

Demokratiske algoritmer

Legitimitet, rettferdighet, og representasjon: 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

Om rapporten	5
Forfattere	5
Referansegruppe	5
Finansiering	5
Sitering	5
Utvidet sammendrag	6
1 Bakgrunn og motivasjon	9
1.1 Byråkratisk omveltning	9
1.2 Beslutningsprosesser, rettferdighet og legitimitet	10
2 Data og metode	13
2.1 Norsk medborgerpanel	13
3 Innbyggernes relasjon til NAV	15
3.1 Erfaring med og kjennskap til NAV	15
3.2 Tillit	16
3.3 Byråkratisk kompetanse	17
3.4 Tiltro til likebehandling i NAV	21
4 Innbyggernes oppfatninger om maskinlæring og KI	24
4.1 Kunnskap om maskinlæring og kunstig intelligens	24
4.2 Bekymring	25
4.3 Blir interesser bedre ivaretatt med maskinlæring?	27
4.4 Forventinger til automatisering i offentlig sektor	28
5 Rettferdighetsoppfatninger	31
5.1 Når er det passende å bruke kunstig intelligens?	31
5.2 Hvilken informasjon anses som passende?	33
5.3 Statistisk paritet	37
6 Representativt byråkrati	42
7 Diskusjon og videre forskning	46
7.1 Deliberativ meningsmåling	47
7.2 Saksbehandlernes rolle	47

Figurliste

1	Antatt årsakssammenheng	11
2	Personlig kontakt med NAV	16
3	Selvrapportert kunnskap om NAV	16
4	Tillit til NAV	16
5	Tillit til NAV for ulike nivåer av selvrapportert kjennskap	17
6	Tillit til NAV for de som har eller ikke har hatt personlig kontakt med NAV	17
7	Byråkratisk kompetanse	19
8	Byråkratisk kompetanse covariater	20
9	Byråkratisk kompetanse for ulike nivåer av tillit til NAV	21
10	Fordeling av tilltro til likebehandling i NAV på tvers av saksområder	22
11	Gjennomsnitt av tilltro til likebehandling for ulike saksområder	22
12	Tilltro til likebehandling i NAV for ulike nivåer av byråkratisk kompetanse	23
13	Selvrapportert kunnskap om maskinlæring og kunstig intelligens	25
14	Bekymring for maskinlæring	26
15	Bekymring for maskinlæring for ulike nivåer av selvrapportert kunnskap om maskinlæring	26
16	Forventninger om interesser ivaretas bedre med maskinlæring i NAV	27
17	Forventninger om interesser ivaretas bedre med maskinlæring i NAV for ulike nivåer av selvrapportert kunnskap om maskinlæring	28
18	Generelle forventninger til automatisering i offentlig sektor	29
19	Spesifikke forventninger til automatisering i offentlig sektor	29
20	Effekt av å få vite konkrete eksempler på forventninger til automatisering i offentlig sektor	30
21	Repiterbarhet etter selvrapportert kunnskap	32
22	Repiterbarhet etter selv plassering på politiske skala	32
23	Gjennomsnitt av hvor passende det oppfattes å bruke variabelen	35
24	Fordeling for hver variabel av hvor passende den oppfattes å bruke	35
25	Gjennomsnitt av hvor passende det oppfattes å bruke variabelen for ulike utdanningsnivå	36
26	Gjennomsnitt av hvor passende det oppfattes å bruke variabelen for ulike nivå av selvrapportert kjennskap til maskinlæring	37
27	Preferanse for statistisk paritet	39
28	Preferanse for statistisk paritet etter hvilken gruppe som får skjeivt utfall (treatment)	40

29	Preferanse for statistisk paritet etter hvilken gruppe som får skjeivt utfall (treatment) og respondentens kjønn	41
30	Preferanse for statistisk paritet etter hvilken gruppe som får skjeivt utfall (treatment) og respondentens kjønn	41
31	Representasjon: Gjennomsnitt av hvor viktig det oppfattes at saksbehandler har samme [egenskap] for å ivareta interesser, delt opp etter hvorvidt saksbehandleren bruker maskinlæring	44

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.

Sitering

Ved referering til rapporten, vennligst bruk følgende sitering: Arnesen, Sveinung, Mikael P. Johannesson (2022): Demokratiske algoritmer. Norce-rapport #..., Bergen, mars 2022. Url: <https://bookdown.org/sveinungarnesen78/NAV/>.

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 ivaretagelsen 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 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. Kunstig intelligens kan brukes både i automatiserte beslutningsprosesser, og som beslutningsstøtte for saksbehandlere. 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 maskinlæring og kunstig intelligens kan benyttes i NAV er blant annet for å beregne sannsynlighet for at den arbeidsledige trenger bistand fra NAV; til å bestemme om en person i sykefravær skal kalles inn til oppfølgingsmøte fra 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 maskinlæring og kunstig intelligens i saksbehandlingen. Datagrunnlaget for denne rapporten er samlet inn i Norsk medborgerpanels i 2021, med et representativt utvalg av innbyggerne i Norge på 2000 respondenter. 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. 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 organisasjonen. 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 har rimelig høy tillit til NAV. Tilliten er 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.

Vi finner videre økt støtte for *representativt byråkrati* når forvaltningen tar i bruk maskinlæring og kunstig intelligens. 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 maskinlæring og kunstig intelligens 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. Det er i alle fall klart at maskinlæring og kunstig intelligens er fremmed for et flertall av innbyggerne i Norge: Mer enn seks av ti innbyggerne i Norge har liten eller ingen kjennskap til temaet.

Innbyggerne er delt i oppfatningen om bruken av maskinlæring og kunstig intelligens i forvaltningen fobedrer tjenestene, og om bruken er noe å bekymre seg over. De som oppfatter at de har god kunnskap om tematikken er mer positive til at det brukes i forvaltningen. Innbyggere som plasserer seg mot midten av det politiske spekteret er mer positive enn de som plasserer seg til mot endene.

Et viktig spørsmål knyttet til bruk av maskinlæring og kunstig intelligens 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 tilhørighet. Den konkrete saken gjelder bruk av maskinlæring og kunstig intelligens 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 er 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 maskinlæring og kunstig intelligens er hvilke data det oppfattes som passende å bruke. I mange brukstilfeller vil det finnes et bredt spektrum av informasjon tilgjengelig, 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. Vi har derfor spurt innbyggere hvor passende de synes en rekke variabler er i et konkret eksempel. Eksempelet vi trekker fram gjelder å bruke maskinlæring og kunstig intelligens 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 så 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 maskinlæring og kunstig intelligens. Vi er fortsatt i en tidlig fase, hvor de fleste innbyggerne har liten kjennskap til tematikken, og hvor svært få har personlige erfaringer med saksbehandling hvor dette benyttes. 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 maskinlæring 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 viktig å ta hensyn til de politiske dimensjonene knyttet til bruk av maskinlæring og kunstig intelligens i forvaltningen. Representasjon av interessegrupper, medvirkning i utformingen av modellene, og politisk ansvarliggjøring er demokratiske verktøy som virker konflikt-dempende 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. Allerede i dag jobber NAV på en måte som åpner for innspill fra flere hold når maskinlæringsmodellene utvikles, og det er grunn til å anta at dette styrker legitimiteten til bruken av dem. Vi anbefaler å videreutvikle innspillsprosessen 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.

1 Bakgrunn og motivasjon

1.1 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 ivaretagelsen 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 stordateknologi 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 et al., 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. Mulige områder hvor maskinlæring og kunstig intelligens kan benyttes i NAV er blant annet for å beregne sannsynlighet for at den arbeidsledige trenger bistand fra NAV; til å bestemme om en person i sykefravær skal kalles inn til oppfølgingsmøte fra 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.

1.2 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

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 et al., 2012; Arnesen, 2017; Christensen et al., 2020), hvem beslutningstakerne er, og hvor godt disse beslutningstakerne gjenspeiler befolkningen med tanke på sosial bakgrunn (Arnesen and Peters, 2018; Clayton et al., 2019). Videre er et gjennomgående funn at dersom utfallet går imot ens egne ønsker, blir prosessen diskreditert (Esaiasson et al., 2016).



11

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 “algoritmestyrte 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 overordnede spørsmål om maskinlæring og kunstig intelligens i forvaltningen.

2 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 nøkkelementer. 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.

2.1 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 kodefor 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).

3 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 ivare tatt 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.

3.1 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.

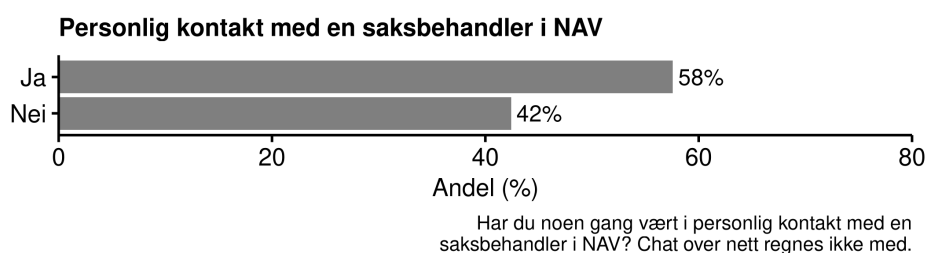


Figure 2: Personlig kontakt med NAV

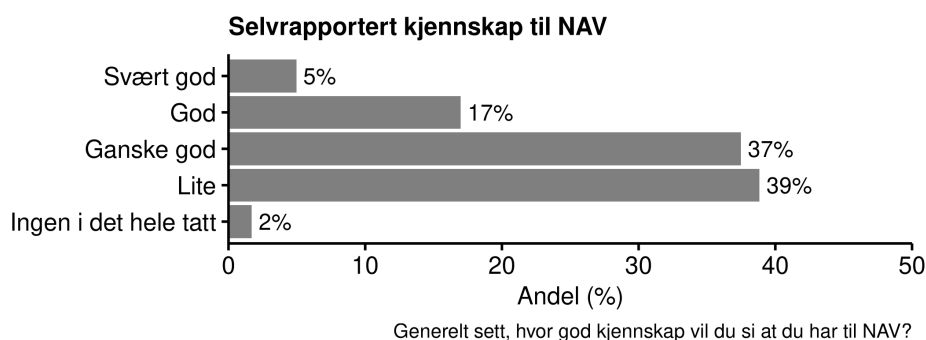


Figure 3: Selvrapportert kunnskap om NAV

3.2 Tillit

Tillit er noe som tar lang tid å bygge opp, men som fort kan rives ned. Norge er kjent som et land der myndighetene nyter høy tillit i befolkningen. Dette bekreftes også i vår undersøkelse for NAV sin del. 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 er et godt utgangspunkt at NAV nyter tillit i befolkningen, og ikke desto viktigere at denne relasjonen ivaretaes parallelt med at organisasjonen endrer seg og utvikler morgendagens forvaltning. Hvordan NAV ligger i forhold til andre forvaltningsorganer har vi ikke data på.

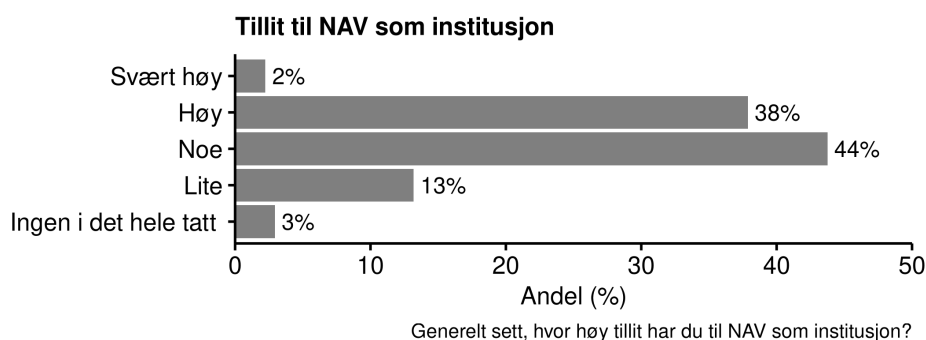


Figure 4: Tillit til NAV

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 er har noe lavere tillit. Forskjeller er relativt liten men statistisk signifikant.

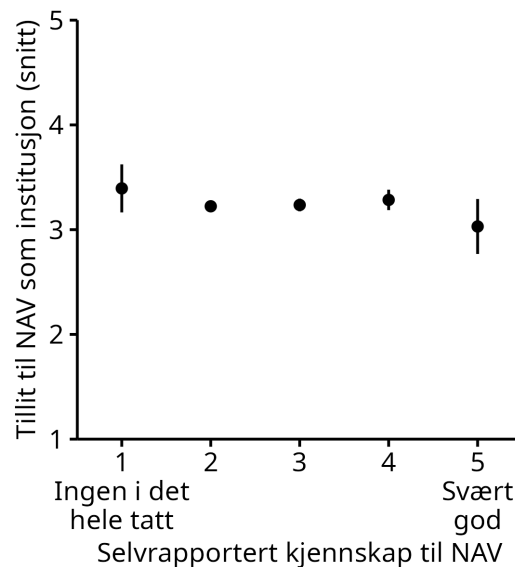


Figure 5: Tillit til NAV for ulike nivåer av selvrapportert kjennskap

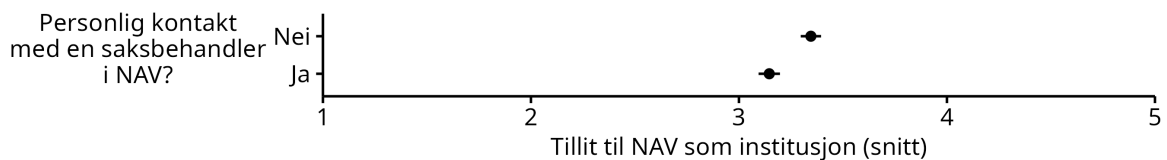


Figure 6: Tillit til NAV for de som har eller ikke har hatt personlig kontakt med NAV

3.3 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* - intern og ekstern. 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.

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).

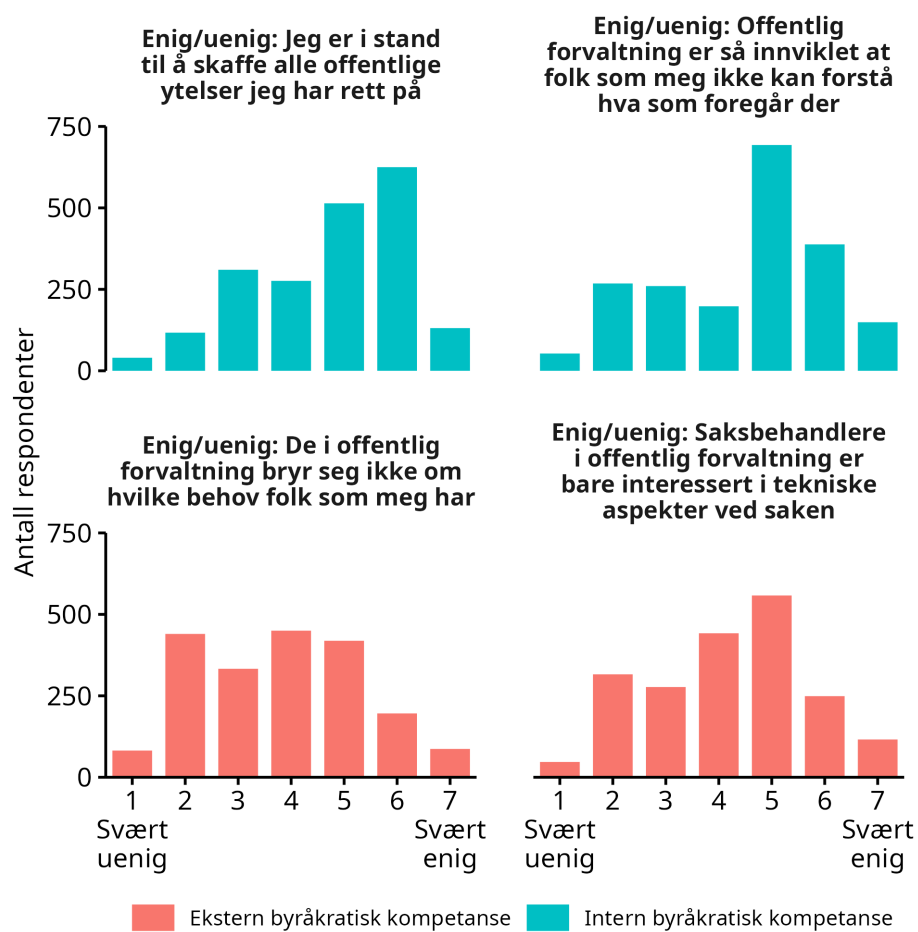


Figure 7: Byråkratisk kompetanse

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.

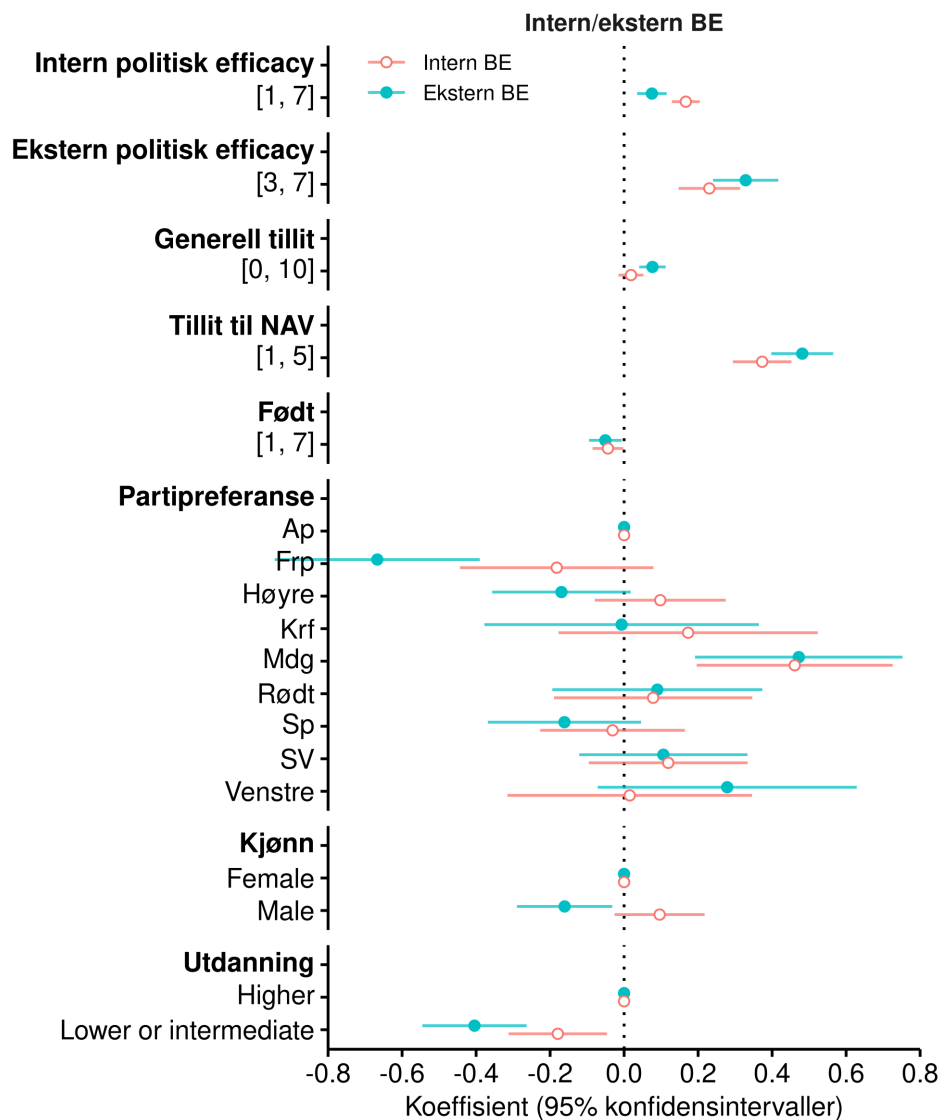


Figure 8: Byråkratisk kompetanse covariater

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

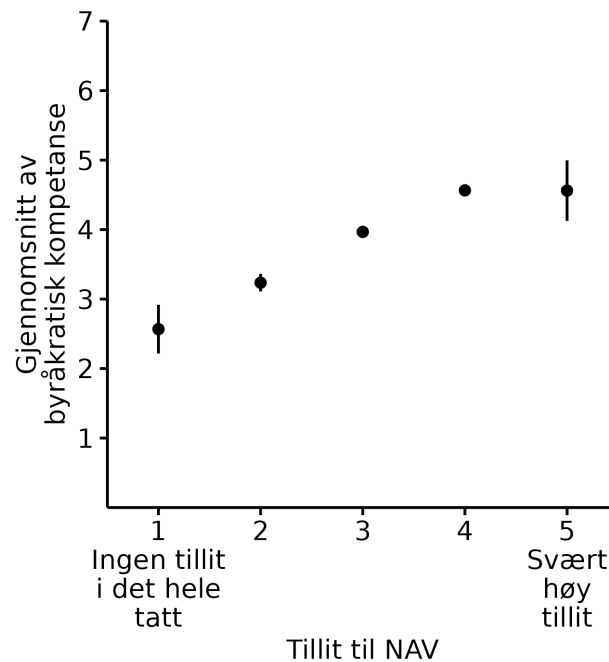


Figure 9: Byråkratisk kompetanse for ulike nivåer av tillit til NAV

3.4 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.

Figuren under viser også at byråkratisk kompetanse samvarierer med oppfatningen om at

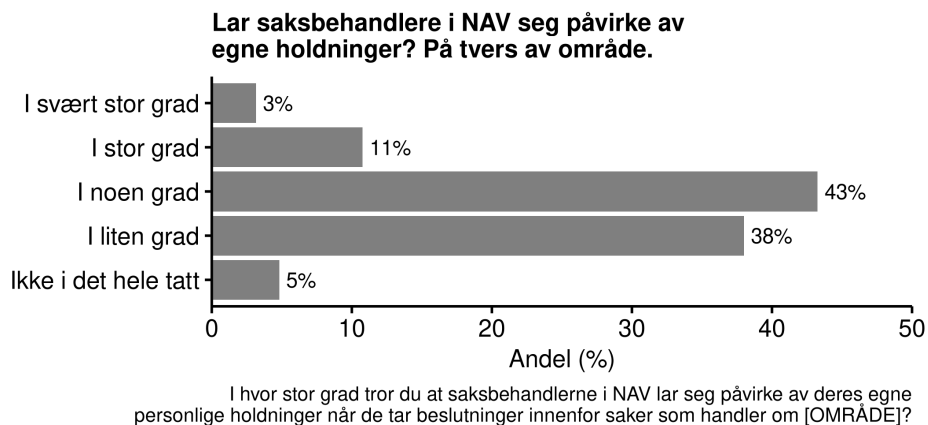


Figure 10: Fordeling av tilltro til likebehandling i NAV på tvers av saksområder

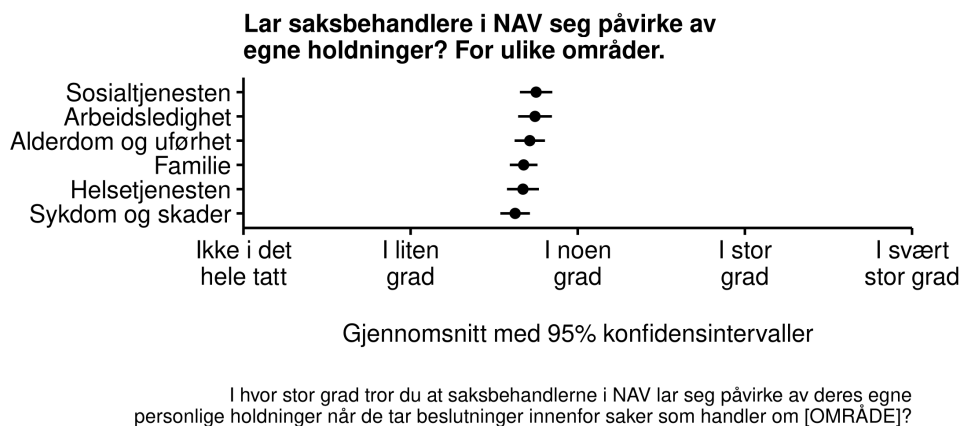


Figure 11: Gjennomsnitt av tilltro til likebehandling for ulike saksområder

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.

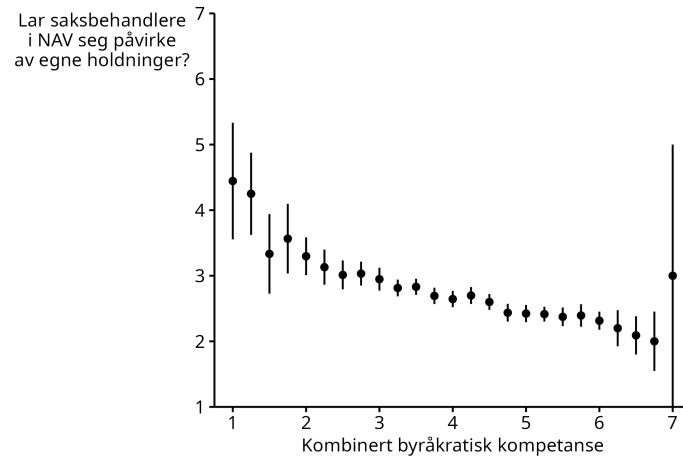


Figure 12: Tilltro til likebehandling i NAV for ulike nivåer av byråkratisk kompetanse

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.

4 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 en 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 selv plassering 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.

4.1 Kunnskap om maskinlæring og kunstig intelligens

Kunnskap rundt maskinlæring, kunstig intelligens, algoritmer, etc., utgjør 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 så 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 of-

fentlige 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 tredjedeler sier at de har liten eller ingen kjennskap til det i det hele tatt.

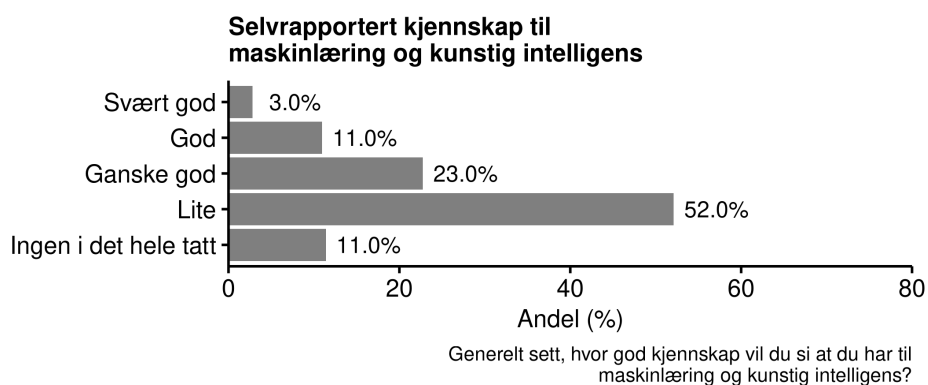


Figure 13: Selvrapportert kunnskap om maskinlæring og kunstig intelligens

4.2 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.

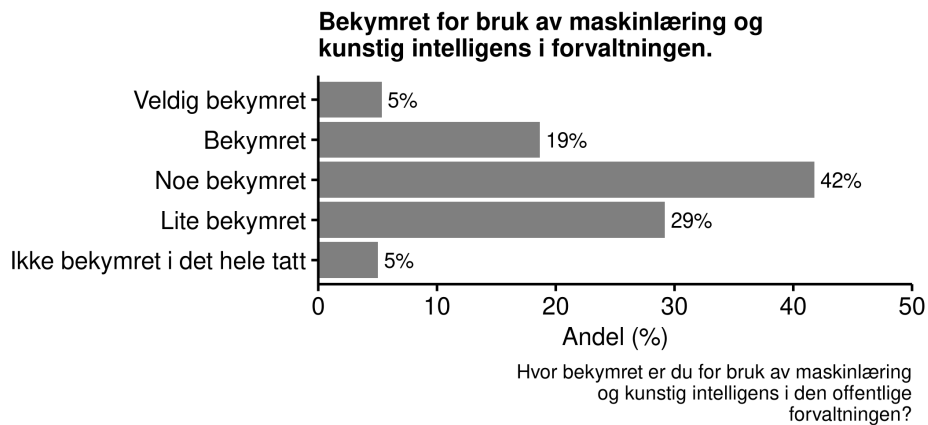


Figure 14: Bekymring for maskinlæring

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.

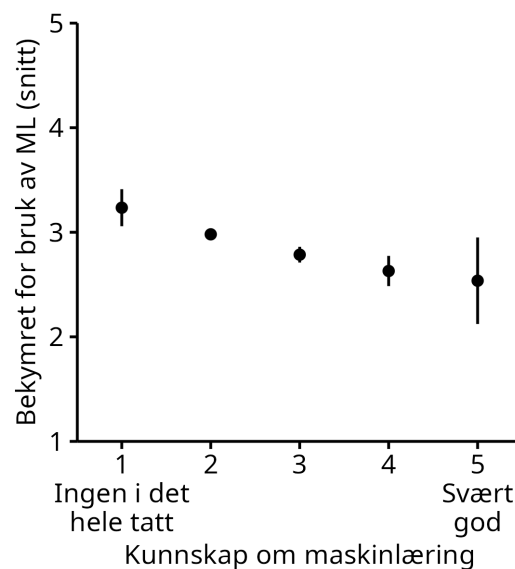


Figure 15: Bekymring for maskinlæring for ulike nivåer av selvrapportert kunnskap om maskinlæring

4.3 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.

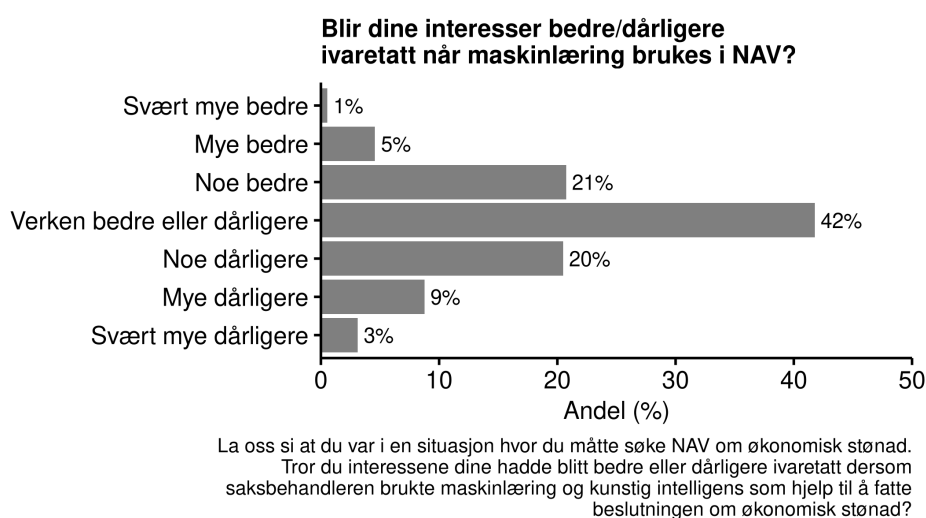


Figure 16: Forventninger om interesser ivaretas bedre med maskinlæring i NAV

Den neste figuren viser hvordan dette varierer 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.

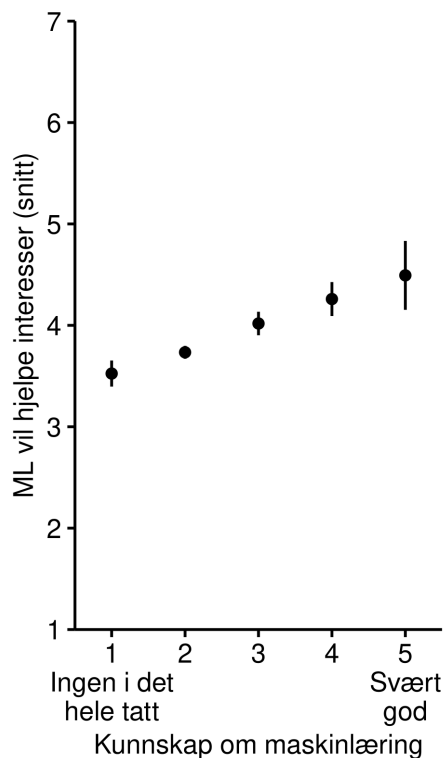


Figure 17: Forventninger om interesser ivaretas bedre med maskinlæring i NAV for ulike nivåer av selvrapportert kunnskap om maskinlæring

4.4 Forventinger til automatisering i offentlig sektor

Vi har studert befolkningens forventninger til automatisering i offentlig sektor ved en tidligere spørreundersøkelse. Resultatene under er basert på spørsmål stilt i 2018 (runde 13) i Norsk Medborgerpanel. 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 konkrete eksempler. Deretter ble de spurt om 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 konkrete eksempler. Her ser vi at respondentene hadde delvis ulike forventning for de ulike punktene. De er relativt jevnfordelt når det gjelder generelle forventninger om det vil føre til forbedring eller forverring av offentlig tjenester. Samtidig tror majoriten av innbyggere at det vil føre til mindre etterprøvbarhet og legitimitet, men mer upartiskhet.

Resultat: Folk har forventninger, og gjennomsnittet ligger like under 4.

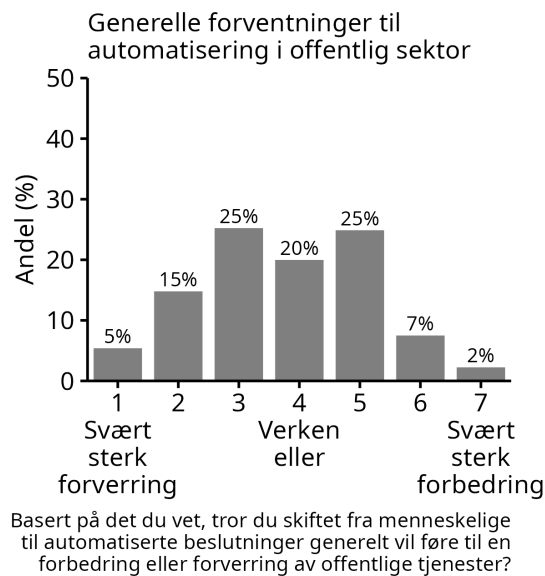


Figure 18: Generelle forventninger til automatisering i offentlig sektor

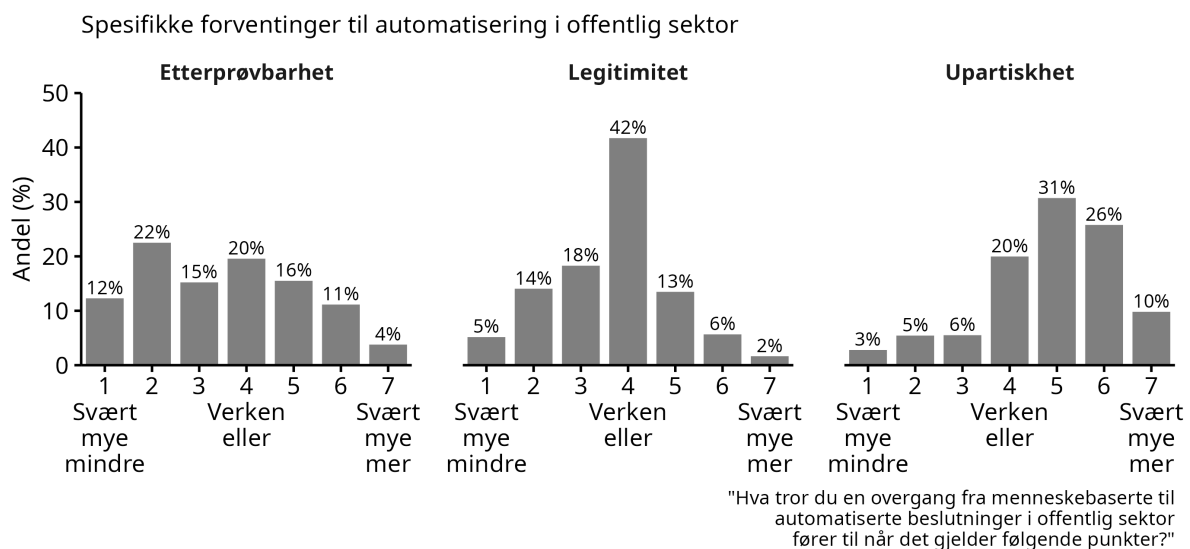


Figure 19: Spesifikke forventninger til automatisering i offentlig sektor

Figuren under viser gjennomsnittet delt opp etter hvorvidt de fikk oppgitt konkrete eksempler i spørsmålet om generelle forventninger. 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 eksempler. Effekten er svak men statistisk signifikant.

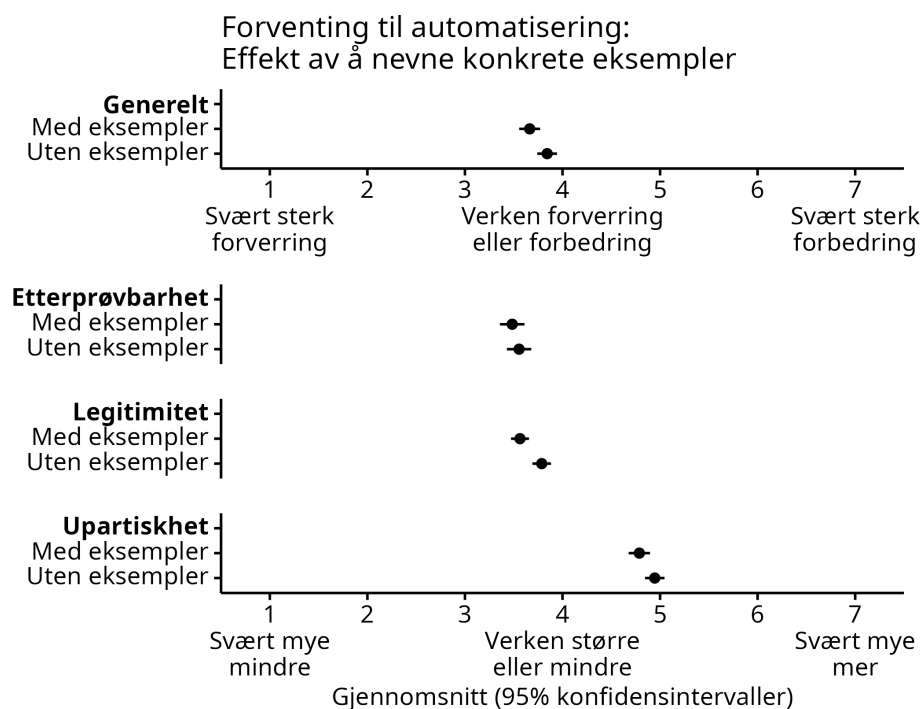


Figure 20: Effekt av å få vite konkrete eksempler på forventninger til automatisering i offentlig sektor

5 Rettferdighetsoppfatninger

5.1 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 spesifikke 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

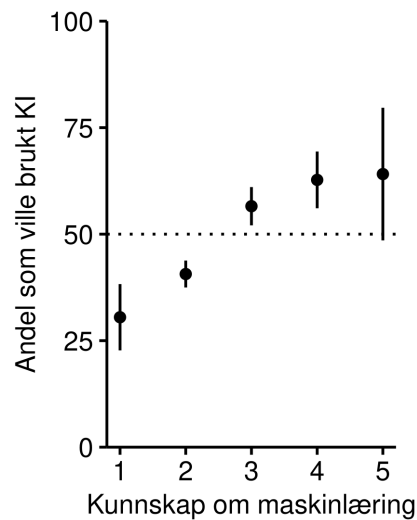


Figure 21: Repiterbarhet etter selvrapportert kunnskap

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

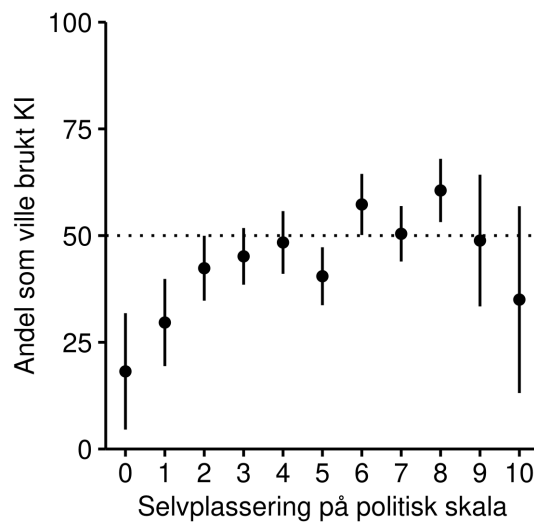


Figure 22: Repiterbarhet etter selvplassering på politiske skala

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.

5.2 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 tilgjengelig, 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 jobbbrette tiltak NAV skal tilby en jobbsøker. Tilgangen til jobbbrettede tiltak er behovsbasert og jobbsøkeren har ikke anledning til fritt å velge hvilke tiltak hun eller han ønsker seg. 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 urettferdig.

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 jobbbrettede 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 landsbakgrunn 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

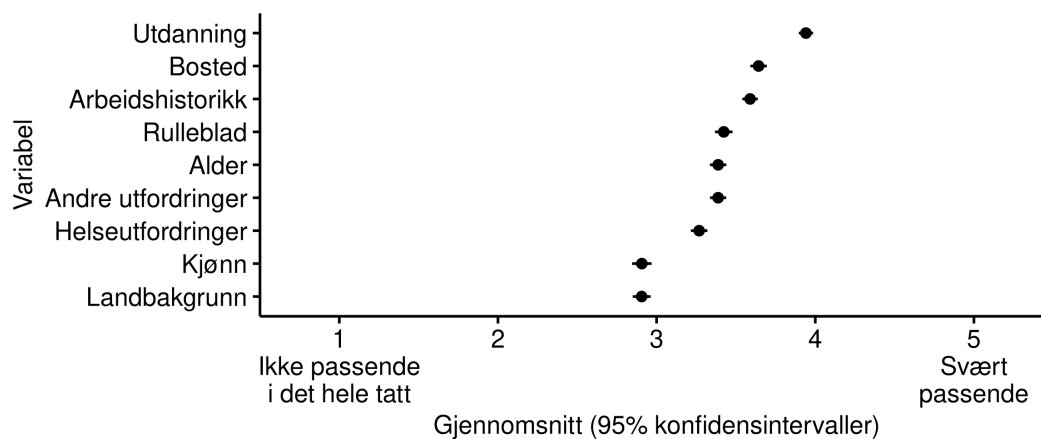


Figure 23: Gjennomsnitt av hvor passende det oppfattes å bruke variabelen

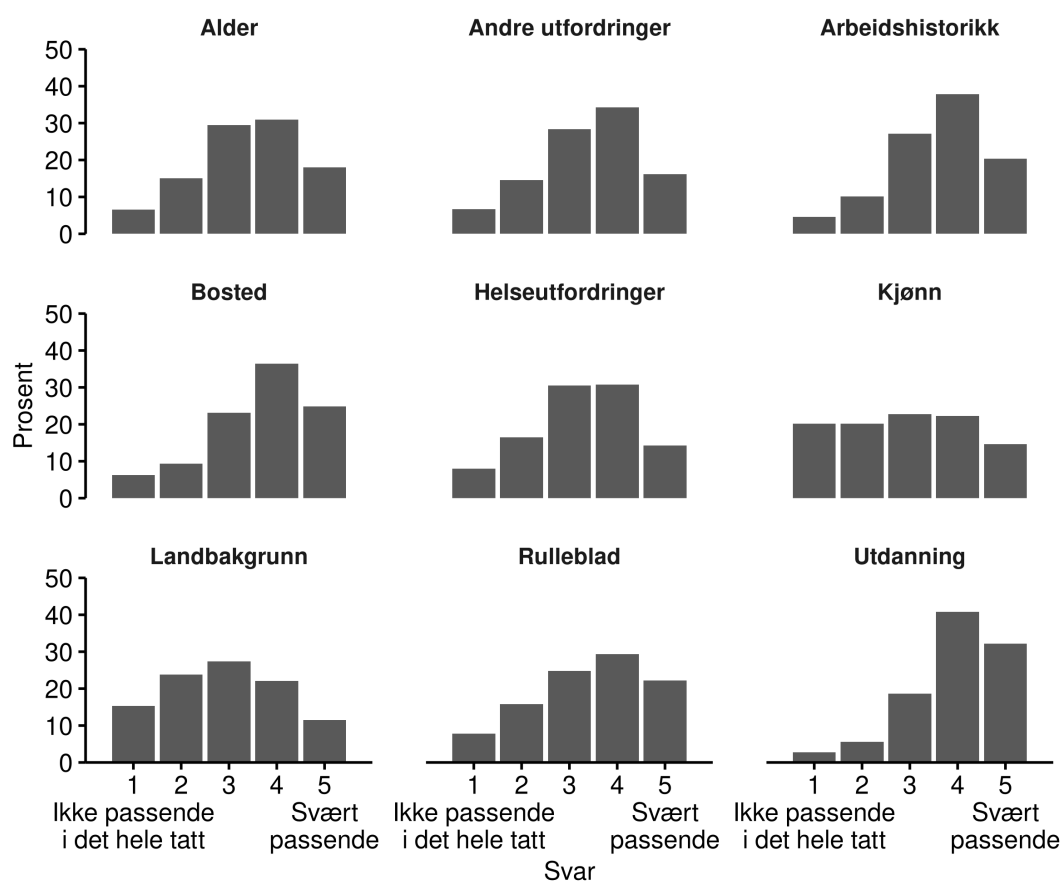


Figure 24: Fordeling for hver variabel av hvor passende den oppfattes å bruke

med ulik utdanning har noe ulikt syn på ulike variablene, og da særskilt når det kommer 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 årsaken 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.

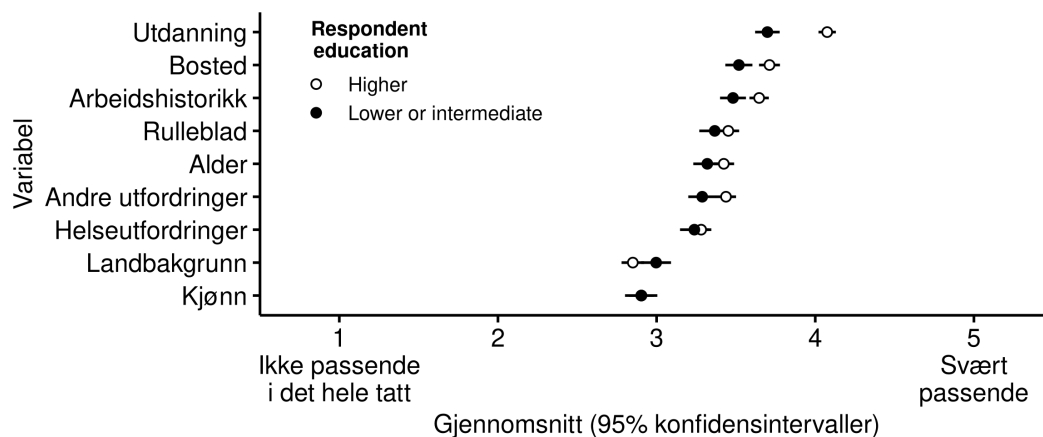


Figure 25: Gjennomsnitt av hvor passende det oppfattes å bruke variabelen for ulike utdanningsnivå

I den neste figuren er det differensiert etter hvor mye selverklært kunnskap respondene 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.

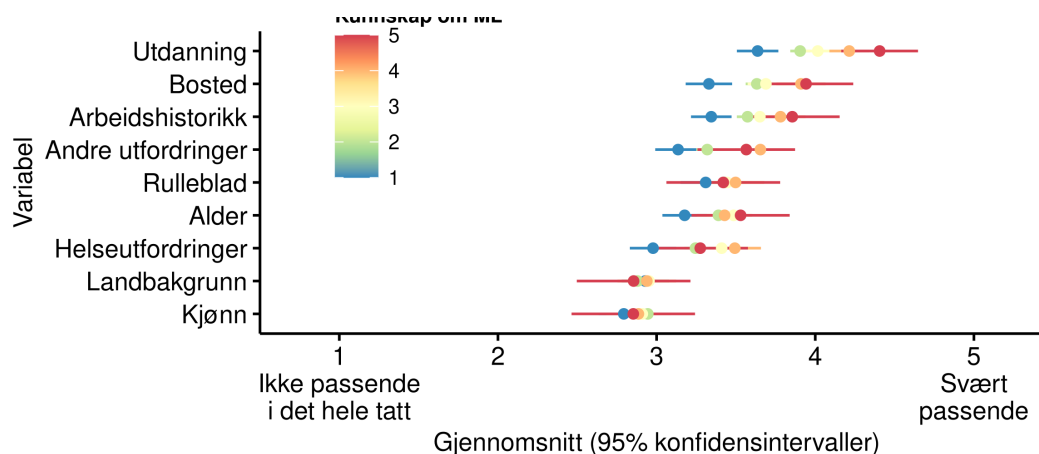


Figure 26: Gjennomsnitt av hvor passende det oppfattes å bruke variabelen for ulike nivå av selvrapportert kjennskap til maskinlæring

5.3 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 negative (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 innbyggenes 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 estimerer 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 påny 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 prioriteringen 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 sykemeldte 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 rettfærdig? 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.

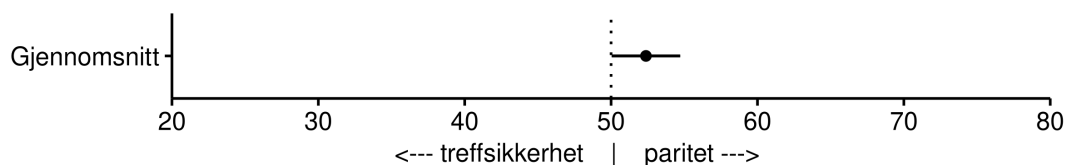


Figure 27: Preferanse for statistisk paritet

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.

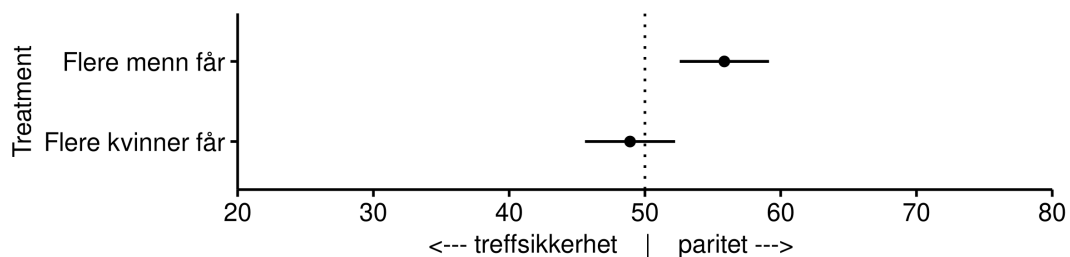


Figure 28: Preferanse for statistisk paritet etter hvilken gruppe som får skjeivt utfall (treatment)

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.

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

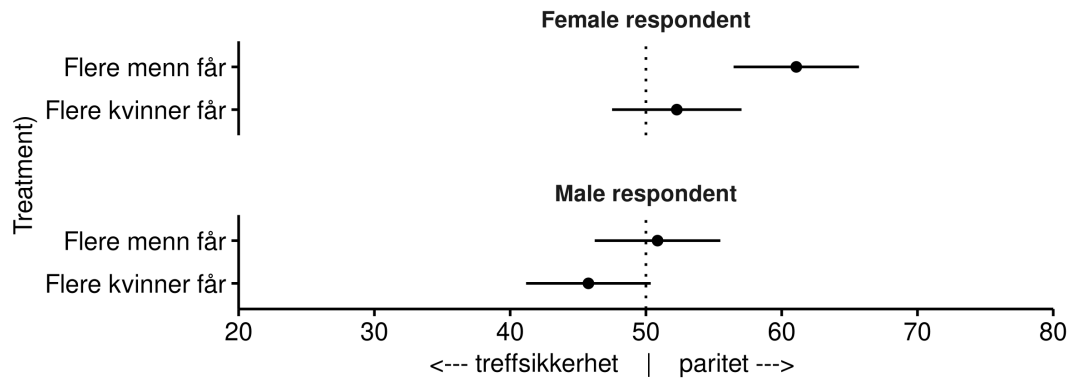


Figure 29: Preferanse for statistisk paritet etter hvilken gruppe som får skjeivt utfall (treatment) og respondentens kjønn

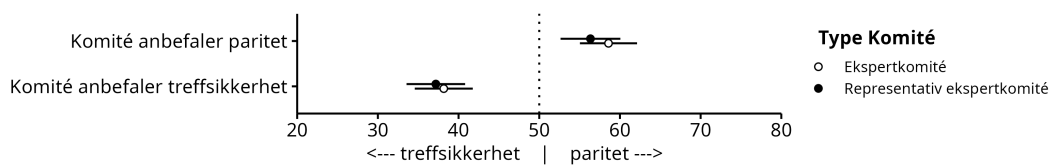


Figure 30: Preferanse for statistisk paritet etter hvilken gruppe som får skjeivt utfall (treatment) og respondentens kjønn

En 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. Ekspertene 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.

6 Representativt byråkrati

Den dominerende 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; Christensen et al., 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?

¹Deskriptiv 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 avhengig 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 innbyggers 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.

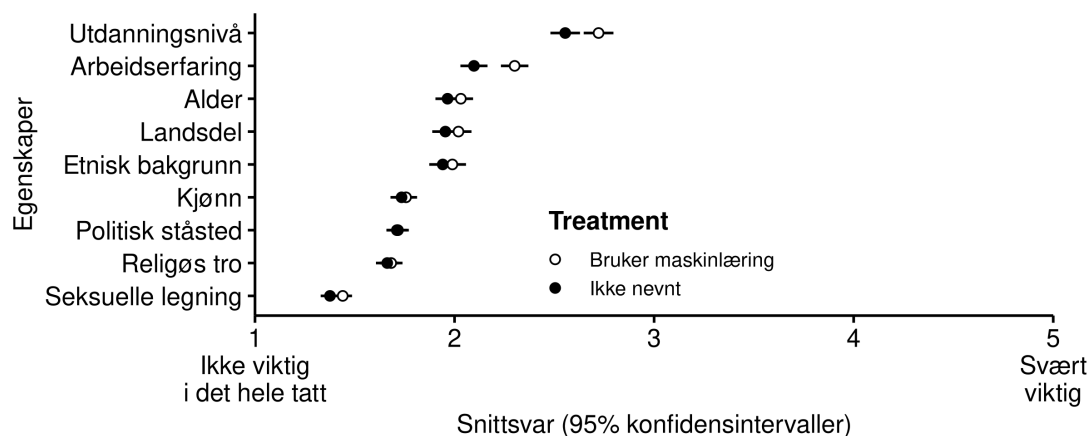


Figure 31: Representasjon: Gjennomsnitt av hvor viktig det oppfattes at saksbehandler har samme [egenskap] for å ivareta interesser, delt opp etter hvorvidt saksbehandleren bruker maskinlæring

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 gjennom-siktig 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 tas 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.

7 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.

Gjennom dialog med sentrale aktører i NAV har vi inntrykk av at det er grundige prosesser på plass for å validering og kvalitetssikring av modeller før de tas i bruk. For å ytterligere styrke legitimiteten til denne prosessen mener vi det er viktig å også ta med på råd de som til syvende og sist påvirkes av beslutningene. Det kan bety å invitere relevante interessegrupper inn på et tidlig stadium for å hjelpe med å identifisere potensielle konfliktpunkter. Det kan også bety å

gi befolkningen medvirkning og invitere til en bred diskusjon om etiske og politiske aspekter knyttet til bruk av maskinlæring og kunstig intelligens. 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.

7.1 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 fagekspertene, 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.

7.2 Saksbehandlingens rolle

Et viktig funn i denne rapporten er at de med mye (selverklært) kunnskap om maskinlæring tenderer til å ha andre oppfatninger 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.

Referanser

- Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. and it's biased against blacks. *ProPublica*.
- Arnesen, S. (2017). Legitimacy from decision-making influence and outcome favourability: Results from general population survey experiments. *Political Studies*, 65(1_suppl):146–161.
- Arnesen, S. and Peters, Y. (2018). The legitimacy of representation: How descriptive, formal, and responsiveness representation affect the acceptability of political decisions. *Comparative Political Studies*, 51(7):868–899.
- Barocas, S. and Selbst, A. D. (2016). Big data's disparate impact.
- Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., and Shadbolt, N. (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*, pages 1–14.
- Broderstad, T. S. (2022). Democratic reflections: To what extent do representatives mirror their constituents, and how does it affect the challenges modern, representative democracy are facing?
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2):153–163.
- Chouldechova, A., Benavides-Prado, D., Fialko, O., and Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on Fairness, Accountability and Transparency*, pages 134–148.
- Christensen, H. S., Himmelroos, S., and Setälä, M. (2020). A matter of life or death: A survey experiment on the perceived legitimacy of political decision-making on euthanasia. *Parliamentary Affairs*, 73(3):627–650.
- Christensen, T., Lægreid, P., and Zuna, H. R. (2001). Profesjoner i regjeringsapparatet 1976-1996. *Makt- og demokratiutredningens rapportserie*. ISSN 1501-3065.

- Clayton, A., O'Brien, D. Z., and Piscopo, J. M. (2019). All male panels? representation and democratic legitimacy. *American Journal of Political Science*, 63(1):113–129.
- De Fine Licht, J., Naurin, D., Esaiasson, P., and Gilljam, M. (2014). When does transparency generate legitimacy? experimenting on a context-bound relationship. *Governance*, 27(1):111–134.
- de Fine Licht, K. and de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. *AI & society*, 35(4):917–926.
- Duwe, G. and Rocque, M. (2017). Effects of automating recidivism risk assessment on reliability, predictive validity, and return on investment (roi). *Criminology & Public Policy*, 16(1):235–269.
- Easton, D. (1965). A systems analysis of political life.
- Esaiasson, P., Gilljam, M., and Persson, M. (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, P., Persson, M., Gilljam, M., and Lindholm, T. (2016). Reconsidering the role of procedures for decision acceptance. *British Journal of Political Science*, pages 1–24.
- Golder, M. and Stramski, J. (2010). Ideological congruence and electoral institutions. *American Journal of Political Science*, 54(1):90–106.
- Hansen, H.-T., Lundberg, K., and Syltevik, L. J. (2018). Digitalization, street-level bureaucracy and welfare users' experiences. *Social Policy & Administration*, 52(1):67–90.
- Krislov, S. (2012). *Representative bureaucracy*. Quid Pro Books.
- Lægreid, P. and Olsen, J. P. (1978). Byråkrati og beslutninger (bureaucracy and decisions). *Bergen: Universitetsforlaget*.
- Lim, H.-H. (2006). Representative bureaucracy: Rethinking substantive effects and active representation. *Public administration review*, 66(2):193–204.

- Lind, E. A. and Tyler, T. R. (1988). *The social psychology of procedural justice*. Springer Science & Business Media.
- Miller, S. M. and Keiser, L. R. (2021). Representative bureaucracy and attitudes toward automated decision making. *Journal of Public Administration Research and Theory*, 31(1):150–165.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Pitkin, H. F. (1967). *The concept of representation*. University of California Press.
- Rosanvallon, P. (2011). *Democratic legitimacy*. Princeton University Press.
- Rothstein, B. (2009). Creating political legitimacy: Electoral democracy versus quality of government. *American Behavioral Scientist*, 53(3):311–330.
- Schwarz, S. and Coppock, A. (2018). What have we learned about gender from candidate choice experiments? a meta-analysis of 67 factorial survey experiments.
- Tyler, T. R. (2003). Procedural justice, legitimacy, and the effective rule of law. *Crime and justice*, 30:283–357.
- Tyler, T. R. (2021). *Why people obey the law*. Princeton university press.
- Verma, S. and Rubin, J. (2018). Fairness definitions explained. In *2018 ieee/acm international workshop on software fairness (fairware)*, pages 1–7. IEEE.
- Weber, M. (2009). *The theory of social and economic organization*. Simon and Schuster.
- Zarsky, T. (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):118–132.