid også risikoer, deriblant risiko for feil, urettferdig eller diskriminerende resultater.

Vårt oppdrag har vært å fremskaffe *et bedre kunnskapsgrunnlag for arbeidet med å forebygge diskriminerende effekter ved bruk av kunstig intelligens i offentlig virksomhet*. Studien har kartlagt bruk av KI i offentlig sektor, mulige risikomoment, oppmerksomhet rundt risiko for diskriminering samt hva som gjøres for å forebygge slik risiko. Vi har gjennomført en spørreundersøkelse og dybdeintervjuer for å kartlegge status i arbeidet med KI og for å identifisere relevante KI-prosjekter og aktiviteter i offentlig sektor. Vi har primært fokusert på KI-prosjekter som involverer persondata, som kan danne grunnlag for diskriminering.

Formålet med KI-prosjektene vi har kartlagt varierer fra virksomhet til virksomhet. Et fåtall av KI-systemene som er i bruk, involverer persondata. Disse prosjektene har ulike formål, fra å forbedre kvalitet på datagrunnlaget til et KI-prosjekt, avdekke mistenkelige mønstre i et KI-system, og til prediksjon av brukeratferd og støtte til saksbehandlere. De fleste KI-prosjektene er forankret i toppledelsen, og de fleste gjennomføres med vekt på tverrfaglighet, altså ikke i et teknisk miljø alene. Majoriteten av respondentene er skeptiske til at KI kan ta beslutninger uten menneskelig innblanding. Flere virksomheter har teknisk kompetanse om kunstig intelligens, og flertallet har kompetanse om personvern og kunnskap om retningslinjer og lovverk tilknyttet KI-prosjekter. Samtidig er det usikkerhet knyttet til lovverkenes relevans og handlingsrommet de gir for KI. Spørreundersøkelsen og intervjuene viser sammen at det er varierende grad av kunnskap om risiko for diskriminering, og kun 3 % mener at KI kan *øke* slik risiko.

Det er flere utfordringer på reisen fra ide til produksjon i et KI-prosjekt. Det er bred enighet om at ledelsesnivået har behov for kompetanseheving om KI. Store utfordringer er knyttet til data, fra tilgang til tillatelser og kvalitet – en rekke kritiske spørsmål for offentlige virksomheter som er underlagt et særlig strengt regime for bruk av persondata. Bred og tverrfaglig kompetanse er kritisk for å lykkes med KI-prosjekter, og særlig for små virksomheter kan dette by på utfordringer.

Basert på studien har vi utviklet et *rammeverk for å identifisere risiko for diskriminering når KI tas i bruk i offentlig sektor*. Her har vi lagt vekt på at *den vanligste situasjonen* i offentlig sektor i 2022, var *å ikke ha et KI-system som inkluderer persondata i bruk*. Rammeverket henvender seg bredt til ulike kompetansebehov og inkluderer tekniske, kulturelle, politiske og juridiske vurderinger som må håndteres fra start til slutt i et KI-prosjekt. Rapporten avsluttes med *anbefalinger for å forebygge diskriminerende effekter når KI tas i bruk i offentlig sektor*.

English Summary

Artificial intelligence (AI) has made significant progress in recent years. It is expected that AI will have a major impact on solving small and large social challenges, and that it will affect private as well as public sector and individuals. AI is considered an important tool for improving and making the Norwegian public sector more efficient. Adopting AI, however, also entails risks, including the risk of incorrect, unfair or discriminatory results.

The purpose of this study was to provide *a better knowledge base for preventing discriminatory effects of AI in public sector*. The study has mapped AI projects in the public sector, possible risk factors of AI in public sector, and efforts to prevent such risk. We have conducted a survey and in-depth interviews to map the status of AI projects and activities in the public sector, with a focus on AI projects involving information about individuals and the use of personal data.

The purpose of the AI projects we have mapped varies, and we found a small number of AI systems already in use that also involve personal data. These projects have different purposes, from improving the quality of data for an AI project, uncovering suspicious patterns, and predicting user behavior and support for decision making and case management. Most AI projects are anchored in top management, and most involve interdisciplinary groups, i.e. they are not isolated to a technical environment. The majority of respondents are skeptical to AI making decisions without human intervention. Some organisations have technical expertise in artificial intelligence, and the majority have expertise in privacy and knowledge of guidelines and legislation associated with AI projects, while many are uncertain of how well current legislation fits the conditions of AI. There is a varying degree of knowledge about the risk of discrimination, and only 3% believe that AI can increase such risk.

We identified several challenges on the journey from idea to production in an AI project. There is broad agreement that managers in the sector need more competence about AI, while major challenges are related to data, from access to permissions and quality – involving critical issues for public sector organisations that are subject to a strict regime when dealing with personal data. Broad and interdisciplinary expertise is identified as critical for successfull AI projects, which can present challenges, particularly for small organisations.

Based on the study, we have developed a *framework to identify the risk of discrimination when AI is used in the public sector*. We have emphasized the most common situation in the public sector in 2022, which was *not having an AI system involving personal data in use*. The framework addresses a wide range of competences needed in AI development and includes technical, cultural, political and legal issues that must be handled from start to finish in an AI project. The report concludes with *recommendations to prevent discriminatory effects when AI is used in the public sector*.

1. Innledning

Det er store forventninger til at digitalisering og nye teknologier skal gi positive effekter for samfunn, næringsliv og offentlig sektor i Norge så vel som internasjonalt. Digitalisering skjer i dag på tvers av bransjer og sektorer, og vil ifølge *digitaliseringsstrategien for offentlig sektor* bidra til "en mer effektiv offentlig sektor, mer verdiskaping i næringslivet og en enklere hverdag for folk flest".[1] Det er særlig store forventninger knyttet til at kunstig intelligens (KI) kan fornye og effektivisere en rekke samfunnsområder og -oppgaver. Regjeringen ønsker økt bruk av kunstig intelligens i offentlig forvaltning, understreket i *Nasjonal strategi for kunstig intelligens*, og foreslår at "KI-systemer skal legge til rette for inkludering, mangfold og likebehandling".[2] Mens strategien oppfordrer til å ta i bruk KI som beslutningsstøtte for offentlig saksbehandling, i samhandlingen mellom det offentlige og borgerne og som del av et offentlig digitalt tjenestetilbud, påpekes det også at bruk av KI involverer utfordringer og "vanskelige spørsmål". Erfaring fra Europa og USA har også vist at KI som beslutningsstøtte kan ha svært uheldige diskriminerende effekter for individer og grupper.[3]

Utfordringen vi diskuterer i denne rapporten er nettopp risikoen for at KI skal ha diskriminerende effekter når teknologien tas i bruk i offentlig sektor i Norge. Mens prinsipper for KI som teknologi kan være de samme i både privat og offentlig sektor, bringer KI noen særlige utfordringer for offentlig sektor. Det offentlige er underlagt strenge lover og regler for håndtering av persondata, er

[1] Kommunal- og moderniseringsdepartementet (2019), *Én digital offentlig sektor: Digitaliseringsstrategi for offentlig sektor 2019–2025*, https://www.regjeringen.no/no/dokumenter/en-digital-offentlig-sektor/id2653874/

[2] Kommunal- og moderniseringsdepartementet (2020), *Nasjonal strategi for kunstig intelligens*, https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/

[3] Suresh, H., & Guttag, J. (2021), A framework for understanding sources of harm throughout the machine learning life cycle. *EAAMO 2021 – Equity and access in algorithms, mechanisms, and optimization* (1-9).

pliktig til å sikre likestilte offentlige tjenester overfor en mangfoldig befolkning, og borgere kan ikke velge en annen tjenestetilbyder. Det betyr at risiko ved å bruke KI er kritisk å håndtere for offentlig sektor, for å unngå at KI skal forsterke eksisterende diskriminering, skape nye diskrimineringsgrunnlag eller redusere borgeres tillit til offentlig sektor.

Både Europarådet og EU er opptatt av å utvikle regulerende rammeverk for å møte de utfordringene som kunstig intelligens innebærer, inkludert å beskytte mot diskriminerende effekter.[4, 5] Norge følger EU og forslag til forordning om kunstig intelligens, og KI-forordningen som kommisjonen la frem i april 2021, har vært på høring i Norge.[6]

1.1 Om oppdraget

Denne rapporten presenterer en studie som er gjennomført på oppdrag fra Barne-, ungdoms- og familiedirektoratet (Bufdir), med formål å fremskaffe et bed*re kunnskapsgrunnlag for arbeidet med å forebygge diskriminerende effekter ved bruk av kunstig intelligens i offentlig virksomhet*.

Studien inkluderer en spørreundersøkelse og dybdeintervjuer med offentlige virksomheter, med mål å besvare følgende problemstillinger:

- I hvilken grad har offentlige virksomheter tatt i bruk, eller planlegger å bruke, kunstig intelligens?
- Hvilke risikoer for diskriminering finnes i det offentliges bruk av kunstig intelligens?
- I hvilken grad er offentlige instanser oppmerksomme på diskrimineringsrisiko, og hva gjør de for å forebygge slik risiko?

[4] The Artificial Intelligence Act, EU, https://artificialintelligenceact.eu/
[5] Council of Europe, Council of Europe and Artificial Intelligence, https://www.coe.int/en/web/artificial-intelligence
[6] Regjeringen (2021), *Forslag til forordning om kunstig intelligens (KI-forordningen)*, https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2021/juni/forslag-til-forordning-om-kunstig-intelligens-ki-forordningen/id2884935/

Basert på disse spørsmålene samt annen kunnskap om disse utfordringen, skal vi skissere anbefalinger for hvordan offentlige virksomheter kan forebygge diskriminerende effekter når kunstig intelligens tas i bruk.

Et stort antall statlige og kommunale virksomheter fra ulike sektorer (fra helse- og omsorgssektoren, utdanningssektoren, arbeids- og velferdsforvaltningen, skatteetaten til toll og politi) ble invitert til å svare på en spørreundersøkelse for å rapportere bruk og planer om bruk av kunstig intelligens. For å favne bredt ble deltakerne oppfordret til å svare i forhold til alle relevante områder, inkludert forvaltning (saksbehandling/kontroll), digital kommunikasjon til befolkningen, tjenesteytelse, undervisning og andre aktuelle formål. For å fange risiko for diskriminering ble deltakerne deretter sortert i forhold til hvorvidt persondata inngikk i bruken av KI. Undersøkelsen la til grunn likestillings- og diskrimineringslovens diskrimineringsgrunnlag for å etablere en forståelse av diskriminering.[7] Virksomheter som hadde KI-prosjekt (eller planer) som omfattet persondata, ble så invitert til dybdeintervju.

I dette kapitlet vil vi starte med å definere henholdsvis kunstig intelligens og diskriminering, før vi ser nærmere på den internasjonale og nasjonale konteksten for regulering av KI. Til sist i kapitlet presenteres metode og gjennomføring av prosjektet. Kapittel 2 presenterer hovedtrekkene og et overblikk over status for KI i offentlig sektor slik det kommer frem gjennom undersøkelsen. Kapittel 3 drøfter utfordringer generelt, mens kapittel 4 drøfter risiko for diskriminering spesielt. Her blir et rammeverk for å forstå risiko for diskriminering ved bruk av KI i offentlig sektor presentert, slik dette ser ut i dag. Basert på studien presenterer kapittel 5 anbefalinger til tiltak for å håndtere risiko for diskriminering ved bruk av KI i offentlig sektor.

[7] § 1.Formål: «Lovens formål er å fremme likestilling og hindre diskriminering på grunn av kjønn, graviditet, permisjon ved fødsel eller adopsjon, omsorgsoppgaver, etnisitet, religion, livssyn, funksjonsnedsettelse, seksuell orientering, kjønnsidentitet, kjønnsuttrykk, alder og andre vesentlige forhold ved en person.» https://lovdata.no/dokument/NL/lov/2017-06-16-51

1.2 Kunstig intelligens

Den nasjonale strategien for kunstig intelligens definerer kunstig intelligens (KI) slik:

> *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.*[8]

Kunstig intelligens (KI) hadde fra sin opprinnelse på 1950- og 1960-tallet som mål å simulere menneskelig intelligens gjennom datasystemer. KI blir dermed gjerne forstått som bruk av algoritmer og matematiske instruksjoner for å få "*computers to do the sorts of things that [human] minds can do*".[9] Et KI-system er ofte definert som et sett med algoritmer med evne til å analysere og tolke enorme mengder data for å oppnå bestemte mål. KI brukes i økende grad både i privat næringsliv og i offentlig sektor og på tvers av ulike felt som transport, utdanning, helsevesen, nasjonal sikkerhet, sosiale medier og mer. KI-systemer som mange av oss møter på daglig basis inkluderer systemer som søkemotorer, ansiktsgjenkjenningssystemer og stemmeassistenter. KI kan også være bygget inn i maskinvare, for eksempel autonome biler, roboter og "Internet of Things" (IoT) applikasjoner.[10]

Den nasjonale KI-strategien bygger på definisjonen av KI foreslått i Europakommisjonens (EU) ekspertgruppe for KI, som ble utarbeidet til

[8] KMD (2020), Nasjonal strategi for kunstig intelligens
[9] Boden, M. A. (2018), Artificial Intelligence: A Very Short Introduction. Oxford: Oxford University Press.
[10] European Commission, High-Level Expert Group on Artificial Intelligence (2019), A Definition of AI: Main Capabilities and Disciplines, https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

forslaget for et felles europeisk rammeverk. EU foreslår en vid definisjon, som inkluderer en rekke ulike algoritmiske metoder:[11]

- Maskinlæringsmetoder: veiledet, ikke veiledet og forsterket, og dyplæring
- Logikk- og kunnskapsbaserte metoder: inklusiv induktiv programmering, kunnskapsbasert systemer og ekspertsystemer
- Statistiske metoder: metoder for estimering, søk og optimalisering.

Et viktig prinsipp ved KI som bidrar til å aktualisere spørsmål om diskriminering, er at KI innebærer effektive metoder for å analysere og tolke store datasett på en slik måte at algoritmen "*lærer*" sammenhenger. Datasett og algoritmer er derfor begge like viktige deler av et KI-system - uten data finnes ikke KI.[12] Store mengder data er nødvendig for å trene KI-systemer og for at slike systemer skal fungere godt. Dette reiser mange spørsmål om risiko for å produsere urettferdige eller partiske resultater samt risiko for å forsterke eksisterende eller produsere nye former for diskriminering. Det har allerede blitt avdekket en rekke slike eksempler i ulike vestlige land,[13] både fra privat og offentlig sektor. Blant de mest kjente er COMPAS, et KI-system som estimerer gjentakelsesfare for lovbrudd, brukt i straffeutmåling i USA, og som diskriminerte personer med afroamerikansk bakgrunn.[14] Andre eksempler er hvordan manglende demografisk mangfold i datasett, eller manglende sammenfall mellom treningsdata og populasjonen som KI-systemet skal brukes på, har gitt diskriminerende resultater, for eksempel at biometriske systemer lettere gjenkjenner hvite, mannlige ansikter.[15]

[11] Se KMD (2020) *Nasjonal strategi for kunstig intelligens* for utdypende forklaring av ulike typer kunstig intelligens.
[12] Gröger, C. (2021), There Is No AI Without Data. *Communications of the ACM*, 64(11), 98-108. doi: 10.1145/3448247
[13] Broomfield, H., & Reutter, L. M. (2021), Towards a Data-Driven Public Administration: An Empirical Analysis of Nascent Phase Implementation. *Scandinavian Journal of Public Administration*, 25(2), 73-97.
[14] Suresh, H., & Guttag, J. (2021)
[15] Klare, B. F., M. J. Burge, J. C. Klontz, R. W. Vorder Bruegge and A. K. Jain, (2012), Face Recognition Performance: Role of Demographic Information. *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1789-1801, Dec. 2012, doi: 10.1109/TIFS.2012.2214212. https://ieeexplore.ieee.org/document/6327355

En kritisk risiko for at KI produserer diskriminerende resultater for sosiale grupper og individer henger sammen med bruk av *persondata* i et KI-system. Dette inkluderer ikke bare sensitive persondata, men enhver form for informasjon om enkeltpersoner, fordi det alltid er en mulighet for at også slike data gjenspeiler *mønstre av institusjonalisert diskriminering*.[16] Risikoen for diskriminering øker med skjevheter i data, for eksempel historisk skjevhet i data som gjenspeiler forskjeller i menn og kvinners yrkesprofiler.[17] Dette har blant annet resultert i assosiasjoner mellom spesifikke typer yrker eller stillinger som kan knyttes til menn eller kvinner (f.eks. leder med menn, sykepleier med kvinner). Dermed er det ikke bare risiko for diskriminering ved bruk av KI når det er skjevhet i datasettet, men også når datasettet er helt faktisk korrekt, fordi data reflekterer en virkelighet der diskriminering forekommer.

1.2.1 Etikk i kunstig intelligens - åpenhet, forklarbarhet og rettferdighet

Kunstig intelligens blir i stadig større grad tatt i bruk i offentlig forvaltning og i næringslivet og borgerne eller kundene kommer i kontakt med tjenester og produkter som baserer seg på bruk av kunstig intelligens stadig oftere. Med økende bruk har vi også sett økende oppmerksomhet knyttet til at KI kan produsere uønskede resultater, som i verste fall har skadelig effekt. Betydningen av tillit til KI systemer har blitt satt på agendaen.

Både i akademia og blant ledende leverandører innen kunstig intelligens er det generelt høy oppmerksomhet knyttet til "AI ethics" og mange av initiativene har fokus på å bidra til at man utvikler KI systemer som brukere kan ha tillit til. For eksempel har IBM etablert et eget *AI Ethics Board* og Microsoft har en omfattende satsning på *Responsible AI*.

> *AI is embedded in everyday life, business, government, medicine and more. At IBM®, we are helping people and*

[16] Connell, R. W. (2002), *Gender*, Cambridge: Polity.
[17] Broomfield, H., & Reutter, L. M. (2021).

organizations adopt AI responsibly. Only by embedding ethical principles into AI applications and processes can we build systems based on trust.[18] *(IBM)*

We are committed to the advancement of AI driven by ethical principles that put people first.[19] *(Microsoft)*

Feltet KI-etikk tar opp utfordringer knyttet til hvordan KI kan utvikles ansvarlig, transparent, forklarbar og rettferdig. Ledende kompetansemiljøer er opptatt av at økt bruk av KI skjer på ansvarlig vis. Det innebærer å integrere etiske prinsipper i etablering, utvikling og bruk av KI, slik at vi kan bygge systemer som brukerne kan ha tillit til. Risiko for diskriminering er imidlertid i liten grad nevnt eksplisitt i denne forbindelse.

Forklarbar KI er et annet begrep som får mye oppmerksomhet, og særlig det motsatte: KI som ikke så lett kan forklares: "Enkelte dyplæringsalgoritmer kan sammenlignes med en 'sort boks', der man ikke har innsyn i modellen som kan forklare hvorfor en gitt inndataverdi gir et gitt resultat," forklarer KI-strategien (s.12). Forklarbar KI tar sikte på å gjøre "sort boks"- ("black box") algoritmer forståelige. Dette er ikke det samme som å publisere koden bak algoritmen, eller å gi innsyn i fullstendige treningsdatasett, da en slik tilnærming vil kunne bryte med både opphavsrett og personvern. Forklarbar KI kan i stedet analysere hvilke data som har hatt *betydning* for resultatet, og hvor stor betydning de ulike elementene har hatt, for slik å forklare logikken bak resultatet.[20]

Prinsippene om åpenhet og rettferdighet fremgår av det som Gjerdsbakk kaller for "*personvernets grunnlov*: personvernforordningens[21] grunnleggende prinsipper". Disse prinsippene styrer hva som er lovlig behandling av

[18] IBM (2022), AI ethics, https://www.ibm.com/artificial-intelligence/ethics
[19] Microsoft (2022), Responsible AI, https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6
[20] KMD (2020), Nasjonal strategi for kunstig intelligens
[21] Personvernforordningen, https://lovdata.no/lov/2018-06-15-38/gdpr

personopplysning.[22] Som "teknologinøytral" omfatter personvernforordningen også personopplysninger som håndteres ved hjelp av KI.[23] KI underlegges derfor både formkrav og innholdskrav for åpenhet og rettferdighet som er etablert i denne forordningen, slik Gjerdsbakk beskriver det i tidsskriftet Lov & Data:

> *Kravet om åpenhet er viktig for at den registrerte skal forstå hva som skjer med sine opplysninger. Åpenhetskravet er todelt. For det første forutsetter åpenhet at behandlingsansvarlig gir informasjon om behandlingen av personopplysninger (innholdskrav). For det andre må informasjonen være forståelig for den registrerte (formkrav). Åpenhet er essensielt for å skape tillit til behandlingsprosessen og er en forutsetning for at den registrerte kan vurdere om behandlingen tilfredsstiller personvernforordningens krav.*[24]
>
> *Ordet "rettferdig" i personvernforordningen art. 5 (1) a)*[25] *tilsier at behandlingen må være rimelig og ligge innenfor de moralske og etiske normene som følger av lover og regler, men også innenfor samfunnets rettferdighetsoppfatning. Utover dette er rettferdighet et vidt begrep med et bredt anvendelsesområde. Rettferdighet innebærer blant annet krav til lovlighet, åpenhet, fravær av uønsket forskjellsbehandling og at det tas hensyn til asymmetrisk maktbalanse. Videre spiller etikk og moral, kultur og kontekst inn på hva som oppfattes som rettferdig.*[26]

I møter med referansegruppen og ulike fageksperter (se nedenfor) møtte vi en rekke motstridende synspunkt på utfordringsbildet når det gjelder KI i offentlig

[22] Gjerdsbakk, T.C.G., 2022. Åpen og rettferdig kunstig intelligens, i *Lov & Data* nr. 150 – hefte 3/2022, https://lovdata.no/artikkel/apen_og_rettferdig_kunstig_intelligens/4139
[23] Gjerdsbakk, T.C.G., 2022.
[24] Gjerdsbakk, T.C.G., 2022.
[25] Personvernforordningen art. 5, https://lovdata.no/lov/2018-06-15-38/gdpr/a5
[26] Gjerdsbakk, T.C.G., 2022.

sektor. Mens manglende oppdatering av lovverket for å reflektere økt digitalisering ble nevnt av noen, ble det også pekt på at likestillings- og diskrimineringsloven skal gi like godt vern mot diskriminering som skjer gjennom bruk av KI som i andre kontekster. Mens noen mente at det alltid er mulig å identifisere hvorvidt diskriminering forekommer i et KI-system, mente andre at det ikke var mulig å etablere sikker kunnskap om dette ved bruk av de mest kompliserte "black box"-algoritmene. Med hensyn til pågående arbeid med å lage et regulerende rammeverk for trygg bruk av KI i Norge som i EU, fant vi enighet i viktigheten av dette arbeidet. Samtidig ble det påpekt at nettopp EU sitt forslag til en KI-forordning har et innebygget kompromiss mellom hva som er ønskelig (ingen risiko) og hva som er mulig (noe risiko). Vi møtte blant annet synspunktet at dette kompromisset er nødvendig for i det hele tatt å muliggjøre KI-utvikling. Spennet mellom disse ulike synspunktene er en utfordring, ikke bare på det politiske eller juridiske nivå, men også for enkeltvirksomheter, påpekte en av våre fageksperter; selv når alle forholdsregler er tatt, kan KI fortsatt innebære en risiko for feil eller diskriminerende resultater.

En rekke utfordringer og "vanskelige spørsmål", som KI-strategien kaller det, gjenstår å finne gode løsninger på når kunstig intelligens nå dukker opp i stadig flere sammenhenger. Vårt mål har imidlertid vært avgrenset til å utrede hvilke utfordringer knyttet til *diskriminering* som oppstår når offentlig sektor tar i bruk KI, og nedenfor ser vi nærmere på hvordan vi forstår begrepet diskriminering.

1.3 Diskriminering

Med begrepet diskriminering i likestillings- og diskrimineringslovens forstand referer vi til *usaklig eller urettmessig forskjellsbehandling* av individer eller grupper på bakgrunn kjønn, graviditet, permisjon ved fødsel og adopsjon, omsorgsoppgaver, etnisitet, religion, livssyn, funksjonsnedsettelse, seksuell

orientering, kjønnsidentitet, kjønnsuttrykk, alder, eller kombinasjoner av disse.[27]

Forskjellsbehandling blir regnet som diskriminering hvis individer eller grupper blir behandlet dårligere enn andre, eller når de blir behandlet likt, men utfallet fører til at de blir stilt dårligere enn andre. Den første kalles direkte forskjellsbehandling og siste er indirekte forskjellsbehandling.[28]

Som mange har hevdet er diskriminering notorisk vanskelig å måle vitenskapelig, og det har preget både politikk og forskning.[29] Vi operasjonaliserte begrepet diskriminering i både spørreundersøkelsen og intervjuene gjennom likestillings- og diskrimineringsloven.

Mens risiko for diskriminering ofte diskuteres i lys av likestillings- og diskrimineringslovens, vil bruk av KI aktualisere en vid definisjon av personopplysninger. I den forbindelse er det nyttig å se på likestillings- og diskrimineringsloven diskrimineringsgrunnlag (§6) og personvernforordningens definisjon av personopplysninger.

Personvernombudet definerer personopplysninger som "enhver informasjon om enkeltpersoner som kan gi grunnlag for å identifisere individet".[30] Personopplysninger omfatter mange ulike typer av informasjon, fra navn, adresse og personnummer til lyd og bilde, biometri, opplysninger om adferdsmønster (hva vi handler, hva du ser på TV, osv.) til registreringsnummer på bil og IP-adresse. Noen opplysninger defineres som sensitive personopplysninger i Personvernforordningen, som:

[27] Lov om likestilling og forbud mot diskriminering (likestillings- og diskrimineringsloven) – Lovdata, https://lovdata.no/dokument/NL/lov/2017-06-16-51

[28] Bufdir, Begreper: https://bufdir.no/Statistikk_og_analyse/Etnisitet/begreper_og_kunnskapsgrunnlag/begreper/

[29] Midtbøen, A. H. (2015). Etnisk diskriminering i arbeidsmarkedet. *Tidsskrift for samfunnsforskning*, *56*(1), 4-30.

[30] Datatilsynet, Personvernforordningen, https://www.datatilsynet.no/rettigheter-og-plikter/personopplysninger/, Personvernforordningen, artikkel 4.

> *opplysninger om etnisk opprinnelse, politisk oppfatning, religion, filosofisk overbevisning, fagforeningsmedlemskap, genetiske opplysninger, biometriske opplysninger med det formål å entydig identifisere noen, helseopplysninger, opplysninger om seksuelle forhold, opplysninger om seksuell legning.*[31]

Et noe annet og delvis overlappende sett av personopplysninger spesifiseres i likestillings- og diskrimineringsloven, som viser til at det er forbudt å diskriminere på grunnlag av:

> *kjønn, graviditet, permisjon ved fødsel eller adopsjon, omsorgsoppgaver, etnisitet, religion, livssyn, funksjonsnedsettelse, seksuell orientering, kjønnsidentitet, kjønnsuttrykk, alder eller kombinasjoner av disse grunnlagene er forbudt.*[32]

Samfunnet og individers muligheter og valg, reflektert i representasjon, deltakelse, tilstedeværelse eller fravær fra visse arenaer og situasjoner, er også en refleksjon av hvordan en rekke av diskrimineringsgrunnlagene over tid har blitt institusjonalisert og "stivnet" i samfunnet.[33] Det vil si at det ikke bare er diskrimineringsgrunnlagene i likestillings- og diskrimineringsloven som kan danne grunnlag for diskriminering mellom personer og grupper, men at også økonomiske, kulturelle og sosiale identiteter (jf. personvernforordning art. 4) kan være knyttet til diskrimineringsgrunnlagene. Det er derfor viktig å ta utgangspunkt i en vid definisjon av personopplysninger for å kunne fange opp når og hvor risiko for diskriminering kan oppstå ved bruk av kunstig intelligens.

[31] Datatilsynet, Personvernforordningen
[32] Datatilsynet, Personvernforordningen
[33] Friedman, B., & Nissenbaum, H. (1996), Bias in computer systems. *ACM Transactions on information systems (TOIS),* 14(3), 330-347.

1.4 Europeiske og norske strategier og lovverk for å fremme og regulere bruk av KI

Offentlig sektor i Norge jobber målrettet for å bidra til at det offentlige tjenestetilbudet kan utvikles og forbedres på en effektiv og bærekraftig måte. Digitalisering og bruk av ny teknologi er et viktig virkemiddel for å få det til, og her vil kunstig intelligens kunne stå for et vesentlig bidrag. Bruken av nye teknologier er med på å skape offentlige digitale tjenestetilbud som er tilpasset brukernes behov.[34] Brukt på sitt beste kan teknologien generelt, og KI spesielt, bidra til effektivisering og kvalitetsheving av offentlige tjenester. Det finnes gode eksempler på bruk av KI i Norge, for eksempel i helsesektoren,[35] og både ønske om og potensiale for ytterligere bruk, er sterkt til stede.

Nasjonal strategi for kunstig intelligens[36] peker på at Norge har en rekke forutsetninger som gjør at vi kan lykkes med KI. Norge kjennetegnes ved høy grad av tillit til offentlig sektor, høy grad av digital kompetanse i befolkning, en godt utbygd teknologisk infrastruktur og en offentlig sektor som har kommet langt i utvikling av digital forvaltning. Særlige fordeler i arbeidet med KI er gode registerdata og lange tidsserier, noe som muligens kan gi viktig tilgang til data for å utvikle KI. KI-strategien påpeker også at offentlige virksomheter "har kapasitet og kompetanse til å eksperimentere med nye teknologier", noe som kan være avgjørende for at det offentlige skal kunne ta i bruk ny teknologi, som KI.

KI-strategien peker imidlertid også på utfordringer knyttet til bruk av KI, for eksempel:

[34] KMD (2019), Én digital offentlig sektor: Digitaliseringsstrategi for offentlig sektor 2019–2025, https://www.regjeringen.no/no/dokumenter/en-digital-offentlig-sektor/id2653874/

[35] Se Kunstig intelligens i norsk helsetjeneste (KIN) sin liste over prosjekter i helsesektoren: https://www.helsedirektoratet.no/tema/kunstig-intelligens/kompetanse-og-erfaringsdeling

[36] KMD. (2020), Nasjonal strategi for kunstig intelligens, https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/

> *Hvem har ansvaret for konsekvensene av en beslutning som er truffet av KI? Hva skjer når autonome systemer tar egne beslutninger som vi ikke er enige i og som i verste fall fører til skade? Og hvordan sørger vi for at teknologien ikke viderefører og forsterker bevisst og ubevisst diskriminering og forutinntatthet?*

Dette er spørsmål som gjør KI særlig utfordrende for offentlig sektor, som sammen med øvrige sektorer og aktører må følge regjeringens retningslinjer for KI i Norge, som forventer at den skal "bygge på etiske prinsipper, og respektere menneskerettighetene og demokratiet [...] bidra til ansvarlig og pålitelig kunstig intelligens" og "ivareta den enkeltes integritet og personvern" (KI-strategien s. 6).

KI-strategien reiser også spørsmål knyttet til sikkerhet og kontroll, blant annet ved å foreslå at: "digital sikkerhet skal bygges inn i utvikling, drift og forvaltning av løsninger" for KI, og at "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" (KI-strategien s. 6).

KI-strategien viser altså til eksisterende tilsynsmyndigheter, mens en særlig tilsynsmyndighet for KI ikke er etablert i Norge. Derimot ble det i 2020 opprettet en *regulatorisk sandkasse* i regi av Datatilsynet, med mål å stimulere til utvikling av KI ved å hjelpe virksomheter å følge regelverk og å utvikle "personvernvennlige løsninger".[37]

Et særlig viktig prinsipp for bruk av KI i offentlig sektor er kravene til offentlig saksbehandling, og vi deler derfor et lengre sitat fra KI-strategien[38]:

> *Saksbehandlingen i offentlig sektor er i stor grad regelstyrt, med større eller mindre elementer av skjønnsmessige vurderinger i prosessen. Det betyr at en løsning ikke behøver*

[37] Datatilsynet, *Sandkassesiden*, https://www.datatilsynet.no/regelverk-og-verktoy/sandkasse-for-kunstig-intelligens/
[38] KMD (2020), Nasjonal strategi for kunstig intelligens

å være enten manuell eller automatisert. […] Allerede i dag er det mye saksbehandling i offentlig sektor som er automatisk. Det finnes saksbehandlingssystemer med integrert søknads-dialog som gir muligheter for å fatte automatiserte vedtak umiddelbart.

Felles for dagens automatiserte sakbehandlingsløsninger (sic) er at de er regelbaserte. Regelverket er programmert inn i løsningen, og dermed er det mulig å begrunne vedtak. Forvaltningsloven krever at alle enkeltvedtak begrunnes. Denne plikten til å gi en begrunnelse er viktig for å ivareta innbyggernes mulighet til å etterprøve og kontrollere beslutninger som fattes om dem.

Det er et stort potensial for økt bruk av kunstig intelligens i offentlig saksbehandling – både i form av regelstyrte systemer og maskinlæring. Forvaltningslovutvalget peker på at automatisering kan bidra til mer likebehandling og konsistent gjennomføring av regelverket. Samtidig må det være en forutsetning ved innføring av saksbehandlings-løsninger med elementer av KI at algoritmenes vurderinger er minst like gode og tillitvekkende som det menneskelige skjønnet de erstatter. For å kunne være sikre på at det er slik, må vi ha systemer som er transparente og forklarbare. (KI-strategien, s. 26)

Når anvendelser av KI omtales i media vil det i enkelte tilfeller være krevende for journalister å ha tilstrekkelig kompetanse til å skille mellom de regelbaserte løsningene som helt transparent automatiserer et regelverk og således vil kunne fatte vedtak som kan begrunnes på den ene siden, og på den andre siden løsninger som bruker maskinlæring eller "black box" algoritmer og som i dag

ikke er generelt transparente og forklarbare, og således ikke kan brukes til å fatte enkeltvedtak.[39, 40]

Også internasjonalt har utfordringer knyttet til KI blitt satt på agendaen de siste årene. Europarådet vil foreslå til ministerrådet en strategisk agenda for å inkludere KI-lovgivning som en av de viktigste virkemidler for å etablere en "fair balance" mellom gevinster av teknologisk utvikling og beskyttelse av fundamentale verdier.[41]

Europakommisjonen sendte i april 2021 ut forslag til et regulatorisk rammeverk, *Artificial Intelligence Act, eller KI-forordningen,* for å sikre at bruk av KI i EU er trygg, lovmessig og i henhold til EUs fundamentale rettigheter.[42] Denne er også sendt på høring Norge.[43]

Det overordnede målet er å stimulere bruk av etisk troverdig KI i EUs økonomi, og å gjøre:

> *EU best i verden når det gjelder utvikling og bruk av sikker, pålitelig og menneskesentrert KI, fordi tillit til KI-systemer er viktig og nødvendig for at for at det sosiale og økonomiske potensialet i KI kan utnyttes fullt ut.*[44]

Med en "risikobasert tilnærming" deler forslaget KI inn i fire risikokategorier: uakseptabel risiko, høyrisiko, begrenset risiko, og minimal risiko, og det

[39] Nickelsen, T. (2019) Roboter er på full fart inn i jussen, Forskning.no, https://forskning.no/juridiske-fag-roboter/roboter-er-pa-full-fart-inn-i-jussen/1588380
[40] Nilsen, C. M. (2020) Begynnelsen på slutten for IB?, Khrono, https://khrono.no/begynnelsen-pa-slutten-for-ib/504277
[41] Council of Europe, *Council of Europe and Artificial Intelligence*, https://www.coe.int/en/web/artificial-intelligence
[42] European Commission (2021), Artificial Intelligence Act, https://artificialintelligenceact.eu/ https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF
[43] Kunstig intelligens-forordningen på høring i Norge: https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2021/juni/forslag-til-forordning-om-kunstig-intelligens-ki-forordningen/id2884935/
[44] Regjeringen (2021), Forslag til forordning om kunstig intelligens (KI-forordningen), https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2021/juni/forslag-til-forordning-om-kunstig-intelligens-ki-forordningen/id2884935/

ledende prinsippet er at "jo høyere risiko bruken av KI utgjør, desto strengere bør bruken reguleres."[45] Beskrivelse av de ulike risikokategoriene er sitat fra regjeringens posisjonsnotat til EU-kommisjonen:[46]

Uakseptabel risiko skal være forbudt og omfatter "KI-systemer som ansees å være uakseptabelt fordi de strider med grunnleggende verdier i EU." Slike systemer inkluderer (a) KI-systemer med et betydelig *potensial for å manipulere personer ved subliminale teknikker*, (b) systemer som *utnytter sårbarheter hos en gruppe* på grunn av alder eller psykisk eller fysisk funksjonsnedsettelse, (c) systemer som *vurderer troverdigheten til personer på grunnlag av sosial oppførsel*, og (d) systemer som *benytter sanntids biometriske identifikasjonssystemer* plassert i offentlig tilgjengelige rom.

Høyrisiko KI-systemer er underlagt "særskilte regler", og omfatter "KI-systemer som innebærer høy risiko for personers helse, sikkerhet eller grunnleggende rettigheter. [...] I tråd med Kommisjonens risikobaserte tilnærming vil disse KI-systemene være lovlige når systemet er i tråd med reglene" angitt i forordningen.

Begrenset risiko omfatter "systemer som skal samhandle med fysiske personer. Disse systemene må være innrettet slik at personen blir informert om at de samhandler med et KI-system, hvis ikke dette framgår av sammenhengen av bruken." Slike KI-systemer er underlagt en "transparensforpliktelser."

Minimal risiko omfatter de fleste KI-systemer som er i bruk, og ifølge KI-forordningen representerer disse "minimal risiko for enkeltpersoner og samfunnet som helhet. De KI-systemene som ikke er omfattet av de tre første kategoriene, anses som systemer med minimal risiko, og er derfor ikke underlagt særskilte forpliktelser."

[45] Regjeringen (2021)
[46] Regjeringen (2021)

KI-forordningen har primært som mål å regulere "høyrisiko"-bruk av KI, ettersom denne i størst grad anses å utgjøre "en stor potensiell trussel mot samfunnet og enkeltpersoner."[47]

1.4.1 Kunstig intelligens-forordning fra EU til Norge

Forordningen har vært på høring og høringsperioden ble avsluttet i august 2021. Den norske regjeringen har sendt et tilsvar til EU-forslaget.[48] I sitt tilsvar sier regjeringen seg enig med en vid og åpen definisjon av KI fordi dette handler om teknologi i rask utvikling. Det blir også uttrykt støtte til måten å dele forslaget inn i risikonivåer, fordi det unngår å håndtere spesifikke teknologier med ikke-intenderte konsekvenser for utviklingen. Balansen mellom regulering og risiko i 4-nivå modellen støttes således av den norske regjeringen. Regjeringen støtter også det overordnede i nivå 2: høy-risiko, men har også innspill og betenkeligheter til forslaget, bl.a. knyttet til å bestemme grensetilfeller på de øvre risikonivåene og til regulering av nivået "høyrisiko", som primært er basert på egen-evaluering. Her stiller regjeringen spørsmål om det er nødvendig med en tredjepart for å bistå denne oppgaven.

Det blir også stilt spørsmål ved om det er tilstrekkelig å kreve at data er "fri for feil" og "fullstendig", fordi dette er "en umulig oppgave" (vår oversettelse), ifølge regjeringen. En mildere formulering kan bidra til at det blir mulig å finne en balanse mellom *best mulig* og *mulig å gjennomføre*. Videre uttrykkes bekymring for at det ikke er tilstrekkelig krav om åpenhet når KI inngår i administrasjon av retts- og demokratiprosesser, og regjeringen foreslår et tydeligere krav her, og det påpekes behov for at KI-regulering ses i forhold til og samkjøres med andre lover, som personvernforordningen (GDPR). Heri inngår en bekymring for at nye KI-reguleringer vil *overkjøre* eksisterende

[47] Regjeringen (2021)

[48] Kommunal og moderniseringsdepartementet (2021), 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), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665314_en

reguleringer bl.a. i arbeidslivet, der det i Norge legges stor vekt på ansattes påvirkningsrett når nye teknologier skal introduseres. Regjeringen flagger også planer om å etablere et *nasjonalt tilsyn for KI*, i tråd med forordningen.

Flere av regjeringens bekymringer knyttet til EUs forslag til en felles Europeisk KI-forordning blir også reflektert i vår undersøkelse i offentlig sektor, og noe kan også finnes igjen i rammeverket i kapittel 4.

1.5 Metode

For å løse oppdraget har prosjektet:

- gjennomført en spørreundersøkelse blant virksomheter i statlig og kommunal sektor
- gjennomført 19 dybdeintervjuer med virksomheter med KI prosjekt som behandler persondata
- gjennomført analyse av risiko for diskriminering
- utarbeidet anbefalinger for å håndtere risiko for diskriminering
- etablert kontakt med eksterne gjennom en egen referansegruppe for prosjektet og kontaktet Digitaliseringsdirektoratets fagforum for Kunstig intelligens i offentlig sektor og presentert prosjektet for dem

1.5.1 Kunnskapskartlegging gjennom referansegruppe og fageksperter

Prosjektet etablerte tidlig en referansegruppe med nasjonal og internasjonal eksperter på ulike perspektiver knyttet til KI, for å kvalitetssikre innholdet og anbefalingene i prosjektet. I begynnelsen av prosjektet hadde vi innledende møter både med referansegruppen og andre KI-eksperter. Målet var å få et innblikk i ulike perspektiver på utfordringer med KI of offentlig sektors arbeid med KI, spesielt sett i sammenheng med diskrimineringsperspektivet. Det var i tillegg viktig å snakke med faglige eksperter, både fra det teknologiske, samfunnsvitenskapelige og juridiske feltet.

Referansegruppen har gjennom hele prosjektet bistått med viktige tilbakemeldinger og innspill i utformingen av intervjuguiden, validering av funn og kvalitetssikring av anbefalinger.

Referansegruppen har bestått av Heather Broomfield fra Digitaliseringsdirektoratet, professor Nicola Marsden fra Hochschule Heilbronn, Kari Laumann fra Datatilsynet, Kathinka Theodore Aakenes-Vik fra Likestillings- og diskrimineringsombudet og professor Ingunn Ikdahl fra juridisk fakultet, Universitetet i Oslo.

I tillegg har vi hatt samtaler om prosjektet med Alex Moltzau, Birte Hansen og Klas H. Pettersen fra NORA (Norwegian Artificial Intelligence Research Consortium), professor Morten Goodwin fra Universitetet i Agder, IBM Fellow og leder for IBMs *AI Ethics Global* Francesca Rossi og Frank Vevle fra Bufdir.

Prosjektteamet har også presentert prosjektet for Fagforum for Kunstig intelligens i offentlig sektor, som arrangeres i regi av Digitaliserings-direktoratet.

1.5.2 Spørreundersøkelse

I forbindelse med prosjektet ble det gjennomført en online kvantitativ spørreundersøkelse. Undersøkelsen ble sendt på epost til 491 offentlige organisasjoner, herunder bedrifter og kommuner. Hovedformålet med spørreundersøkelsen var å kartlegge status for offentlig sektors arbeid med kunstig intelligens. Vi var spesielt interesserte i å avdekke hvor mye aktivitet det var i sektoren når det gjaldt kunstig intelligens, i form av KI-aktivitet og KI-prosjekter, særlig prosjekter som involverte persondata og dermed kunne vurderes i forhold til diskrimineringsrisiko.

Utformingen av spørreskjemaet skjedde i dialog med referansegruppen. Spørreundersøkelsen ble gjennomført elektronisk gjennom Rambølls spørreundersøkelsesverktøy SurveyXact, og respondentene besvarte spørsmålene individuelt i perioden januar og februar 2022. Til sammen fikk vi en svarprosent på 41 %.

Spørsmålene i spørreundersøkelsen varierer i svarprosent av flere grunner. Respondenter som svarte at de ikke hadde konkrete planer om å bruke kunstig intelligens, og at persondata eller individdata ikke ble behandlet i det aktuelle

prosjektet, gikk ikke videre i undersøkelsen. Men i resultatene markeres også disse svarene som "gjennomført", og gir totalt 200 gjennomførte svar. Av disse var det 60 respondenter som hadde konkrete planer om å bruke kunstig intelligens, og kun 39 av disse behandler persondata/individdata i sine prosjekter. Analysen av spørreundersøkelsen er derfor basert på de 39 respondentene som har besvart alle eller mesteparten av spørsmålene i undersøkelsen vår.

1.5.3 Kvalitative dybdeintervjuer

Prosjektet gjennomførte i perioden februar til april 2022, 19 intervju med representanter fra virksomheter som i spørreundersøkelsen hadde svart at:

- de hadde pågående prosjekt som bruker KI
- de behandlet persondata/individdata i prosjektet
- de hadde problemstillinger som illustrerer bredden i det som fremkom i spørreundersøkelsen

Spørreundersøkelsen var den primære kilden for å identifisere relevante informanter til kvalitative dybdeintervjuer. Spørreundersøkelsen ble også brukt som rekrutteringskanal, ved å legge inn følgende tekst avslutningsvis: *"Undersøkelsen vil spørre deg om din kontaktinformasjon slik at vi kan ta kontakt med deg ved et senere tidspunkt dersom din virksomhet blir valgt ut til oppfølgingsintervju. Å oppgi kontaktinformasjon i undersøkelsen er frivillig"*. Enkelte ble også rekruttert gjennom direkte kontakt dersom vi hadde informasjon om at virksomheten var aktive på KI-feltet.

1.5.4 Gjennomføring

Intervjuguider ble utformet sammen med referansegruppen. I intervjuene var det alltid en intervjuer og en som gjorde notater som ble renskrevet etter intervjuet, og disse utgjorde analysematerialet. Intervjuer ble tatt opp dersom informantene samtykket til dette, for at vi skulle ha mulighet til å kontrollere hva som ble sagt dersom notatene var mangelfulle.

I rapporten er alle informanter anonymisert, og vi identifiserer kun om en informant kommer fra statlig eller kommunal sektor, med stat eller kommune, når vi gjengir sitater.

1.5.5 Rammeverk for analyse og anbefalinger av tiltak

Rammeverket for å kunne diskutere og analysere diskriminering i prosjektet har fremkommet ved gjennomgang av relevant litteratur, innsikt gjennom spørreundersøkelsen, og innsikt fra dybdeintervjuene. Ved å sammenholde innsikt fra disse tre kilden vil vi i kapittel 4 skissere et rammeverk for å vurdere risiko for diskriminering ved bruk av KI i offentlig sektor. Dette rammeverket dekker pre-prosjektfasen, planleggingsfasen, utviklingsfasen, test- og evalueringsfasen og produksjonsfasen av et KI-prosjekt. Denne danner også et utgangspunkt for anbefalinger av tiltak for å forhindre diskriminering.

I neste kapittel skal vi presentere undersøkelsens bilde av status for offentlig sektors arbeid med kunstig intelligens i ulike prosjekter og aktiviteter.

2. Kartlegging av KI-prosjekter i offentlige organisasjoner

I dette kapittelet presenterer vi en status på offentlig sektors arbeid med kunstig intelligens i ulike prosjekter og aktiviteter. Vi presenterer resultat fra spørreundersøkelsen og de kvalitative intervjuene på tvers, og ser på funnene i forhold til hverandre. Som vi vil vise her, finnes det en rekke konkrete KI-prosjekter og KI-aktiviteter i offentlig sektor, mens det er et mindretall av disse som faktisk er i drift eller brukes som del av virksomhetens operative drift. Status for KI-aktivitetene viser tegn på sektoren ikke er moden i forhold til å ta i bruk KI, og at virksomhetene opplever flere uklarheter knyttet til teknologien. Samtidig er aktørene ikke negative til å satse på KI når de får innsikt i potensiale ved denne teknologien.

2.1 KI-prosjekter og formål

Spørreundersøkelsen hadde som mål å kartlegge bruk av KI i offentlig sektor, særlig der persondata eller individdata inngår i KI-prosjekter. Det var totalt 200 organisasjoner som gjennomførte undersøkelsen av totalt 491 organisasjoner som ble invitert til å delta. 60 organisasjoner som deltok i undersøkelsen oppga at de hadde prosjekter (eller aktiviteter) som bruker KI. Like mange svarte at de hadde konkrete planer om KI-prosjekter . Videre, var det kun 39 av disse respondentene som oppga at de behandler persondata eller individdata i det aktuelle KI-prosjektet.

Disse 39 organisasjonene jobber med KI-prosjekter av ulik natur og formål, og alle disse har svart på hele undersøkelsen vår (se kapittel 1 for nærmere informasjon om svarprosenten). Blant disse er det flere som har chatboter på sine nettsider. Flere oppgir at KI brukes eller planlegges for å effektivisere interne arbeidsprosesser og systemer i virksomheten. Videre brukes eller planlegges KI brukt for å predikere virksomhetsrelaterte fenomener, for

eksempel sykefravær. Det er også nevnt at KI anvendes innen rekrutteringsprosesser for å matche kandidater til riktig jobb.

Det er tre elementer som tydelig karakteriserer KI-prosjektene til organisasjonene som har besvart undersøkelsen:

1. Maskinlæring/dyplæring
2. KI som støtte, der beslutning tas av saksbehandler
3. Automatisering eller regelstyrte beslutninger

Figur 2.1. Hvilke elementer karakteriserer KI-prosjektet i organisasjonen?

Ser vi videre på hva slags data virksomhetene bruker til maskinlæring, opplyser respondentene at de ofte bruker nyere data fra egen virksomhet eller historiske data fra egen eller tilsvarende virksomhet. Syntetiske data og særlig tilrettelagte testdata brukes også, men i mindre grad enn de ovennevnte data.

På spørsmålet om KI har skapt helt nye innsikter eller prosesser i respondentenes organisasjoner, finner vi ikke en entydig tendens. Majoriteten er hverken enige eller uenige i denne påstanden.

2.2 Et fåtall KI-prosjekter i bruksstadiet

I intervju med 17 av disse organisasjonene kom det frem at mange av de ovennevnte KI-prosjektene er under utforsking og utvikling, og at et fåtall er i bruk. KI-prosjekter settes i gang for å avdekke hvilke muligheter teknologien kan bringe for organisasjonene, i tillegg til at man ønsker å prøve ut KI-

modeller. Vi fant en tendens til at KI-prosjekter blir stoppet av ulike grunner i ulike faser, noe som fører til at et fåtall KI-prosjekter har nådd fram til bruksstadiet. Omtrent halvparten av respondentene svarer at KI-prosjektet deres ikke har nådd målet og derfor har blitt avsluttet eller avviklet.

KI-prosjektene vi fikk innsikt i gjennom undersøkelsen kan kategoriseres på bakgrunn av bruksområde. Vi har kategorisert dem på følgende måte:

1. Forbedre kvaliteten til datagrunnlaget
2. Avdekke mistenkelige mønstre i systemet
3. Avdekke feil eller manglende informasjon i systemet
4. Prediksjon av behov i virksomheten
5. Prediksjon av brukeratferd
6. Behandling av pasienter
7. Behandling av syntetiske test data
8. Støtte saksbehandlere i saksbehandlingsprosessen

I flere prosjekter blir KI brukt for å **forbedre kvaliteten til datagrunnlaget**. Her blir KI *ikke* brukt til behandling av data, men heller å søke og rette opp feil i datagrunnlaget. Datagrunnlaget kan inneholde persondata. Regelstyrt KI blir brukt for dette formålet. En rekke prosjekter bruker også KI for å **avdekke feil eller manglende informasjon** i systemet. Formålet i disse KI-aktivitetene er å forbedre brukernes interaksjon med tjenesten ved å gjøre prognoser som gjør at tjenesten aktivt hjelper brukere å unngå å gjøre feil når de skal bruke tjenesten. Datagrunnlaget kan inneholde individdata omformulert som beskrivelse av hendelser. Her blir forklarbar KI brukt.

I en rekke andre prosjekter blir KI brukt for **prediksjon av behov i virksomheten**, blant annet som prediksjon av fraværsfrekvens. Formålet i disse prosjektene er effektivisering av systemet. Her kan datagrunnlaget inneholde individdata, og forklarbar KI blir brukt.

I noen prosjekter blir KI-modeller brukt til **å avdekke mistenkelige mønstre i systemet**. Her er formålet å oppdage misbruk, og "black box" KI er nevnt å bli brukt. Datagrunnlaget kan innholde personopplysning og kontaktinformasjon.

I andre prosjekter blir KI brukt for **å predikere atferden til brukere av** velferdstilbud. Formålet er å gjøre velferdstjenester mer tilgjengelig for brukere, og redusere feil bruk av velferdstilbudet. Her er individdata omformulert som beskrivelse av hendelser, og forklarbar KI blir brukt. I noen prosjekter blir KI brukt i **behandling av pasienter**, som for eksempel bildediagnostikk. Datagrunnlaget inneholder individdata, og her blir blant annet maskinlæring brukt.

I et prosjekt blir KI brukt i **behandling av syntetiske test data**. Her er syntetiske data formet som beskrivelser av hendelser, og maskinlæring er brukt i prosjektet. KI blir også brukt til **å støtte saksbehandler i saksbehandlingsprosessen,** og her brukes også forklarbar KI.

2.3 KI sitt bidrag i saksbehandling

I undersøkelsen stiller vi en rekke spørsmål om hvorvidt ulike påstander knyttet til rollen KI kan ha i saksbehandling, stemmer for respondentenes organisasjon. Majoriteten av respondentene er skeptiske til at KI kan ta beslutninger uten menneskelig innblanding. I intervjuene kommer det frem at i KI-prosjekter som er i drift, bidrar KI til beslutningsstøtte der vedtak angår individer og en beslutning kan gi eller ta vekk en rettighet. I slike tilfeller er det saksbehandler som tar den siste avgjørelsen, ifølge respondentene. For eksempel kan en informant fortelle at kunstig intelligens ikke alene kan hjelpe dem med å finne feilutbetalinger, og at saksbehandler er viktig for å kvalitetssikre resultatet som foreslås av KI. Samtidig er det uklart hvorvidt og på hvilken måte saksbehandler bør forholde seg til KI som beslutningsstøtte.

I enkle regelbaserte vedtak der beslutningen automatisk er positiv hvis alle krav er oppfylt, tar KI beslutningen. Men i intervjuene oppdager vi at i slike tilfeller benyttes regelbasert KI som automatiserer en prosess, og ikke maskinlæringsbaserte algoritmer. Undersøkelsen viser at begrepet KI omfatter både regelstyrte prosesser som *kan brukes til beslutning*, og maskinlæringsalgoritmer som *kun skal bruke som beslutningsstøtte*. Denne

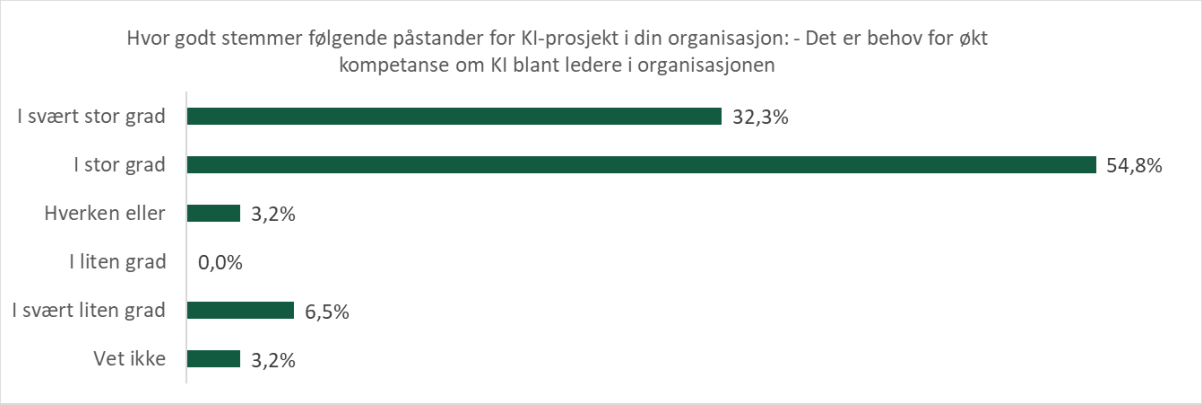
dobbeltheten gjør at forståelsen av hvilken rolle KI kan og bør spille i saksbehandlingen, blir uklar.

2.4 Forankring av KI-prosjekter og eksterne samarbeid

De fleste KI-prosjektene som er identifisert gjennom spørreundersøkelsen er forankret i toppledelsen, og de fleste prosjektene har også avklart finansiering og budsjetter.

Til tross for at de fleste sier at KI-prosjektene er forankret i toppledelsen, mener flertallet at lederne mangler kompetanse om KI. Det er et klart flertall som mener at det i stor grad er behov for å øke kompetanse om KI blant ledere i egen organisasjon. Dette er et interessant funn, med tanke på at flere toppledere også er respondenter i undersøkelsen.

Figur 2.2. Hvor godt stemmer følgende påstander for KI-prosjekt i din organisasjon: - Det er behov for økt kompetanse om KI blant ledere i organisasjonen



KI-prosjektene jobber sjelden isolert i tekniske miljøer, men er snarere forankret i en kombinasjon av fagområder. Tverrfaglighet fremheves som en suksessfaktor for å lykkes med KI-prosjekter. Dette kan tyde på noe mer tverrfaglighet i prosjektteamene, som vi også finner i intervjuene.

Undersøkelsen viser at de fleste virksomhetene samarbeider med eksterne miljøer om utfordringer knyttet til KI. I intervjuene ser vi stor variasjon i hvordan ulike organisasjoner benytter samarbeid med eksterne. Mens store statlige organisasjoner over tid har skaffet seg kompetanse internt, sliter små organisasjoner, og spesielt små kommuner, med å skaffe seg nok kompetanse

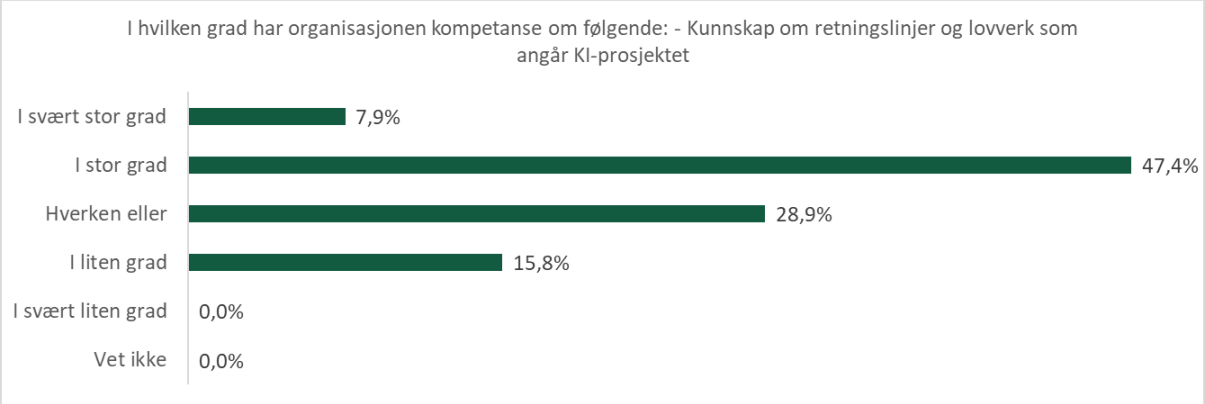
innen IT generelt, og KI spesielt. Disse organisasjonene benytter ofte enten eksterne samarbeidspartnere i IT bransjen, eller kjøper algoritmer som "hyllevare" fra eksterne miljøer. En rekke både store og små organisasjoner nevner at de samarbeider med forskningsmiljøer i Norge.

2.5 Kompetanse om kunstig intelligens og diskriminering

Spørreundersøkelsen viser at flere virksomheter bekrefter at de har **teknisk kompetanse** om kunstig intelligens enn de som mener at de ikke har dette, men forskjellen mellom disse gruppene er ikke signifikant. I tillegg er det en andel som ikke kan besvare om de besitter teknisk kompetansen eller ikke.

Når vi ser på respondentenes **kunnskap om retningslinjer og lovverk** som angår sitt KI-prosjekt er det imidlertid en betydelig større andel som har denne kunnskapen, enn de som ikke har det. Samtidig fremkommer det i intervjuene at det er en del usikkerhet rundt hvilke lovverk som er relevante for KI-systemer, hvorvidt lovverket er tilpasset nye digitaliserte verktøy, og hvilke handlingsrom disse lovene gir for KI-prosjekter i offentlig sektor.

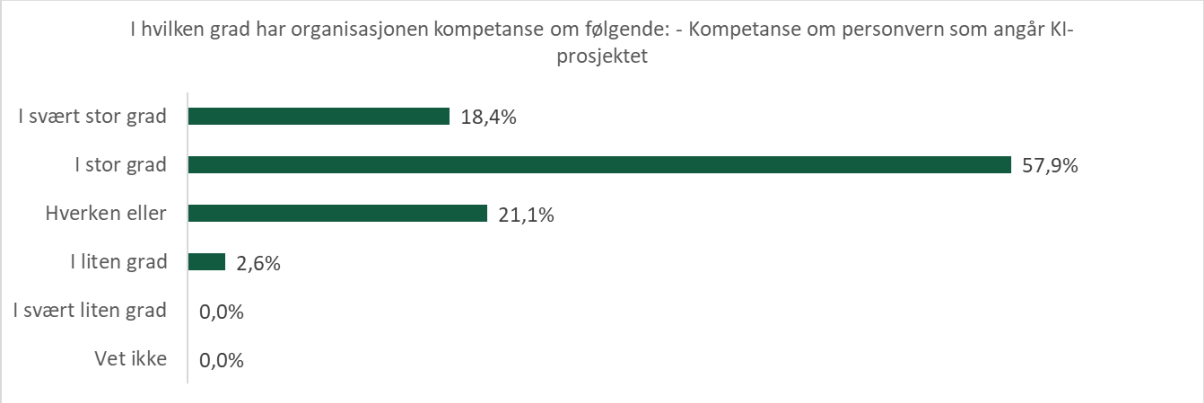
Figur 2.3. I hvilken grad har organisasjonen kompetanse om følgende: - Kunnskap om retningslinjer og lovverk som angår KI-prosjektet



På spørsmålet om KI-prosjektet blir utsatt i påvente av juridiske regelverk, svarte majoriteten at prosjektene deres i liten grad hadde blitt utsatt. Samtidig er det ikke en tydelig tendens her, fordi det fortsatt er en markant andel som stilte seg nøytral til spørsmålet om KI-prosjekt hadde blitt utsatt i påvente av juridiske regelverk. Majoriteten av informantene frykter heller ikke at juridiske regelverk skal utgjøre barrierer for at KI-muligheter forfølges.

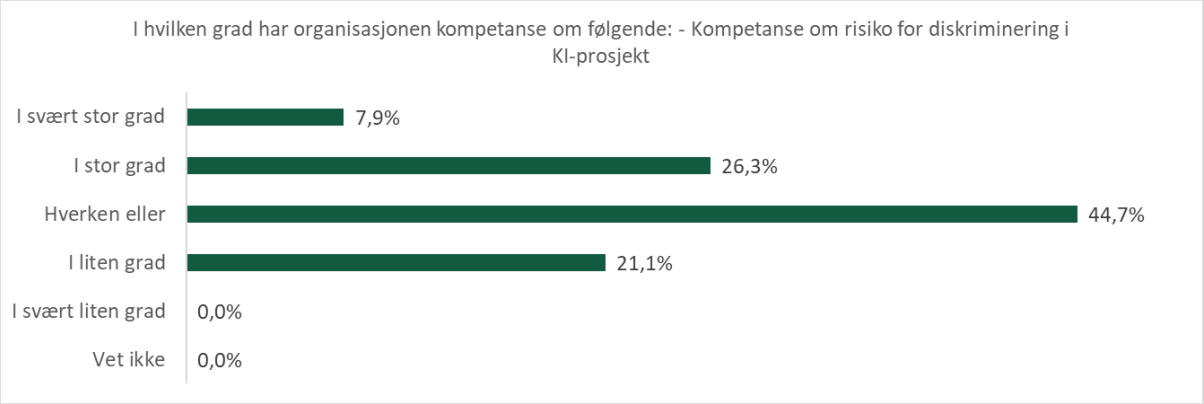
58 % mener de i stor grad har **kompetanse om personvern**. *Innenfor kompetanse er det dette området respondentene svarer at de har mest kompetanse om*. Personvern og GDPR har vært høyt på agendaen i flere år, noe som tyder på at dette er et område som virksomheter har blitt modne på (selv om denne modenheten ikke er målt i undersøkelsen).

Figur 2.4. I hvilken grad har organisasjonen kompetanse om følgende: - Kompetanse om personvern som angår KI-prosjektet



Opp mot halvparten av respondentene mener at organisasjonen hverken har eller ikke har **kompetanse om risiko for diskriminering i KI prosjekter**. En femtedel av respondenter svarer at organisasjonen i liten grad har kompetansen, mens nesten 1 av 3 svarer at de i stor grad eller svært stor grad har slik kompetanse. Mens kun en av fem respondenter i spørreundersøkelsen altså mener at organisasjonen *ikke har tilstrekkelig kompetanse* om risiko for diskriminering i KI-prosjekter, viser våre samtaler med informanter et annet bilde. Intervjuene viser at forståelsen av *diskriminering i KI* ofte ikke er entydig og bringer med seg mange ulike assosiasjoner. Mange assosierer risiko for diskriminering i KI med personvern, og enda flere "bytter ut" begrepet "diskriminering" med andre begrep, som etikk, forklarbar KI, åpenhet, med flere. I flere tilfeller svarte informanter at de *ikke hadde tenkt på risiko for diskriminering*, eller at dette temaet ikke var en del av prosjektets nåværende fase, men skulle adresseres senere. I disse tilfeller var KI-prosjektene i en utviklingsprosess, og det viser en tendens til at diskriminering i likestillings- og diskrimineringslovens forstand ikke alltid inngår i planlegging av KI-prosjekt.

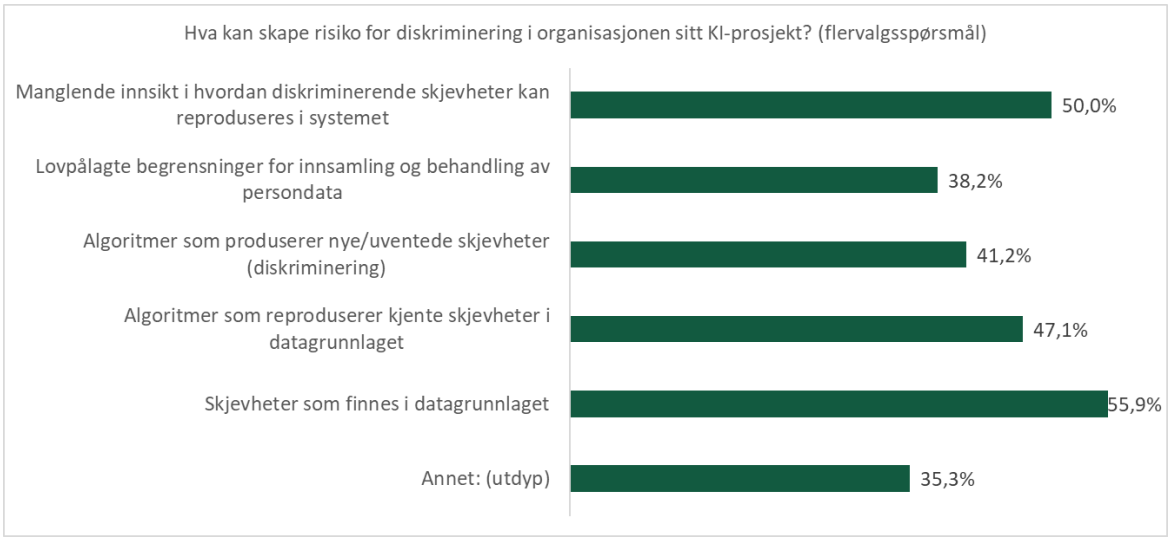
Figur 2.5. I hvilken grad har organisasjonen kompetanse om følgende: - Kompetanse om risiko for diskriminering i KI-prosjekt



2.6 Risiko for diskriminering i KI-prosjekter

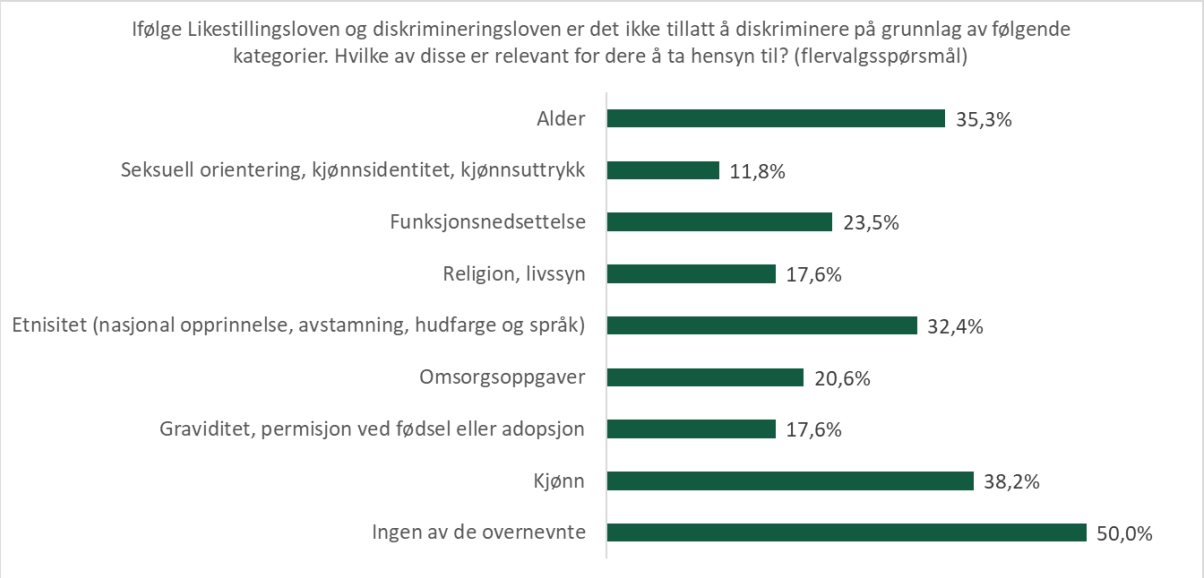
I spørreundersøkelsen stilte vi spørsmål om hva som kan *skape risiko* for diskriminering i organisasjonen sitt KI-prosjekt, der respondentene kunne velge mellom flere ulike svar, som vist i figuren nedenfor. *Skjevheter i datagrunnlaget, manglende innsikt i hvordan diskriminerende skjevheter kan reproduseres i systemet* og *algoritmer som reproduserer kjente skjevheter i datagrunnlaget* trekkes fram som de tre mest sentrale risikoaspektene knyttet til diskriminering i organisasjonene sine KI-prosjekt. Respondentene fikk også mulighet til å utdype dersom de hadde andre risikofaktorer. Her nevnte noen at risiko for diskriminering i KI-prosjektet er en ukjent eller ikke relevant problemstilling for dem.

Figur 2.6. Hva kan skape risiko for diskriminering i organisasjonen sitt KI-prosjekt?



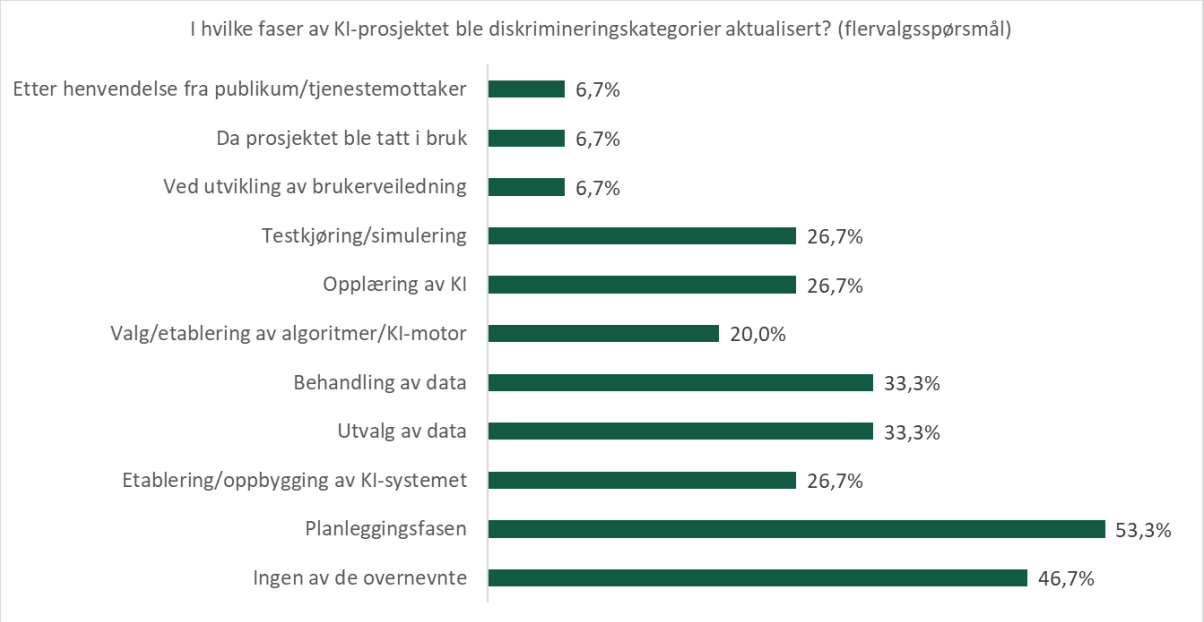
Ifølge likestillings- og diskrimineringsloven er det ulovlig å diskriminere på grunnlag *av kjønn, graviditet, permisjon ved fødsel eller adopsjon, omsorgsoppgaver, etnisitet, religion, livssyn, funksjonsnedsettelse, seksuell orientering, kjønnsidentitet, kjønnsuttrykk, alder og andre vesentlige forhold ved en person*. I spørreundersøkelsen stilte vi spørsmål om hvilke av kategoriene i loven som er mest relevante for organisasjonene å ta hensyn til. Halvparten av respondentene mente at ingen av disse kategoriene var relevante for eget prosjekt. Blant øvrige var kjønn, etnisitet og alder vektlagt som de mest relevante kategoriene å ta hensyn. Samtidig viser figuren nedenfor at *alle diskrimineringsgrunnlagene er relevant for offentlig sektor*.

Figur 2.7: Ifølge Likestillingsloven og diskrimineringsloven er det ikke tillatt å diskriminere på grunnlag av følgende kategorier. Hvilke av disse er relevant for dere å ta hensyn til?



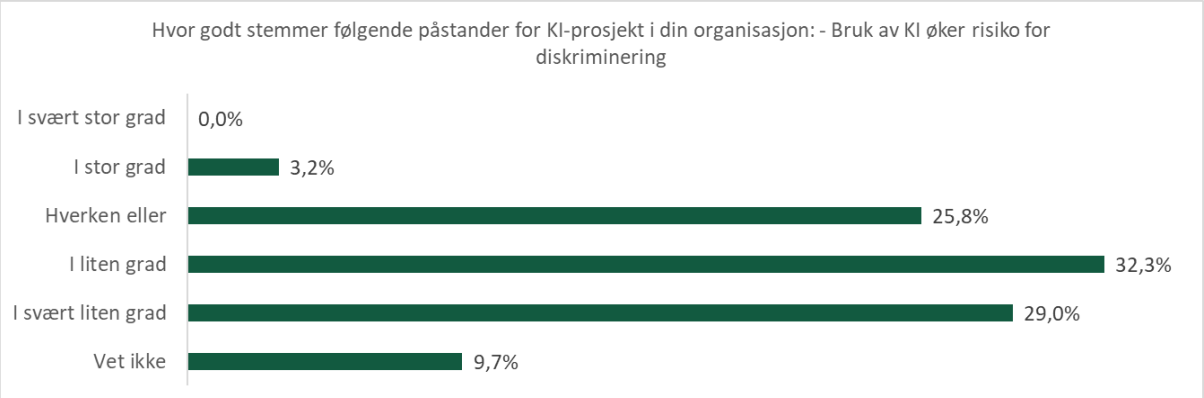
Videre stilte vi spørsmål om i hvilke faser i organisasjonens KI-prosjekt diskrimineringskategorier hadde blitt aktualisert i (se figur 2.8). Hele 53 % av respondentene våre svarer at diskrimineringskategoriene er aktuelle i planleggingsfasen, mens 47 % mener at ingen av de nevnte fasene aktualiserer diskrimineringskategoriene i deres KI-prosjekt. 33 % av respondentene vektlegger at diskrimineringskategoriene også aktualiseres når data velges ut og når data behandles.

Figur 2.8: I hvilke faser av KI-prosjektet ble diskrimineringskategorier aktualisert?



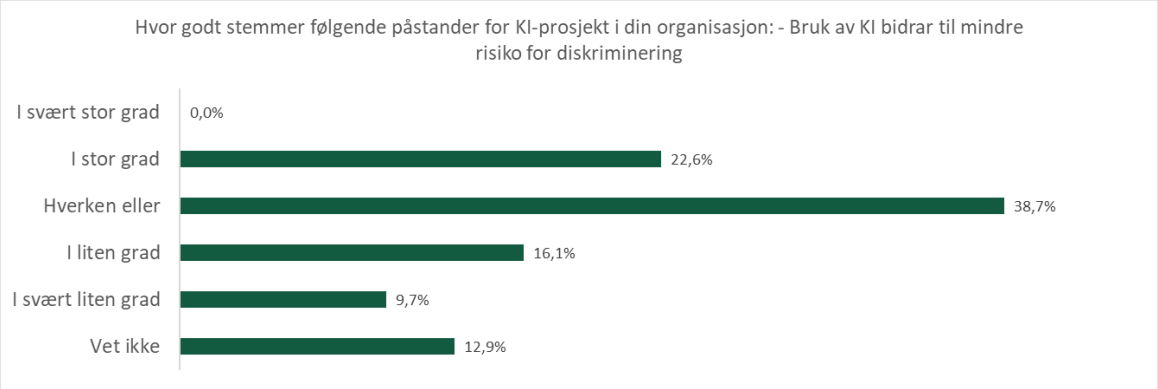
I undersøkelsen spurte vi om hvor enig eller uenig virksomhetene er i påstanden om at bruk av KI øker risiko for diskriminering (se figur 2.9). Majoriteten av respondentene er uenige i påstanden. Kun 3 % av respondentene er enig i at KI øker risiko for diskriminering, mens 26 % av respondenter svarer hverken eller på påstanden. Dette antyder at mange ikke har en klar formening eller tilstrekkelig innsikt i hvorvidt KI kan medføre økt risiko for diskriminering.

Figur 2.9. Hvor godt stemmer følgende påstander for KI-prosjekt i den organisasjon: - Bruk av KI øker risiko for diskriminering



Vi spurte også om respondentene mente at bruk av KI kan bidra til å *redusere* risiko for diskriminering (se figur 2.10). Svarene på dette spørsmålet viser ikke en tydelig tendens. Majoriteten svarer at de hverken er enige eller uenige i påstanden og 13 % av respondentene svarer at de ikke vet. Graden av respondenter som ikke har en formening i retning enig eller uenig er høyere i dette spørsmålet enn under spørsmålet om KI *øker* risiko for diskriminering.

Figur 2.10. Hvor godt stemmer følgende påstander for KI-prosjekt i din organisasjon: - Bruk av KI bidrar til mindre risiko for diskriminering



I intervjuene identifiserte vi flere tilnærmingsmåter når virksomhetene reflekterer over spørsmålene knyttet til diskriminering:

- Bias som potensial for diskriminering
- Reliabilitet og validitet
- Uklart skille mellom differensiering og diskriminering
- Ingen refleksjon rundt risiko for diskriminering

I likhet med svar fra spørreundersøkelsen som viser at mange er engstelige for hvordan skjeve data kan føre til diskriminering, nevner flest informanter **bias som potensial for diskriminering**. Bias eller feilslutninger kan oppstå når algoritmen lærer seg feil eller utilsiktede relasjoner ved å ikke vurdere (ikke ha tilgang til) tilstrekkelig data, eller når data inneholder skjevheter.

> *Jeg mener at det ikke finnes datasett som ikke har skjevhet i bunn. Fordi det enten gjenspeiler menneskets atferd som er skjev, eller vurderinger av andres atferd, som vil være skjev. (stat)*

Mange informanter reflekterer rundt spørsmålet om risiko for diskriminering gjennom **reliabilitet og validitet** av forskning. For disse informantene er risiko for diskriminering et resultat av svekket pålitelighet og gyldighet.

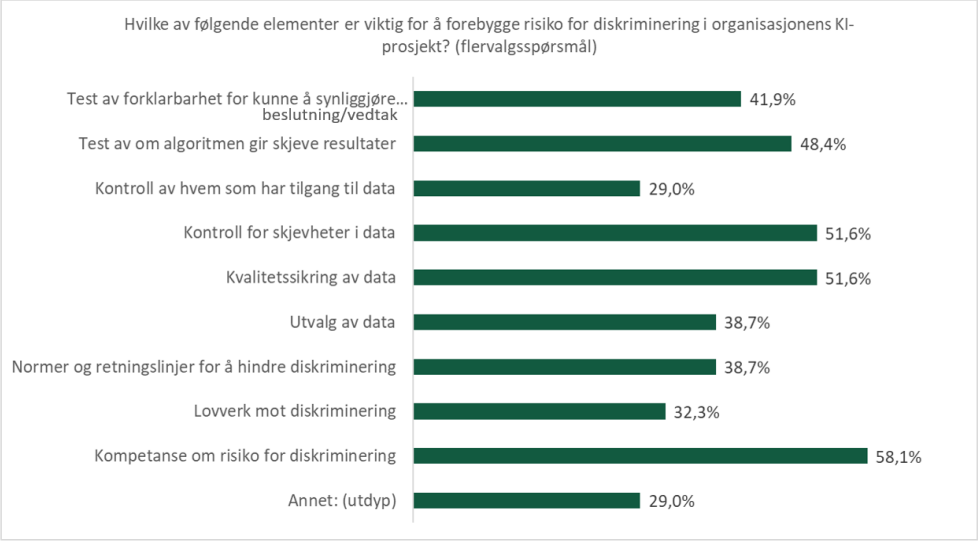
Enda flere forveksler begrepene **differensiering og diskriminering**. De nevner for eksempel at diskriminering, her i betydningen "differensiering", mellom mann og kvinne, eller mellom gravide og ikke-gravide er naturlig.

Noen informanter tolker spørsmål om *diskriminering* som utfordringer knyttet til *personvern* i GDPR-loven, snarere enn til likestillings- og diskrimineringsloven. Til sist er det flere informanter som forteller at de ikke har brukt tid på å vurdere diskriminering.

2.7 Tiltak for å unngå diskriminering i KI-prosjekter

Hvilke elementer er viktig for å *forebygge risiko for diskriminering* i organisasjonens KI-prosjekt, spurte vi i spørreundersøkelsen (se figur 2.11). Majoriteten av respondentene beskriver **kompetanse om risiko for diskriminering** som et sentralt tiltak for å unngå diskriminering i KI prosjekter. Videre er både **kvalitetssikring av data** og **kontroll for skjevheter i data** relevante tiltak som trekkes frem av 52 % av respondentene, og 4 8% mener at **testing av algoritmer som gir skjeve resultater** er viktig.

Figur 2.11. Hvilke av følgende elementer er viktig for å forebygge risiko for diskriminering i organisasjonens KI-prosjekt?



I intervjuene trekker flere informanter frem **forklarbarhet og etterprøvbarhet** som de viktigste elementer for å redusere risiko for diskriminering. For eksempel, det å teste variabler som er brukt i KI-modellen kan avdekke nye former for diskriminering som ikke var mulig å forutse i begynnelsen av prosjektet. Det kan i tillegg redusere "systematisk bias" som finnes i data, mener en informant.

Flere av informantene reflekterte rundt diskriminering i forbindelse med bias og pålitelighet i forskning.

> *Jeg tenker at hvis vi skiller mellom algoritmer, så er det to grupper; vi har de nevrale nettverkene – man kan egentlig ikke etterprøve hva algoritmen kan gjøre, de er bare trent på et stort datasett og kan vise gode resultater. Så har vi de algoritmene som er veiledet, og man kan ta ut og vekte ulike variabler. De er transparente, og vi kan publisere det som åpen kildekode og diskutere det i samfunnet, og hvilke gode resultater de gir sammenlignet med våre kompetente og nøyaktig medarbeidere som korrigerer manuelt. (stat)*

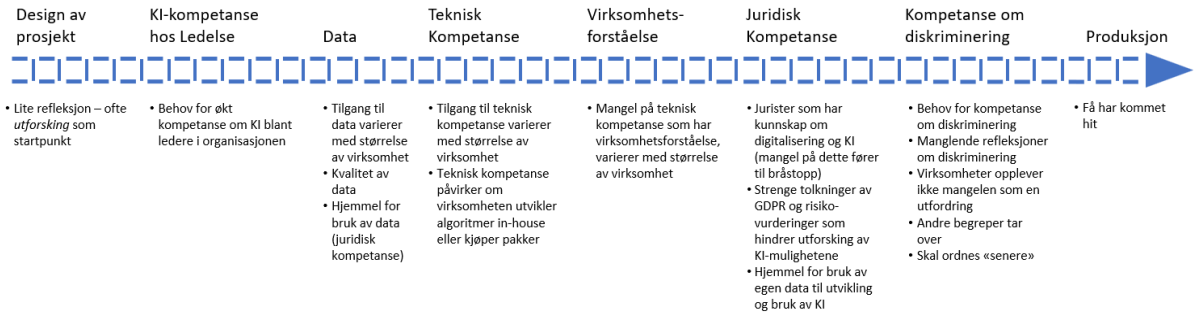
Informanter som anerkjente at de ikke hadde reflektert omkring diskriminering i startfasen av prosjektet, mente likevel at de kan håndtere diskriminering etter at de begynner med prosjektet.

> *Diskriminering var ikke et tema i det hele tatt, virkelig ikke. Men det hadde vi kanskje kommet til på et eller annet tidspunkt hvis vi måtte tenke på kjønn, alder. Hvis vi hadde kommet lenger i prosjektet. Men ikke på det stadiet som vi var. (stat)*

3. Hvilke utfordringer er knyttet til bruk av KI i offentlig sektor?

Mange virksomheter i offentlig sektor er i ferd med å utforske hvordan KI kan bli brukt for å gjøre offentlige tjenester bedre og mer effektive. Samtidig viser undersøkelsen vår at reisen fra ide til produksjon også byr på mange utfordringer som ofte fører til at KI-prosjekter stoppes midlertidig eller avsluttes. Figuren nedenfor illustrerer et KI-prosjekt som en linje med en rekke elementer som alle må være på plass for å komme trygt fra start og "design av prosjekt" til slutt og "produksjon". Selv om de ulike elementene ikke representerer en perfekt kronologisk utvikling, så har vi i analysen diskutert dette som en "toglinje" der de ulike elementene har likhetstrekk ved stoppesteder på en hop-on-hop-off reise: Det var ikke alle KI-prosjektene vi hørte om som hadde begynt med hverken en overordnet design- eller ledelsesstrategi. Og det var heller ikke alle som kom i mål med "produksjon". Ved bruk av denne figuren ser vi i dette kapitlet på utfordringer som ble identifisert i spørreundersøkelsen og i intervjuene.

Figur 3.1 En togreise med KI fra start til produksjon



3.1 KI-kompetanse hos ledelse

I lys av kunstig intelligens er data "den nye oljen vi enda ikke har pumpet opp", hevder Norwegian Cognitive Center og Bergen Næringsråd i en rapport fra 2022. De to organisasjonene har i fellesskap kartlagt hvordan det står til med kunnskap om KI blant ledere i næringslivet i nær 400 virksomheter. Et av hovedfunnene er at *"Toppledere forstår i vesentlig mindre grad enn sine medarbeidere og mellomledere hvor kritisk kunstig intelligens og data er for framtidens næringsliv".*[49] Vår undersøkelse viser at også i offentlig sektor er dette en utfordring. 87 % av respondentene i spørreundersøkelsen var enige i at det er behov for økt kompetanse om KI blant ledere i offentlig sektor.

Det er ikke bare kompetanse, men også det å finne tid til i det hele tatt til å ta opp KI som et tema, som kunne være utfordrende.

> *Det å få ledelse i en kommune til å tenke utover egen drift er vanskelig. Det har vært et tungt løft fordi mange ikke klarer å bidra, er for opptatt med sin daglige drift, så mye kompetanse er hentet fra utsiden, gjennom investeringsbudsjett, så leverandøren av plattformen har bidratt inn i dataprosjektet. (kommune)*

Mens ledelsens involvering i liten grad ble tematisert i intervjuene, ser vi samtidig tendenser til at *fravær av toppstyring* gjør KI til et spørsmål om *hvem som vil* i virksomheten:

> *Grov kategorisert har vi en veldig stor gruppe som ikke har viljen og sier "Nei, vi vil ikke", og en grei gruppe som sier "Joda, det kunne vært noe, men vi har ikke data". Så har vi en supersnever gruppe som sier "Oj, det må prøve å få til". De*

[49] Norwegian Cognitive Center og Bergen Næringsråd (2022) Digital Modenhet på Vestlandet. Delrapport 1: Kunstig intelligens, rapport

sitter typisk på økonomitjenesten. Det er fordi de har orden på tallene, og de har viljen. (kommune)

Ettersom vår undersøkelse har fokusert på de virksomhetene i stat og kommune som *har pågående KI-aktivitet*, antyder dette og flere eksempler fra intervjuene at KI-utviklingen fortsatt er *ganske umoden i store deler av offentlig sektor*. Utviklingen ser ut til å være relativt *lite lederstyrt*, og i større grad *styrt av interesse og vilje til å prøve ut teknologien*. Dertil kommer et *politisk påtrykk* for å bruke KI i offentlig sektor,[50, 51] som også ble omtalt en rekke ganger i intervjuene. Ser vi samlet på de tre nivåene ser det dermed slik ut:

- *Politisk nivå:* Strategier for digitalisering i offentlig sektor og for KI i Norge skaper forventninger til bruk av KI.
- *Ledernivå i offentlig sektor:* Lavt kunnskapsnivå, begrenset fokus på å etablere eller styre utvikling av KI.
- *Virksomhetsnivå*: Enkeltpersoner og enheter innen den enkelte virksomhet som har "vilje" bidrar til å fremme KI-utviklingen.

3.2 Data

Et KI-system er resultatet av at en algoritmisk modell analyserer data for å identifisere sammenhenger, og et KI-system har behov for en betydelig mengde data å trene på.[52] Det er mange utfordringer knyttet til data, fra å identifisere, høste, tillatelse til å bruke, dele osv., og en del av disse utfordringene blir diskutert i *Nasjonal strategi for kunstig intelligens*.[53] Som nevnt over, har undersøkelsen fokusert på virksomheter som har KI-prosjekt som involverer persondata eller individdata – altså alle former for data som omhandler individer og som dermed kan gi opphav til diskriminerende resultater.

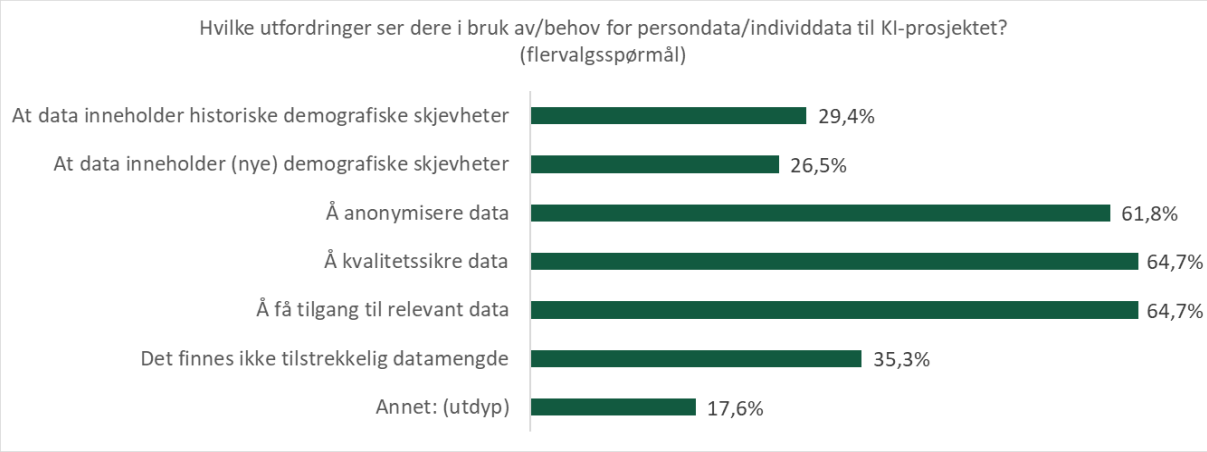
3.2.1 Persondata/individdata

Det er mye oppmerksomhet blant KI-utviklere og -forskere om hvordan mangler og skjevheter i data som brukes for å utvikle et KI-system, kan gi risiko

[50] KMD (2019), *Én digital offentlig sektor: Digitaliseringsstrategi for offentlig sektor 2019–2025*
[51] KMD (2020), *Nasjonal strategi for kunstig intelligens*,
[52] Gröger, C. (2021), There Is No AI Without Data. *Communications of the ACM*, 64(11), 98-108. doi:10.1145/3448247
[53] KMD (2020)

for skjevheter i resultater, ofte omtalt som *bias*, men også som diskriminering, urettferdig eller skadelig resultat.[54, 55, 56] I spørreundersøkelsen stilte vi spørsmål om hvilke utfordringer offentlige virksomheter ser i forhold til persondata/individdata i KI-prosjekter. Som figuren nedenfor viser finner nesten 2 av 3 respondenter at *tilgang til relevant data* samt *kvalitetssikring og anonymisering av data* er utfordrende. Om lag en tredjedel har utfordring med at det *ikke finnes tilstrekkelig data*, fulgt av utfordring med at *data inneholder demografiske skjevheter*.

Figur 3.1. Hvilke utfordringer ser dere i bruk av/behov for persondata/individdata til KI-prosjektet? (flere alternativer mulig)



3.2.2 Tilgang på data

Utfordringer rundt tilgang til data var også en gjenganger i intervjuene:

> *Tilgang til data er ofte en utfordring. Få tak i data, hvordan få dem tilgjengeliggjort, hvordan anonymisere, få nødvendige godkjenninger for å få tilgang. Vi snakker om personsensitive data som det finnes lovreguleringer rundt. Vi er umodne på dette fortsatt, spesielt når vi snakker om stordata.*

[54] Srinivasan, R., & Chander, A. (2021), Biases in AI systems. *Communications of the ACM*, 64(8), 44-49.
[55] Friedman, B., & Nissenbaum, H. (1996), Bias in computer systems. *ACM Transactions on information systems (TOIS),* 14(3), 330-347.
[56] Belenguer, L. (2022), AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. *AI and Ethics*. doi:10.1007/s43681-022-00138-8

Det er mange uavklarte spørsmål knyttet til data som gjør at KI fortsatt fremstår som en umoden teknologi for offentlig sektor. Utfordringer knyttet til data varierer også med en virksomhets størrelse, noe som også gjelder for offentlig sektor.

> *Anbud, leverandører, og tilgang til data er generelt et stort problem. Fordi mange av våre systemleverandører mener at dataene slett ikke er våre. Eller de vil ikke gi innsyn i datamodellen. Det er derfor jeg endte opp med å bruke data fra kvalitet på asfalt. Fordi det er data som vi har selv. (kommune)*

Mens store virksomheter ofte har god tilgang til store datamengder som kan muliggjøre det å utforske mulighetene i bruk av KI, strever små virksomheter med tilgang til nok data. Små virksomheter har både mindre tilgang til egne data og mindre kapasitet til å innhente treningsdata for å utvikle et KI-system.

3.2.3 Kvalitet på data

Ikke bare tilgang til data, men også god kvalitet på data er avgjørende for at KI-prosjekter blir satt i gang. Mange virksomheter jobber med å forbedre kvaliteten på egne data. I flere virksomheter kan det dreie seg om at grunnlagsdata er generert ut fra informasjon som er produsert, brukt, lagret eller distribuert på ulike måter på tvers av ulike fagmiljøer og IT-systemer i samme virksomhet.

> *Når en jobber med informasjon er det viktig at informasjon forstås på samme måte på tvers av de miljøene der informasjonen spiller en rolle. Man må ha noen klare spilleregler for det. Har man ikke god nok kontroll på dette så vil datakvaliteten være varierende. Og da klarer man kanskje ikke å sette sammen rapportene. Da må man korrigere det før det presenteres – endre på dataene, og da blir det vanskelig å føre tilbake. Da forsvinner kunnskapen om*

hvordan man får disse usammenhengende dataene til å henge sammen i en rapport. (stat)

3.2.4 Rett til å bruke data

Rett til å bruke virksomhetens data til utvikling av KI-system er den tredje utfordringen som peker seg ut i intervjuene. Bruk av virksomhetens data til KI reiser nye og andre spørsmål både om behov og rett til å bruke persondata enn det som dekkes av virksomhetens (tradisjonelle) hjemler for å håndtere slike data.

> *For å unngå personvern problemstillingen jobber jeg en del med energidata. (…) Her er heller ikke bias involvert. (kommune)*

Også dette punktet kan resultere i at KI-prosjekt stopper opp helt eller midlertidig:

> *Vi prøvde å sladde, og bruke materialet vårt med å ta ut personopplysninger og sensitive opplysninger. Det var en vei inn, for da tenkte vi selvsagt ikke å innhente samtykke. Men personvernombudet vårt var strenge med oss og vi kom til den konklusjonen at hvis vi skulle sladde og gjøre så mye med materialet så vil det ta så mye tid at det var ikke overkommelig for oss.*

Prosjektet stoppet opp når avveiningene mellom nytte (effektivisering) og kostnad (personvern) av å ta i bruk KI ble synliggjort av personvernombudet.

NAV er en av virksomhetene som har deltatt i Datatilsynets *Regulatoriske sandkasse*. I sluttrapporten ble nettopp denne skvisen mellom KI og juridiske vurderinger synliggjort. Mens NAV har det rettslige grunnlaget for å bruke KI som beslutningsstøtte for saker som angår enkeltindivider, finner Datatilsynet det "usikkert om det rettslige grunnlaget åpner for å bruke personopplysninger

til å utvikle selve algoritmen".[57] I tillegg "oppstår en spenning mellom personvern og rettferdighet" når det å avdekke og motvirke diskriminering krever *mer* behandling av personopplysninger. Saken illustrerer hvordan ulike hensyn trekker i ulike retninger; mens bruken av KI i seg selv kan bidra positivt i virksomheten, er det juridiske rammeverket ikke tilpasset behovet for (store mengder) data for å trene opp et KI-system.

3.3 Teknisk kompetanse

3.3.1 Intern VS ekstern kompetanse

Har det offentlige Norge teknisk kompetanse knyttet til KI? For det store flertallet er svaret nei, og særlig for kommune-Norge. På vårt spørsmål om kommuner kan skaffe nødvendig kompetanse for å utvikle KI svarer en representant for sektoren: "*Nei, jeg ser ikke for meg at vi skal klare å knytte til oss den type kompetanse for å drive med det sjøl*". For de fleste betyr det en avhengighet av ekstern kompetanse:

> *Vil vel si at så langt har vi støttet oss på eksterne ressurser. Vi har noen halv-lyse hoder her hos oss også, men vi er ikke spesialister, så til [KI-prosjektet] har vi et eksternt firma som har hjulpet oss med ideer og hvordan ting skal løses. (kommune)*

3.3.2 Tilgang til kompetanse

Tilgang til teknisk kompetanse varierer også med virksomhetens størrelse. Små virksomheter, slik som små og middels store kommuner, har ikke samme tilgang til teknisk kompetanse som større kommuner:

> *Jeg vil si at de 10 største kommunene har en stor IT-tjeneste som kan gape opp over det. Jeg er den eneste kommuneansatt*

[57] Datatilsynet (2022). Sluttrapport fra sandkasseprosjektet med NAV: Temaer: rettslig grunnlag, rettferdighet og forklarbarhet, https://www.datatilsynet.no/contentassets/ebd705b85bbc4bfc8b13638a28863e10/nav_sluttrapport.pdf

som er datascientist og jeg er faktisk ikke leid inn. Det er fordi denne kommunen er stor nok til å gjøre det. Utav 356 kommuner har godt over 200 færre enn 5000 innbyggere og som har dertil liten IT-tjeneste. De har null sjanse til å ha den (nødvendige) kompetansen på det nivået. (kommune)

Offentlig sektor har ikke kapasitet til å gjøre grunnforskning på KI-løsningene, sier noen. Samarbeid med forskningsmiljøer er viktig grunnlag for utvikling av løsninger og overføring av kunnskap. Mange ser heller ikke behov for å lage **interne algoritmer** og vil heller **kjøpe algoritmer** – altså KI som "hyllevare" som allerede er tilgjengelig, og det gjelder enheter både i stat og kommune-sektor. KI som "hyllevare" reiser flere utfordringer, blant annet i forhold til innsikt i hvordan modellen er utviklet, hvilke data modellen er basert på og hvilken risiko for diskriminering som kan følge med bruken av modellen. Samarbeid med eksterne reiser også spørsmål i forhold til neste punkt på listen: virksomhetsforståelse.

3.4 Virksomhetsforståelse

Mens flere små virksomheter bruker leverandører til å dekke manglende teknisk kompetanse, viser flere til at de også mangler tverrfaglig kompetanse som inkluderer både teknisk kompetanse og virksomhetsforståelse, og dette dekkes ikke alltid like godt av eksterne leverandører. Behovet for å bruke ekstern kompetanse på de tekniske sidene av KI-prosjekt, og manglende virksomhetsforståelse hos ekstern leverandør på den andre, oppleves som et problem for mange. For noen har kunnskapsgapet som da oppstår, ført til betydelige utfordringer og utgjort en trussel mot KI-prosjektet, blant annet fordi det har skapt usikkerhet om hvorvidt valgte KI-løsninger overholder relevante regelverk. Flere virksomheter satser også på at de etter hvert vil overføre teknisk kompetanse fra eksterne samarbeidspartnere til interne medarbeidere som har den nødvendige virksomhetsforståelsen.

Disse eksemplene illustrerer ikke bare behovet for tverrfaglig kompetanse i KI-prosjekt, men at behovet for tverrfaglighet ikke bare gjelder innad i et team,

men også for personene som deltar. Behovet det pekes på av flere er at den som har teknisk kompetanse *også* bør ha forståelse for virksomhetens mål og målgrupper, at den som har juridisk kompetanse *også* forstår personvern eller diskriminering, osv. En måte å oppnå slik spisset tverrfaglig kompetanse kan være offentlig sektor-ph.d., altså en doktorgrad som gjennomføres i en offentlig sektor virksomhet og med fokus på virksomhetens utfordringer, noe som flere av virksomhetene var oppmerksomme på.

3.5 Juridisk kompetanse og personvern

Behov for juridisk kompetanse er avgjørende for å finne ut om, og hvordan, persondata kan brukes i KI-prosjekter. Flere virksomheter understreker at **jurister som har god kunnskap om digitalisering og KI** er avgjørende for å kunne jobbe med KI prosjekter. I noen tilfeller har virksomheter begynt med KI prosjekter, men på grunn av mangel på *tverrfaglig kunnskap* i *KI og juridiske spørsmål* knyttet til tolkning av GDPR-loven og til virksomhetens hjemmel for å håndtere persondata, har aktiviteter knyttet til utforsking og bruk av KI stoppet opp forholdsvis raskt etter oppstart.

Vi ser store variasjoner i når og hvordan kunnskap om personvern blir involvert i KI-prosjekter. I noen tilfeller blir kompetanse om personvern involvert for å vurdere risikoene i tidlig fase av et KI-prosjekt:

> *Vi gjennomfører risiko-sårbarhetsanalyse og de som er ansvarlig må komme med alle mulige scenarier for hva som kan gå galt og rating-system for hvor sannsynlig og hvor alvorlig det er, i en matrise, i en grafisk framstilling. Så må vi beskrive alle risiko-mitigerende tiltak, og vi har et miljø som passer på at alle tiltak blir fulgt opp. Vi er ganske strenge. Vi har et stort miljø rundt personvern og informasjonssikkerhet, så det fokuseres veldig mye på det, og vi kjører alltid sårbarhetsanalyse, og hvis det behandler persondata på en eller annen måte så er personvern på med en gang. (stat)*

For noen har fokus på personvern blitt en *stopp-melding* for et påbegynt KI-prosjekt, som vi så over, og flere påpeker konflikten i å oppfylle ulike behov. På grunn av de krav som stilles til offentlig sektor er det ofte risiko knyttet til KI som til sist vinner. I slike tilfeller oppleves rutiner noen ganger som *for **strenge***, og dermed som **hinder for utforsking av mulighetene** som KI kan representere for virksomheten.

> *Av og til føles det som det går på bekostning av kreativiteten, men det er bedre det enn at det kommer et avisoppslag om at vi har gjort noe feil, så det er nesten litt smertelig at vi ikke kan gjøre mer. (stat)*

I spørreundersøkelsen mente 35 % at frykt for konflikt med juridiske regelverk førte til at muligheter ved KI ikke ble fullt utnyttet. 29 % mente også at KI-prosjekt har blitt utsatt i påvente av et bedre og mer oppdatert juridisk regelverk som kunne dekke problemstillinger som oppstår med KI.

Samtidig nevner flere i intervjujene at de føler på kravet til offentlig sektor om å være ressurseffektiv, for eksempel understreket i nasjonale strategier for digitalisering og KI, som fremhever at offentlig sektor har mange fordeler å hente, for eksempel ved at teknologi bidrar til å "effektivisere ressursbruken gjennom styrket samordning på tvers av forvaltningsnivåer og sektorer, og systematisk uthenting av gevinster fra digitalisering".[58] Forventning til at offentlig sektor forvalter sine ressurser godt gjør at noen også spør; *kan vi la være å bruke KI når det kan effektivisere oppgavene våre?* I praksis opplever altså flere at de blir skvist mellom en *forventning* til å ta i bruk KI, som åpenbart kan bidra til effektivisering og store innsparinger for offentlig sektor på den ene siden, og på den andre siden usikkerhet, advarsler eller til og med *stopp-meldinger* på grunn av utfordringer knyttet til bruk av persondata i KI-system.

Noen virksomheter klarer imidlertid å *fortsette* arbeidet med KI selv etter å ha støtt på utfordrende spørsmål om personvern. En fellesnevner for disse er at de har et nettverk som har god kunnskap om juridiske spørsmål knyttet til KI.

[58] KMD (2019)

> *Vi prøver å jobbe etter en litt smidig tilnærming. Ofte vil det være en jurist som er 20 % med i prosjektet som følger statusmøter, endringsforslag osv., for å få det perspektivet tidlig i løpet. (stat)*

I andre tilfeller ble det å definere arbeidet som et *internt prosjekt* oppfattet som en løsning når de støtte på utfordringer knyttet til spørsmål om persondata, personvern og rettigheter til å bruke data:

> *Spillereglene er forskjellige om du gjør noe internt eller eksternt. Når det er slik at kommuner vil sette i gang KI-prosjekter internt, har man større fullmakter. (kommune)*

Ulike måter å løse juridiske spørsmål rundt bruk av persondata i slike tilfeller inkluderte pseudonymisering, altså å endre/vaske data slik at data ikke lenger kan knyttes til en bestemt person, eller aggregerte data som innebærer en sammenslåing av data for å unngå personidentifikasjon, eller å trene algoritmer ved å bruke data som ikke involverer persondata.

3.6 Kompetanse om diskriminering

Et av elementene som i liten grad påvirket "togreisen" var kompetanse om diskriminering. I spørreundersøkelsen var det et fåtall som hadde klare strategier for å håndtere diskrimineringsrisiko. Som vist i forrige kapittel var det bare 3 % i spørreundersøkelsen som mente at KI øker risiko for diskriminering i virksomhetens KI prosjekter. På spørsmål om når diskrimineringskategorier ble aktualisert, svarte nesten halvparten "aldri", mens resten mente det kunne være en aktuell problemstilling i planleggingsfasen (53 %), i utvalg (33 %) og behandling av data (33 %). Et langt vanligere svar på spørsmål som "hvordan håndterer dere risiko for diskriminering?", var at dette ikke hadde vært gjenstand for et spesielt fokus i prosjektet. Intervjuene illustrerte også at noen ikke hadde tenkt på problemstillingen overhodet:

> *Nei, må ærlig innrømme at det [vurdere diskriminering] har vi ikke gjort. Men det var interessant at denne undersøkelsen kom ut, for det er også noe vi må tenke på. (kommune)*

Noen mente at prosjektet ikke hadde kommet langt nok til å vurdere diskriminering, mens andre igjen mente det ikke var en utfordring, fordi data var anonymisert. De svært ulike holdningene til problemstillingen illustrerer ikke bare utfordringer med å håndtere eventuelle risikoer for diskriminering, men samlet peker datamaterialet på at *gapet* mellom de som er oppmerksomme på risikoen og de som ikke oppfatter det som en aktuell problemstilling, er stor i offentlig sektor.

Gjennom intervjuene finner vi at en årsak til de ulike forståelsene av hvorvidt diskriminering utgjør en risiko ved KI, er et tilsvarende stort sprik i forståelse av hva diskriminering i seg selv handler om. Mens vi stilte spørsmål om *diskriminering i lys av likestillings- og diskrimineringsloven* og de ulike diskrimineringsgrunnlagene denne verner, erfarte vi i mange intervjuer at svaret handlet om noe annet: *om forklarbarhet, åpenhet, transparens, etikk, rettferdighet eller bias*. Eller at ordet diskriminering ble brukt, men med en annen mening: differensiering som et *naturlig skille* mellom folk, for eksempel gravide vs. ikke-gravide. Mens dette handler om en ønsket form for differensiering for å kunne gi den enkelte de tjenester de har behov for, handler diskriminering i likestillingslovens forstand om *uønskede* måter å gjøre forskjell på individer og grupper. Omskrivingen av spørsmålet om diskriminering til å handle om noe annet enn likestillings- og diskrimineringslovens tolkning, viser et stort behov for synliggjøring, bevisstgjøring og økt kunnskap om diskriminering.

Det er rimelig å tolke dette kunnskapshullet som et resultat av de *nye utfordringene* som KI reiser, som en av informantene uttrykte det:

> *Ja, akkurat nå er den største risikoen for diskriminering frykten for risiko for diskriminering. Den største risikoen i forhold til maskinlæring og diskriminering er frykten for diskriminering. Det er bare å komme i gang. Man må være*

åpen på problemet, være åpen på prosessen, ha dyktige folk, ikke være redd, det er ingen i Norge som har gjort det før. (kommune)

En tilsvarende utfordring er vist i sammenheng med digitalisering i andre former, nemlig et økt behov for tverrfaglig kompetanse, fordi digitalisering ikke skjer i en digital silo, men skal samspille med en rekke ulike sosiale, kulturelle, politiske, juridiske faktorer og spilleregler.[59, 60] Risiko for diskriminering blir altså i liten grad oppfattet som sentralt i utfordringsbildet som informantene fra offentlig sektor tegner. Begrenset kompetanse om diskriminering og lite fokus på diskriminering forsterker hverandre på en slik måte at virksomhetene i liten grad opplever at de mangler kompetanse om diskriminering.

3.7 Paradokser, prioriteringer og balanse

Jeg tror at å bruke maskiner i større grad i saksbehandling vil være en bra ting, ikke minst med tanke på ressursbruk i offentlig sektor. Samtidig er det en reell risiko knyttet til diskriminering, men kan vi si at mennesker er bedre enn maskiner heller enn omvendt? Jeg tror ikke det: menneskelig skjønn betyr i praksis at en og samme sak kan håndteres ulikt av to personer. (stat)

Mange av utfordringene vi har pekt på i dette kapitlet handler om frustrerende paradokser som oppstår når KI blir introdusert i offentlig sektor. For det første, at offentlig sektor har andre krav til å være korrekt i tjenestene enn privat sektor, fordi kundene (innbyggerne) ikke kan velge vekk offentlig sektor sine tjenester. For det andre, er regelverk og det juridiske rammeverk ikke tilpasset den digitale realiteten, og dermed forblir mange spørsmål uavklart, eller har

[59] Gerards, J. and Xenidis, R. (2021) Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law. European Commision.
[60] Bartoletti, I. & Xenidis, R. (October 2022). Preliminary draft Council of Europe study on the impact of artificial intelligence, its potential for promoting equality, including gender equality, and the risks to non-discrimination, The Gender Equality Commission (GEC) and the Steering Committee on Anti-Discrimination, Diversity and Inclusion (CDADI), The Council of Europe, https://rm.coe.int/gec-2022-9-study-on-ai-211022/1680a8ad89

flere og motstridende tolkninger. For det tredje, er de ulike regelverkene som krysser hverandre på dette feltet i konflikt med hverandre. Det overordnede prinsippet i GDPR-lovgivningen er å lagre "minst mulig data i korteste mulig periode", mens KI krever mye data, og i noen tilfeller *mer* data enn det opprinnelige systemet dersom diskriminering skal utelukkes. Slike paradoks blir gjenstand for vurdering og prioritering: Hva veier tyngst av utvikling og effektivisering på den ene siden, og risiko for at feil skal skje på den andre siden?

4. Risiko for diskriminering når offentlig sektor tar i bruk kunstig intelligens

Norge har, sammenlignet med andre land, en svært stor offentlig sektor, som forvalter offentlige ressurser, midler, tilbud og tjenester som ofte ikke mottas gjennom privat eller tredje sektor. Offentlig sektor i Norge har en særlig forpliktelse til å gi et godt og rettferdig tjenestetilbud.[61] Derfor har det offentlige *også en unik tilgang til data om borgerne.* Dette er en tilgang til data om befolkningen som offentlig sektor har etablert over tid,[62] og som omfatter alt fra sensitive persondata, helsedata, til data om arbeid, utdanning og økonomi helt ned til detaljer fra dagligvarehandel.[63] I tillegg har den norske befolkningen, i internasjonal sammenheng, en høy grad av *tillit* til offentlig sektor.[64] Riktig og god bruk av teknologi, herunder kunstig intelligens (KI), er viktig for tillit til offentlig sektor, og i siste instans, for demokratiet.[65] Fageksperter har uttrykt bekymring for at det offentliges tilgang til store mengder data sammen med økt digitalisering "kan gi et press for å bruke data på nye og inngripende måter" samt at det er en risiko for at utviklingen peker i retning større kontroll ved at for eksempel KI i økende grad brukes til å

[61] KMD (2019)
[62] Broomfield, H., & Reutter, L. M. (2021)
[63] Statistisk sentralbyrå (2022), Leveranse av bongdata fra dagligvarekjedene Rema 1000, NorgesGruppen, Coop og Bunnpris, (SSB 6. mai 2022), https://www.ssb.no/omssb/ssbs-virksomhet/kost-nyttevurdering/leveranse-av-bongdata-fra-dagligvarekjedene-rema-1000-norgesgruppen-coop-og-bunnpris
[64] OECD (2022), Drivers of Trust in Public Institutions in Norway, *Building Trust in Public Institutions*, OECD Publishing, Paris, https://doi.org/10.1787/81b01318-en.
[65] Andreasson, U., & Stende, T. (2019), Nordiske kommuners arbeid med kunstig intelligens: Nordic Council of Ministers.

"avdekke juks og svindel".[66, 67] Mens negative effekter av KI ofte er utilsiktet,[68] er klare mål og strategier også nødvendige for å unngå en utvikling i retning av "offentlig overvåkingskapitalisme".[69, 70] I forrige kapittel presenterte vi status for KI i offentlig sektor og diskuterte utfordringsbildet som har kommet frem gjennom spørreundersøkelsen og intervjuene med aktører fra statlig og kommunal sektor. I dette kapitlet skal vi med utgangspunkt i innsikt fra den nasjonale KI-strategien, KI-forordningen samt faglitteratur om risiko for diskriminering og tilstøtende problematikk, ta et tilbakeblikk på utfordringene knyttet til diskriminering som undersøkelsen og intervjuene identifiserte. Kapitlet peker på ulike former for risiko for diskriminering ved bruk av KI i offentlig sektor, til sist i kapitlet presentert som et rammeverk for å forstå og å identifisere slik risiko.

4.1 Mennesker, maskiner, feilbarlighet og tillit

Er det verre at teknologi gjør feil enn at mennesker (saksbehandlere) gjør feil, er et spørsmål vi hørte en rekke ganger i denne studien. Med dette spørsmålet har mange også stilt spørsmål om hvorfor vi forsker på risiko for diskriminering ved bruk av KI, og hvorfor det tilsynelatende stilles høyere krav til teknologi enn mennesker i denne sammenheng. Vårt svar har vært at KI har potensial til å gjøre feil i stort omfang og hurtig tempo samt at feil i offentlig forvaltning kan ha store konsekvenser for den enkelte det eventuelt rammer. I tillegg krever forvaltningsloven at alle enkeltvedtak skal begrunnes. Dette gjør det ekstra kritisk at KI-systemer i offentlig sektor er transparente og forklarbare, og altså

[66] Broomfield, H., & Lintvedt, M. N. (2022), Is Norway Stumbling into an Algorithmic Welfare Dystopia? *Tidsskrift for velferdsforskning*, 25(3), 1-15. Doi: 10.18261/tfv.25.3.2.

[67] Alston, P. (2019). Report of the Special Rapporteur on extreme poverty and human rights. UN
General Assembly A/74/493. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/312/13/PDF/N1931213.pdf?OpenElement

[68] Redden, J. (2018). Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society, 5*(2). https://doi.org/10.1177%2F2053951718809145.

[69] Broomfield & Lintvedt 2022.

[70] Jørgensen, R. F. (2021). Data and rights in the digital welfare state: the case of Denmark. *Information, Communication & Society*. https://doi.org/10.1080/1369118X.2021.1934069

ikke produserer feil.[71] I undersøkelsen fant vi at slike spørsmål kompliseres av at det ikke er alle beslutninger i offentlig sektor hvor det er ønskelig å gi fullt innsyn, særlig i forhold til det offentliges kontrollfunksjoner. Dette skaper et rom for fortolkning og diskusjon, og illustrerer at regulatoriske retningslinjer og regelverk ikke gir en garanti for at alle deler samme forståelse.

4.2 Diskriminering er ikke på agendaen

I kapittel 3 så vi at spørsmål knyttet til diskrimineringsrisiko ikke var spesielt synlig i utfordringsbildet. Dels ble spørsmål om diskriminering byttet ut med andre begreper, noe som reflekterer det nasjonale og internasjonale bildet av KI-utvikling, der begreper som etikk, rettferdighet (fairness), åpenhet og transparens dominerer bildet. Diskriminerings-begrepet fungerer som det teoretikerne Laclau og Mouffe kaller en "myte",[72] i betydningen et begrep vi kan snakke om sammen, men uten at vi har blitt enige om en konkret definisjon av begrepet. "*Bias*" er ofte diskutert, som regel som *ubevisst bias*, mens "diskriminering" som begrep i mindre grad er brukt i dette feltet. Bruk av "bias" som begrep har en faglig forankring i KI-feltets utforsking av hvordan algoritmer kan gi opphav til systematisk forskjellsbehandling av ulike sosiale grupper.[73, 74] Både fokus på "bias", som *noe som bare eksisterer i samfunnet,* og manglende fokus på diskriminering, kan ha negative konsekvenser for diskrimineringsrisikoen innen KI-feltet. Hvis *diskriminering* ikke oppfattes som et problem, vil det heller ikke bli håndtert som en utfordring.

4.3 Kunnskap om diskriminering

Vi fant et stort sprik mellom *noen få virksomheter med god kunnskap og klare strategier* for å motvirke diskriminering i KI-systemer og andre virksomheter ikke hadde vurdert temaet ennå, eller hadde satt det på vent mens KI-prosjektet

[71] KMD (2020).
[72] Laclau, E., & Mouffe, C. (1985), *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics*, London: Verso.
[73] Srinivasan, R. og Chander, A. (2021)
[74] Belenguer, L. (2022)

ble utviklet. Utfordringene er altså mangfoldige, og mens risikoene er omtrent de samme for ulike KI-prosjekt, øker risikoen omvendt proporsjonalt med hvor stort fokus det er på temaet. Manglende kunnskap om diskriminering gjør det mindre sannsynlig at temaet blir integrert i KI-arbeidet.

4.4 Likestillings- og diskrimineringsloven

Risiko for diskriminering øker også når kunnskap om diskriminering i likestillings- og diskrimineringslovens betydning ikke inngår i prosjektet fra starten. Undersøkelsen og intervjuene viser at det er lite oppmerksomhet og mangel på kunnskap om diskriminering i lovens forstand. I spørreundersøkelsen svarte 50 % av respondentene som bekreftet at de behandler personopplysninger med KI, at ingen av diskrimineringsgrunnlagene fra loven var relevant for dem å vurdere. Som nevnt over opplevde vi en rekke ganger at "diskriminering" ble byttet ut med andre tilstøtende begreper. I tillegg erfarte vi noen ganger at "diskriminering" ble forstått som en ren "differensiering" som et naturlig skille mellom personer. Dette bidrar til å ta fokus vekk fra likestillings- og diskrimineringslovens forståelse av *uønsket* diskriminering.

4.5 Tverrfaglig og mangfoldig kunnskap

Tverrfaglig og mangfoldig kunnskap må inkluderes helt fra starten av KI-prosjekter. Studien viser at ofte er *teknologisk nysgjerrighet* en drivkraft, og det kan føre til at kun teknisk kunnskap om KI tas i betraktning når et prosjekt skal startes. Utviklingsteamet må ha tverrfaglig kompetanse, men også fagfolk bør ha en *tverr- og flerfaglig* innsikt, særlig *domenekunnskap* og kunnskap om virksomhetens hjemmel for bruk av personopplysninger samt om personvern og diskriminering. Fordi risiko for diskriminering kan "designes inn" i alle faser av et KI-prosjekt, må kunnskap om diskriminering være til stede i teamet fra starten av prosjektet. Manglende fokus på mangfold i utviklingsteam kan potensielt gi diskriminerende resultat i teknologiutvikling.

4.6 Kompetansefellesskap

Det er en mulig økt risiko i små team og dermed også i små virksomheter på grunn av mangel på kunnskap om diskriminering. I tillegg ser vi også en tendens til at små enheter, både kommuner og andre, har mindre tilgang på gode kompetansenettverk for utvikling av KI.

4.7 Kjønnet teknologi

Kjønn kan også bygges inn i teknologi, for eksempel i chatbots i framtreden og i responser. Kommune-Kari er ett eksempel, mens andre kjente eksempler er smart-telefoner og hverdags-apper som er designet med en bestemt forestilling om en bruker som enten overdriver, eller overser, forskjellige behov hos kvinner og menn. Slik teknologi kan være diskriminerende når input eller responser er tett knyttet til eller basert på særtrekk eller stereotypier om visse grupper. Har det noe å si at det er "Kommune-Kari" som svarer på spørsmål fra innbyggere? Og kan en chatbot svare på spørsmål som angår både kvinner og menn?

4.8 Data – til trening og i bruk

Når persondata håndteres, må det forstås i forhold til likestillings- og diskrimineringslovens diskrimineringsgrunnlag samt i forhold til personvernforordningens definisjon av personopplysninger. Fordi persondata kan reflektere institusjonalisert diskriminering, vil KI basert på slike data alltid kunne innebære risiko for diskriminering (se innledning). Ulike former for skjevhet i data kan gi opphav til diskriminering. Det gjelder for eksempel skjevhet i historiske data (f.eks. mannsdominans blant teknologer som førte til at Amazon sin rekrutteringsalgoritme valgte bort kvinner[75]), og representasjons- eller utvalgsskjevhet i forhold til KI-systemets tiltenkte populasjon. Risiko i forhold til data som reflekterer tidligere (institusjonalisert)

[75] Koshiyama A., Kazim E. et al. (2021). Towards algorithm auditing: a survey on managing legal, ethical and technological risks of AI, ML and associated algorithms. *Soc Sci Res Netw*. https://doi.org/10.2139/ssrn.3778998 (SSRN Scholarly Paper ID 3778998).

diskriminering, øker for eksempel i lange dataserier basert på en mer homogen befolkning enn dagens situasjon. En KI-algoritme kan dermed identifisere en *objektivt* riktig prediksjon om personer eller grupper, som likevel vil være *diskriminerende* i dagens samfunn. Et enkelt eksempel er algoritmer som identifiserer den "beste" leder som hvit mann, fordi statistisk sett er det flere hvite menn som har hatt en slik posisjon. Slik kan tidligere og eksisterende diskriminering forsterkes ved bruk av KI. En særlig utfordring i Norge er at vi har samlet persondata digitalt i lang tid, og eksisterende data kan dermed ikke bare reprodusere eksisterende eller tidligere diskriminering, men kan også innføre diskriminering mot sosiale grupper eller kategorier som ikke er tilstrekkelig representert i dataserien.[76]

Skjevheter i data kan også handle om utvalg av variabler som innebærer skjevheter, ettersom både *for mye* og *for lite* informasjon kan bidra til diskriminering. Et kjent eksempel på dette er KI-systemet COMPAS, brukt til å estimere risiko for gjentatt lovbrudd, og som hadde diskriminerende effekter for afroamerikanere.[77]

Språk og begrepsbruk i registrering av data kan inneholde diskriminerende begrepsbruk, og upresist språk kan gjøre data vanskelig å sammenligne.

Dersom treningsdata ikke representerer den endelige populasjonen, har mangler i forhold til relevans eller partiskhet, eller ikke er oppdatert, kan KI-systemet *innlære feil koblinger*, noe som kan bidra til diskriminering når systemet tas i bruk. "Bias in, bias out", advarer Xenidis og Senden,[78] men påpeker samtidig at selv når KI-algoritmer etter beste evne er antatt å *ikke* diskriminere, så har det likevel blitt identifisert diskriminerende resultat. Det er derfor verdt å merke seg at forestillinger om at relativt enkle grep, som "vasking av data" eller bruk av proxy-variabler, kan fjerne

[76] Broomfield, H., & Reutter, L. M. (2021)
[77] Belenguer, L. (2022)
[78] Xenidis, R., & Senden, L. (2020), EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination. In U. Bernitz, X. Groussot, J. Paju, & S. A. de Vries (Eds.), *General Principles of EU law and the EU Digital Order* (151-182): Kluwer Law International.

diskrimineringsrisiko, også i seg selv kan være en risiko ved at det skaper en *falsk trygghet* for at diskrimineringsrisiko (allerede) er eliminert.

Korrekte data er et åpenbart mål, men vi ønsker å peke på at det også kan være en kilde til falsk trygghet i forhold til risiko for diskriminering ettersom selv helt "korrekte" data kan bidra til diskriminering.

4.9 Valg av algoritme

Prediktive algoritmer som foretar sannsynlighets-beregning av personers framtidige handlinger inkluderer en risiko for diskriminering fordi de ofte tar utgangspunkt i særtrekk ved en sosial gruppe, og dermed bidrar til å generalisere eller naturalisere negative trekk ved gruppen. Maskinlæringsalgoritmer (alle typer) der algoritmen lærer koblinger mellom opplysninger, øker risiko for diskriminering. "Black box"-algoritmer gjør det vanskelig å forklare resultatet og å teste om resultatet innebærer diskriminering. Dette kan bidra til å forsterke gamle eller lage nye koblinger til diskrimineringsgrunnlag. Valg av algoritmer som innebærer utfordring i forhold til forklarbarhet gir også en ekstra utfordring for krav til redegjørelse for vedtak i offentlig sektor.

4.10 Brukere og publikums opplevelse av KI i offentlig sektor

Risiko for diskriminering kan oppstå når personer som søker om offentlige ytelser eller tjenester, tvinges til å utlevere, eller får behandlet private og sensitive opplysninger i større grad enn befolkningen generelt. KI kan oppleves krenkende og invaderende og bidra til generelle negative holdninger til KI og mistillit til teknologi, offentlige tjenester og demokratiske prinsipper.

4.11 Generell digital kompetanse

Digital kompetanse kan påvirke tilgang til tjenester i offentlig sektor. Personer med lav digital kompetanse kan ha utfordringer med å motta tjenester de har

krav på fra det offentlige hvis de ikke mestrer de digitale løsningene som tilbys, og som i økende grad overtar for skranketjenester.

Problemstillingen gjelder digitale tjenester generelt, men også for KI spesielt, fordi digital kompetanse også følger flere av de tradisjonelle diskrimineringskategoriene. Det har blitt pekt på at personer kan miste rettighetsbaserte tjenester når disse digitaliseres, og at digitale tjenester kan medføre brudd på pålegg om å gi individuelt tilpassede tjenester.[79]

4.12 Rigging av et KI-prosjekt: team, kunnskap og forståelse av prosjektet

Risiko for diskriminering kan "designes inn" i alle faser av et KI-prosjekt; i forståelse av utfordringsbildet som KI skal håndtere, og i utforming av prosjektets mål og risikobilde. Risikoen øker når kunnskap om diskriminering ikke inngår i prosjektet fra starten. KI-forskere har advart mot at KI-utviklere alene blir gjort ansvarlige for alle aspekter av et KI-system.[80] For eksempel, i test og evaluering kan utvikler eller evaluator, ved å stole mer på egne antakelser eller stereotypier enn på modellen som er utviklet, bidra til å innføre diskriminerende effekter.

KI-utvikling krever tverrfaglig kompetanse og innsikt i virksomhetens tilstand og behov samt relevante regelverk, og det er derfor en rekke ulike kompetanser som må på plass for at et KI-prosjekt skal lykkes i offentlige sektor: teknologisk kompetanse, juridisk kompetanse, virksomhetsforståelse og personvern er noen av dem som oftest nevnes. Undersøkelsen vår har vist at kunnskap om diskriminering i liten grad blir nevnt som en naturlig del av KI-prosjekt. Utfordringen økes ved at det finnes ulike forståelser (eller misforståelser) av hva diskriminering kan bety i et KI-prosjekt, og av at små team, ofte i små virksomheter, mangler den flerfaglige kompetansen som bør inngå. Mangfold i en prosjektgruppe kan også i seg selv fungere som en motvekt mot risiko for

[79] Korsvik, T. R., Hulthin, M., & Sæbø, A. (2020). *Kunstig intelligens og likestilling: En kartlegging av norsk forskning*. Kilden: Kjønnsforskning.no.
[80] Srinivasan, R., & Chander, A. (2021)

diskriminering, ved å øke oppmerksomhet for variasjon og forskjeller blant målgruppen for teknologien som skal utvikles.[81]

4.13 Ansvar

Det er til en viss grad et uavklart forhold mellom KI-resultater og saksbehandler i forhold til hvordan den endelige beslutningen skal foretas. Siden mange av prosjektene vi undersøkte fortsatt er i utvikling, er dette for mange en teoretisk utfordring, men ikke mindre viktig å håndtere. Er det slik at saksbehandler alltid eller aldri stoler på KI-resultatet? Risiko for diskriminering kan skjule seg i begge alternativene. I dag er normen at KI ikke skal "fatte vedtak" i offentlig sektor, men vil det være slik i framtiden? Og hvor går grensen for hva som er KI og hva som bare er digitalisering?

4.14 Det politiske nivået

Fremvoksende teknologier som kunstig intelligens er rammet inn av nasjonale og internasjonale politiske signaler, holdninger og vedtak, og dette utgjør en viktig ramme når offentlig sektor skal ta i bruk KI. I lys av dagens politiske bilde er det særlig to forhold som kan bidra til risiko for at diskriminering oppstår.

Det første er den **politiske forventningen** om å ta i bruk KI i offentlig sektor for å effektivisere, øke kvalitet eller muliggjøre nye oppgaver, formidlet gjennom nasjonale strategier for digitalisering[82] og kunstig intelligens.[83] Mens både KI-strategien og EU sitt forslag til KI-forordning vektlegger at KI må unngå å diskriminere, viser undersøkelsen at disse i liten grad gir detaljer som kan styres etter på virksomhetsnivå. Videre viser spørreundersøkelsen tendenser til at den politiske drivkraften og forventningen til å bruke nye teknologier, kan bidra til å sette i gang KI-prosjekter i organisasjoner som fortsatt er *umodne* i

[81] Schiebinger, L. & Klinge, I. (2020). *Gendered Innovations 2: How Inclusive Analysis Contributes to Research and Innovation*. Luxembourg: Publications Office of the European Union.
[82] KMD (2019), Digitaliseringsstrategi for offentlig sektor 2019–2025
[83] KMD (2020), Nasjonal strategi for kunstig intelligens,

forhold til oppgaven.[84] I undersøkelsen så vi eksempler på dette, der nødvendig kompetanse og ressurser ikke var på plass ved prosjektstart. Dette bekrefter bildet fra NOKIOS sin *Teknologiradar,* der 90 % av respondentene fra offentlig sektor mente at KI kan ha verdi i egen virksomhet, tett på 100 % mente at teknologien var moden, og 60 % mente at virksomheten var moden. Derimot hadde bare litt over 10 % prøvd ut KI. Blant utfordringene som stopper satsing på feltet, er mangel på kompetanse. Vår undersøkelse har vist at virksomheters *umodenhet* eller *manglende ressurser* ved oppstart av KI-prosjekt ofte innebærer at kunnskap om og hensyn til diskrimineringsrisiko ikke inngår i prosjektet. KI-utviklere "skulle hatt et bindende profesjonelt etisk ansvar", sier Inga Strümke fra Norsk råd for digital etikk (NORDE) i et intervju med *Kode24*,[85] og peker dermed på noe av utfordringen med KI, nemlig behovet for tverrfaglig kompetanse som også inkluderer kunnskap om diskriminering.

Det andre forholdet som kan bidra til å øke risiko for diskriminering, er en **retorikk** rundt KI som et "objektivt" redskap og om datagrunnlag som kan korrigeres til å bli helt "korrekt og fullstendig", eller som i KI-forordningen, å finne en balanse (eller kompromiss) mellom *best mulig* og *mulig å gjennomføre*. Ettersom persondata alltid innebærer en viss risiko for diskriminering, kan denne retorikken skape falsk trygghet. Blant annet feministiske perspektiver på KI har utfordret tanken om at KI kan være et objektivt redskap fritt for diskriminerende strukturer i samfunnet.[86, 87]

4.15 Verdighet, tillit, demokrati

Ny teknologi, som KI, er vanskelig å forstå for folk flest. Det kan være en ytterligere utfordring å forstå hvordan *informasjon om oss* blir til *data* som brukes i KI. Slik usikkerhet kan skape negative holdninger til KI, og særlig

[84] Jf. Alston 2019 sin bekymring.
[85] Kode24 (2022), Ber norske utviklere få på plass etisk regelverk for kunstig intelligens, 21. november 2022, https://www.kode24.no/artikkel/ber-norske-utviklere-fa-pa-plass-etisk-regelverk-for-kunstig-intelligens/77760836.
[86] Amrute, S. 2019. Of Techno-Ethics and Techno-Affects. *Feminist Review*, Vol. 123, No. 1, pp. 56–73. DOI: 10.1177/0141778919879744
[87] UNESCO. (2020), Artificial intelligence and gender equality: key findings of UNESCO's Global Dialogue, Division for Gender Equality, UNESCO

dersom det er knyttet usikkerhet til hva som er grunnlag for et vedtak, hvilke data om oss som blir brukt, hvordan data blir delt, osv. Noen av de formelle og juridiske sidene av slike utfordringer håndteres gjennom regelverk som personvernlovgivning (GDPR). Andre sider av dette utfordringskomplekset angår offentlig sektor sitt omdømme når det oppfattes som uavklart, misforstått eller til og med feilaktig bruk av KI og persondata i offentlig sektors regi. Vi har for eksempel flere steder, i enkle rapporter og nyhetsartikler, møtt påstanden om at offentlig sektor *foretar beslutninger basert på KI*. Regelverket sier derimot at digitalt baserte beslutninger *utover rene regelstyrte systemer*, ikke skal forekomme i offentlig sektor. Beslutninger som gjelder enkeltpersoner skal altså ikke foretas av KI i offentlig sektor, og i komplekse vurderinger skal KI kun være støttende. KI-forskere advarer mot at denne problematikken i sin ytterste konsekvens kan skape mistillit ikke bare til offentlig sektor, men til selve det demokratiske byggverket.[88] Her vil vi peke på at også manglende forståelse av hvordan KI brukes samt uklare skiller mellom KI og andre former for (regelstyrt) digitalisering (som også i noen tilfeller defineres som KI, jf. KI-strategien s. 10), kan utfordre tilliten til digitale offentlige tjenester.

4.16 Et rammeverk for å identifisere risiko for diskriminering ved bruk av KI i offentlig sektor

Tabell 4.1 nedenfor oppsummerer de ulike risikoene vi har omtalt her og presenterer et rammeverk for å identifisere risiko for diskriminering når KI tas i bruk i offentlig sektor. I utforming av rammeverket har vi lagt vekt på at *den vanligste situasjonen* i offentlig sektor da undersøkelsen ble gjennomført fram mot juni 2022, var *å ikke ha et KI-system som inkluderer persondata i bruk*. Målet er derfor at rammeverket skal kunne veilede et KI-utviklingsprosjekt fra idé til bruk. Rammeverket henvender seg bredt til ulike kompetansebehov og ulike hensyn, fra tekniske til kulturelle, politiske og juridiske vurderinger, som må håndteres fra begynnelsen av et KI-prosjekt. Det viser også til utfordringer i

[88] OECD (2022)

ulike *faser*, fra før prosjektet, til planlegging, utvikling, test og evaluering, til KI-systemet er i bruk, eller produksjon. Tabellen nedenfor oppsummerer hvor og når risiko kan oppstå og hva risikoen består i, og viser til at noen typer risiko for diskriminering kan forekomme i flere faser av et KI-prosjekt.

Tabell 4.1. Risiko for diskriminering ved bruk av kunstig intelligens i offentlig sektor

Hvor og når oppstår risiko?	Hva består risikoen i?
1. *Pre*-prosjektfasen: forståelse av KI formes politisk og i samfunnet	
Politisk press for å bruke KI	Premature prosjekter kan skape risiko i KI-prosjekt
KI som "objektivt" redskap	Skaper falsk trygghet for at diskriminering kan unngås i KI
Data som "faktiske/rene"	Skaper falsk trygghet for at data kan "vaskes" for diskrimineringsrisiko
2. Planleggingsfasen: når prosjektet rigges	
Utviklingsteam	Mangel på tverrfaglig kunnskap, særlig om personvern og diskriminering i likestillings- og diskrimineringslovens forstand
Mål- og problemdefinisjon	Manglende fokus på risiko for diskriminering
Definisjon av diskriminering	Når diskriminering i likestillings- og diskrimineringslovens forstand byttes ut med andre begreper gir det risiko for å overse og dekke over diskrimineringsrisiko
Persondata	Bruk av persondata (sensitive personopplysninger og enhver informasjon om en person) har en iboende risiko for diskriminering
Algoritme	Valg av algoritme kan påvirke risiko samt muligheten for å identifisere og forklare diskriminerende resultater
3. Utviklingsfasen: når teori, data og praksis settes sammen	
Treningsdata	Skjevhet i treningsdata øker risiko for diskriminering
Språk og merking av data	Språk og begrepsbruk kan gi diskriminerende resultater
"Vasking" av data	Retorikken om persondata som feilfri og objektiv kan gi falsk trygghet
Bruk av alternative verdier ("Proxy")	Å bytte beskyttede personopplysninger med åpne opplysninger er ingen garanti for fravær av diskriminering
Algoritme og metode for maskinlæring	Risiko øker ved bruk av maskinlæringsalgoritmer, og ytterligere ved grad av kompleksitet i valg av metode, særlig når det ikke inngår en "human-in-the-loop" - en person som kan kontrollere KI-resultat underveis
Feil i algoritmen	Kan før til feilkoblinger med risiko for diskriminering
"Kjønnet" teknologi	Kjønn og andre sosiale kategorier kan bygges inn i teknologiens framtreden
4. Test- og evalueringsfasen: når KI-systemet kontrolleres	
Data	Skjevheter i datasett kan gi feil og risiko for diskriminering
Menneskelig feil	Evaluator kan innføre feil pga. manglende oversikt, egne "bias" eller manglende fokus på diskriminering
5. I produksjon: når KI-systemet går inn i daglig bruk	
Vedtak, beslutningsmyndighet og åpenhet	Uklar forståelse av regler og retningslinjer for hvordan vedtak gjøres av saksbehandler med støtte i KI-system
Prediktive (ML) algoritmer	Innebærer risiko for å diskriminer og å produsere nye diskrimineringsgrunnlag
Innhenting av personopplysninger	Økt innhenting av private og sensitive opplysninger for å oppnå visse offentlige ytelser eller tjenester kan oppleves krenkende og diskriminerende
Digital kompetanse hos brukere	Mangel på digital kompetanse eller skepsis, tvil eller frykt for hva KI bidrar med, kan reflektere enkelte diskrimineringsgrunnlag

5. Anbefalinger for å forebygge diskriminerende effekter av kunstig intelligens i offentlig sektor

Det finnes et økende antall veiledninger, rammeverk og modeller for å motvirke at kunstig intelligens gir skjeve, urettferdige eller skadelige resultater,[89] og flere er på vei mens vi skriver. Mange av disse henvender seg til teknologer som har ansvar for den teknologiske utviklingen av et KI-system[90]. En av "advarslene" fra KI-forskere er imidlertid at KI-utviklere alene ikke må bli gjort ansvarlige for det brede settet av vurderinger som må gjøres i forhold til risiko for diskriminering.[91] Videre antyder undersøkelsen vår at utfordringer knyttet til risiko for diskriminering når et bredt og variert sett av enheter innen offentlig sektor skal ta i bruk KI, ikke kan, og trolig heller ikke bør, løses innen den enkelte virksomhet, men i stedet trenger en god felles styring.[92] Våre anbefalinger for å forebygge diskriminerende effekter av kunstig intelligens i offentlig sektor henvender seg derfor til *flere nivåer* og ulike aktører:

- Det nasjonale og politiske nivået
- Offentlig sektor overordnet
- Den enkelte virksomhet i sektoren
- Utdanningssektoren

Anbefalingene er formulert med utgangspunkt i at KI i offentlig sektor må håndteres holistisk og som en utfordring der teknologi, ledelse, politikk og en tverrfaglig kompetanseprofil må gå hånd i hånd.

[89] Di Noia, T., Tintarev, N., Fatourou, P., & Schedl, M. (2022), Recommender systems under European AI regulations. *Communications of the ACM*, 65(4), 69-73.
[90] Belenguer, L. (2022)
[91] Srinivasan, R., & Chander, A. (2021)
[92] Broomfield, H., & Lintvedt, M. N. (2022)

5.1 KI-isfjellet

Gjennom undersøkelsen fikk vi bekreftet en forståelse av KI som et isfjell:[93] det er bare den øverste toppen som er synlig i diskusjoner, nemlig det punktet der KI som et teknologisk produkt møter samfunn, normer, kultur og politiske hensyn. Den nederste delen av isfjellet – der selve programmeringen av algoritmer pågår, ble ikke spesielt synlig i undersøkelsen, selv om vi både inviterte og diskuterte med denne kompetansegruppen. Undersøkelsen understreker imidlertid at utfordringene knyttet til risiko for diskriminering ved bruk av KI i offentlig sektor ikke må håndteres alene som et teknisk problem, og at mange av de tidligste vurderingene og utfordringene handler om det øverste nivået av isfjellet, der en rekke ulike kompetanser må involveres for å forstå, planlegge, tilrettelegge og gjennomføre et KI-prosjekt i offentlig sektor. Vi ønsker ikke å nedvurdere betydningen av god teknologisk kompetanse, men vil også understreke at dette alene ikke er tilstrekkelig når risiko for diskriminering skal håndteres. Undersøkelsen viser imidlertid også at det å skaffe tilstrekkelig teknologisk kompetanse er utfordrende i små enheter. Det antyder at en god strategi for offentlig sektor bør ta hensyn til dette, og at teknologisk kompetanse ofte blir tilført utenfra, med de utfordringer som dette bringer i forhold til virksomhets- og domenekompetanse.

Dernest ser vi at noen utfordringer kan møtes med tiltak allerede i dag, mens andre utfordringer krever større og gjerne aktive inngrep, og disse må håndteres over tid. Vi deler derfor anbefalingene våre i *langsiktige tiltak* og en *sjekkliste med tiltak som kan iverksettes straks* og direkte i den enkelte virksomhet.

5.2 Anbefalinger for langsiktige tiltak

Flere ganger i løpet av undersøkelsen hørte vi at lover og regler ikke er oppdatert i forhold til utfordringer og problemstillinger som oppstår når offentlig sektors oppgaver og tjenester skal digitaliseres. Dette gjør også at noen

[93] Inspirert av Norwegian Cognitive Center/Digital Norway sitt kurs i kunstig intelligens, rettet mot næringsliv (Sogndal, september 2022).

blir sittende "på gjerdet" og vente på at lovverket skal bli oppdatert. Men vi møtte også det motsatte synspunkt i forhold til risiko for diskriminering i KI: forbudet mot diskriminering som nedlegges i likestillings- og diskrimineringsloven gjelder også for kunstig intelligens.

Trolig har begge synspunktene rett.

En av utfordringene er at KI kan gi ikke-intenderte diskriminerende resultater, og det kan fortsatt være en risiko for diskriminering selv når "data er korrekt og fullstendig", fordi diskriminering potensielt finnes i datagrunnlag som inkluderer persondata.

En annen utfordring som ble synlig i undersøkelsen, er at mens det er stor oppmerksomhet rundt personvernforordningen, GDPR, er det langt mindre oppmerksomhet rundt likestillings- og diskrimineringsloven. Loven gir kanskje like godt vern innen KI som i andre kontekster, men det kan oppstå et gap mellom lovens intensjon og praksis når ikke loven blir integrert i KI-arbeid fra begynnelse til slutt. Uavhengig av lovens egnethet for KI som kontekst, viser dermed undersøkelsen at en stor utfordring handler om samhandling med lovverket, snarere enn mangler i lovverket.

5.2.1 Diskurser om KI må inkludere en forståelse av diskriminering

Undersøkelsen viste at diskriminering ofte blir byttet ut med andre begreper. Det er imidlertid viktig at diskriminering basert på den forståelsen som legges til grunn i likestillings- og diskrimineringsloven, blir inkorporert i lover, regler, reguleringer for å understreke at også dette perspektivet er viktig i KI.

Politiske retningslinjer som beskriver KI som "objektiv" og data som mulig å gjøre "korrekt" og "fullstendig", kan bidra til å skape falsk trygghet ved å signalisere at risiko for diskriminering allerede er håndtert.

5.2.2 Én standard for å håndtere risiko for diskriminering ved bruk av KI

Undersøkelsen dokumenterte en rekke ulike tolkninger av diskriminering, manglende kunnskap om diskriminering, og ofte var temaet ikke på agendaen i KI-prosjekt. En tydeligere standard for hvordan diskriminering må integreres i KI-prosjekt, kan bidra til å sette temaet på agendaen og til langt mindre usikkerhet om hvordan dette skal håndteres. Mens det å sette diskriminering på agendaen også er tatt med i tiltak som kan settes i gang straks, ser vi her til hvordan, for eksempel, GDPR har bidratt til en *holdningsendring* og gjort det til en selvfølge at personvern er et tema som skal inn i KI-prosjekt. Vi etterspør en tilsvarende holdningsendring for diskriminering.

Likestillings- og diskrimineringslovens §24, *Offentlige myndigheters aktivitets- og redegjørelsesplikt*, sammen med arkivlovens §6, setter klare krav til at offentlige virksomheter må *iverksette nødvendige tiltak* for å unngå diskriminering, kunne *redegjøre for hvilke tiltak* som er iverksatt i så henseende, samt *arkivere* og *dokumentere* hvordan man har jobbet med dette.

5.2.3 Revisjon av KI-prosjekt i offentlig sektor

Mens EU sitt forslag til en felles europeisk KI-forordning foreslår et system for å gjennomføre *revisjon* ("audit") for KI-system, peker de også på at denne kompetansen er helt i startfasen.[94] Mens sporbarhet og åpenhet nevnes i EU sitt forslag, savner vi et bredere fokus på revisjon hvor også risiko for diskriminering når persondata er involvert, inngår i revisjonsgrunnlaget. I samsvar med foregående anbefalinger er det også her viktig at kompetanse om diskriminering i tråd med likestillings- og diskrimineringsloven er involvert, og en organisasjon som Likestillings- og diskrimineringsombudet (LDO), som allerede har en god innsats på feltet, bør involveres.

[94] European Commission, Artificial Intelligence Act, april 2021, s. 14.

5.2.4 Tilgang på rådgivningstjeneste og kompetansenettverk

Vårt prosjekt har tydelig vist behovet for å involvere tverrfaglig kompetanse ved utvikling av KI. Det er også slik at ulike virksomheter i offentlig sektor har svært ulik tilgang på kompetanse og ulik tilgang på kompetansenettverk. Tilgang på en rådgivningstjeneste og kompetansenettverk vil være viktig for at virksomheter i offentlig sektor skal kunne etablere og ta i bruk KI med tilfredsstillende hensyn til risiko for diskriminering. Dette vil være viktig for høy-risiko prosjekter der virksomheten selv ikke har tilstrekkelig spisskompetanse og tverrfaglig kompetanse.

Begge deler må håndteres på et overordnet nivå i offentlig sektor, særlig for å hjelpe og rigge KI-prosjekter riktig hos den store andelen av offentlig sektor virksomheter som ikke har tilstrekkelig bred kompetanse i egen organisasjon.

Det bør også sendes et signal til utdanningsinstitusjoner i forhold til behov for at utdanninger som leder til KI-arbeid også tar opp i seg behovet for kompetanse om risikoer involvert i KI.

5.2.5 Nasjonal sertifisering av KI til offentlig sektor

Kompetanseutfordringer gjør mange offentlig sektor virksomheter avhengig av ekstern kompetanse og "hyllevare". Nasjonal strategi for KI peker på at kommunal sektor har like oppgaver og kan samarbeide, men det er også slik at ulik størrelse gir store ulikheter i tilgjengelige ressurser og kompetanse. Nasjonal sertifisering av KI, vil kunne bidra til å redusere ulempen med ulikheter og begrensninger i kompetanse. Dette kan særlig være aktuelt for kommunal sektor, der mange lignende enheter har tilsvarende oppgaver.

Sertifisering vil være et virkemiddel for "hyllevare"-KI som øker trygghet i å ta i bruk KI, særlig for de mindre virksomhetene i offentlig sektor.

5.2.6 KI-systemet bør sikte mot høyest mulig grad av transparens og forklarbarhet

På grunn av krav om å kunne redegjøre for beslutninger i offentlig sektor bør bruk av "black box"-algoritmer kombinert med persondata gjøres med stor varsomhet. Strategien "Stopp og tenk! " er viktig, og her kan også samarbeid på tvers i offentlig sektor bidra til større klarhet omkring god praksis.

5.3 Start nå! Sjekkliste for diskriminering, med forslag til den enkelte virksomhet

1. Diskriminering på agendaen: basert på likestillingsloven – ikke forveksle diskriminering med personvern.
2. Tenk på diskriminering allerede i planleggingsfasen.
3. Tverrfaglige team: kunnskap om diskriminering må inngå.
4. Vær bevisst på at all bruk av persondata kan føre til diskriminering.
5. Når alt er riktig, er det fortsatt risiko for diskriminering.
6. "Human-in-the-loop": inkluder kompetanse om diskriminering.
7. Knytt kontakter til kompetansemiljøer: for eksempel Fagforum for KI i offentlig sektor, KIN (KI i norsk helsetjeneste), NORA, og forskningsmiljøer.
8. Bruk toglinjen som veileder til hvilke problemstillinger man må tenke på.
9. Forklarbare algoritmer: "black box" algoritmer er mer utfordrende hva gjelder risiko for diskriminering enn andre typer algoritmer.
10. "Hyllevare"-algoritmer: still spørsmål til hvordan den er utviklet, be om dokumentasjon av modellen, hvilke data den er basert på, osv.

6. Avsluttende refleksjoner

Dette prosjektet har vært en spennende oppgave for Vestlandsforsking og Rambøll Management Consulting. Prosjektet har gitt oss anledning til å komme i dialog med miljø i offentlig sektor som jobber med å utrede og ta i bruk KI løsninger samt ulike kompetansemiljø i Norge og internasjonalt som er opptatt av kunstig intelligens og diskriminering.

Vi erfarte at offentlig sektor har et stort fokus på å forbedre kvalitet og effektivitet i sine prosesser og tjenester ved hjelp av digitalisering. Det er særlig store forventninger til at kunstig intelligens skal bidra positivt til dette.

Vårt prosjekt peker på at utvikling av KI-løsninger stiller krav til involvering av mangfoldig og tverrfaglig kompetanse. Vi ser at mange av organisasjonene vi har snakket med, ikke har like høy oppmerksomhet om dette og heller ikke tilgang til all relevant kompetanse. Det er i særlig grad tilfelle med kompetanse om diskriminering.

Som en av våre informanter uttrykte det: "Diskriminering, det har vi ikke tenkt på. Takk for at dere minnet meg på det!"

Som mange peker på er det slik at både mennesker og maskiner gjør feil. Sannsynligvis gjør maskiner mindre feil enn mennesker, men vi aksepterer ikke at maskiner gjør feil, fordi i vår forståelse skal ikke maskiner gjøre feil:

- Maskiner skal være ufeilbarlige
- Hvis de gjør feil er det fordi de har en "bug"
- Det er mer akseptert at mennesker gjør feil
- Mennesker gjør feil usystematisk og deres feil rammer ikke så mange
- Hvis maskiner gjør feil, gjør de systematisk feil
- Kunstig intelligens har potensiale til å feile raskt og i større omfang enn mennesker

I det private næringsliv finnes det en rekke drivere for å sørge for at selskapene forholder seg til etisk og ansvarlig KI, som for eksempel konkurranse, tillit blant

kunder, omdømme osv. Offentlig sektor har ikke helt de samme driverne, og har derfor større behov for lovgivning og regelverk. Som en av våre eksperter påpekte: "Reguleringer har samme oppgave for offentlig sektor som markedskrefter har for privat sektor".

Til sist er det mange spørsmål som skaper utfordringer eller barrierer når offentlig sektor tar i bruk KI, slik som "toglinjen" (figur 3.1) illustrerer. Det er ikke alle utfordringene som er direkte knyttet til diskriminering, selv om noen også kan ha indirekte diskriminerende effekter. Disse utfordringene må også håndteres, og for spørsmål om diskriminering er kanskje den største utfordringen at dette er problemstillinger som kan ta oppmerksomheten vekk fra diskriminering. Her nevner vi eksempler på *erfarte utfordringer* som vi har møtt i undersøkelsen:

- Tilgang til tilstrekkelig og relevant data
- Et lovverk lite tilpasset digitalisering generelt og bruk av KI spesielt
- Usikkerhet knyttet til regelverk for å bruke virksomhetsdata
- Små enheter har begrenset med ressurser til å starte KI-prosjekt og til å etablere flerfaglige team
- Antakelsen i nasjonal strategi for KI, om at kommuner enkelt kan samarbeide fordi de har like oppgaver, stemmer ikke alltid med (kommune-)kartet
- Stor avstand mellom noen få offentlige aktører med høy KI-kompetanse og flertallet, deriblant mange små aktører, som både mangler kompetanse og tilgang til kompetansenettverk
- Forvirring knyttet til KI som oppstår på grunn av et uklart skille mellom KI og enkle former for regelbaserte systemer, og dette truer både forståelse av KI i sektoren og omdømme til KI i offentlig sektors tjeneste, særlig knyttet til hvorvidt KI "tar beslutninger"
- Når KI-tjenester kjøpes eksternt er det fare for at viktig domenekunnskap, som virksomhetens behov, styringsregler osv., ikke er tydelig kjent for KI-produsent
- Usikkerhet oppstår når KI-system med begrenset grad av åpenhet tas i bruk, for eksempel når eksterne utviklere ikke deler tilstrekkelig informasjon om KI-system

6.1 KI setter punktum for rapporten om KI

I denne rapporten har vi satt søkelys på utfordringer ved bruk av KI, særlig risiko for diskriminering når KI blir en del av offentlig sektor sine tjenester. KI er imidlertid i rask utvikling, og nesten hver dag kan vi lese om nye områder der KI gjør store og raske framskritt. De siste ukene har en chatbot – altså en

KI-slektning av chatboten *Kommune-Kari*, fått mye oppmerksomhet med Open AI sin invitasjon til å teste deres chatbot (ChatGPT).[95] Den har sjokkert og gledet i omtrent likt monn, på grunn av evnen til å gi tilsynelatende gode tekstlige svar som kan bestå en eksamen,[96] men som også tildekker "fake news" og presenterer overtro og løgn like overbevisende som korrekte påstander.[97] Vi har utfordret Open AI sin ChatGPT til å avslutte rapporten, med instruksjonen: *Kan du skrive det avsluttende avsnittet i en rapport som allerede har analysert data og gjennomgått forskning om diskriminerende effekter av KI i offentlig sektor?*

Open AI ChatGPT sine sluttord for en rapport om KI i offentlig sektor:

> ***Til slutt kan det konkluderes med at bruk av KI har muligheten til å forsterke og forlenge eksisterende fordommer og diskriminering i vårt samfunn. Det er viktig at vi anerkjenner og tar tak i disse problemene mens vi fortsetter å utvikle og integrere KI i våre liv. Vi må prioritere mangfold og inkludering i design og implementering av KI-systemer, og jobbe for at de er rettferdige og likeverdige for alle medlemmene av vårt samfunn. Kun ved å gjøre dette kan vi virkelig utnytte KI's fulle potensial for å forbedre våre liv og skape en bedre fremtid for alle.[98]***

[95] Open AI ChatGPT https://chat.openai.com/chat
[96] Arnesen, M. (2022), Chatbot løste eksamens-oppgave på få sekunder. Fikk karakter B, *Khrono* (8. desember 2022). https://khrono.no/chatbot-loste-eksamensoppgave-pa-fa-sekunder-fikk-karakter-b/740911
[97] Medieprofessor: Kunstig intelligens vil ta jobben frå mange forskarar, *Khrono* (9. desember 2022), https://khrono.no/medieprofessor-kunstig-intelligens-vil-ta-jobben-fra-mange-forskarar/741517
[98] Open AI ChatGPT (https://chat.openai.com/chat), tekst generert på norsk 15. desember 2022.

Referanser

Alston, P. (2019). Report of the Special Rapporteur on extreme poverty and human rights. UN General Assembly A/74/493. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N19/312/13/PDF/N1931213.pdf?OpenElement

Amrute, S. 2019. Of Techno-Ethics and Techno-Affects. *Feminist Review*, Vol. 123, No. 1, pp. 56–73. DOI: 10.1177/0141778919879744

Andreasson, U., & Stende, T. (2019), Nordiske kommuners arbeid med kunstig intelligens: Nordic Council of Ministers.

Arnesen, M. (2022) Chatbot løste eksamens-oppgave på få sekunder. Fikk karakter B, *Khrono* (8. desember 2022). https://khrono.no/chatbot-loste-eksamensoppgave-pa-fa-sekunder-fikk-karakter-b/740911

Bartoletti, I. & Xenidis, R. (October 2022). Preliminary draft Council of Europe study on the impact of artificial intelligence, its potential for promoting equality, including gender equality, and the risks to non-discrimination, The Gender Equality Commission (GEC) and the Steering Committee on Anti-Discrimination, Diversity and Inclusion (CDADI), The Council of Europe, https://rm.coe.int/gec-2022-9-study-on-ai-211022/1680a8ad89

Belenguer, L. (2022), AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. *AI and Ethics*. doi:10.1007/s43681-022-00138-8

Boden, M. A. (2018), Artificial Intelligence: A Very Short Introduction. Oxford: Oxford University Press.

Broomfield, H., & Lintvedt, M. N. (2022), Is Norway Stumbling into an Algorithmic Welfare Dystopia? *Tidsskrift for velferdsforskning*, 25(3), 1-15. Doi: 10.18261/tfv.25.3.2.

Broomfield, H., & Reutter, L. M. (2021), Towards a Data-Driven Public Administration: An Empirical Analysis of Nascent Phase Implementation. *Scandinavian Journal of Public Administration*, 25(2), 73-97.

Barne-, ungdoms- og familiedirektoratet (Bufdir), Begreper: https://bufdir.no/Statistikk_og_analyse/Etnisitet/begreper_og_kunnskapsgrunnlag/begreper/

Connell, R. W. (2002), *Gender*, Cambridge: Polity.

Council of Europe, *Council of Europe and Artificial Intelligence*, https://www.coe.int/en/web/artificial-intelligence

Datatilsynet (2022). *Sluttrapport fra sandkasseprosjektet med NAV*: Temaer: rettslig grunnlag, rettferdighet og forklarbarhet, https://www.datatilsynet.no/contentassets/ebd705b85bbc4bfc8b13638a28863e10/nav_sluttrapport.pdf

Datatilsynet, *Personvernforordningen*, https://www.datatilsynet.no/rettigheter-og-plikter/personopplysninger/

Datatilsynet, *Sandkassesiden*, https://www.datatilsynet.no/regelverk-og-verktoy/sandkasse-for-kunstig-intelligens/

Di Noia, T., Tintarev, N., Fatourou, P., & Schedl, M. (2022), Recommender systems under European AI regulations. *Communications of the ACM*, 65(4), 69-73.

European Commission (2021), Artificial Intelligence Act, https://artificialintelligenceact.eu/, https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF

European Commission, High-Level Expert Group on Artificial Intelligence (2019), A Definition of AI: Main Capabilities and Disciplines, https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

Regjeringen (2021), *Forslag til forordning om kunstig intelligens (KI-forordningen)*, https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2021/juni/forslag-til-forordning-om-kunstig-intelligens-ki-forordningen/id2884935/
Friedman, B., & Nissenbaum, H. (1996), Bias in computer systems. *ACM Transactions on information systems (TOIS),* 14(3), 330-347.
Gerards, J. and Xenidis, R. (2021) Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law. European Commision.
Gjerdsbakk, T.C.G., 2022. Åpen og rettferdig kunstig intelligens, i *Lov & Data* nr. 150 – hefte 3/2022, https://lovdata.no/artikkel/apen_og_rettferdig_kunstig_intelligens/4139
Gröger, C. (2021), There Is No AI Without Data. *Communications of the ACM*, 64(11), 98-108. doi: 10.1145/3448247
IBM (2022), AI ethics, https://www.ibm.com/artificial-intelligence/ethics
Jørgensen, R. F. (2021). Data and rights in the digital welfare state: the case of Denmark. *Information, Communication & Society*. https://doi.org/10.1080/1369118X.2021.1934069
Klare, B. F., M. J. Burge, J. C. Klontz, R. W. Vorder Bruegge and A. K. Jain, (2012), Face Recognition Performance: Role of Demographic Information. *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1789-1801, Dec. 2012, doi: 10.1109/TIFS.2012.2214212. https://ieeexplore.ieee.org/document/6327355
Kode24 (2022), Ber norske utviklere få på plass etisk regelverk for kunstig intelligens, 21. november 2022, https://www.kode24.no/artikkel/ber-norske-utviklere-fa-pa-plass-etisk-regelverk-for-kunstig-intelligens/77760836.
Kommunal- og moderniseringsdepartementet (KMD) (2019), *Én digital offentlig sektor: Digitaliseringsstrategi for offentlig sektor 2019–2025*, https://www.regjeringen.no/no/dokumenter/en-digital-offentlig-sektor/id2653874/
Kommunal- og moderniseringsdepartementet (KMD) (2020), *Nasjonal strategi for kunstig intelligens*, https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/
Kommunal- og moderniseringsdepartementet (2021), 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), https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665314_en
Korsvik, T. R., Hulthin, M., & Sæbø, A. (2020). *Kunstig intelligens og likestilling: En kartlegging av norsk forskning*. Kilden: Kjønnsforskning.no.
Koshiyama A., Kazim E. et al. (2021). Towards algorithm auditing: a survey on managing legal, ethical and technological risks of AI, ML and associated algorithms. *Soc Sci Res Netw*. https://doi.org/10.2139/ssrn.3778998 (SSRN Scholarly Paper ID 3778998).
Kunstig intelligens-forordningen på høring i Norge: https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2021/juni/forslag-til-forordning-om-kunstig-intelligens-ki-forordningen/id2884935/
Laclau, E., & Mouffe, C. (1985), *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics*, London: Verso.
Lov om likestilling og forbud mot diskriminering (likestillings- og diskrimineringsloven) – Lovdata, https://lovdata.no/dokument/NL/lov/2017-06-16-51
Microsoft (2022), Responsible AI, https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6
Midtbøen, A. H. (2015). Etnisk diskriminering i arbeidsmarkedet. *Tidsskrift for samfunnsforskning*, *56*(1), 4-30.
Nickelsen, T. (2019) Roboter er på full fart inn i jussen, Forskning.no, https://forskning.no/juridiske-fag-roboter/roboter-er-pa-full-fart-inn-i-jussen/1588380
Nilsen, C. M. (2020) Begynnelsen på slutten for IB?, *Khrono*, https://khrono.no/begynnelsen-pa-slutten-for-ib/504277

Norwegian Cognitive Center og Bergen Næringsråd (2022) Digital Modenhet på Vestlandet. Delrapport 1: Kunstig intelligens, rapport
OECD (2022) Drivers of Trust in Public Institutions in Norway, *Building Trust in Public Institutions*, OECD Publishing, Paris, https://doi.org/10.1787/81b01318-en
Open AI ChatGPT https://chat.openai.com/chat
Personvernforordningen, https://lovdata.no/lov/2018-06-15-38/gdpr
Redden, J. (2018). Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society, 5*(2). https://doi.org/10.1177%2F2053951718809145.
Regjeringen (2021), Forslag til forordning om kunstig intelligens (KI-forordningen), https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2021/juni/forslag-til-forordning-om-kunstig-intelligens-ki-forordningen/id2884935/
Schiebinger, L. & Klinge, I. (2020). *Gendered Innovations 2: How Inclusive Analysis Contributes to Research and Innovation*. Luxembourg: Publications Office of the European Union.
Kunstig intelligens i norsk helsetjeneste (KIN), https://www.helsedirektoratet.no/tema/kunstig-intelligens/kompetanse-og-erfaringsdeling
Srinivasan, R. og Chander, A. (2021) Biases in AI systems. *Communications of the ACM*, 64(8), 44-49.
Statistisk sentralbyrå (2022), Leveranse av bongdata fra dagligvarekjedene Rema 1000, NorgesGruppen, Coop og Bunnpris, (SSB 6. mai 2022), https://www.ssb.no/omssb/ssbs-virksomhet/kost-nyttevurdering/leveranse-av-bongdata-fra-dagligvarekjedene-rema-1000-norgesgruppen-coop-og-bunnpris
Suresh, H., & Guttag, J. (2021), A framework for understanding sources of harm throughout the machine learning life cycle. *EAAMO 2021 – Equity and access in algorithms, mechanisms, and optimization* (1-9).
Svendsen, N.V. (2022), Medieprofessor: Kunstig intelligens vil ta jobben frå mange forskarar, *Khrono* (9. desember 2022), https://khrono.no/medieprofessor-kunstig-intelligens-vil-ta-jobben-fra-mange-forskarar/741517
The Artificial Intelligence Act, EU, https://artificialintelligenceact.eu/
UNESCO. (2020), Artificial intelligence and gender equality: key findings of UNESCO's Global Dialogue, Division for Gender Equality, UNESCO
Xenidis, R., & Senden, L. (2020), EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination. In U. Bernitz, X. Groussot, J. Paju, & S. A. de Vries (Eds.), *General Principles of EU law and the EU Digital Order* (151-182): Kluwer Law International.