Demokratiske algoritmer

Fremtidsutsikter til automatiserte beslutningsprosesser i offentlig forvaltning i befolkningens øyne

Sveinung Arnesen og Mikael Poul Johannesson

Sluttrapport skrevet for NAV FOU.
Denne versjonen: 9 Mars, 2022

Innholdsfortegnelse

Figurliste

Om rapporten

Forfattere

Sveinung Arnesen er Forsker I og faglig leder for Demokrati og innovasjon ved NORCE, og førsteamanuensis II ved Institutt for administrasjons- og organisasjonsvitenskap, UiB. PhD-graden ble avlagt ved Institutt for sammenliknende politikk, UiB. Arnesen er Norges nasjonale koordinator for Den europeiske samfunnsundersøkelsen (ESS). ORCID. Github. Google Scholar.

Mikael P. Johannesson er forsker III ved NORCE og PhD-kandidat ved Institutt for Sammenliknende Politikk, UiB. Han har bred erfaring med eksperimentelle metoder, maskinlæring (inkludert deep learning), og surveyforskning. Johannesson har utviklerkompetanse i statistikkprogrammet R, samt erfaring med Python (inkludert TensorFlow og Keras). Github. Google Scholar.

Referansegruppe

Anne Lise Fimreite er professor ved Institutt for administrasjons- og organisasjonsvitenskap, UiB. Hun har tidligere ledet den forskningsrådsfinansierte evaluering av NAV-reformen og har arbeidet mye med styringsutfordringer i flernivåsystem. Hun har også nylig vært medlem av det offentlige utvalget som i 2019 leverte forslag til ny forvaltningslov (NOU 2019:5) og har egen erfaring fra offentlig forvaltning som prorektor ved UiB i fire år fra 2013 til 2017.

Jacob Aars er professor ved Institutt for administrasjons- og organisasjonsvitenskap, UiB, og har ledet den NFR-finansierte evalueringen av NAV-reformen (tok over da Fimreite gikk inn i rektoratet ved UiB). Han har blant annet forsket på lokaldemokrati og tilfredshet med offentlige tjenester.

Finansiering

Forskningsrapporten er finansiert av NAV Forskning og Utvikling. Finansieringen er bidragsfinansiert.

Utvidet sammendrag

Den pågående automatiseringen av beslutningsprosesser i offentlig forvaltning representerer en omveltning innenfor byråkratisk myndighetsutøvelse. Tilgang på store mengder relevant digital data og økende muligheter for å behandle informasjonen gjør at oppgaver som tidligere måtte behandles manuelt kan overlates til hel- eller halvautomatiserte prosesser med vesentlig redusert menneskelig inngripen. På den ene siden gir denne utviklingen store effektiviseringsmuligheter og potensial for offentlige besparelser. På den andre siden er ivaretakelsen av forvaltningens legitimitet i befolkningen et risikoaspekt i denne utviklingen.

NAV er ledende i utviklingen av digitale tjenester og verktøy, og utvikler systemer som kan nyttiggjøre seg framskritt som gjøres innenfor databehandling og analyse. Beslutningsprosesser som benytter seg av maskinlæring (ML) og kunstig intelligens (KI) vil være en del av løsningen for at NAV skal oppnå samfunnsoppdraget sitt om å bidra til at flere kommer i arbeid og færre på stønad, og samtidig sørge for at de som trenger det, får rett ytelse til rett tid gjennom en pålitelig og effektiv forvaltning. For NAV vil det i første omgang dreie seg om bruk av ML/KI som beslutningsstøtte til veiledere og andre saksbehandlere, og ikke for å gjøre helautomatiserte beslutninger.

Et sentralt kjennetegn ved kunstig intelligens er at slike systemer etterligner, erstatter og utvider menneskelig intelligent handling, og menneskelig beslutningstaking og vurdering. Mulige områder hvor ML/KI kan benyttes som beslutningsstøtte i NAV er blant annet for å beregne sannsynlighet for at den arbeidsledige trenger bistand fra NAV; til å predikere varighet på sykefravær i forbindelse med dialogmøte med NAV; og til å anbefale arbeidsrettede tiltak.

På veien mot bedre tjenester er det viktig at man har med seg brukerne – det vil si innbyggerne i Norge – og lager ansvarlige systemer som gir lik og rettferdig behandling uavhengig av sosial status. Det er viktig å studere rettferdighet fra et statsvitenskapelig perspektiv fordi oppfatninger av rettferdighet antas å påvirke institusjonell legitimitet. Det overordnede målet med denne rapporten er derfor å belyse ut fra et demokratiperspektiv om, og i så fall hvordan, oppfattelsen av NAV som institusjon blant innbyggere i Norge påvirkes av en overgang til økt bruk av ML/KI i saksbehandlingen. Hvordan, om i det hele tatt, endres relasjonen mellom innbyggere og myndigheter når datamaskiner får sterkere innflytelse på beslutninger som gjelder den enkelte borger? I denne rapporten løfter vi noen forskningsspørsmål som hver for seg er ulike, men som har til felles at de kan knyttes til dette spørsmålet. Vi er forsiktige med å trekke vidtrekkende konklusjoner på dette tidspunktet. Vårt for i denne omgang er først og fremst å forsøke å stille noen av de rette spørsmålene som kan sette agendaen på en tematikk som fortsatt er ny og lite utforsket både internasjonalt og ikke minst i Norge. Pågående arbeid bygger videre på resultatene fra prosjektet som denne rapporten presenterer.

Datagrunnlaget for rapporten er samlet inn i Norsk medborgerpanels i 2021, med et representativt utvalg av innbyggerne i Norge på 2000 respondenter. Undersøkelsen er utviklet av forfatterne, og relevante fagpersoner i NAV har fått tilgang til og kommentert på undersøkelsen før den gikk i felten. Vi presenterer også noen resultater fra relevante spørsmål som vi stilte i 2018, i medborgerpanelets runde 13.

For å kartlegge konteksten denne studien gjøres i, inkluderer undersøkelsen noen generelle spørsmål om innbyggernes forhold til NAV om kontakt med, kjennskap til og oppfatninger om NAV. Verdt å trekke fram fra svarene er at over halvparten av innbyggerne har vært i personlig kontakt med saksbehandler i NAV, men mange oppgir likevel at de har liten kjennskap til organisasjonen. Flertallet av innbyggerne opplever videre at de er i stand til å få de tjenestene de

har krav på fra det offentlige. Dette på tross av at mange opplever forvaltningen som krevende å forstå. Når det gjelder tillit til NAV er den høyere blant de som opplever at de får ytelsene de har rett på; som føler at de forstår de byråkratiske prosessene; som tenker at de som jobber i offentlig sektor bryr seg om folks behov; og som oppfatter at saksbehandlerne ikke bare forholder seg til tekniske aspekter i saksbehandlingen. Motsvarende er tilliten lavere blant de som har et annet syn på forvaltningen. Den samme forskjellen observerer vi når spørsmålet dreier seg om hvorvidt saksbehandlerne oppfattes som upartiske i sin myndighetsutøvelse. Innbyggerne mener i liten til noen grad at saksbehandlere i NAV lar seg påvirke av egne holdninger, men denne oppfattelsen varierer sterkt etter hvilken oppfatning de har om forvaltningen og tilliten de har til NAV.

Et av forskningsspørsmålene rapporten fokuserer på er oppfatninger om såkalt *representativt byråkrati* når forvaltningen tar i bruk ML/KI. Vi finner i våre analyser at oppslutningen om dette øker. Begrepet representativt byråkrati springer ut fra tanken om at forvaltningen skal gjenspeile befolkningen og slik hindre at sosiale grupper blir forskjellsbehandlet. Eventuelle bias saksbehandlere måtte ha vil i så fall utjevnes ved at deres bakgrunn er variert og representativ for befolkningen samlet sett. Innbyggerne blir mer opptatt av at saksbehandlerne deler deres sosiale bakgrunn når slike kunstig intelligens brukes i saksbehandlingen, og dette gjelder spesielt når det kommer til utdanningsnivå og arbeidserfaring.

En hypotese er at bruk av ML/KI leder til økt fremmedgjøring, og at behovet øker for saksbehandlere som forstår den enkeltes situasjon og kan gripe inn i tilfeller hvor den maskinelle vurderingen ikke tar tilstrekkelig hensyn til kontekst. I så fall vil det i framtiden være behov for enda sterkere fokus på å ha et representativt byråkrati i de deler av forvaltningen som utvikler og benytter seg av ML/KI som beslutningsstøtte, med mindre slik bruk blir mindre fremmed for befolkningen i framtiden enn den er i dag. Det er i alle fall klart at ML/KI er fremmede begreper for et flertall av innbyggerne i samtidens Norge: Mer enn seks av ti innbyggerne i Norge har liten eller ingen kjennskap til temaet.

Et viktig spørsmål knyttet til bruk av ML/KI er hvordan modellene kan ivareta oppfattelsen av at beslutninger som tas er rettferdige. Det finnes imidlertid ulike definisjoner av hva rettferdighet er, og det er vanskelig – for ikke å si umulig – å oppfylle alle definisjonene på samme tid. Vi har derfor i undersøkelsen tatt for oss et konkret tilfelle som er relevant for NAVs tjenester, hvor vi måler innbyggernes støtte til det som kalles *statistisk paritet*.

Denne rettferdighetsdefinisjonen innebærer at man sikrer lik fordeling av et gode blant bestemte undergrupper i samfunnet, ofte valgt ut basert på sosial eller etnisk tilhøriget. Den konkrete saken gjelder bruk av ML/KI for å understøtte en beslutning om hvilke personer blant de sykmeldte som skal få tilbud om dialogmøte med NAV. I vårt tilfelle har vi spurt hvilken modell man foretrekker av en som er mer treffsikker totalt sett, men skeivfordeler på kjønn, eller en som er mindre treffsikker totalt sett, men sikrer at like mange sykmeldte fra hvert kjønn får tilbud om dialogmøte.

Vi finner at befolkningen er delt omtrent på midten, med en liten overvekt av støtte til å bruke statistisk paritet. Kvinner støtter statistisk paritet noe mer enn menn i dette konkrete tilfellet. Både menn og kvinner støtter statistisk paritet i sterkere grad dersom det er kvinner som blir forfordelt, dog er denne effekten sterkest blant kvinner. Befolkningen er med andre ord ikke samstemt om hvilket rettferdighetskriterie som skal gjelde i dette tilfellet, noe som samsvarer med andre studier som viser at hva som er rettferdig er avhengig både av kontekst og av øyet som ser.

Et annet viktig spørsmål knyttet til bruk av ML/KI er hvilke data det oppfattes som passende å bruke som input i modellene som skal gjøre prediksjoner som er relevante for beslutninger som skal tas om enkeltindivider. I mange brukstilfeller vil det finnes et bredt spektrum av informasjon tilgjenglig, men det vil sannsynligvis variere hvor passende innbyggere faktisk mener det er å bruke de ulike typene informasjon – uavhengig av om de gjør prediksjonen mer treffsikker. Vi har derfor spurt innbyggere hvor passende de synes en rekke variabler er i et konkret eksempel.

Eksempelet vi trekker fram gjelder å bruke ML/KI for å foreslå hvilke arbeidsrettede tiltak en jobbsøker skal få tilgang til. Vi finner at ingen variabler blir sett på som utvilsomt passende eller upassende, men samtidig at det er tydelige forskjeller mellom variabler. For eksempel blir informasjon om kjønn og landbakgrunn sett på mindre passende enn andre variabler, og utdanning sett på som mer passende enn andre variabler. Samtidig finner vi også viktige systematiske forskjeller mellom ulike undergrupper i befolkningen for hvordan de gjør denne vurderingen.

Mye forskning gjenstår for å kunne trekke vidtrekkende konklusjoner om hvordan tillit og legitimitet best kan bevares i overgangen til økt bruk av ML/KI. Vi er fortsatt i en tidlig fase, hvor de fleste innbyggerne har liten kjennskap til tematikken, og hvor bruken fortsatt er på design- og utprøvingsstadiet.

Problemstillingene kan være komplekse, og av og til kan det være vanskelig for den jevne innbygger å ta stilling til spørsmål som de ikke har tenkt mye over. Samtidig er det nyttig å allerede nå merke seg at befolkningen er delte i mange av spørsmålene om ML/KI i forvaltningen, både når det gjelder bekymring for bruk og hva som er rettferdig framgangsmåte. Innbyggerne er mer følsomme for spørsmål om bruk av ML/KI under omstendigheter hvor bruken knyttes opp mot sosiale bakgrunnsvariabler som ellers i samfunnet er politisk ladete. Også internasjonalt ser vi at bruk av ML/KI når offentlighetens søkelys i de tilfellene hvor marginaliserte grupper opplever at de blir forskjellsbehandlet.

Det er derfor viktig for NAV og andre myndighetsorganer å ta hensyn til de politiske dimensjonene knyttet til bruk av ML/KI i forvaltningen. Opplevd urettferdig behandling er aldri tillitsbyggende, men kanskje ekstra skadelig hvis uretten kan tilskrives "kode-diskriminering". Veien er i disse tilfellene kort til å trekke slutninger om systematisk, strukturell urettferdighet mot bestemte sosiale grupper.

Representasjon av interessegrupper og medvirkning i utformingen av modellene er demokratiske verktøy som virker konfliktdempende i andre sammenhenger, og som det er grunn til å anta vil virke også i en overgang til mer automatisert forvaltning. I eksperimentet med statistisk paritet så vi at det hadde en positiv effekt å opplyse respondentene om at modellene som ble brukt hadde blitt anbefalt av en komite som på forhånd hadde vurdert modellene. Det er behov for mer forskning, men vi ser allerede med det vi har lært fra dette prosjektet at det er fornuftig å skynde seg sakte på dette feltet. Under innfasing av ML/KI som beslutningsstøtte i saker som berører enkeltpersoner bør det være grundige innspillsprosesser slik at innbyggere og berørte parter blir involvert allerede i designfasen og slik på et tidlig stadium kan medvirke til å identifisere etiske dilemma, interessekonflikter, og andre potensielle konfliktsaker som kan oppstå senere.

Bakgrunn og motivasjon

Byråkratisk omveltning

Den pågående automatiseringen av beslutningsprosesser i offentlig forvaltning representerer en omveltning innenfor byråkratisk myndighetsutøvelse. Tilgang på store mengder relevant digital data og økende muligheter for å behandle informasjonen gjør at oppgaver som tidligere måtte behandles manuelt kan overlates til hel- eller halvautomatiserte prosesser med vesentlig redusert menneskelig inngripen (Zarsky 2016). På den ene siden gir denne utviklingen store effektiviseringsmuligheter og potensial for offentlige besparelser (Duwe and Rocque 2017). Den representerer også en mulighet for å utvikle bedre, evidensbaserte beslutninger, som i sin tur kan bidra til å bevare tilliten og legitimiteten til offentlig forvaltning. På den andre siden er nettopp ivaretakelsen av forvaltningens legitimitet i befolkningen også et risikoaspekt i denne utviklingen. En frykt er at feil bruk kan lede til utfall som negativt forskjellsbehandler svakerestilte grupper i samfunnet, som igjen underminerer systemtilliten.

En arbeidsgruppe oppnevnt av den tidligere amerikanske president Barack Obama publiserte rapporter hvor de uttrykte bekymring for "kode-diskriminering» i automatiserte beslutninger, hvor diskriminering av sosiale grupper oppsto som en utilsiktet følge av måten stordatateknologi er strukturert og brukes. Dystopiske skildringer av "svart boks"-samfunn maler et skremmende bilde av et framtidssamfunn der innbyggernes skjebner blir bestemt av skjulte, upresise, og diskriminerende automatiske beslutningsprosesser (Barocas and Selbst 2016; Pasquale 2015). I de tilfeller oppmerksomheten når ut til allmennheten, har fokus tendert å handle om hvordan beslutningene slår ulikt ut sosiale grupper. Det amerikanske nyhetsmagasinet ProPublica viste hvordan prediksjonsmodeller som brukes til å forutsi gjentakelsesfare for lovbrudd blant fengselsinnsatte systematisk kategoriserte svarte insatte oftere enn hvite feilaktig som personer med høy risiko for å begå en ny forbrytelse når de løslates fra fengselet (Angwin et al. 2016).

Opplevd diskriminering fra myndighetenes side mot sosiale grupperinger er ikke noe nytt, og spesielt ikke i USA hvor automatiserte beslutninger har fått mest oppmerksomhet til nå. I Norge har vi mindre forskjeller mellom folk, både økonomisk, politisk og sosialt. Norge har også en høyt kompetent og effektiv offentlig forvaltning som jevnt over nyter høy tillit i

befolkningen. I overgangen til økt automatisering i forvaltningen er det viktig at tilliten og legitimiteten til offentlig forvaltning opprettholdes.

NAV er ledende i utviklingen av digitale tjenester og verktøy(Hansen, Lundberg, and Syltevik 2018), og utvikler systemer som kan nyttiggjøre seg framskritt som gjøres innenfor databehandling og analyse. Beslutningsprosesser som benytter seg av maskinlæring og kunstig intelligens vil være en del av løsningen for at NAV skal oppnå samfunnsoppdraget sitt om å bidra til at flere kommer i arbeid og færre på stønad, og samtidig sørge for at de som trenger det, får rett ytelse til rett tid gjennom en pålitelig og effektiv forvaltning. Maskinlæring og kunstig intelligens kan brukes både i helautomatiserte beslutningsprosesser, og som beslutningsstøtte for saksbehandlere. Et sentralt kjennetegn er at slike verktøy etterligner, erstatter og utvider menneskelig intelligent handling, og menneskelig beslutningstaking og vurdering.

For NAV vil det i første omgang dreie seg om bruk av ML/KI som beslutningsstøtte til veiledere og andre saksbehandlere, og ikke for å gjøre helautomatiserte beslutninger. Mulige områder hvor maskinlæring og kunstig intelligens kan benyttes som beslutningsstøtte i NAV er blant annet for å beregne sannsynlighet for at den arbeidsledige trenger bistand fra NAV; til å predikere varighet på sykefravær i forbindelse med dialogmøte med NAV; og til å anbefale arbeidsrettede tiltak. På veien mot bedre tjenester er det viktig at man har med seg brukerne — det vil si innbyggerne i Norge — og lager ansvarlige systemer som gir lik og rettferdig behandling uavhengig av sosial status.

Beslutningsprosesser, rettferdighet og legitimitet

En massiv litteratur på prosedyrerettferdighet med utspring fra sosialpsykologi (Lind and Tyler 1988) har hatt stor innvirkning på hvordan vi forstår relasjonene mellom innbyggere og myndighetene, og hvordan myndighetene bør forholde seg i møte med innbyggerne. Når beslutningsprosessene dreier i retning av mer bruk av maskinlæring og kunstig intelligens, fungerer denne litteraturen som et velegnet rammeverk for å undersøke empirisk om, og i så fall hvordan, relasjonene mellom innbyggerne og myndighetene vil påvirkes av denne utviklingen. Det vi vet fra eksperimentell forskning på politisk atferd er at både aspekter ved

prosessen og utfallet i seg selv påvirker rettferdighetsoppfatningen av beslutningen og i sin tur villigheten til å akseptere beslutningen (se figur under). På kort sikt handler det om å skaffe aksept for enkeltbeslutninger. I et mer overordnet perspektiv dreier det seg om systemstøtte; om å sikre tillit og legitimitet til styresmaktene, og opprettholde tilfredsheten med demokratiet som styresett. Legitimitet forstår vi her som makten til å få noen til villig å føye seg etter en beslutning (Weber 2009), som i sin tur gir myndighetene den autoriteten de trenger for å styre effektivt uten bruk av sanksjoner (Tyler 2021). Demokratisk legitimitet viser til den legitimiteten som vinnes ved at beslutningene utgår fra folkeviljen (Rosanvallon 2011). Både tillit og legitimitet omhandler relasjonen mellom innbyggere og myndighetene, og faller under det bredere konseptet om politisk støtte (Easton 1965).

Prosessrelaterte spørsmål som har blitt studert er blant annet om rettferdighetsoppfatningen og aksepten av beslutningen påvirkes av forhold som er sentrale i demokratiske systemer. Slike forhold kan for eksempel være grad av åpenhet rundt beslutningsprosessen (De Fine Licht et al. 2014), mulighet for direkte påvirkning på en avgjørelse (Esaiasson, Gilljam, and Persson 2012; Arnesen 2017; H. S. Christensen, Himmelroos, and Setälä 2020), hvem beslutningstakerne er, og hvor godt disse beslutningstakerne gjenspeiler befolkningen med tanke på sosial bakgrunn (Arnesen and Peters 2018; Clayton, O'Brien, and Piscopo 2019). Videre er et gjennomgående funn at dersom utfallet går imot ens egne ønsker, blir prosessen diskreditert (Esaiasson et al. 2016).

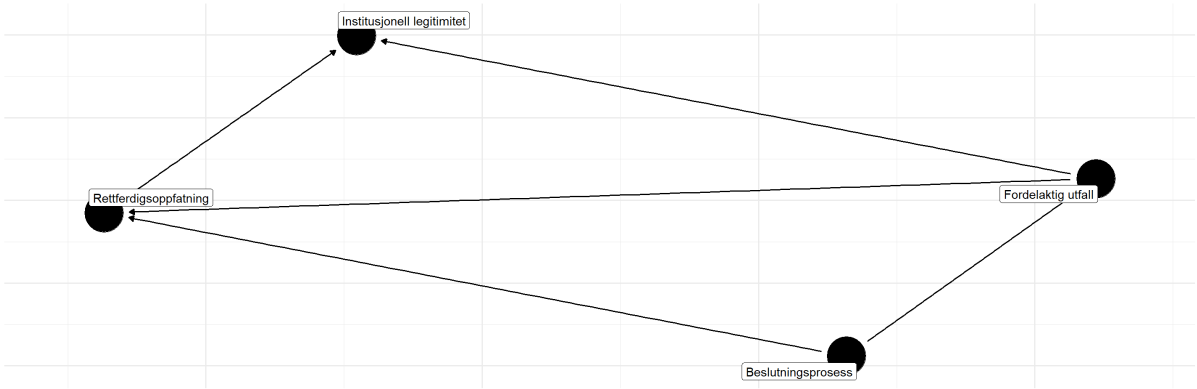

Figure 1: Antatt årsakssammenheng

Demokrati- og opinionsforskere har i liten grad studert hvordan overgangen til økt bruk av maskinlæring og kunstig intelligens i forvaltningens beslutningsprosesser kan påvirke

systemtillit og legitimitet. Unntak er De Fine Licht & De Fine Licht (2020) som studerer rollen som åpenhet når det gjelder hvordan allmennheten oppfatter kunstig intelligens-beslutninger som legitim, og Binns et al (2018) som foretar eksperimentelle studier som undersøker folks oppfatning av rettferdighet i algoritmiske beslutninger. Vår kunnskap om hvordan overgangen vil påvirke innbyggernes oppfatninger av forvaltningen er imidlertid fortsatt begrenset, og behovet for befolkningsrepresentative studier med et demokratiperspektiv er stort. Det er motivasjonen for å lage denne rapporten.

Rapporten presenterer resultatene fra en spørreundersøkelse gjennomført på et befolkningsrepresentativt utvalg av innbyggere i Norge. Undersøkelsen tar for seg generelle holdninger til og kunnskap om maskinlæring og kunstig intelligens i befolkningen. Å bruke maskinlæring innebærer å få datamaskiner til å lære seg å løse oppgaver basert på et datamateriale. Ofte kan datamaskinen bli ekstremt treffsikker, men det krever typisk veldig mye datamateriale. Maskinlæring er i dag grunnlaget for alt fra automatisk stemmegjenkjenning til førerløse biler. I den internasjonale litteraturen om prosesseringen av data med formål å forbedre og effektivisere beslutninger i forvaltningen brukes flere begreper om hverandre, så som "automatiserte beslutninger", "bruk av maskinlæring og kunstig intelligens i forvaltningen", og "algoritmestyrt forvaltning". Vi bruker også begrepene til en viss grad om hverandre i denne rapporten, avhengig av hvilke tidligere studier vi forholder oss til og hvilke spørsmålsformuleringer som har blitt brukt i undersøkelsene vi viser til.

Bredere spørsmål knyttet til relasjonen mellom NAV og innbyggerne – uavhengig av tematikken om maskinlæring og kunstig intelligens – blir også analysert for å kontekstualisere de mer spesifikke spørsmålene og eksperimentene knyttet til bruk av maskinlæring og kunstig intelligens i NAV. Deretter fokuserer vi på holdninger som har relevans for en framtid hvor maskinlæring og kunstig intelligens vil spille en sentral rolle. Dette gjelder problemstillinger knyttet til konkrete, aktuelle situasjoner i NAV, såvel som mer overordnete spørsmål om maskinlæring og kunstig intelligens i forvaltningen.

Data og metode

Denne rapporten er basert på spørreundersøkelse på et befolkningsrepresentativt utvalg av innbyggere i Norge. Undersøkelsen var delt i tre hoveddeler. Den første delen tar for seg spørsmål knyttet til relasjonen mellom innbyggerne og NAV. Den andre delen tar for seg mer generelle holdninger og kunnskap til maskinlæring og kunstig intelligens. I den tredje delen av undersøkelsen bruker eksperimentscenarier som er eller kan bli realistiske i NAV-sammenheng, hvor respondent må gjøre spesifikke avveinger rundt om og hvordan maskinlæring brukes eller implementeres.

Ved flere deler av undersøkelsen benytter vi eksperiment. Eksperiment som metode i samfunnsforskningen defineres av to nøkkelelementer. For det første må det være en datainnsamlingsprosess som samfunnsforskeren setter i gang for å søke svar på et forskningsspørsmål. Eksperimentdata kan ikke være observasjonsdata, altså data som allerede eksisterer i «den virkelige verden» og har skjedd – og ville ha skjedd – uten forskerens innblanding. Et eksperiment må også ha en kontrollgruppe som viser hvordan verdiene på den avhengige variabelen framstår når de ikke er påvirket av den eller de faktorene som undersøkes. Når stimulusgruppene blir introdusert for påvirkning, måler vi nivåene på disse gruppene opp mot kontrollgruppen for å se om det er noen forskjell. Eksperimenter i samfunnsforskningen kan gjennomføres som felteksperimenter, som lab-eksperimenter og som surveyeksperimenter. Sistnevnte har vært mest brukt innenfor studier av demokratiske beslutningsprosesser og er også det vi benytter oss av i denne rapporten.

Norsk medborgerpanel

Data for denne rapporten er samlet inn i Norsk medborgerpanels 22 og 23. Samlet antall respondenter som svarte på deler av denne undersøkelsen var 3971. Typisk antall respondenter for hvert spørsmål er om lag 2000. Vi presenterer også noen resultater samlet inn i 2018 (runde 13).

Norsk medborgerpanel er en internettbasert undersøkelse om nordmenns holdninger til viktige samfunnstema. Panelet drives av samfunnsforskere ved Universitetet i Bergen og NORCE,

og er et non-profit prosjekt utelukkende benyttet til forskningsformål. Deltakerne representerer et tverrsnitt av den norske befolkningen, som noen ganger i året inviteres til å si sin mening i viktige spørsmål om norsk samfunn og politikk. Panelet blir driftet av Digital samfunnsvitenskapelig kjernefasilitet (DIGSSCORE), som er samarbeidspartner i prosjektet.

Metoderapport og kode for data kan lastes ned her. Kodebøker kan lastes ned her. Data er åpent tilgjengelig for forskere, og kan lastes ned ved å kontakte Sikt – Kunnskapssektorens tjenesteleverandør (tidligere kjent som NSD).

Innbyggernes relasjon til NAV

I dette kapitlet ser vi nærmere på relasjonen mellom NAV og innbyggerne i Norge. Kapitlet danner grunnlaget for å forstå konteksten når vi senere fokuserer mer spesifikt på maskinlæring og kunstig intelligens i organisasjon. Vi måler folks tillit til NAV, samt deres erfaring med og kjennskap til NAV. Vi måler også i hvilken grad innbyggerne opplever at de forstår hvordan organisasjonen fungerer, og i hvilken grad de opplever at deres interesser blir ivaretatt i systemet.

Vi finner at

- over halvparten av innbyggerne har vært i personlig kontakt med saksbehandler i NAV, men mange oppgir likevel at de har liten kjennskap til NAV.
- innbyggerne har middels til høy tillit til NAV. Det er få som oppgir å ha svært høy tillit eller ingen tillit i det hele tatt.
- flertallet av innbyggerne opplever at de er i stand til å få de tjenestene de har krav på fra det offentlige. Dette på tross av at mange opplever forvaltningen som krevende å forstå.
- innbyggerne i liten til noen grad mener at saksbehandlere i NAV lar seg påvirke av egne holdninger. Det skiller lite mellom hvilke områder av NAVs ansvarsområde man spør etter.

Erfaring med og kjennskap til NAV

Et flertall av innbyggerne i Norge har hatt befatning med NAV i en eller annen form. Noen kontakter er lite personlige, som for eksempel når man mottar barnetrygd. Andre krever mer kontakt med NAV, og gjerne personlig kontakt med en saksbehandler. I vårt utvalg har godt over halvparten minst en gang vært i personlig kontakt med NAV.

Fire av ti innbyggere i Norge har liten eller ingen kjennskap til NAV, mens seks av ti har ganske god, god, eller svært god kjennskapt til institusjonen.

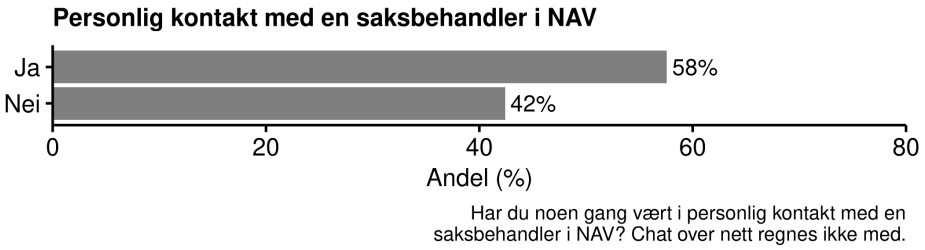

Figure 2: Personlig kontakt med NAV

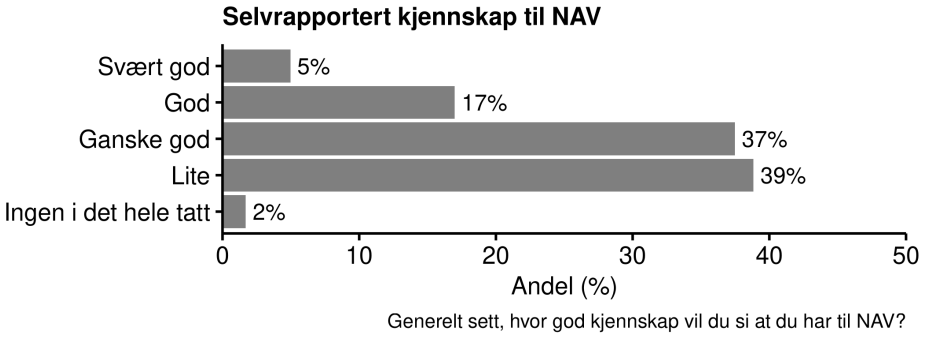

Figure 3: Selvrapportert kunnskap om NAV

Tillit

Tillit er noe som tar lang til å bygge opp, men som fort kan rives ned. Norge er kjent som et land der myndighetene nyter høy tillit i befolkningen. Det gjenspeiles også i våre resultater. Fire av ti innbyggere i Norge har høy eller svært høy tillit til NAV, mens bare en av seks har liten eller ingen tillit. Det fremstår derfor som at NAV nyter høy tillit i befolkningen, og ikke desto viktigere at denne relasjonen ivaretaes parallelt med at organisasjonen endrer seg og utvikler morgendagens forvaltning. Hvor høyt dette er sammenlignet med andre institusjoner i Norge, eller samme type institusjoner i andre land, fokuserer vi ikke på her. Formålet med å måle tillit til NAV i denne undersøkelsen er spesifikt for å kunne sammenligne hvordan de med høy og lav tillit forholder seg til bruk av ML/KI i NAV.

Om vi bryter ned tillit etter kjennskap til NAV, ser vi at det ikke er noen tydelig sammenheng mellom selvrapportert kjennskap og tillit, men at de som har hatt personlig kontakt med en saksbehandler i snitt har noe lavere tillit. Forskjellen er relativt liten men statistisk signifikant. Det er viktig å påpeke her at dette ikke sier noe om hvorvidt kontakten med NAV er det som fører til lavere tillit, siden bakenforliggende faktorer som gjør at man har høyere sannsynlighet for å ha kontakt med NAV også påvirker tillit. Det er imidlertid en viktig deskriptiv forskjell

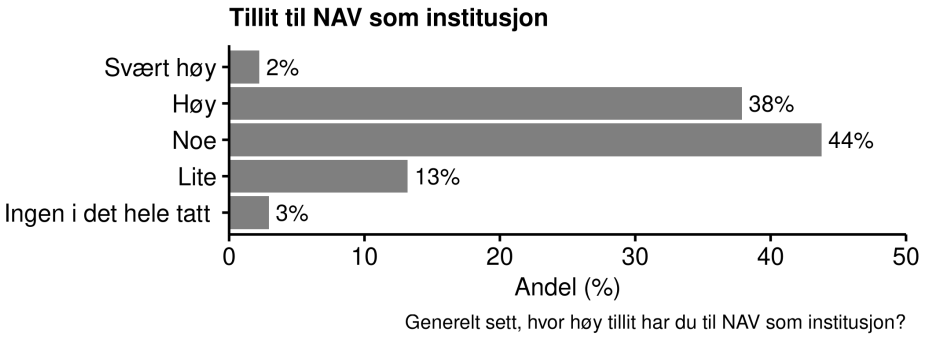

Figure 4: Tillit til NAV

mellom de som har hatt og de som ikke har hatt kontakt med NAV, siden de med høy og lav tillit kan ha forskjellige forventinger til bruk av ML/KI i NAV.

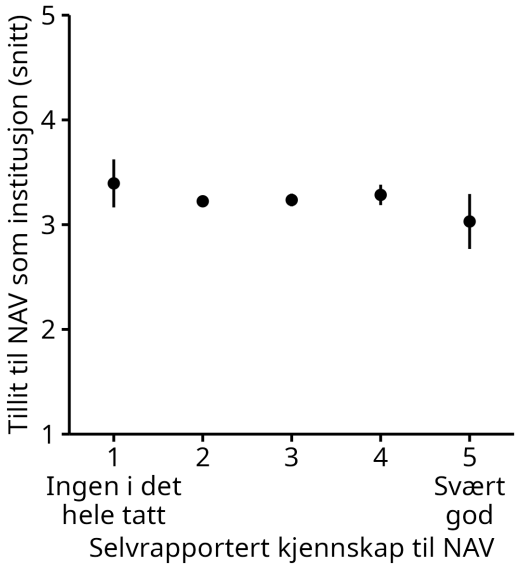

Figure 5: Tillit til NAV for ulike nivåer av selvrapportert kjennskap

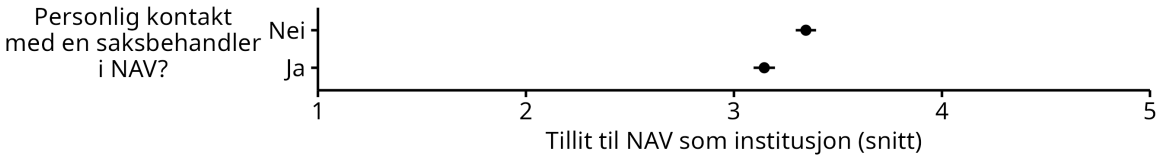

Figure 6: Tillit til NAV for for de som har eller ikke har hatt personlig kontakt med NAV

Byråkratisk kompetanse

I statsvitenskapen opererer man med et begrep som på engelsk heter *political efficacy*. Vi kan gjerne oversette begrepet på norsk til *politisk kompetanse*. Politisk kompetanse viser til en persons selvopplevde evne til å forstå politikk (*internal political efficacy*), og personens

opplevelse av å kunne påvirke politiske prosesser (*external political efficacy*). Vi ser at intern og ekstern politisk kompetanse henger sammen med politisk deltakelse, med tilfredshet med demokratiet, tillit til institusjoner, blant annet.

På samme måte som at innbyggerne har ulike evner til å forstå politikk og påvirke politiske prosesser, har de også ulike evner til å forstå forvaltningen. I motsetning til i politisk arbeid er det ikke et mål at innbyggerne nødvendigvis skal kunne påvirke en byråkratisk prosess, men det er likefullt en kjennsgjerning at noen personer er flinkere til å følge opp saker på vegne av seg selv eller pårørende, og slik sett er bedre i stand til å ivareta sine interesser i saker som angår dem.

For å avdekke hvordan disse ferdighetene fordeler seg i befolkningen har vi spurt respondentene om det vi benevner deres *byråkratiske kompetanse*. Kompetanse dreier seg altså her ikke om byråkratiets kompetanse, men om innbyggerens kompetanse overfor byråkratiet. Det finnes da to types slik kompetanse: intern og ekstern. *Intern* byråkratisk kompetanse omhandler en persons oppfatning av egne evner – hvorvidt de tror de forstår forvaltningen og hvorvidt de tror de kan skaffe de offentlige ytelser, tjenester, etc, de har rett på. Dette er nært knyttet til det som kalles "bureaurcratic compentence" (Gordon 1975) og "administrative literacy" (Döring 2021) i litteraturen. *Ekstern* byråkratisk kompetanse omhandler en persons oppfatning av hvordan de tror byråkratiet forholder seg til deres behov.

Intern byråkratisk kompetanse måler vi ved hjelp av to påstander som de skal si seg enig eller uenig i:

1. Jeg er i stand til å skaffe alle offentlige ytelser, tjenester, og tillatelser som jeg har rett på.
2. Den offentlige forvaltningen er så innviklet at folk som meg ikke kan forstå hva som foregår innad i ulike etater, direktorat, kommuner, og så videre.

Ekstern byråkratisk kompetanse måler vi ved hjelp av to nye påstander:

1. De som jobber i den offentlige forvaltningen bryr seg ikke om hvilke behov folk som meg har.
2. Saksbehandlere i den offentlige forvaltningen er bare interessert i tekniske aspekter ved

saken, ikke hva de det berører faktisk ønsker.

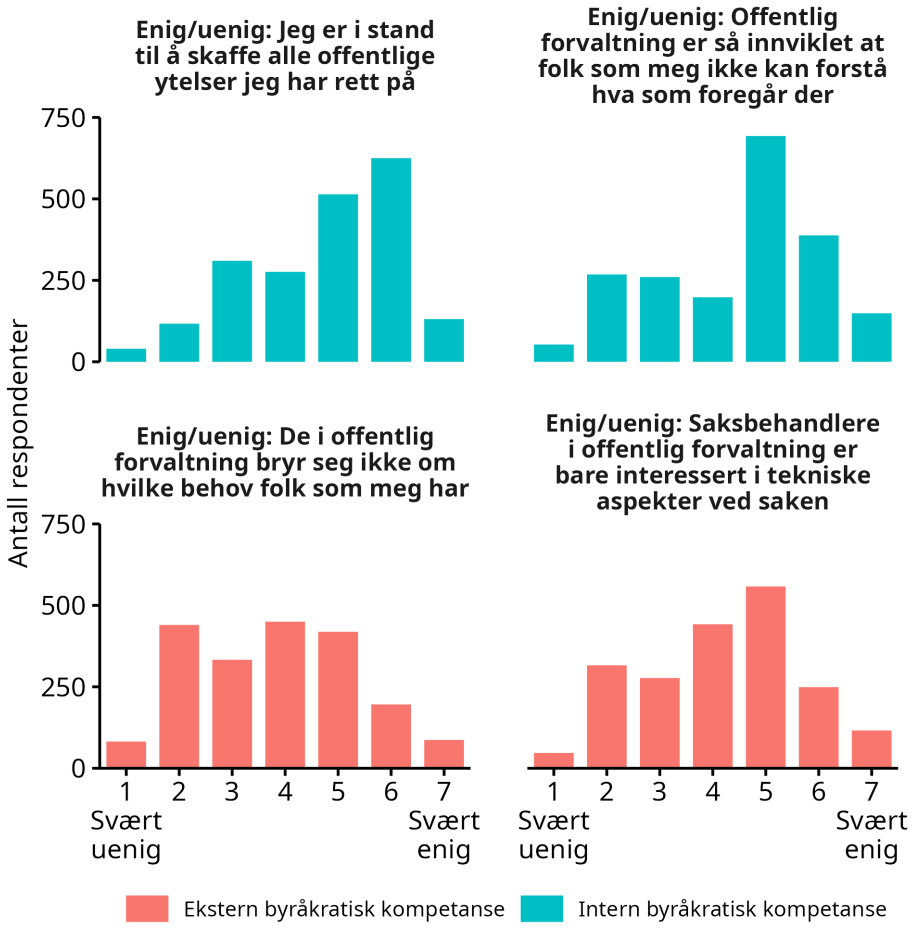

Figure 7: Byråkratisk kompetanse

Figurene viser at det er ganske stor spredning i svarene. Flertallet av innbyggerne er enige i påstanden at de får alle ytelser som de har rett på. Et flertall svarer samtidig at de opplever byråkratiet som vanskelig å forstå. Flertallet er uenige i at de som jobber i offentlig forvaltning ikke bryr seg om folks behov, men det er også et stort mindretall som er enige i denne påstanden. Et lite flertall mener at saksbehandlere bare er interessert i tekniske aspekter ved saken.

Dette er befolkningen i sin helhet. Når vi bryter svarene ned på sosiopolitiske undergrupper ser vi at dem som har lav byråkratisk kompetanse også har lav politisk kompetanse (efficacy). Det vil si at de som har tillit til NAV er de som opplever at de får ytelsene de har rett på, som føler at de forstår de byråkratiske prosessene, som tenker at de som jobber i offentlig sektor bryr seg om folks behov, og som ikke bare forholder seg til tekniske aspekter i saksbehandlingen.

Byråkratisk kompetanse varierer også til dels mye mellom folk når vi deler dem inn i hvilket

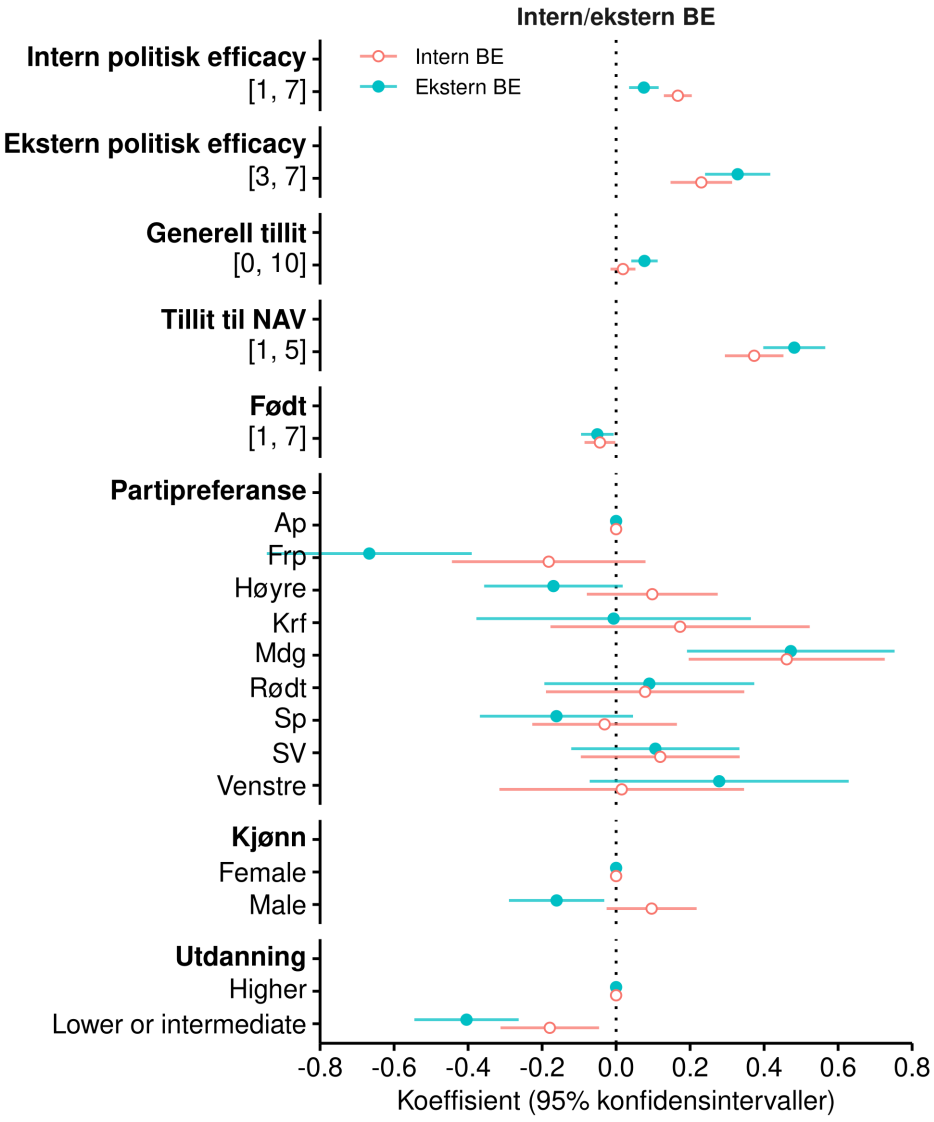

Figure 8: Byråkratisk kompetanse covariater

parti de ville ha stemt på dersom det var stortingsvalg i morgen. Figuren under viser at det er størst forskjell på de som stemmer Fremskrittspartiet og de som stemmer Miljøpartiet de grønne, spesielt når det gjelder ekstern byråkratisk kompetanse.

Menn scorer noe lavere enn kvinner på ekstern byråkratisk kompetanse, og det samme gjør personer uten høyere utdanning sammenliknet med personer med høyere utdanning. De scorer også noe lavere på intern byråkratisk kompetanse, men her er forskjellen mindre (men fortsatt statistisk signifikant).

Tillit til NAV samvarierer også sterkt med byråkratisk kompetanse. Figuren under viser mer dette mer detaljert. Vi observerer en lineær sammenheng mellom de to variablene, som flater ut først blant de som har svært høy tillit til NAV.

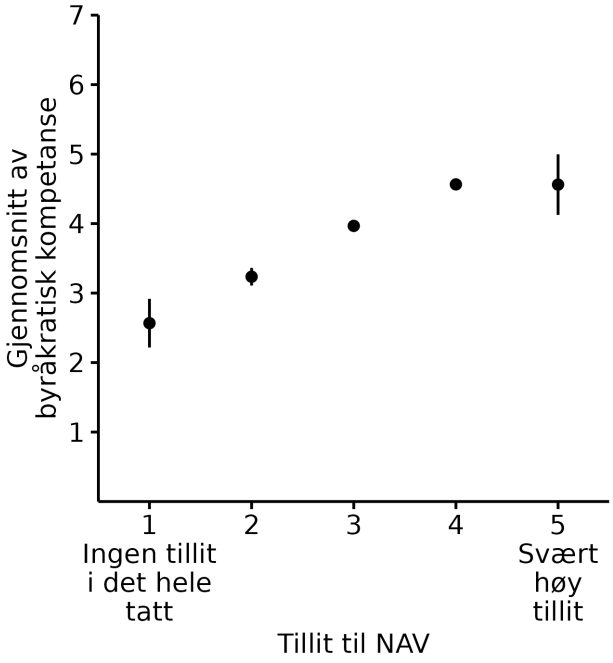

Figure 9: Byråkratisk kompetanse for ulike nivåer av tillit til NAV

Tiltro til likebehandling i NAV

Et styrende prinsipp i norsk forvaltning er nøytralitet: Saksbehandlere skal utøve sitt mandat uten å la egne holdninger komme i veien og påvirke beslutningsprosessen. Slik upartisk behandling av innbyggerne står også sentralt i forskning på hvilke egenskaper ved byråkratiet som underbygger legitimitet og rettferdighetsoppfatninger. Når vi spør innbyggerne i Norge

hvordan de oppfatter at saksbehandlerne i NAV lar seg påvirke av egne holdninger, finner vi at de tror det forekommer i noen grad. Det er liten forskjell på de ulike områdene innenfor NAV.

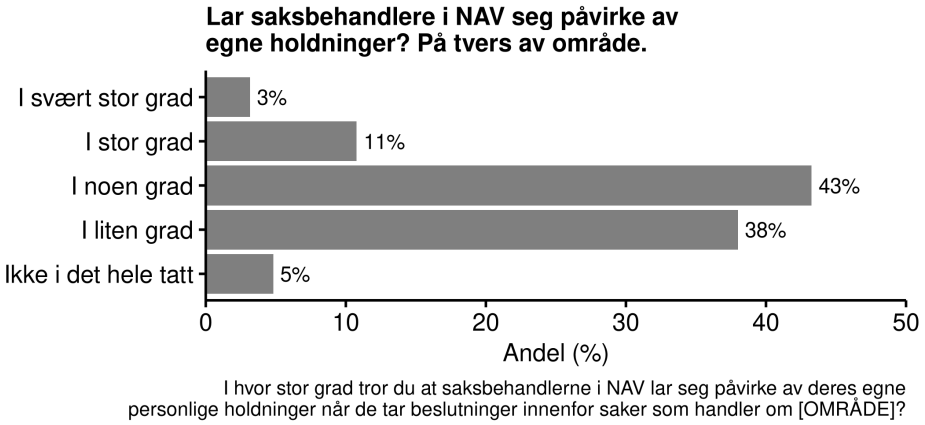

Figure 10: Fordeling av tilltro til likebehandling i NAV på tvers av saksområder

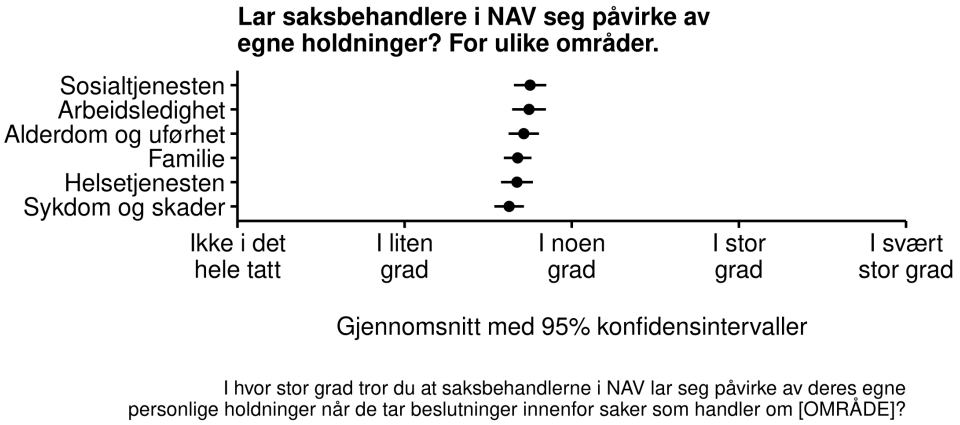

Figure 11: Gjennomsnitt av tilltro til likebehandling for ulike saksområder

Figuren under viser også at byråkratisk kompetanse samvarierer med oppfatningen om at saksbehandlere lar seg påvirke av egne holdninger: Jo høyere byråkratisk kompetanse en innbygger har, desto mindre mener man at saksbehandlerne lar seg påvirke av egne holdninger.

I utvalget mener de som har vært i personlig kontakt med en saksbehandler i NAV i litt større grad at saksbehandlerne lar seg påvirke av egne holdninger. Forskjellen er imidlertid ikke stor.

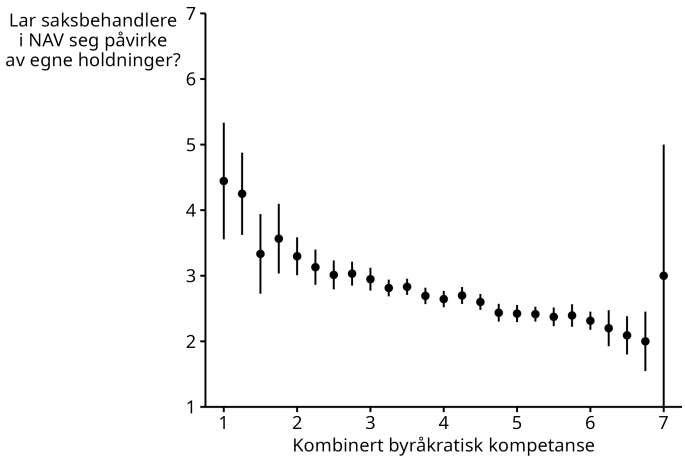

Figure 12: Tilltro til likebehandling i NAV for ulike nivåer av byråkratisk kompetanse

Innbyggernes oppfatninger om maskinlæring og KI

I dette kapitlet ser vi nærmere på hvilke forventinger innbyggerne i Norge har når det gjelder bruk av maskinlæring og kunstig intelligens generelt og i NAV. Respondentene ble først gitt enn introduksjon som kort forklarte maskinlæring og kunstig intelligens og spurte de hvor mye kjennskap de har til temaet. Deretter undersøkte vi deres generelle forventinger rundt dette. Vi finner at

- Mer enn seks av ti innbyggerne i Norge har liten eller ingen kjennskap til maskinlæring og kunstig intelligens
- Innbyggerne er delt i oppfatningen om bruken av maskinlæring og kunstig intelligens i forvaltningen er noe å bekymre seg over
- De som oppfatter at de har god kunnskap om maskinlæring er mer positive til bruk av kunstig intelligens i forvaltningen
- Det er en omvendt U-formet sammenheng mellom selvplassering på politisk høyre/venstre-skala og oppslutning om bruk av kunstig intelligens: Innbyggere som plasserer seg mot midten av det politiske spekteret er mer positive enn de som plasserer seg mot en av endene på skalaen.

Kunnskap om maskinlæring og kunstig intelligens

Kunnskap rundt maskinlæring, kunstig intelligens, algoritmer, m.m., utgjør en viktig brikke for å forstå hvordan befolkningen forholder seg til bruken av slike verktøy. Maskinlæring og kunstig intelligens er avanserte tema som krever spesialkompetanse for å kunne implementere og bruke på en god måte. For å kunne grundig diskutere rammene for hva som utgjør rettferdig eller legitim bruk av slike verktøy må man ha en viss forståelse for hva de rammene er, noe som da krever en viss grunnkunnskap. I undersøkelsen ga vi respondentene en kort forklaring av maskinlæring og spurte dem i hvor stor grad de har kjennskap til dette temaet. Vi ga dem følgende introduksjon til temaet:

> "*Nå ønsker vi å spørre om dine holdninger rundt bruk av maskinlæring i den offentlige forvaltningen. Maskinlæring blir også ofte omtalt som kunstig intelligens.*
>
> *Å bruke maskinlæring innebærer å få datamaskiner til å lære seg å løse oppgaver basert på et datamateriale. Ofte kan datamaskinen bli ekstremt treffsikker, men det krever typisk veldig mye datamateriale. Maskinlæring er i dag grunnlaget for alt fra automatisk stemmegjenkjenning til førerløse biler.*
>
> *Den offentlige forvaltningen, inkludert NAV, bruker i enkelte tilfeller maskinlæring for å hjelpe med å ta beslutninger i saker de har ansvar for. Formålet er å redusere kostnader og behandlingstid, og å gjøre beslutninger bedre og mer treffsikre. Et eksempel kan være å lære en datamaskin å forutsi omtrent hvor lenge en person vil være sykmeldt, basert på informasjon om sykdommen og personen. Det kan en saksbehandler da bruke for å velge passende tiltak.*"

Vi spurte deretter om hvor god kjennskap de har til maskinlæring og kunstig intelligens. Figuren under viser hvordan respondentene fordelte seg på spørsmålet. Det er kun én av syv innbyggere som oppgir at de har god eller svært god kjennskap, mens nesten to tredjeler sier at de har liten eller ingen kjennskap til det i det hele tatt.

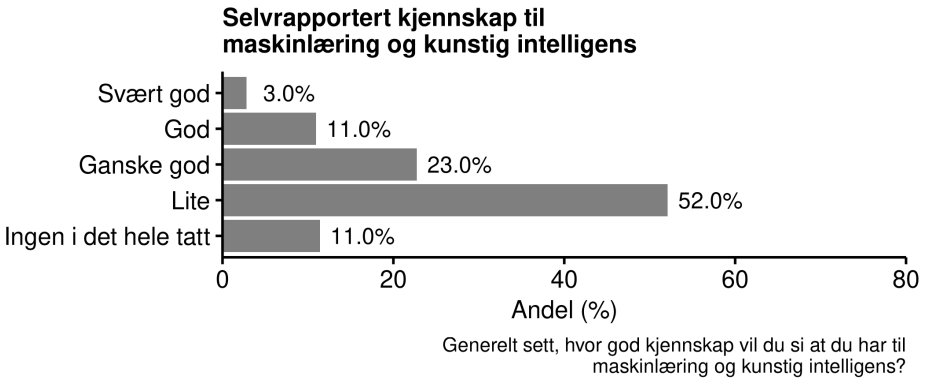

Figure 13: Selvrapportert kunnskap om maskinlæring og kunstig intelligens

Bekymring

Vi spurte deretter om hvor bekymret de er for bruken av slike verktøy i den offentlig forvaltningen. Figurene under viser hvordan respondentene fordelte seg på spørsmålet. Den viser at innbyggerne er delt i synet på grad av bekymring knyttet til bruk av maskinlæring og kunstig intelligens i den offentlige forvaltningen, hvorav over halvparten er bekymret eller noe bekymret. Samtidig er det kun fem prosent som oppgir at de er veldig bekymret.

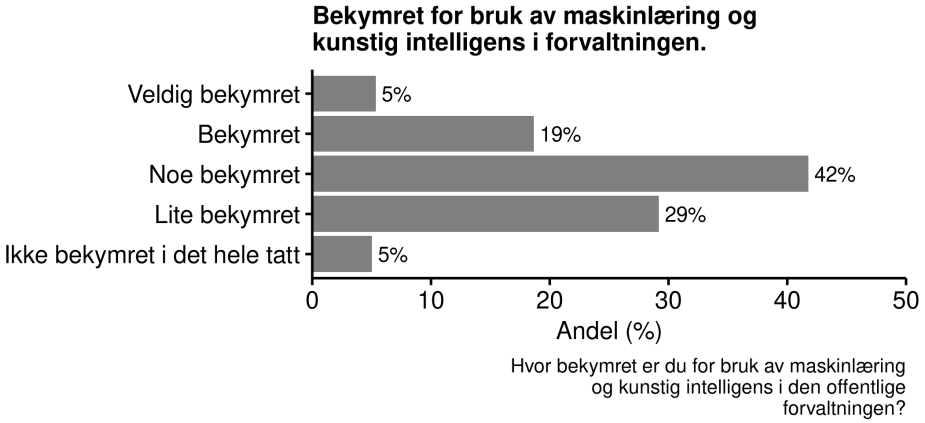

Figure 14: Bekymring for maskinlæring

Den neste figuren viser hvordan gjennomsnitt av bekymring for ulike nivåer av kunnskap. Her ser vi at de med ingen kunnskap er mer bekymret enn de med god eller svært god kunnskap om maskinlæring.

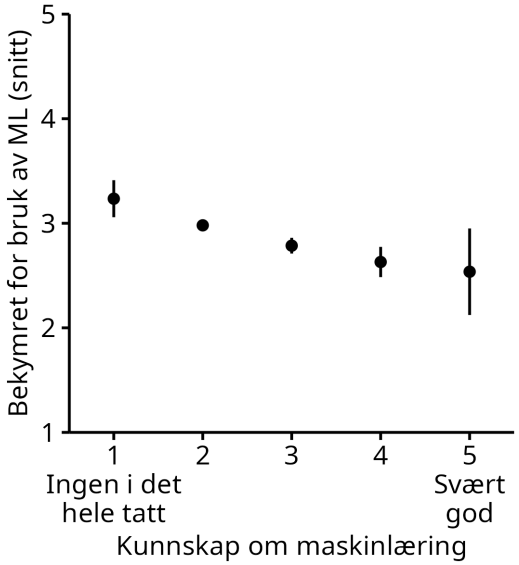

Figure 15: Bekymring for maskinlæring for ulike nivåer av selvrapportert kunnskap om maskinlæring

Blir interesser bedre ivaretatt med maskinlæring?

Vil folk oppleve at det er lettere eller vanskeligere å forstå hvordan byråkratiet fungerer? Vil deres interesser ivaretas bedre eller dårligere når maskinlæring brukes i NAV? For å undersøke dette ba vi dem ta stilling til et noe mer konkret situasjon:

> La oss si at du var i en situasjon hvor du måtte søke NAV om økonomisk stønad. Tror du interessene dine hadde blitt bedre eller dårligere ivaretatt dersom saksbehandleren brukte maskinlæring og kunstig intelligens som hjelp til å fatte beslutningen om økonomisk stønad?

Figuren under viser fordelingen av svar. Det mest vanlige svaret var midtkategorien" 'verken bedre eller dårligere". For øvrig fordelte svarene seg normalt rundt denne midtkategorien. I spørreundersøkelser kan midtkategorier i slike bipolare skalaer ofte skjule at respondentene ikke har noen mening om spørsmålet.

Den neste figuren viser hvordan dette varier etter nivå av selvrapportert kunnskap: De med mye kunnskap har større sannsynlighet for å tro at deres interesser blir bedre ivaretatt enn de med lite kunnskap.

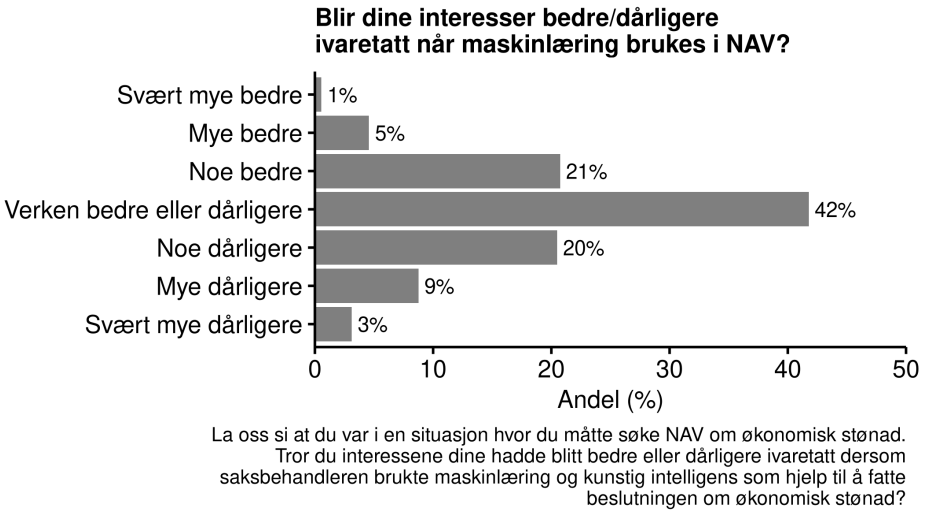

Figure 16: Forventninger om interesser ivaretas bedre med maskinlæring i NAV

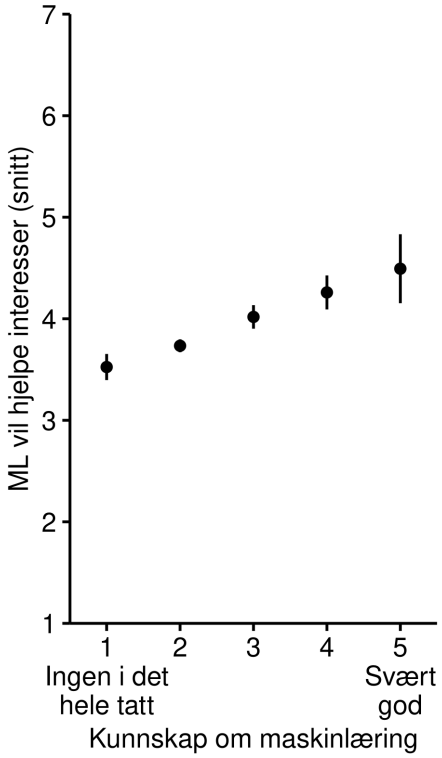

Figure 17: Forventninger om interesser ivaretas bedre med maskinlæring i NAV for ulike nivåer av selvrapportert kunnskap om maskinlæring

Forventinger til automatisering i offentlig sektor

Vi har studert befolkningens forventinger til automatisering i offentlig sektor ved en tidligere spørreundersøkelse. Resultatene under er basert på spørsmål stilt i 2018 (runde 13) i Norsk Medborgerpanel. Dette er ikke direkte relatert til bruk av ML/KI i NAV, siden det ble stilt spørsmål om offentlig sektor generelt – ikke NAV spesifikt – og siden det brukes eksempler på ML/KI som ikke nødvendigvis ligner på hvordan eventuell bruk av slike verktøy vil være i NAV. Med det i mente er disse resultatene fortsatt relevante og interessante, siden det gir et mer overordnet perspektiv om hvilke forventinger befolkningen har til automatisering i offentlig sektor.

Vi spurte dem først om de tror skiftet fra menneskelige til automatiserte beslutninger generelt vil føre til en forbedring eller forverring av offentlig tjenester. Dette var et eksperiment hvor halvparten også fikk oppgitt en rekke konkrekte eksempler. Deretter ble de spurt om de tror økende automatisering i offentlig sektor fører til mer eller mindre av henholdsvis etterprøvbarhet, legitimitet, og upartiskhet. De to figurene under viser fordelingene for alle, altså for både de som fikk og ikke fikk oppgitt konkrekte eksempler. Her ser vi at respondentene hadde delvis ulike forventing for de ulike punktene. De er relativt jevnfordelt når det gjelder generelle forventinger om det vil fære til forbedring eller forverring av offentlig tjenester. Samtidig tror majoriten av innbyggere at det vil føre til mindre etterprøvbarhet og legitimitet, men mer upartiskhet.

Figuren under viser gjennomsnittet delt opp etter hvorvidt de fikk oppgitt konkrete eksempler i spørsmålet om generelle forventinger. Halvparten fikk følgende ekstra tekst:

> Ett konkret eksempel fra Norge er Utlendingsdirektoratet (UDI), som i enkelte tilfeller lar roboten Ada bestemme om enkeltpersoner skal få opphold i Norge. Et par eksempler fra USA er algoritmer som produserer anbefalinger vedrørende hvor politiet bør patruljere for å øke sjansen for å ta kriminelle, samt hvilke bekymringsmeldinger barnevernet bør ta på alvor og hvilke de kan la passere.

Respondentene som fikk vist konkrete eksempler forventer i større grad at automatiseringen vil føre til forverring av offentlig tjenester, samt til mindre legitimitet, enn de som ikke fikk vist

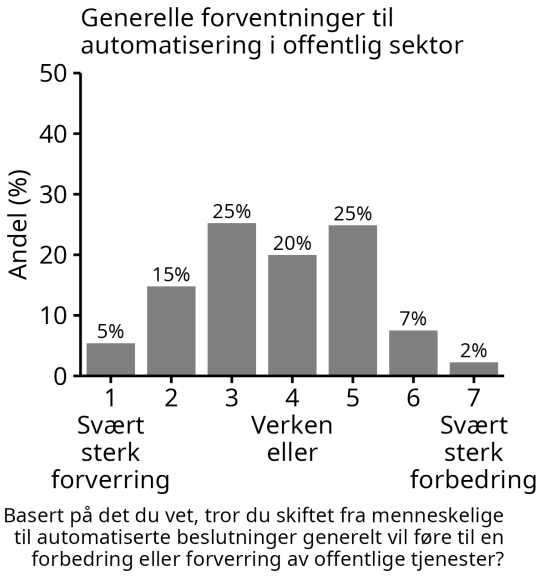

Figure 18: Generelle forventinger til automatisering i offentlig sektor

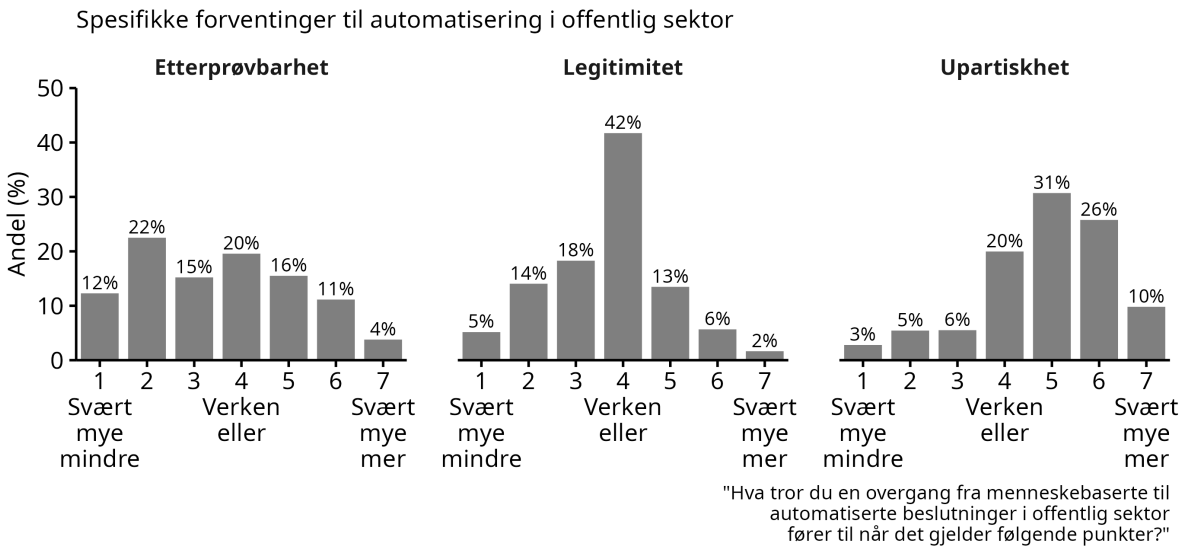

Figure 19: Spesifikke forventinger til automatisering i offentlig sektor

eksempler. Effekten er svak men statistisk signifikant. Det er viktig å påpeke her at resultatene kunne endret seg hvis man hadde brukt andre eksempler, og at disse eksemplene ikke ligner på bruk av ML/KI som NAV har vurdert. Det er likevel et viktig og interessant mønster at når man får oppgitt ekte eksempler på bruk av ML/KI i offentlig sektor, så ser befolkningen i snitt mer negativt på automatisering. Det antyder at kontroversielle eksempler i fremtiden, uavhengig av om det er i NAV, kan være utslagsgivende for hvordan befolkningen forholder seg til bruk av ML/KI i NAV.

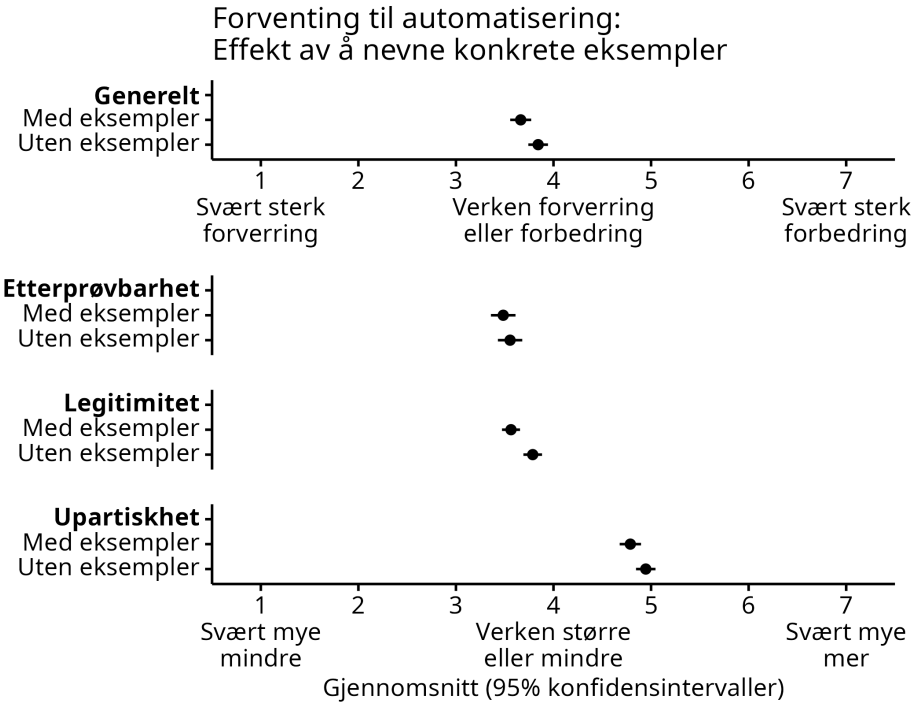

Figure 20: Effekt av å få vite konkrete eksempler på forventinger til automatisering i offentlig sektor

Rettferdighetsoppfatninger

Når er det passende å bruke kunstig intelligens?

I mange beslutninger i forvaltningen må det utvises skjønn basert på en samlet vurdering av den enkelte saken. Om man tar i bruk kunstig intelligens, ved hjelp av maskinlæring, vil beslutningene antakelig bli mer treffsikker, og dermed øke andelen riktige beslutninger. Samtidig kan heller ikke en datamaskin være helt treffsikker. Det er også grunn til å tro at den gjenværende andelen uriktige beslutninger går mer systematisk ut over noen grupper i samfunnet når man bruker maskinlæring og kunstig intelligens. Dette fordi det er stor variasjon mellom hvordan menneskelige saksbehandlere utviser skjønn, mens for en datamaskin er det ingen variasjon.

Med dette som bakgrunn spurte vi respondentene hva foretrekker i slike situasjoner: Enten 1) Bruke kunstig intelligens, som fører til mange flere riktige beslutninger i bytte mot at det alltid er de samme som blir gjenstand for uriktige avgjørelser, eller 2) ikke bruke kunstig intelligens, som fører til mange færre riktige beslutninger i bytte mot at det varierer hvem som blir gjenstand for uriktige avgjørelser. Fordelingen er vist i tabellen under. Respondentene delte seg på midten i dette spørsmålet, hvor rundt 47 prosent foretrakk å bruke kunstig intelligens, mens 53% foretrakk å ikke bruke kunstig intelligenst.

Table 1: Tradeoff mellom generell treffsikkerhet og spesfikke systematiske feil (bruke eller ikke bruke KI)

Svar	Prosent
Bruke KI	47
Ikke bruke KI	53

Figuren under viser at de med lav kjennskap til maskinlæring og kunstig intelligens var mest skeptiske. Det kan altså ha sammenheng med skepsis til det ukjente.

Spørsmålet har også en politisk-filosofisk dimensjon over seg. Premisset som legges til grunn for spørsmålet er at man ved å innføre kunstig intelligens påvirker fordelingen av riktige

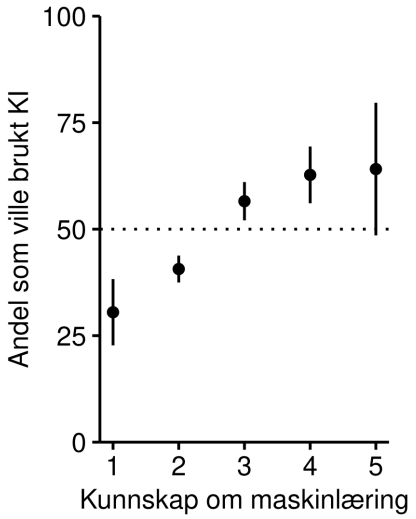

Figure 21: Repiterbarhet etter selvrapportert kunnskap

beslutninger. Det blir da et spørsmål om fordeling av goder, og om man er villig til å ofre et lite antall individer som systematisk forfordeles med uriktige beslutninger, mot at populasjonen som helhet nyter godt av en høyere andel riktige beslutninger.

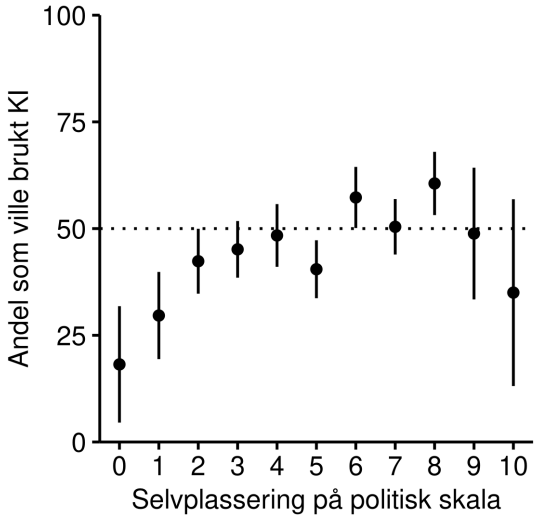

Figure 22: Repiterbarhet etter selvplassering på politiske skala

Ut fra dette perspektivet gir det mening at de som plasserer seg lengst til venstre på den politiske høyre/venstre-skalaen er minst villige til å bruke kunstig intelligens, da man kan anta at borgere som plasserer seg lengst til venstre er mer egalitær enn de som plasserer seg på høyresiden. Vi noterer oss imidlertid også at de som plasserer seg lengst til høyre også er mindre villige til å bruke kunstig intelligens når konsekvensene av bruken presenteres slik som den har blitt gjort i dette konkrete tilfellet. Inntil videre må vi nøye oss med å konstatere at det

er store forskjeller i svarene respondentene gir basert på deres politiske ståsted, og overlate til framtidig forskning å dykke dypere i hva denne forskjellen skyldes.

Hvilken informasjon anses som passende?

Et annet viktig spørsmål knyttet til bruk av maskinlæring og kunstig intelligens er hvilke data det oppfattes som passende å bruke. I mange brukstilfeller vil det finnes et bredt spektrum av informasjon tilgjenglig, men det vil sannsynligvis variere hvor passende innbyggere faktisk mener det er bruke de ulike typene informasjon – uavhengig av om de gjør prediksjonen mer treffsikker. Det er derfor nyttig å ha kunnskap om hvordan befolkning vurderer ulike typer informasjon.

Et realistisk eksempel hvor maskinlæring kan brukes i forvaltninger er hvilke jobbrette tiltak NAV skal tilby en jobbbsøker. Tilgangen til jobbrettede tiltak er behovsbasert og jobbsøkeren har ikke anledning til fritt å velge hvilke tiltak hun eller han ønsker seg. Det er også et begrenset gode der NAV må prioritere. Godt over halvparten av alle jobbsøkere tilbys ingen eller liten bistand fra NAV. Saksbehandler vil på bakgrunn av en individuell vurdering av søkerens behov bestemme innsatsgruppe, og dermed også hvilke tiltak han/hun skal få tilbud om. Tidligere har denne vurderingen blitt gjort av saksbehandleren alene. I dag prøver NAV ut maskinlæring for å bistå saksbehandleren med forslag i denne vurdering.

I prinsippet finnes det et enormt utvalg av mulige variabler som kan være relevant for en slik prediksjon – i den forstand at de kan bidra med å gjøre en prediksjon mer nøyaktig. Disse variablene omhandler et stort og variert utvalg informasjon om den enkelte jobbsøker. Det er derfor et godt eksempel på en situasjon hvor det må gjøres en avveining om hvilke variabler man skal bruke, hvor det er sannsynlig at innbyggere vil oppfatte noen variabler som mer eller mindre passende enn andre. I verste fall kan enkelte variabler bli oppfattet som direkte urettferdige.

For å studere dette spurte vi respondentene hvor passende de synes det er å bruke hver av en liste variabler, med utgangspunkt i at de skal brukes for å foreslå jobbrettede tiltak. I spørsmålet satt vi premisset at hver variabel bidrar med å gjøre forslagene mer nøyaktige. Vi ba dem om å

vurdere hver variabel på en fem-punkts skala, fra "Ikke passende i det hele tatt" (1) til "Svært passende" (5). Vi spurte dem både om variabler som det har vært aktuelt for NAV å bruke ved en eventuell slik implementering og om variabler som det ikke har vært aktuelt å bruke:

- Alder;
- Arbeidshistorikk: Hvorvidt den arbeidssøkende har hatt sammenhengende jobb i 6 av de siste 12 mnd;
- Bosted: Hvor i landet bor brukeren;
- Kjønn;
- Helse: Hvorvidt jobbsøkere opplyser at hen har helseutfordringer;
- Landbakgrunn;
- Rulleblad: Har brukeren blitt dømt for kriminelle handlinger.
- Ufordringer: Hvorvidt jobbsøkere opplyser at hen har andre utfordringer som hindrer dem fra å jobbe;
- Utdanning: Hvorvidt jobbsøkeren har fullført utdanning godkjent i Norge;

Resultatene fra spørsmålet vises i figurene under. Den første figurene viser snittet på skalaen for hver variabel, hvor variablene er rangert nedover etter hvor passende respondentene synes de var i snitt. Den andre figuren viser hele fordelingen for hver variabel. Samlet sett ser vi at ingen av variablene oppfattes som utvilsomt passende eller upassende. De fleste har et gjennomsnitt mellom 3 ("Noe passende") og 4 ("Passende"). Men noen skiller seg ut. På den ene siden skiller kjønn og landsbakgrunn seg spesielt ut ved å ha et gjennomsnitt under 3, substantielt under de andre; på den andre siden skiller utdanning seg ut ved å ha et gjennomsnitt over 4, substantielt over de andre. Hva variasjonen skyldes vet vi ikke med sikkerhet. At kjønn og landbakgrunn blir sett på som mindre passende er gjerne fordi de er mer direkte knyttet til spørsmål og bekymringer om diskriminering. Samlet sett er det tydelig forskjeller i hvor passende ulike typer informasjon ble oppfattet av respondentene.

De to figurene under sammenligner gjennomsnittet for ulike undergrupper. I den første er det differensiert mellom respondenter med og uten høyere utdanning. Vi ser her at respondenter med ulik utdanning har noe ulikt syn på ulike variablene, og da særskilt når det kommer

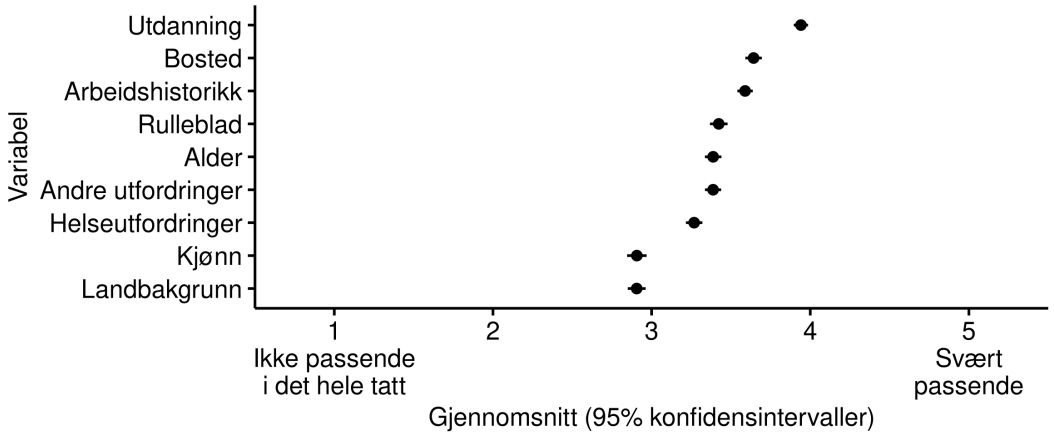

Figure 23: Gjennomsnitt av hvor passende det oppfattes å bruke variabelen

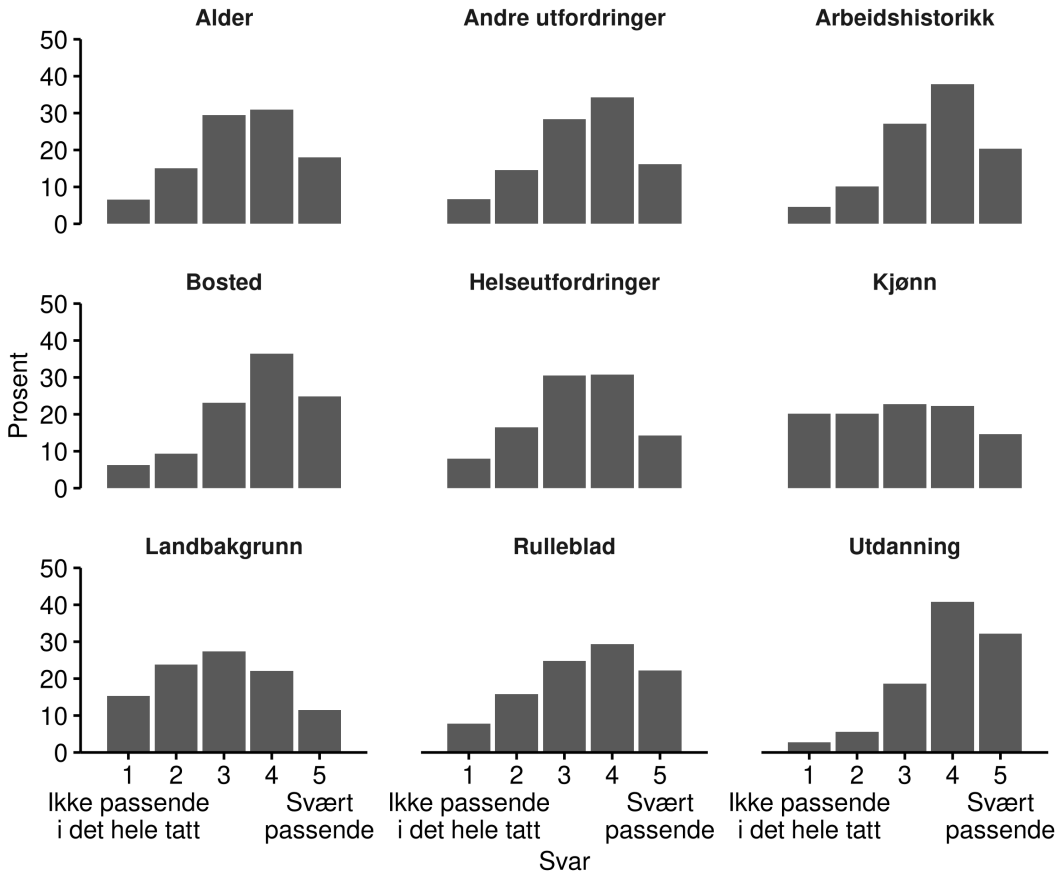

Figure 24: Fordeling for hver variabel av hvor passende den oppfattes å bruke

til nettopp utdanning. De med høyere utdanning synes det er betydelig mer passende enn de uten høyere utdanning å bruke utdanning som variabel. De underliggende årsakene til disse forskjellene kan være mange, men resultatene peker på at grupper med ulik oppfatning, kunnskap, erfaring, etc, kan forholde seg substansielt forskjellig til samme type informasjon.

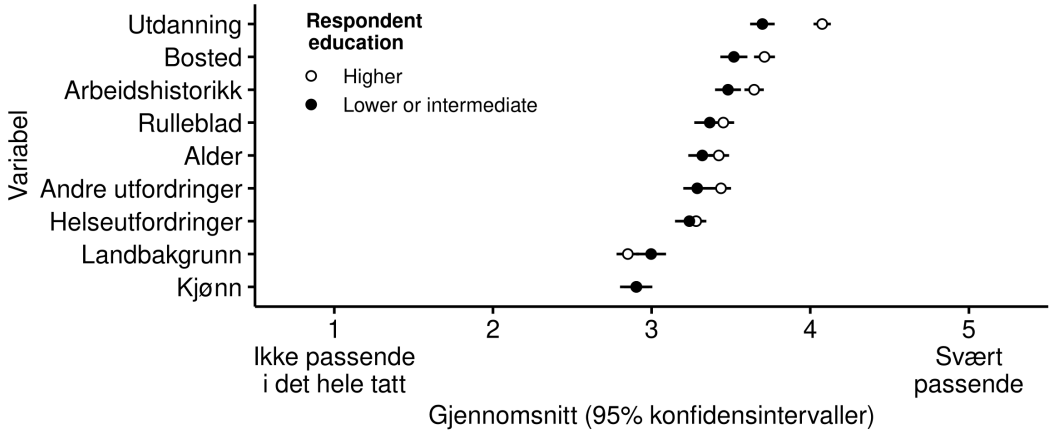

Figure 25: Gjennomsnitt av hvor passende det oppfattes å bruke variabelen for ulike utdanningsnivå

I den neste figuren er det differensiert etter hvor mye selverklært kunnskap respondenene har om maskinlæring. Her ser vi også noen viktige forskjeller: De med mye selverklært kunnskap om maskinlæring synes at de fire variablene på toppen (som generelt blir sett på som mest passende) er substansielt mer passende enn de uten kunnskap. Dette kan være relatert til underliggende sosioøkonomiske forskjeller mellom de med høy og lav selverklært kunnskap, slik som utdanningsnivå, men det kan også være kunnskap i seg selv som endrer denne oppfatningen. En viktig alternativ mekanisme er at de med mye kunnskap om maskinlæring verdsetter økt nøyaktighet mer enn de uten slik kunnskap, som da endrer hvordan de balanserer ulike hensyn når de vurderer hvor passende variablene er.

Sett i helhet antyder resultatene at den generelle befolkning er uenig i hvor passende det er å bruke ulike typer informasjon. Samtidig er det tydelig forskjeller mellom variablene, hvor særskilt kjønn og landbakgrunn blir sett på mindre passende av mange. Resultatene viser også at det er viktige systematiske forskjeller mellom ulike grupper for hvordan de gjør denne vurderingen.

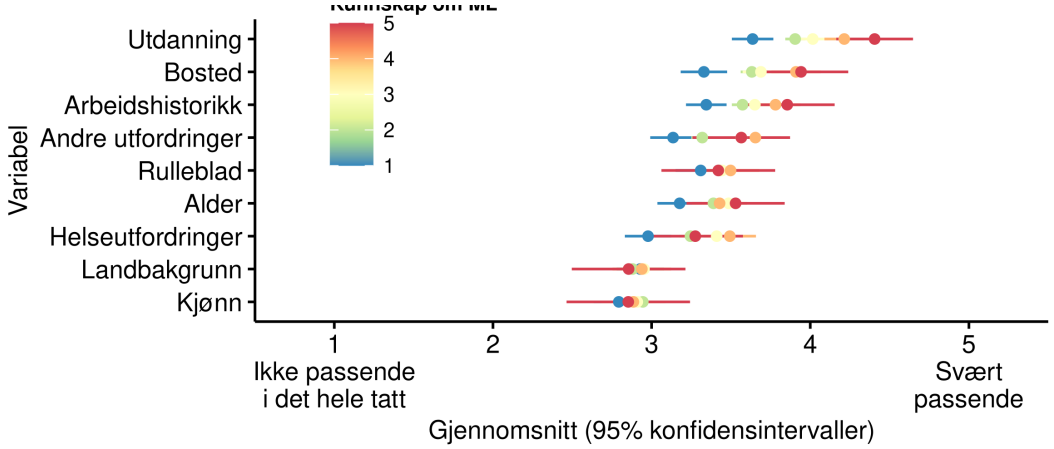

Figure 26: Gjennomsnitt av hvor passende det oppfattes å bruke variabelen for ulike nivå av selvrapportert kjennskap til maskinlæring

Statistisk paritet

Det er viktig å studere rettferdighet fra et statsvitenskapelig perspektiv fordi oppfatninger av rettferdighet antas å påvirke institusjonell legitimitet (Tyler 2003). Dette begrenser seg ikke bare til input-siden av det politiske systemet hvor politikk vedtas, men også output-siden i forvaltningen hvor politikk settes ut i live (Krislov 2012; Rosanvallon 2011; Rothstein 2009).

Flere definisjoner brukes om rettferdighet, og de fleste av dem er basert på forhold mellom sanne/falske positive og sanne/falske negativer (Verma and Rubin 2018). Alexandra Chouldechova (2018) viser teoretisk og empirisk hvordan to velbrukte definisjoner av rettferdighet umulig kan oppnås samtidig i visse tilfeller. Hvilke definisjoner bør prioriteres når man står overfor slike avveininger? Chouldechovas studie viser med tydelighet at det ikke er gjort i en håndvending å lage rettferdige prediksjonsmodeller. Tvert imot er det komplisert utfordring som krever oppmerksomhet om konkret kontekst. Vi vet fortsatt lite om hvilke definisjoner som resonnerer blant innbyggerne, og i hvilken grad de modereres av kontekst eller innbyggernes sosiale bakgrunn eller politiske holdninger. Vi vet fra samfunnsforskning at hva som oppfattes som rettferdig kan variere med sosial identitet og kultur, politiske holdninger, og personlige karaktertrekk, og det er derfor viktig å gjøre konkrete empiriske studier. I dette kapitlet tar vi derfor for oss et realistisk scenario for NAV hvor vi setter opp to motstridende rettferdighetshensyn knyttet til hvilke sykmeldte NAV skal tilby dialogmøte.

Et dialogmøte er en samtale mellom NAV og den sykmeldte som anses som positivt for den

sykmeldtes muligheter for å komme tilbake i arbeid. I prinsippet har alle rett på et dialogmøte, men i praksis foregår det en siling hvor det gjøres en vurdering av hvem som har mest nytte av et slikt møte. Bruk av maskinlæring og kunstig intelligens kan i dette tilfellet bidra til bedre estimater for hvem som er i fare for å bli langtidssykemeldt, og derfor kan ha større nytte av et dialogmøte. Derfor er NAV i innledende stadier på å utvikle maskinlæringsmodeller som predikerer sannsynlighet for at en sykemeldt fortsatt vil være sykemeldt 12 uker fram i tid.

Når man bestemmer innretningen på en modell må man foreta prioriteringer. En prinsipielt viktig prioritering handler om man skal ta i bruk såkalt statistisk paritet som rettferdighetsprinsipp på utvalgte egenskaper ved individene det gjelder. Kjønnsparitet er ett eksempel, men det kan også handle om statistisk paritet etter alder, etnisitet, geografi, med mer. Et kjent eksempel innenfor litteraturen om rettferdig bruk av kunstig intelligens er studien som viser hvordan afro-amerikanske fengselsinnsatte sjeldnere blir tilbudt prøveløslatelse enn hva andelen deres skulle tilsi. Dette skjer når avgjørelsen om prøveløslatelse baserer seg på prediksjonsmodeller om risikoen for at den innsatte blir tatt pånytt for en kriminell handling dersom hen slippes fri (Chouldechova 2017). Å anvende paritetsprinsippet her innebærer å sikre at andelen innsatte som tilbys prøveløslatelse samsvarer med andelen innsatte for hver av de etniske gruppene i fengselet. Fordelen med å bruke statistisk paritet etter etnisitet er at prøveløslatelse blir likt fordelt blant de etniske gruppene, og slik sett kan oppleves som rettferdig fordelt. Utfordringen ved å bruke dette prinsippet er at andre egenskaper ved de innsatte – som for eksempel risikovurderinger om tilbakefall til kriminelle handlinger – blir nedprioritert. Er etnisitet i dette tilfellet så viktig at man bør la det gå på bekostning av risikovurderinger knyttet til tilbakefall?

NAVs tilfelle om dialogmøte er mindre dramatisk enn eksempelet om prøveløslatelse. Samtidig er de prinsipielle problemstillingene de samme. Statistisk paritet innebærer i tilfellet om dialogmøte at modellen sikrer at like mange menn og kvinner skal få tilbud om dialogmøte. Denne prioriteringer vil i så fall gå delvis på bekostning av å prioritere treffsikkerhet med tanke på å invitere de som har størst nytte av et slikt møte.

For å studere respondentenes umiddelbare reaksjoner til et slik etisk dilemma knyttet til rettferdig bruk av kunstig intelligens ber vi respondentene se for seg et valg mellom to alternative

maskinlæringsmodeller for å velge hvem som skal få tilbud om dialogmøte.

Ingen av modellene er perfekte, men de feiler på ulike måter.

Den første modellen er mest *treffsikker*. Det vil si at det totalt sett er flere sykemeldte med behov for dialogmøte som får tilbudet enn tilfellet er for den andre modellen. Samtidig har modellen en bias til fordel for menn, som gjør at det er flere kvinner med behov for dialogmøte som ikke får tilbudet. Andelen som har behov for dialogmøte *uten å få tilbud* er altså større hos kvinner enn menn.

Den andre modellen sikrer statistisk paritet etter kjønn, nemlig at andelen av de sykmeldte som kalles inn til dialogmøte er like stor henholdsvis for kvinner som for menn. Imidlertid er den mindre treffsikker totalt sett, slik at færre som har behov for dialogmøte blir innkalt. Dette gjelder både kvinner og menn.

Hvis det står mellom disse to modellene, hvilken modell synes respondentene virker mest rettferdig? Figuren under viser at et knapt flertall foretrekker en modell som vektlegger statistisk paritet. Det vil si at de ønsker å bruke en modell som sikrer likebehandling av kjønn, selv på bekostning av lavere treffsikkerhet totalt sett.

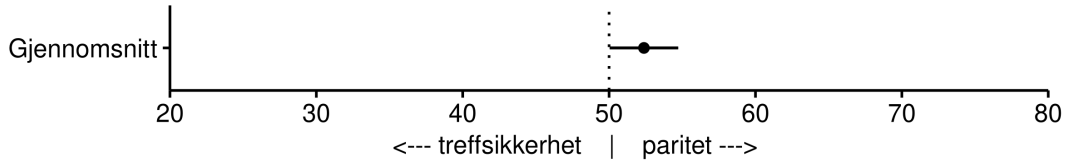

Figure 27: Preferanse for statistisk paritet

I teksten over står det at den mest treffsikre modellen favoriserte menn. For å undersøke om det har noen innvirkning på svarene hvilket kjønn modellen favoriserer veksler vi på denne beskrivelsen. Halvparten av respondentene får vite at modellen har en bias til fordel for menn, mens den andre halvparten av respondentene får vite at modellen favoriserer kvinner. Spiller det noen rolle hvilket kjønn modellen favoriserer? Resultatene viser at det gjør det.

I figuren under ser vi at det er i de tilfeller hvor menn blir fordelaktig behandlet ved bruk av den mest treffsikre modellen at flertallet ønsker å bruke en modell som sikrer likebehandling av kjønn. Det er en signifikant større andel av respondentene som foretrekker paritetsprinsippet

når menn har fordel av den treffsikre modellen enn når kvinner har det.

Hva dette skyldes vet vi ikke. Man ser liknende kjønnseffekter i eksperimenter om politisk representasjon, hvor kvinnelige kandidater jevnt over foretrekkes i noe høyere grad enn mannlige kandidater gjør (Schwarz and Coppock 2018). I den litteraturen pekes det på forklaringer om at folk er motivert ut fra et ønske om å kompensere for historisk underrepresentasjon av kvinner i politiske stillinger. Hvorvidt det ligger liknende strukturelle motivasjoner for våre resultater, psykologiske faktorer, eller andre forhold er et interessant forskningsspørsmål som vi ikke har data til å besvare, og som derfor bør studeres videre.

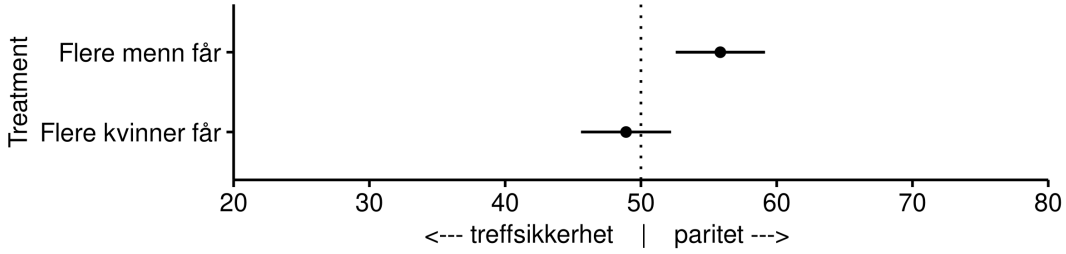

Figure 28: Preferanse for statistisk paritet etter hvilken gruppe som får skjeivt utfall (treatment)

Det vi imidlertid ser, og til forskjell fra litteraturen om politisk representasjon, er at kvinner responderer noe mer på informasjon om hvilket kjønn som kommer best ut av en modell som prioriterer treffsikkerhet. Både menn og kvinner foretrekker paritetsmodellen oftere i de tilfellene kvinnene kommer dårlig ut av treffsikkerhetsmodellen enn i de tilfellene hvor menn kommer dårlig ut av samme modell, men denne effekten er noe sterkere hos kvinner enn menn.

Det er også en generell forskjell blant respondentene i den forstand at kvinner i sterkere grad foretrekker paritetsmodellen enn menn gjør, uavhengig av om det er menn eller kvinner som kommer best ut av det.

Et oppfølgingseksperiment viser hvordan innbyggerne responderer på signaler om hvordan modellene har blitt til. Mens de to modellene i utgangspunktet er tilnærmet like populære blant respondentene, endrer svarene seg markant når vi opplyser om at en ekspertkomité anbefaler den ene modellen framfor den andre. Dette viser med tydelighet at innbyggerne bryr seg om ikke bare hvordan modellene virker, men også hvordan prosessen i forkant har vært.

Eksperter har autoritet i kraft av sin kompetanse som gir legitimitet til bruken av modellene. En

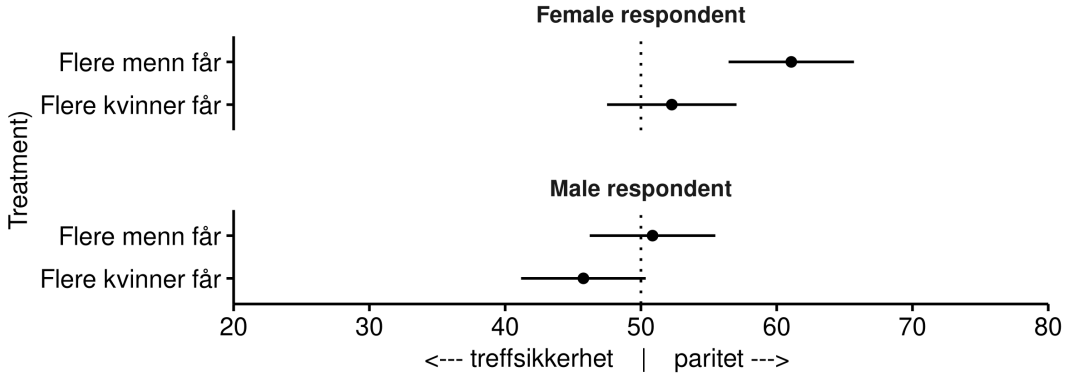

Figure 29: Preferanse for statistisk paritet etter hvilken gruppe som får skjeivt utfall (treatment) og respondentens kjønn

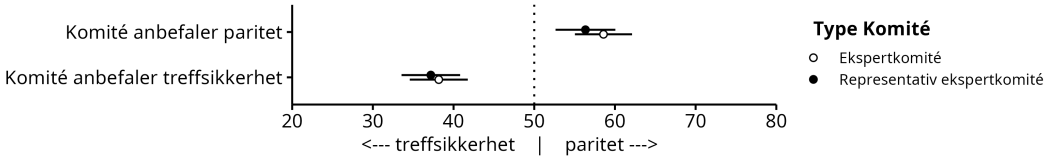

Figure 30: Preferanse for statistisk paritet etter hvilken gruppe som får skjeivt utfall (treatment) og respondentens kjønn

ekspertkomité fungerer som et godkjentstempel – en svanemerking om du vil – på at modellene oppfyller grunnleggende etiske, administrative og politiske krav. At folk responderer på signaler fra eliter er ikke uvanlig i studier av politisk atferd, og spesielt ikke i saker hvor innbyggerne mangler kunnskap eller ikke har sterke oppfatninger på forhånd. Eksperter kan være så mangt, og for å følge opp studiet om byråkratisk representasjon (beskrevet i et annet kapittel) undersøker vi om det spiller noen rolle for innbyggerne at ekspertene gjenspeiler befolkningen, altså er deskriptivt representativ. Denne ekstra informasjonen om ekspertkomiteen gir ingen statistisk signifikant utslag på oppslutningen om modellene. Det betyr enten at innbyggerne ikke er opptatt av representativitet blant eksperter, eventuelt at dette er implisitt antatt slik at eksplisitt informasjon om at ekspertgruppen er representativ ikke tilfører respondentene noen ny informasjon.

Representativt byråkrati

Den dominerende akademiske forståelsen av hva det norske politisk-administrative systemet skal være er den weberianske forestillingen om et byråkrati som ikke skal ha noen selvstendig innflytelse på politikken. Innbyggernes demokratiske innflytelse skjer under utformingen av politikken, mens forvaltningen utfører den vedtatte politikken på en upartisk måte (Rothstein 2009; Rosanvallon 2011). Spørsmål knyttet til politisk representasjon har derfor i hovedsak fokusert på input-siden av det politiske systemet hvor politikk vedtas, heller enn på output-siden hvor politikk gjennomføres.

Et unntak er litteraturen om *representativt byråkrati* (Krislov 2012; Lim 2006), som vektlegger at menneskene som utgjør forvaltningen har en selvstendig påvirkning på hvilken politikk som blir gjennomført. Alle mennesker har systematiske bias som i større eller mindre grad former deres holdninger og atferd. Det er ikke realistisk å anta at saksbehandlere fullt og helt klarer å legge fra seg egne bias i sitt arbeid, selv ikke i profesjoner hvor objektivitet etterstrebes. En måte å utlikne bias er å sørge for at saksbehandlernes bakgrunn reflekterer befolkningen. I det representative byråkratiet skal derfor forvaltningsstaben utgjøre et tverrsnitt av det folket den skal tjene (Lægreid and Olsen 1978; T. Christensen, Lægreid, and Zuna 2001).

Vi har tidligere sett at mange innbyggere tror at saksbehandlere i NAV lar seg påvirke i noen grad av egne holdninger. Med dette som bakteppe er det grunn til å anta at innbyggerne ønsker at saksbehandlerne deler erfaringsbakgrunn med dem selv, slik at de forstår deres situasjon kanskje bedre enn en saksbehandler som har en helt annen bakgrunn. Vi undersøker dette med utgangspunkt i en et design hentet fra en studie om deskriptiv representasjon i politiske beslutningsprosesser (Arnesen and Peters 2018). [1].

Vi spør: Hvilke egenskaper – om noen – ønsker innbyggerne at saksbehandlerne deler med dem?

[1]Deskriptiv representasjon er et viktig konsept innenfor studiet av politisk representasjon, og en av fire former for representasjon slik det ble beskrevet i Hannah Pitkins klassiker *The Concept of Representation* (1967) Deskriptiv representasjon omhandler det å bli representert av kandidater som deler deres sosiale bakgrunn, ikke minst fordi de antar at disse kandidatene deler deres politiske interesser og vil ivareta dem på en god måte. *Byråkratisk representasjon* og *deskriptiv representasjon* er konsepter som i stor grad overlapper hverandre, med unntak av at de har blitt utviklet i forskningstradisjoner som studerer henholdsvis input- og output-siden av det politiske systemet. Vi sidestiller begrepene, og bruker dem om hverandre i denne rapporten

Med vårt fokus på maskinlæring og kunstig intelligens ønsker vi å vite om behovet for representativt byråkrati påvirkes når forvaltningen tar i bruk dette verktøyet. Det er ikke åpenbart på forhånd hvordan det vil slå ut. På den ene siden kan behovet for at saksbehandlerne deler ens sosiale bakgrunn bli mindre viktig, ettersom alle beslutninger blir mer strømlinjeformede og dermed mindre påvirket av saksbehandlernes bakgrunn. På den andre siden kan innbyggerne oppleve at man med denne strømlinjeformingen går glipp av viktige nyanser i hver enkelt avgjørelse, og at det er nettopp i slike situasjoner at man er avhengive av saksbehandlere som har forståelse for innbyggernes situasjon og kan gå inn og korrigere i enkelttilfeller.

Det er gjort lite forskning akkurat på hvordan maskinlæring og kunstig intelligens påvirker innbyggeres preferanser for representativt byråkrati. Ett unntak er en studie av hvite og svarte innbyggere i USA, og deres preferanser for enten videoovervåkning av lyskryss eller å ha politibetjenter til å overvåke lyskrysset for å fange opp bilister som kjører på rødt lys (Miller and Keiser 2021). I deres tilfelle fant man at svarte innbyggere i vesentlig høyere grad foretrakk den automatiserte løsningen med kameraovervåkning heller enn politibetjenter, men kun i de tilfellene hvor politibetjentene var hvite. Denne inngruppeeffekten viser hvordan tillit til myndighetene kan påvirkes av hvilken bakgrunn myndighetspersonene innbyggerne møter har.

I vår studie spør vi altså respondentene rett fram hvor viktig det er for dem med representativt byråkrati, brutt ned på ulike dimensjoner som kan være relevante. Spørsmålsformuleringen er som følger:

> La oss si at du var i en situasjon hvor du måtte søke NAV om økonomisk stønad. Dersom du kunne velge en saksbehandler som skulle ivareta dine interesser hos NAV, hvor viktig tror du at egenskapene under ville vært for denne personen?

Eksperimentdelen av studien innebærer at vi legger til en ekstra setning til halve utvalget hvor vi opplyser om at maskinlæring brukes i saksbehandlingen:

> Som støtte i beslutningsprosessen bruker saksbehandleren kunstig intelligens, basert på maskinlæring, som anbefaler hvem som skal få støtte.

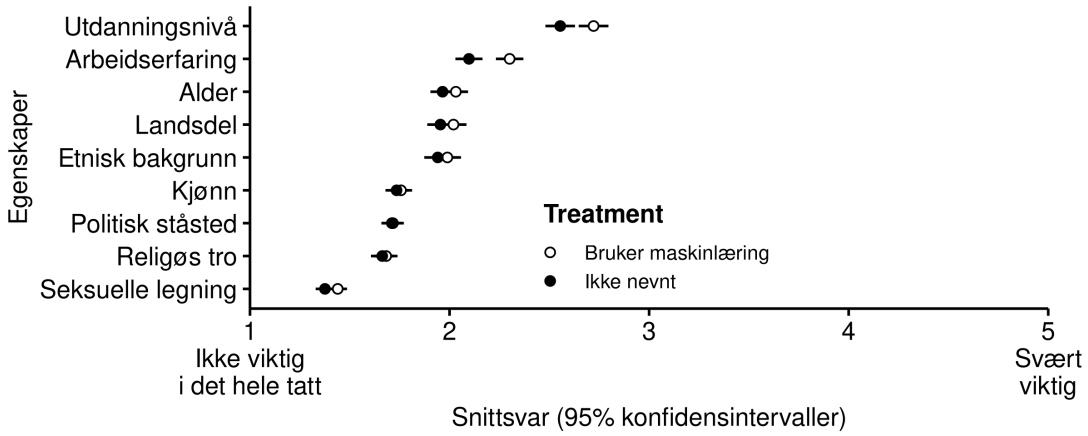

Figure 31: Representajon: Gjennomsnitt av hvor viktig det oppfattes at saksbehandler har samme [egenskap] for å ivareta interesser, delt opp etter hvorvidt saksbehandleren bruker maskinlæring

Resultatene fra eksperimentet viser at folk jevnt over blir mer opptatt av representativt byråkrati når forvaltningen benytter seg av maskinlæring og kunstig intelligens som beslutningsstøtte. Spesielt arbeidserfaring og utdanningsnivå blir viktigere for respondentene. Relevant i denne sammenhengen er et spørsmål vi stilte tilbake i 2018 om man trodde økt grad av automatisering ville gjøre forvaltningen verre eller bedre. Respondente ble bedt om å begrunne svaret i en åpen tekstboks. De som mente det ville lede til dårligere forvaltning trakk typisk fram at beslutningsprosessen ville virke fremmedgjørende på innbyggerne, og at mulighetene for å utvise skjønn ble redusert. En respondent ordla seg slik:

> Jeg tror kunstig intelligens, maskinlæring og annen bruk av teknologi vil gjøre det lettere å ta vanskeligere beslutninger på mange områder. Men det er ikke helt uten ulemper, for eksempel vil det kreve ekspertise hvis man vil undersøke hvilke parametre som ligger bak en beslutning. Og det vil på sikt gi et mindre gjennomsiktig byråkrati. Men totalt sett tror jeg de offentlige tjenestene vil forbedres, som igjen presser fram et følt behov for å ha noen beslutningstakere som kjenner deres situasjon og kan ivareta deres rettigheter og interesser i denne prosessen.

Videre forskning bør fokusere på å forstå mekanismene som gjør at representativt byråkrati blir viktigere for innbyggerne når maskinlæring og kunstig intelligens taes i bruk i forvaltningen, og i hvilke situasjoner innbyggerne er mest opptatt av dette. En hypotese som bør forfølges er da at kunstig intelligens fører til fremmedgjøring av beslutningsprosesser i forvaltningen,

som i sin tur øker behovet for saksbehandlere som deler ens bakgrunn og kan sikre at deres rettigheter ivaretaes på en god måte.

Diskusjon og videre forskning

Bruk av maskinlæring og kunstig intelligens i forvaltningen er per i dag ikke høyt på agendaen hos innbyggere i Norge. Det er mange som oppgir at de har lite kunnskap om temaet, og innbyggerne er delt med tanke på om de er bekymret for utviklingen og om de ser på det som en forbedring eller forverring av forvaltningen. Vi ser også i konkrete norske eksempler at innbyggere har ulike oppfatninger om hvilke som er rettferdig bruk av modeller, noe som peker i samme retning. Erfaringer internasjonalt viser at det ligger konfliktpotensiale ved en rekke aspekter av denne utviklingen, og som det er viktig å være føre var på. Vi finner også i vår studie at innbyggerne oppfatter beslutninger som mindre legitime når man blir konkret på tilfellene hvor maskinlæring og kunstig intelligens har blitt benyttet blant annet i USA. Dette viser at det er grunn til å utvise forsiktighet når det gjelder hvilke områder man benytter seg av modeller basert på maskinlæring og kunstig intelligens. Når det er sagt, er det mange innbyggere som anerkjenner fordelene på et mer generelt plan: Økt automatisering gjør byråkratiet mer effektivt, noe som sparer både tid og penger for samfunnet og enkeltpersoner som er i kontakt med forvaltningen. Dessuten blir det nevnt av flere respondenter at det kan lede til mer likebehandling av beslutninger, siden prosesseringen av data er standardisert og man blir mindre avhengig av den enkelte saksbehandlers bias. Et flertall mener at automatisering gjør saksbehandlingen mer upartisk. Det framstår i denne sammenhengen da kanskje som et paradoks at innbyggerne blir mer opptatt av byråkratisk representasjon i tilfeller hvor maskinlæring og kunstig intelligens blir brukt som beslutningsstøtte for utbetaling av økonomisk stønad. Kanskje handler dette om en oppfattelse av det blir viktigere at saksbehandlerne har tilstrekkelig kjennskap til den enkeltes situasjon, og slik besitter kompetanse til å evaluere og eventuelt overprøve modellene som benyttes i gitte tilfeller.

Problemstillingene kan være komplekse, og av og til kan det være vanskelig for den jevne innbygger å ta stilling til spørsmål som de ikke har tenkt mye over. Samtidig er det nyttig å allerede nå merke seg at befolkningen er delte i mange av spørsmålene om maskinlæring og kunstig intelligens i forvaltningen, både når det gjelder bekymring for bruk og hva som er rettferdig framgangsmåte. Innbyggerne er mer følsomme for spørsmål om bruk av maskin-

læring og kunstig intelligens under omstendigheter hvor bruken knyttes opp mot sosiale bakgrunnsvariabler som ellers i samfunnet er politisk ladete. Også internasjonalt ser vi at bruk av maskinlæring og kunstig intelligens når offentlighetens søkelys i de tilfellene hvor marginaliserte grupper opplever at de blir forskjellsbehandlet.

Det er derfor viktig for NAV og andre myndighetsorganer å ta hensyn til de politiske dimensjonene knyttet til bruk av maskinlæring og kunstig intelligens i forvaltningen. Opplevd urettferdig behandling er aldri tillitsbyggende, men kanskje ekstra skadelig hvis uretten kan tilskrives "kode-diskriminering". Veien er i disse tilfellene kort til å trekke slutninger om systematisk, strukturell urettferdighet mot bestemte sosiale grupper.

Representasjon av interessegrupper og medvirkning i utformingen av modellene er demokratiske verktøy som virker konfliktdempende i andre sammenhenger, og som det er grunn til å anta vil virke også i en overgang til mer automatisert forvaltning. I eksperimentet med statistisk paritet så vi at det hadde en positiv effekt å opplyse respondentene om at modellene som ble brukt hadde blitt anbefalt av en komite som på forhånd hadde vurdert modellene. Det er behov for mer forskning, men vi ser allerede med det vi har lært fra dette prosjektet at det er fornuftig å skynde seg sakte på dette feltet. Under innfasing av maskinlæring og kunstig intelligens som beslutningsstøtte i saker som berører enkeltpersoner bør det være grundige innspillsprosesser slik at innbyggere og berørte parter blir involvert allerede i designfasen og slik på et tidlig stadium kan medvirke til å identifisere etiske dilemma, interessekonflikter, og andre potensielle konfliktsaker som kan oppstå senere.

Denne rapporten presenterer etter det vi kjenner den første studien i norsk sammenheng som involverer den generelle befolkningen i en slik dialog. Den representerer imidlertid bare starten, og må følges opp av videre studier både med tanke på hvilke spørsmål som taes opp og med tanke på innretningen av en slik dialog. Under presenterer vi to tilnærminger til videre opinionsforskning, hvor den første er metodisk og den andre er tematisk.

Deliberativ meningsmåling

Spørreundersøkelser har mange fordeler, ikke minst det at man når et representativt og relativt stort antall innbyggere på kort tid. Samtidig er det behov for å gå mer i dybden, hvor innbyggerne får anledning til å sette seg grundigere inn i det som ofte er komplekse spørsmål. De involverte forskerne i denne rapporten skal derfor gjennomføre en såkalt deliberativ meningsmåling sommeren 2022, hvor et representativt utvalg av innbyggerne i Norge gjennom en hel dag skal diskutere og uttrykke sine meninger om tematikken. Deliberativ meningsmåling viser til en bestemt prosedyre for å invitere innbyggere til diskusjon og meningsutveksling om forhåndsbestemte politiske spørsmål. I korte trekk går prosessen ut på at man inviterer et representativt utvalg typisk på hundre eller flere innbyggere til å sette av en dag for å diskutere politikk med sine medborgere. De tar stilling til forslag, blir presentert for for- og motargumenter, og diskuterer sakene i mindre grupper. Etter diskusjonen får de i en plenumssesjon anledning til å stille spørsmål til fageksperter, før de mot slutten av arrangementet svarer de på en spørreundersøkelse om deres holdninger til de politiske temaene som var på agendaen. Denne typen forskning komplementerer standard spørreundersøkelser ved at man får vite hva folk mener etter de har fått tid til å tenke seg om. Noen av temaene i denne NAV-rapporten vil bli del av den deliberative meningsmålingen.

Saksbehandlernes rolle

Et viktig funn i denne rapporten er at de med mye (selverklært) kunnskap om maskinlæring tenderer til å ha andre oppfattinger enn de uten mye kunnskap. Slike forskjeller i oppfatning fører til at borgere har ulik oppfatning av byråkratiske prosesser, men gjelder dette også for saksbehandlerne i NAV?

Saksbehandlere – og byråkrater generelt – har spesiell kunnskap, erfaring, og ekspertise som skiller dem fra den generelle befolkningen. Denne ekspertisen gjør at saksbehandlere har bedre forutsetninger enn befolkningen generelt til å fatte NAV-relaterte beslutninger, men også at de sannsynligvis vektlegger andre hensyn i utredninger og vurdering enn hva personer uten slik ekspertise ville gjort. Det kan gi utslag når det kommer til hvilke kriterier de legger til grunn for

å oppfatte implementering og bruk av maskinlæring og kunstig intelligens i deres daglige virke som legitimt og rettferdig. Det kan handle både om når det burde brukes og hvor mye man burde vektlegge maskinanbefalinger der hvor det er tilgjengelig. Selv om man kan forvente at saksbehandlere i stor grad følger samme holdningsmønster, er det sannsynlig at det finnes spesifikke tilfeller hvor det ikke er samsvar i holdninger. Dette kalles for kongruens (Golder and Stramski 2010), og mer spesifikt prosedyrekongruens for samsvar i holdninger knyttet til beslutningsprosesser (Broderstad 2022). Hvis det finnes tilfeller av inkongruens på tvers av saksområder, kan det påvirke tilliten til NAV og forvaltning som helhet negativt. Eventuelle systematiske forskjeller vil belyse fallgruver hvor implementering eller bruk som oppfattes som legitim av forvaltningen vil oppfattes som illegitim av den generelle befolkningen, og som derfor utilsiktet kan føre til redusert systemtillit og tillit til NAV. Framtidig forskning bør derfor søke å identifisere likheter og forskjeller i holdninger til maskinlæring og kunstig intelligens for henholdsvis saksbehandlere og den generelle befolkningen.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks." *ProPublica*. https://www.propublica.org/article/machine-bias-riskassessments-in-criminal-sentencing.

Arnesen, Sveinung. 2017. "Legitimacy from Decision-Making Influence and Outcome Favourability: Results from General Population Survey Experiments." *Political Studies* 65 (1_suppl): 146–61.

Arnesen, Sveinung, and Yvette Peters. 2018. "The Legitimacy of Representation: How Descriptive, Formal, and Responsiveness Representation Affect the Acceptability of Political Decisions." *Comparative Political Studies* 51 (7): 868–99.

Barocas, Solon, and Andrew D Selbst. 2016. "Big Data's Disparate Impact."

Binns, Reuben, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. "'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions." In *Proceedings of the 2018 Chi Conference on Human Factors in Computing Systems*, 1–14.

Broderstad, Troy Saghaug. 2022. "Democratic Reflections: To What Extent Do Representatives Mirror Their Constituents, and How Does It Affect the Challenges Modern, Representative Democracy Are Facing?" The University of Bergen. https://hdl.handle.net/11250/2839299.

Chouldechova, Alexandra. 2017. "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments." *Big Data* 5 (2): 153–63.

Chouldechova, Alexandra, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan. 2018. "A Case Study of Algorithm-Assisted Decision Making in Child Maltreatment Hotline Screening Decisions." In *Conference on Fairness, Accountability and Transparency*, 134–48.

Christensen, Henrik Serup, Staffan Himmelroos, and Maija Setälä. 2020. "A Matter of Life or Death: A Survey Experiment on the Perceived Legitimacy of Political Decision-Making on Euthanasia." *Parliamentary Affairs* 73 (3): 627–50.

Christensen, Tom, Per Lægreid, and Hans Robert Zuna. 2001. "Profesjoner i Regjeringsapparatet 1976-1996." *Makt- Og Demokratiutredningens Rapportserie*.

Clayton, Amanda, Diana Z O'Brien, and Jennifer M Piscopo. 2019. "All Male Panels? Represen-

tation and Democratic Legitimacy." *American Journal of Political Science* 63 (1): 113–29.

De Fine Licht, Jenny, Daniel Naurin, Peter Esaiasson, and Mikael Gilljam. 2014. "When Does Transparency Generate Legitimacy? Experimenting on a Context-Bound Relationship." *Governance* 27 (1): 111–34.

Döring, Matthias. 2021. "How-to Bureaucracy: A Concept of Citizens' Administrative Literacy." *Administration & Society* 53 (8): 1155–77.

Duwe, Grant, and Michael Rocque. 2017. "Effects of Automating Recidivism Risk Assessment on Reliability, Predictive Validity, and Return on Investment (ROI)." *Criminology & Public Policy* 16 (1): 235–69.

Easton, David. 1965. "A Systems Analysis of Political Life."

Esaiasson, Peter, Mikael Gilljam, and Mikael Persson. 2012. "Which Decision-Making Arrangements Generate the Strongest Legitimacy Beliefs? Evidence from a Randomised Field Experiment." *European Journal of Political Research* 51 (6): 785–808.

Esaiasson, Peter, Mikael Persson, Mikael Gilljam, and Torun Lindholm. 2016. "Reconsidering the Role of Procedures for Decision Acceptance." *British Journal of Political Science*, 1–24.

Fine Licht, Karl de, and Jenny de Fine Licht. 2020. "Artificial Intelligence, Transparency, and Public Decision-Making." *AI & Society* 35 (4): 917–26.

Golder, Matt, and Jacek Stramski. 2010. "Ideological Congruence and Electoral Institutions." *American Journal of Political Science* 54 (1): 90–106.

Gordon, Laura Kramer. 1975. "Bureaucratic Competence and Success in Dealing with Public Bureaucracies." *Social Problems* 23 (2): 197–208.

Hansen, Hans-Tore, Kjetil Lundberg, and Liv Johanne Syltevik. 2018. "Digitalization, Street-Level Bureaucracy and Welfare Users' Experiences." *Social Policy & Administration* 52 (1): 67–90.

Krislov, Samuel. 2012. *Representative Bureaucracy*. Quid Pro Books.

Lægreid, Per, and Johan P Olsen. 1978. "Byråkrati Og Beslutninger (Bureaucracy and Decisions)." *Bergen: Universitetsforlaget*.

Lim, Hong-Hai. 2006. "Representative Bureaucracy: Rethinking Substantive Effects and Active Representation." *Public Administration Review* 66 (2): 193–204.

Lind, E Allan, and Tom R Tyler. 1988. *The Social Psychology of Procedural Justice*. Springer Science & Business Media.

Miller, Susan M, and Lael R Keiser. 2021. "Representative Bureaucracy and Attitudes Toward Automated Decision Making." *Journal of Public Administration Research and Theory* 31 (1): 150–65.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pitkin, Hanna Fenichel. 1967. *The Concept of Representation*. University of California Press.

Rosanvallon, Pierre. 2011. *Democratic Legitimacy*. Princeton University Press.

Rothstein, Bo. 2009. "Creating Political Legitimacy: Electoral Democracy Versus Quality of Government." *American Behavioral Scientist* 53 (3): 311–30.

Schwarz, Susanne, and Alexander Coppock. 2018. "What Have We Learned about Gender from Candidate Choice Experiments? A Meta-Analysis of 67 Factorial Survey Experiments."

Tyler, Tom R. 2003. "Procedural Justice, Legitimacy, and the Effective Rule of Law." *Crime and Justice* 30: 283–357.

———. 2021. *Why People Obey the Law*. Princeton university press.

Verma, Sahil, and Julia Rubin. 2018. "Fairness Definitions Explained." In *2018 Ieee/Acm International Workshop on Software Fairness (Fairware)*, 1–7. IEEE.

Weber, Max. 2009. *The Theory of Social and Economic Organization*. Simon; Schuster.

Zarsky, Tal. 2016. "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making." *Science, Technology, & Human Values* 41 (1): 118–32.