



Kunstig intelligens i helsesektoren, og risiko for kjønnsbias

En rapport utarbeidet av PA Consulting
og CARE Norge



**Bringing
Ingenuity
to Life.**



Innholdsfortegnelse

Forord	5
Sammendrag	6
1: Innledning	11
2: Kunstig intelligens, kjønnsbias og etikk	15
3: Kunstig intelligens i norsk helsetjeneste	25
4: utfordringer knyttet til kjønnsbias i KI i norsk helsetjeneste	37
5: Hvordan bedre forstå og motvirke uønsket kjønnsbias i KI	47
Appendiks 1 Bidragsyttere til rapporten	54
Appendiks 2 Biastyper	55
Appendiks 3 Sentrale etiske rammeverk	56
Kilder	58



FORORD →

PA Consulting og CARE Norge deler målsetningen

PA Consulting og CARE Norge deler målsetningen om økt likestilling i norsk offentlig sektor og næringsliv, og i lys av dette ser vi på bruken av kunstig intelligens i norsk helsetjeneste.

Begge parter stiller seg bak FNs bærekraftsmål, og ønsker å rette søkelyset mot bærekraftsmål 3 ("Sikre god helse og fremme livskvalitet for alle, uansett alder") og bærekraftsmål 5 ("Oppnå likestilling og styrke jenters og kvinners stilling i samfunnet")¹.

CARE Norge er en utviklings- og nødhjelpsorganisasjon som jobber for å styrke jenters og kvinners rettigheter og bekjempe fattigdom. CARE Norge belyser urettferdighet og foreslår løsninger for å fremme likestilling og styrke kvinners rettigheter i samarbeid med myndigheter, næringslivet og sivilsamfunn.

PA Consulting jobber for å bruke teknologi til å gjøre samfunnet bedre ved å digitalisere og effektivisere offentlig sektor og norsk næringsliv. De har et kontinuerlig fokus på rettferdig og bærekraftig utvikling.

PA Consulting og CARE Norge har gjennomført en studie om bruken av kunstig intelligens i den norske helsetjenesten og de tilhørende etiske problemstillingene. Rapporten er et resultat av intervjuer med eksperter og klinikere, med mål om å undersøke dagens bruk av kunstig intelligens (KI). I tillegg til en vurdering av dagens håndtering av bias, og å kartlegge hvor det kan være risiko for bias knyttet til kjønn. Disse risikoene er viktige å kartlegge da bruk av KI-løsninger potensielt kan forsterke kjønnsforskjeller, men også redusere disse dersom algoritmer utformes riktig.

Lanseringen av ChatGPT i starten av 2023 har vært en katalysator for diskusjonen om bruken av KI.

Med denne rapporten ønsker vi å bidra med ny innsikt og konkrete anbefalinger for å fremme trygg, rettferdig og ikke-diskriminerende bruk av KI i den norske helsetjenesten. Vi håper at rapporten vil bidra til økt bevissthet, og inspirere til kunnskapsdeling og konkret handling. Sammen kan vi akselerere utviklingen mot en mer likestilt, mer effektiv og mer bærekraftig helsetjeneste for alle.



Grete Kvernland-Berg
Norgessjef
PA Consulting



Kaj-Martin Georgsen
Generalsekretær
CARE Norge

SAMMENDRAG →

Kunstig intelligens (KI)

Mulighetene med kunstig intelligens (KI) er store og kan ha en positiv innvirkning på samfunnet. Riktig bruk av KI-løsninger kan hjelpe enkeltpersoner med å forbedre sin livskvalitet, og samtidig bidra til å løse viktige samfunnsutfordringer.

Det kan spille en viktig rolle i å nå FNs bærekraftsmål om å oppnå likestilling mellom kjønnene, effektiv bruk av ressurser og forbedret helse for alle.

Innen helsetjenesten i Norge byr bruken av KI på enorme muligheter til både å forbedre tjenester og effektivisere ressursbruk. I dag finnes det en rekke prosjekter som forsker på hvordan man kan benytte ny teknologi gjennom hele pasientløpet - eksempelvis til forebygging og sykdomsdiagnostikk for norske pasienter. KI-løsninger er ennå ikke utbredt i den norske helsesektoren, og de fleste prosjekter er i dag på forskningsstadiet. Dette skyldes en rekke faktorer, men en sentral årsak er at få prosjekter oppnår CE-merking¹⁰¹ og blir godkjent for klinisk bruk. Som igjen skyldes mangel på tilgang til gode data, og at det finnes risikofaktorer knyttet til bias, diskriminering og tillit til KI-løsningene.

Dersom norsk helsesektor skal kunne bidra til å nå FNs bærekraftsmål om god helse og livskvalitet og likestilling mellom kjønn, må man ta i bruk nye verktøy som KI på en god og hensiktsmessig måte. Vi må forstå hvordan utilsiktede konsekvenser kan oppstå, og sette inn tiltak for å håndtere disse.

Vi vet i dag at datasett og algoritmer kan inneholde en rekke uønskede bias mot enkeltgrupper. Dette gir igjen en risiko for diskriminerende eller urettferdig pasientbehandling. Dette kan skyldes bias fra datasett så vel som menneskelige bias fra de som er involvert i prosjektene. I flere land har det blitt avdekket tilfeller av KI-anvendelser som har ført til uønsket forskjellsbehandling basert på kjønn, og det er grunn til å tro at dette kan skje også i norsk helsetjeneste dersom ikke disse problemstillingene håndteres ordentlig.

På den ene siden kan KI-løsninger forsterke eller opprettholde eksisterende kjønnsforskjeller hvis algoritmer utvikles uten å fjerne skjevheter. På den annen side har KI-løsninger potensial til å redusere ulikheter ved å innlemme kjønnsforskjeller på en effektiv måte i helsesektoren, men da må disse utformes riktig. Gevinsten ved utviklingen av presise KI-systemer i helsesektoren vil være at vi på sikt kan muliggjøre persontilpasset behandling som tar i betraktning forskjeller hos individer, samtidig som man unngår diskriminerende skjevheter.

Formålet med denne rapporten er å fremme god og trygg bruk av KI i norsk helsesektor ved å øke bevisstheten rundt bias (og særlig kjønnsbias), og hvordan dette blir håndtert i norsk helsetjeneste. Rapporten har kartlagt sentrale risikofaktorer for kjønnsbias, og basert på disse risikofaktorene har vi identifisert seks tiltak vi mener vil redusere risikoen.

Vi mener det må være **“Mangfoldig og tverrfaglig kompetanse i team som utvikler og leverer KI-løsninger”**, så risikoen håndteres tidlig i prosessen. For å kontinuerlig videreutvikle kompetansen på områder mener vi **“Kunnskapsdeling og bevisstgjøring om kjønnsbias”** bør baseres på erfaringer nasjonalt og internasjonalt. For å sørge for faktisk bruk av gode rammeverk som er etablert på området, mener vi det er viktig med **“Tilgjengeliggjøring av etiske rammeverk gjennom praktiske guider”**. For løsningene som utvikles ser vi **“Økt grad av forklarbarhet, sporbarhet og transparens i KI-løsningene som et konkurransefortrinn”**. Dette for å både bygge tilliten til verktøyet hos helsepersonell, samt å bidra til god pasientsikkerhet gjennom forløpet. Dette oppnås gjennom å sørge for at beslutninger om videre behandling og utredning skjer med medvirkning fra pasienten. Gjennom arbeidet ble det tydelig for oss hvordan CE-merking fungerer, og hva som faktisk er dekket gjennom CE-merking ikke er bredt kjent. Vi anbefaler derfor **“Tydeligere forklaring av hva et CE-merket produkt innebærer, og hvilke krav som stilles til den som skal ta det i bruk”**.

Avslutningsvis mener vi at det er viktig å sette nasjonale rammer gjennom **“Biastesting som del av risikovurderingen og utarbeidelse av retningslinjer”**.

Gjennom implementering av tiltakene over vil det bygges kompetanse om bias, og tilrettelegges for utvikling og implementering av KI i norsk helsetjeneste som vil bidra til god helse og livskvalitet for alle.

Rapporten presenterer seks hovedfunn knyttet til bruk av KI i den norske helsesektoren:

1

Det er en manglende bevissthet og forståelse av kjønnsbias blant leverandører av KI-løsninger til norsk helsesektor.

2

For leverandører av KI-løsninger er det manglende insentiver til å prioritere biastesting (inkl. kjønnsbias) utover krav satt av CE-merkingen.

3

Få bruker etiske rammeverk i utviklingen av KI-løsninger.

4

Innkjøpere av KI-løsninger tar til en viss grad for gitt at bias har blitt håndtert gjennom CE-merkingen.

5

Det er uklart hvor ansvaret for validering ligger i dag, og hvordan ansvarsfordelingen ideelt sett bør være.

6

Algoritmene til KI-løsninger mangler forklarbarhet, tolkbarhet og transparens. Dette kan hindre oppdagelsen av kjønnsbias.

Følgende seks anbefalinger vil bidra til å fremme en trygg, rettferdig og ikke-diskriminerende helsetjeneste ved å håndtere bias bedre innen KI:



Økt grad av forklarbarhet, sporbarhet og transparens i KI-løsningene gir et konkurransefortrinn.



Mangfold og tverrfaglig kompetanse i team som utvikler og leverer KI-løsninger.



Kunnskapsdeling og bevisstgjøring om kjønnsbias.



Tilgjengeliggjøring av etiske rammeverk gjennom praktiske guider.



Biastesting og retningslinjer som en standardisert del av risikovurderingen.



Tydeligere forklaring på hva et CE-merket produkt innebærer, og hvilke krav som stilles til den som skal ta det i bruk.





Innledning

Bakgrunn, formål og metodikk

Nasjonal helse- og sykehusplan retter fokus på flere utfordringer som vil prege helse- og omsorgssektoren fremover. Utfordringene kommer som en konsekvens av den demografiske, teknologiske og økonomiske utviklingen i det norske samfunnet. Vi blir stadig flere eldre, og antall årsverk i den norske helsesektoren må øke med **35%** innen 2035 for å dekke det voksende behovet for helsepersonell. Dagens helsetjenester effektiviseres, og kvaliteten på tjenestene må økes for å holde tritt med samfunnets utvikling. Kunstig intelligens (KI) kan effektivisere og heve kvaliteten på helsetjenester ved å bidra til mer presise diagnoser, høyere kvalitet på behandlinger og bedre tilrettelegging. Det er med andre ord et enormt potensial for bruk av KI i praktisk klinisk bruk.

Det finnes også en rekke begrensninger og risikoer knyttet til KI-løsninger som bør vurderes nøye. Det er flere grunner til at KI-løsninger ennå ikke er utbredt i helsektoren i Norge i dag.

For det første er det mangel på KI-løsninger i operativt bruk. De fleste prosjekter er fortsatt på forskningsstadiet. Dette skyldes blant annet at de ikke oppnår CE-merking, en krevende forutsetning for å kunne tas i bruk i markedet⁵.

For det andre er det manglende tilgang på helsedata. I utviklingen av KI er det nødvendig med tilgang på relevante data av god nok kvalitet om den aktuelle pasientgruppen. Flere norske utredninger peker på at tilgang til data er en av de sentrale utfordringene som må løses for å oppnå effektiv bruk av KI i helsetjenesten i Norge⁶.

En tredje sentral faktor er at innføring av KI reiser en rekke utfordringer knyttet til bl.a. rettferdighet, ansvar, etiske avveininger og juridiske aspekter⁷. Dette er de problemstillingene rapporten primært vil analysere, og med et særlig blikk på bias, ikke-diskriminering og tillit knyttet til KI-løsninger.

Til tross for økt satsning på KI-anvendelser og økt oppmerksomhet rundt det positive potensialet, har det blitt avdekket flere hendelser hvor KI diskriminerer og rammer enkelte grupper, eller enkeltindivider uheldig. Flere eksempler på kjønnsbias har blitt oppdaget. Det er likevel grunn til å tro at disse ikke blir godt nok forstått og hensyntatt av dagens KI. Med det som bakgrunn undersøker denne rapporten den norske helsesektoren for å kartlegge utbredelsen av KI-løsninger i dag, og for å vurdere risiko for kjønnsbias knyttet til bruk av KI.

Utgangspunktet for rapporten er følgende spørsmål:

- Hva vet vi om kunstig intelligens, likestilling og kjønnsbias i dag?
- Er det god nok bevissthet og kompetanse blant dem som jobber med dette i dag?
- Hva kan gjøres for å håndtere kjønnsbias bedre enn i dag?

Gitt den raske teknologiske utviklingen og mulighetsrommet som KI-løsninger skaper, mener vi at dette er spørsmål det er nødvendig å reflektere over. For å opprettholde pålitelig og effektiv pasientbehandling med ny teknologi, må norsk helsetjeneste være oppmerksom på dagens verktøy og metoder samt ta nødvendige grep for å sikre god og trygg bruk i fremtiden. For å utarbeide rapporten har vi gjennomført et litteraturstudie innen norsk og internasjonal forskning på kunstig intelligens, helse, og utfordringer knyttet til kjønn. Vi har deretter gjennomført 25 semistrukturerte intervjuer med personer med innsikt om bruken av KI i norsk helsetjeneste. Intervjuobjektene representerer nasjonal helseforvaltning, helseforetak og sykehus, akademia og KI-eksperter eller KI-forskere. En stor takk rettes til alle som har bidratt med innspill til rapporten. Disse er listet opp i Appendiks 1.

Vi har gjennomført noen faglige avgrensninger. Bias innen KI er et fagområde som omfavner en rekke ulike problemstillinger. I denne rapporten fokuseres det i stor grad på kjønnsbias. Det vil si uønskede skjevheter mot ett av kjønnene, som gir risiko for uønsket variasjon i leveranse av helsetjenester. Dette betyr ikke at andre former for bias (slik som bias knyttet til etnisitet, legning eller sosiale grupper) er mindre viktige. Det bør også forskes videre på disse typene av bias. Uønsket kjønnsbias kan naturligvis ramme begge kjønn. Denne rapporten har derimot ikke inkludert tilfeller der menn rammes av bias. Vi presiserer derfor at flere funn og anbefalinger gjør seg like gjeldende for andre typer bias, og for kjønnsbias både mot menn og kvinner, men oppfordrer til mer forskning for å belyse problematikken. I tillegg mener vi at datakvalitet er helt sentralt for problematikken rundt KI. Det er derfor behov for ytterligere kartlegging av tilstanden, tilgjengeligheten og kvaliteten på norske helsedata, og identifisering av problemområder som helse-Norge eller myndighetene bør ta tak i.

Formålet med rapporten er å fremme god og trygg bruk av KI i norsk helsesektor ved å øke bevisstheten rundt bias (og særlig kjønnsbias), og hvordan dette blir håndtert i norsk helsetjeneste. Rapporten bidrar med kunnskap om sentrale risikofaktorer for kjønnsbias, og tilbyr anbefalinger til hvordan aktørene i helsetjenesten kan handle for å best mulig håndtere disse.

Vi anbefaler at rapporten sees i sammenheng med publikasjoner og konkrete tiltak fra Direktoratet for e-helse, Helsedirektoratet, Nasjonalt senter for e-helseforskning, Teknologirådet og andre som bidrar med viktig kunnskap innen disse fagområdene.



2

Kunstig intelligens, kjønnsbias og etikk

Definisjon av kunstig intelligens (KI)

Kunstig intelligens (KI) er et begrep med flere ulike definisjoner. Denne rapporten tar utgangspunkt i følgende definisjoner som er utarbeidet av EUs ekspertgruppe og som også brukes i Nasjonal strategi for kunstig intelligens^{8,9}:

KI handler om at systemene kan lære seg å gjennomføre handlinger ved å analysere og behandle strukturert og ustrukturert data, og finne nye sammenhenger når det blir tilført ny data. Det som gjør KI mulig, er dermed tilgang på betydelige mengder data. Dette kan for eksempel være data fra ulike sensorer, kameraer, tekst, bilder, video, mikrofoner eller andre datakilder. Det er også nødvendig med regneressurser og utvikling av algoritmer for å utvikle KI-systemene. Kunstig intelligente systemer kan gjennomføre kompliserte oppgaver som før ble gjennomført av mennesker, på en effektiv måte. Dette skaper store muligheter for verdiskapning og utvikling av bedre tjenester¹².

“Sterk” og “svak” kunstig intelligens

Innen KI skilles det ofte mellom “sterk” og “svak” KI. Sterke KI-systemer, også kalt kunstig generell intelligens, er systemer som kan lære å gjennomføre oppgaver på samme måte som mennesker. Systemene som er utviklet og tas i bruk i dag vil kategoriseres som svak KI. Dette er systemer som er utviklet for å gjennomføre spesifikke oppgaver. Innen helse er svak KI for eksempel bildeanalyser eller mønstergjenkjenning i datasett for å predikere sykdom¹³. Denne rapporten fokuserer hovedsakelig på “svak” KI da dette er mest aktuelt for helsesektoren i nær fremtid.



Kunstig intelligente systemer utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål. Enkelte KI-systemer kan også tilpasse seg gjennom å analysere og ta hensyn til hvordan tidligere handlinger har påvirket omgivelsene¹⁰.

EUs ekspertgruppes definisjon av KI

Regelstyrt og datadrevet kunstig intelligens

Et annet viktig skille innenfor kunstig intelligens trekkes mellom alle intelligente systemer, hvor man kategoriserer i to hovedskiller: regelbaserte systemer og datadrevne systemer. Tidligere ble regelstyrte systemer mest brukt og utviklet, men utover 2000-tallet har KI utviklet seg til å bli mer data- og statistikkdrevet. Dette har gjort at datamaskinene har kunnet lære uten at de var forhåndsprogrammerte¹⁴.

Regelbaserte KI-systemer, også kalt ekspertsystemer, er systemer som på forhånd er definert hvordan de skal oppføre seg ved at de er bygget på regler som utvikles fra menneskelig ekspertise. Ekspertsystemet er bygget opp av to hovedkomponenter bestående av en kunnskapsbase og et sett med regler som systemet skal følge. Ekspertkunnskapen kan for eksempel lagres på måten; "hvis betegnelse x er oppfylt kan man trekke slutning y eller gjøre handling z"¹⁵. Datadrevne KI-systemer er bygget opp av at dataprogrammet selv utvikler regler ved å analysere datasettet. Reglene utvikles ved hjelp av algoritmer som er bygget opp på data systemet trenes på. Dagens KI-løsninger er for det meste basert på datadrevet KI og maskinlæring, og vil dermed være i fokus i rapporten.

Maskinlæring

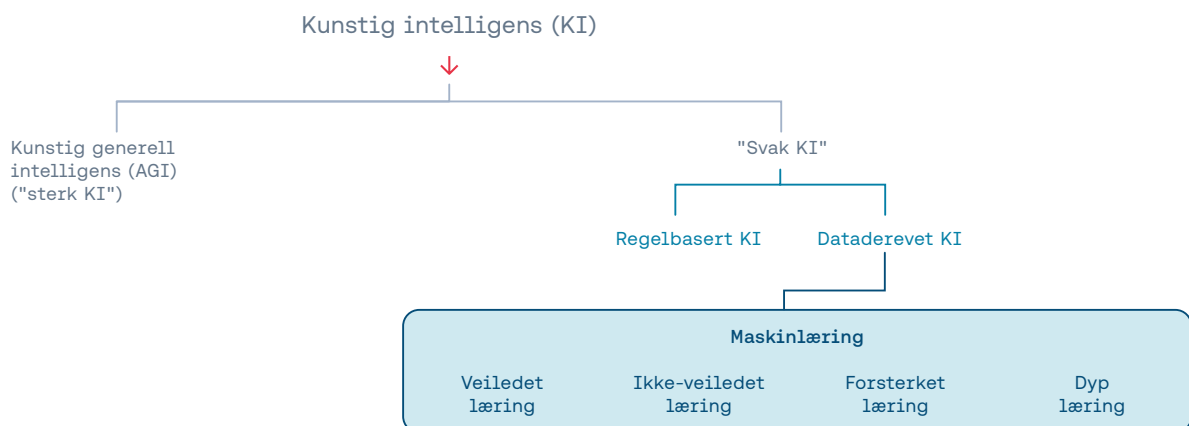
Maskinlæring er en metode innenfor KI som handler om å bruke statistiske metoder så datamaskiner finner mønstre i data¹⁶. Dette innebærer at datamaskinen blir trent opp til å finne ulike mønstre ved å få erfaring gjennom treningsdata, og med det utvikler nye mønstre som det tas beslutninger fra. Det vil da være maskinlæringsalgoritmer som bygger matematiske modeller ut fra eksempelvis data eller treningsdata.

Algoritmene som maskinlæringssystemene er utviklet på, kan lære gjennom tre måter: veiledet, ikke-veiledet og forsterket læring. Veiledet læring innebærer at datasettene algoritmen trenes på inneholder merkede data slik at algoritmen trenes til å gi nøyaktige klassifiseringer og predikeringer. Det brukes treningsdata som inkluderer det riktige resultatet som gjør at algoritmen kan lære over tid. Dette brukes for å bygge modellen slik at den senere kan ta beslutninger selv kun basert på data den blir tilført. KI-modeller basert på veiledet læring kan brukes til blant annet klassifisering, og slik vurdere resultater fra hudbilder som for eksempel "godartet" eller "ondartet"¹⁷. I ikke-veiledet læring får algoritmen kun datasett med "fasit", og ut fra dette må den selv finne mønstre for videre å kunne ta beslutninger. Dette brukes for å finne mønstre i dataen, og det kan eksempelvis brukes ved klyngeanalyser der man vil finne spesielle karakteristikk for en gruppe. Det kan også brukes dersom man ønsker å undersøke om sykdommer har en sammenheng med hverandre¹⁸. Den siste metoden kalles for forsterket læring. Det vil si at algoritmene læres opp ved at modellen får tilbakemelding fra brukeren av systemet om den foreslåtte beslutningen er god eller dårlig, og tilbakemeldingen brukes til å forbedre modellen¹⁹.

Dyplæring

Innenfor maskinlæring ligger dyplæring som er en viktig del i bildebehandling, talegjenkjenning, behandling av naturlig språk og medisinske bilder samt avviksdeteksjon. Dyplæring kan brukes til å prosessere og strukturere data som dokumenter, bilder og tekst²⁰. Dyplæring skiller seg fra maskinlæring ved dataen det er utviklet på og måten systemet lærer på. Der maskinlæring krever en egen form for prosessering av dataen, vil dyplæring selv prosessere dataen²¹. Dyplæring baserer seg på kunstig nevralt nettverk som prøver å etterligne menneskehjernen. De nevralt nettverkene er bygget opp av tre eller flere lag som er koblet sammen, og hvor hvert lag bygger på det forrige. Koblingene mellom disse kan være sterke og svake, og det er styrken på disse bindingene som endres ettersom modellen lærer²².

Figur 1: Forenklet oversikt over kunstig intelligens



Modellen er basert på Direktoratet for e-helse sin rapport "Utredning om bruk av kunstig intelligens i helsesektoren" (2019), intervjuer gjennomført for rapporten, samt PA Consulting og Finansforbundets rapport "Brytningstid: En studie av kunstig intelligens i norske banker og forsikringsselskap" (2022).

Kjønnsbias i kunstig intelligens

Ifølge Oxford Languages kan bias beskrives som “fordommer mot en person eller gruppe – særlig på en måte som oppfattes som urettferdig”²³. På norsk refererer ordet bias gjerne til “skjevheter”, “forutinntatthet” eller “fordommer”, selv om disse begrepene ikke klarer å fange hele ordets betydning. I denne rapporten refererer bias til hvordan mennesker eller KI-løsninger i ulike tilfeller, bevisst eller ubevisst, kan ha fordommer eller diskriminere mot en gruppe av befolkningen basert på karakteristikk som for eksempel kjønn, hudfarge, religion og etnisitet.

De siste årene har den sosiale bevisstheten om bias i teknologi økt. Bias er spesielt bekymringsfullt på områder hvor det er mangel på variasjon i treningsdataen til algoritmen, og helse trekkes frem som et område hvor det er spesielt relevant. Bias kan oppstå på flere måter ved utvikling av KI-systemer. Det er derfor viktig at KI-utviklere er klar over bias når det velges hvilke maskinlæringsteknologier/prosedyrer de bruker for å trene algoritmen og kvalitet, og mangfold i datasettet som brukes i programmeringen²⁴.

Bias vil alltid finne sted i et KI-system, og det er derfor viktig å forstå disse problemstillingene og hvordan de kan påvirke ulike pasientgrupper. Bare da kan vi sørge for at KI-systemene ikke inneholder bias som kan diskriminere eller svekke kvaliteten på tjenestene til de ulike målgruppene²⁵.

Ønskelige og ikke-ønskelige bias

Til tross for at ordet "bias" ofte har en negativ konnotasjon og gjerne forbindes med urettferdighet eller forskjellsbehandling, skilles det gjerne mellom ønskelige og ikke-ønskelige bias. Forskjellen på disse i et helseperspektiv er hvordan de påvirker pasientens behandling og sikkerhet. Ønskelig bias i et helseperspektiv vil for eksempel handle om å forstå hvilke forskjeller som finnes mellom kjønnene. For visse tilstander vil det være ulikhet i hvordan diagnose settes og sykdommen behandles. Ønskelig bias innebærer derfor å ta hensyn til kjønnsforskjeller for å stille en presis diagnose, og anbefale en tilpasset og dermed mer effektiv behandling for hver pasient²⁶. Uønskede bias handler derimot om diskriminering og bruk av ikke-representative grupper som kan resultere i mindre god og effektiv behandling for en gruppe sammenlignet med andre.²⁷

Bias kan være bevisste eller ubevisste. Det er grunn til å tro at mange uønskede bias vil oppstå ubevisst av den grunn at man ikke har tilgang på data som er representativ nok for alle pasientgrupper, og at man ikke har store nok datamengder. I tillegg kan ignoranse være en faktor ved at det er for lite kunnskap og bevissthet om hvilke bias som finnes, og at dette da ikke hensyntas i datainnhenting eller i opplæringen av algoritmer.

Representasjonsbias

Med bakgrunn i at maskinlæringssystemer baserer seg på data, vil systemene kun være så pålitelige og nøytrale som dataene. Den vanligste formen for bias er den som kalles 'representasjonsbias' - at en gruppe er underrepresentert i treningsdatasettet til algoritmen som gir opphav til en iboende bias i algoritmen. Skjevheter kan oppstå dersom det benyttes begrensede, ikke representative data av lav kvalitet²⁹. I et helseperspektiv vil det for eksempel i et KI-verktøy bygget opp på data som er skjevt fordelt mellom ulike pasientgrupper avhengig av alder, kjønn eller etnisitet, kunne utgjøre en risiko for at pasienter feildiagnostiseres og feilbehandles på grunn av mindre presis predikering. Under opplæringen av algoritmer er det derfor viktig at data som brukes er representativ for gruppen systemet skal brukes på. Et eksempel på hvor bias kan oppstå er dersom kliniske beslutningsalgoritmer utvikles på elektroniske helsejournaler og disse inneholder manglende data med ubalansert representasjon av for eksempel kjønn²⁸.

Et annet sentral bias er 'historisk bias'. Historiske data eksisterer i den kontekst de ble samlet inn. Dette innebærer at selv om data kan være korrekt utvalgt og innhentet, kan KI-løsninger basere seg på et verdensbilde som gir uønskede resultater. Et eksempel hvor bruk av historiske data har ført til diskriminering var da Amazon utviklet en algoritme for å analysere jobbsøknader for å kunne skille ut de beste kandidatene. Algoritmen ble trent opp på jobbsøknader fra de siste tiårene hvor mannlige søknader var overrepresentert. Dette førte til at algoritmen favoriserte mannlige kandidater og rangerte kvinner lavere²⁹. Til tross for at det i denne rapporten primært vies oppmerksomhet til representasjonsbias, finnes det også flere andre biastyper som kan gjøre seg gjeldende. De står beskrevet i appendiks 1.

Kjønnsbias

I tillegg til ulike typer bias, finnes det også flere kategorier for hvem som kan rammes av uønskede bias. Kjønnsbias er et av disse, og handler om fordommer som favoriserer det ene kjønn over det andre. Språkrådet definerer kjønnsbias som “forutinntatt mening eller holdning om et kjønn på måter som blir oppfattet som uriktige eller urettferdige, nedfelt i språk, sosial praksis eller institusjoner³⁰”. Kjønnsbias kan gjenspeiles i KI-systemer ved at systemene ofte er en refleksjon av de som har laget systemet, og dataen det er bygget på. Det finnes en rekke årsaker til at kjønnsbias oppstår, og de neste avsnittene vil se på noen av de mest sentrale.

Internasjonalt finnes det en rekke eksempler på kjønnsbias innen KI, som viser de potensielt alvorlige konsekvensene:

- **Det er høyere sannsynlighet for å overse leversykdom hos kvinner enn menn.** En studie publisert i BMJ Health & Care Informatics viste at et nytt KI-system som skulle predikere leversykdom ut fra blodprøver var to ganger mer sannsynlig til å overse sykdom hos kvinner enn hos menn³¹.
- **Det er mindre sannsynlig å oppdage hjerteproblemer hos kvinner enn menn.** Hjerte- og karsykdommer er den vanligste dødsårsaken i verden i dag, og til tross for at det er flere menn enn kvinner som rammes, er dødeligheten høyere hos kvinner. En studie viste at kun **17%** av kardiologene identifiserte at kvinner hadde større risiko for hjerteinfarkt. Det finnes flere årsaker til at det er lettere å oppdage symptomer hos menn enn kvinner, og en av årsakene kan være at leger er mest opplært i å se symptomer som oftere vises hos menn. Det vil si at dersom det utvikles algoritmer på tilgjengelige data fra diagnostiserte tilfeller vil disse påvirkes av kjønnsforskjeller³².
- **KI-systemer er bedre trent på å se etter symptomer for Parkinson for menn.** I en studie på Parkinson hvor digitale biomarkører skulle påvise sykdom, inneholdt datasettet algoritmen som ble trent på kun **18,6%** kvinner. Konsekvensen av at mannlige pasienter var overrepresentert var at systemet ble mer nøyaktig til å påvise Parkinson sykdom hos menn enn hos kvinner fordi menn og kvinner ofte har ulike symptomer på sykdommen³³.

Disse eksemplene understreker at det er et spekter av risikoer involvert i bruk av KI-løsninger, og at det derfor er behov for en bevisst holdning til de etiske og sosiale aspektene. Det finnes også mange grunner til at kjønnsbias forekommer:

- **Maskinlæringssystemer utvikles på datasett med overvekt av menn.** Maskinlæringssystemer er ofte utviklet på datasett fra vestlige land (og noen fra Kina) med en overvekt av menn. Dette gjør at systemene kan fungere dårligere for kvinner eller andre etnisiteter³⁴.
- **Kvinner har historisk sett vært underrepresentert i kliniske studier.** Fordi KI bruker historiske data for å trene opp algoritmene, vil disse utvikles til å gjenspeile fortiden³⁵. Tidligere har ikke kvinner blitt inkludert i medisinsk forskning i like stor grad som menn, og dette har ført til at det i dag er et gap mellom kliniske data på menn og kvinner^{36,37}. En årsak til dette er blant annet at menstruasjonssyklusen er blitt trukket frem til å kunne påvirke resultatene fra studiene³⁸, og at det var ønskelig å beskytte kvinner i fruktbar alder og under graviditet³⁹. Dette kunnskapsgapet gjenspeiles derfor i medisinske data⁴⁰. I 2001 kom Den nasjonale forskningsetiske komité for medisin og helsefag (NEM) med retningslinjer for inkludering av begge kjønn i all medisinsk forskning⁴¹. Dersom det er enkelte prosjekter, hvor det ikke er relevant å inkludere begge kjønn må det begrunnes etisk og vitenskapelig.

Fordi kvinner historisk sett og i dag i mindre grad er inkludert i medisinske studier, har vi mindre kunnskap om kvinnehelse enn for menn⁴². Det gjør også at diagnoser, bivirkninger og medisiner som er basert på data med lavt kvinnegrunnlag derfor vil være mangelfulle og ikke-representative for kvinner⁴³. Det vil si at dersom datasettene i KI-systemene utvikles på den samme kunnskapen, vil forskjellene videreføres⁴⁴.

- **Det er en mannlig dominans blant utviklere av KI.** KI-systemer inneholder bias fordi de er utviklet av mennesker. Systemet blir påvirket av hvem som er på teamet og utvikler det fordi det er mennesker som generer, samler inn, og velger hvilke data og variabler som skal benyttes. I tillegg til å velge hvilke regler algoritmene skal følge⁴⁵. I Norge er åtte av ti som jobber med IT menn, og kun 12% av de som jobber med programvareutvikling er kvinner⁴⁶. Dette kan føre til at bias blir en del av systemene.

Etikk i kunstig intelligens

Til tross for at KI i helsetjenester har et stort positivt potensial, knyttes det også etiske utfordringer til bruken. Dersom etiske prinsipper og menneskerettigheter ikke prioriteres av de som utvikler, regulerer eller bruker KI for helse, kan det få store konsekvenser⁴⁷.

Det finnes flere etiske rammeverk for å utvikle gode KI-systemer som er trygge å bruke. Dette er ikke prinsipper KI-leverandører som utvikler løsningene er lovpålagt å følge, men som bør ligge til grunn fordi bruken av KI-løsningene kan få konsekvenser for enkeltmennesker eller grupper. Rammeverkene kan brukes som veiledning for utviklere, brukere og leverandører av KI for å sikre at etiske hensyn blir vurdert. De kan også brukes som referanseverktøy for å oppnå felles forståelse for problemstillingene. EU har utviklet etiske retningslinjer for utvikling av pålitelige KI-systemer. Disse består av tre punkter som bør etterkommes i hele KI-systemets livssyklus:

- **Lovlig** - systemet bør følge alle gjeldende lover og regler
- **Etisk** - det bør være etisk, og i tråd med etiske prinsipper og verdier
- **Robust** - systemene bør være robuste både fra et teknisk og sosialt perspektiv med gode intensjoner⁴⁸.

Innen utvikling av etiske KI-systemer er det utviklet fire etiske prinsipper:

- **Respekt for menneskelig autonomi:** Mennesker som samhandler med KI må kunne beholde selvbestemmelse og kunne delta i demokratiske prosesser. Systemet må ikke prøve å lure mennesker. KI bør utvikles for å styrke menneskets kognitive, sosiale og kulturelle egenskaper.
- **Forebygge skade:** KI skal ikke påvirke mennesker negativt, og det skal beskytte menneskers verdighet, mental og fysisk integritet. KI og området det opererer i, skal være trygge og systemene skal være teknisk sterke slik at de ikke kan misbrukes.
- **Rettferdighet:** Utvikling og bruk av KI må være rettferdig, og alle bør ha lik tilgang til teknologien. Rettferdighet ses på som å både ha en materiell og en prosessuell dimensjon. Den materielle dimensjonen handler om å sikre lik og rettferdig fordeling av fordeler og kostnader og sikre at individer ikke utsettes for urettferdige skjevheter, diskriminering eller stigmaer. Prosess-siden av rettferdighet handler om at det må være mulig å identifisere hvem som er ansvarlig for beslutningene fra KI-systemet, og beslutningsprosessen til systemet må være forklarbare. Dersom urettferdige bias unngås, kan KI skape mer rettferdighet i samfunnet⁴⁹.
- **Forklarbarhet:** Prinsipper bygger på at det er nødvendig med forklarbarhet for at brukere av KI skal ha tillit til systemet. Dette innebærer at systemet er transparent og at formålet og bruksområdet til KI-systemene er tydelige for de som KI brukes på, både direkte og indirekte.

Et viktig punkt som trekkes frem av både WHO og EU er ansvarlighet for utvikling av etiske KI-systemer. Det handler om hvem som har ansvaret for at de riktige beslutningene tas og hvem som har ansvaret dersom et KI-system gjør eller brukes feil. I medisinsk behandling er det et klart ansvarsforhold rundt medisinske råd og behandlinger ved at helsepersonell står ansvarlige for disse. Når man går over til å bruke kunstig intelligente systemer for å ta beslutninger dukker spørsmålet om hvem som er ansvarlige for disse beslutningene opp. KI-løsningene som denne rapporten baserer seg på gir anbefalinger, og tar ikke selvstendige beslutninger om pasientenes behandling. Allikevel blir det stadig mer krevende å skille mellom hvem som fatter avgjørelsen når det brukes KI. Helsepersonell får anbefalinger fra KI-løsningen, og må deretter vurdere hvordan det skal brukes. Imidlertid kan anbefalingene ta en førende rolle når det er tidspress og etablert tillit mellom helsepersonell og systemet⁵⁰.

KI er ikke alltid forklarbart og en utfordring med dette er hvordan pasientdata brukes, og at det må gis samtykke for å bruke pasientdata. Utfordringen handler om i hvilken grad helsepersonell må informere pasientene om kompleksiteten og bruken av KI. Når KI består av svarte bokser der helsepersonell ikke kan forklare alle resonnementer bak resultatet, blir spørsmålet hvor mye de må kunne forklare til pasientene. Ifølge pasient og brukerrettighetsloven har alle pasienter rett til å få forklaring av faguttrykk. Spørsmålet blir da hva som skal gjøres dersom pasienter er villig til å bruke KI de ikke kan få en full forklaring på og hvordan bruken av trygge og effektive KI-systemer skal balanseres med pasientens rett til forklaring og personvern.⁵¹ Et spørsmål oppstår rundt motsetningene mellom å lage systemer med en høy funksjonalitet og mye data, som kan være vanskelig å forklare og å utvikle systemer med mindre funksjonalitet, men som kan forklares på en enkel måte.

For at KI skal fungere som gode beslutningsstøtteverktøy er det nødvendig at helsepersonell har tillit til resultatene disse produserer, og det er derfor viktig at de forstår hvordan systemene kommer fram til beslutninger. KI-systemer kan være komplekse, og det er ikke alltid mulig for helsepersonell å forstå nøyaktig hvorfor systemet foreslår ulike behandlinger eller diagnoser, og dette kan da påvirke tillitten til systemene⁵². Dersom man ikke vet hvordan systemene er bygget opp vil det også være vanskelig å sikre at det er utviklet et rettferdig system ettersom man ikke alltid har innsyn i hvordan hele modellen er bygget opp⁵³.



3

Kunstig intelligens i norsk helsetjeneste

Status for bruk av KI i norsk helsetjeneste

Politisk ambisjon og vilje

Den norske regjeringen presenterte i 2020 Nasjonal strategi for kunstig intelligens, og viste med dette politisk vilje og ambisjon for å satse på KI. Strategien beskriver at regjeringen vil legge til rette for at Norge skal ha infrastruktur for kunstig intelligens i verdensklasse, og det trekkes frem at et digitaliseringsvennlig regelverk og datadeling er forutsetninger for å lykkes med dette.

I tillegg løfter strategien frem viktigheten av etiske prinsipper for god bruk av KI. I strategien blir det understreket at KI skal “bygge på etiske prinsipper, og respektere menneskerettighetene og demokratiet”, og at forskning, utvikling og bruk av KI skal “bidra til ansvarlig og pålitelig kunstig intelligens” og “ivareta den enkeltes integritet og personvern”. I tillegg presiseres det at “tilsynsmyndigheter skal føre kontroll med at systemer basert på kunstig intelligens på sitt tilsynsområde opererer innenfor prinsippene for ansvarlig og pålitelig bruk av KI”.

“

“Kunstig intelligens som utvikles og brukes i Norge skal bygge på etiske prinsipper, (..) respektere menneskerettighetene, (..) bidra til ansvarlighet og pålitelighet, (...) og ivareta den enkeltes integritet og personvern⁵⁴.”

Nasjonal strategi for kunstig intelligens (2020)

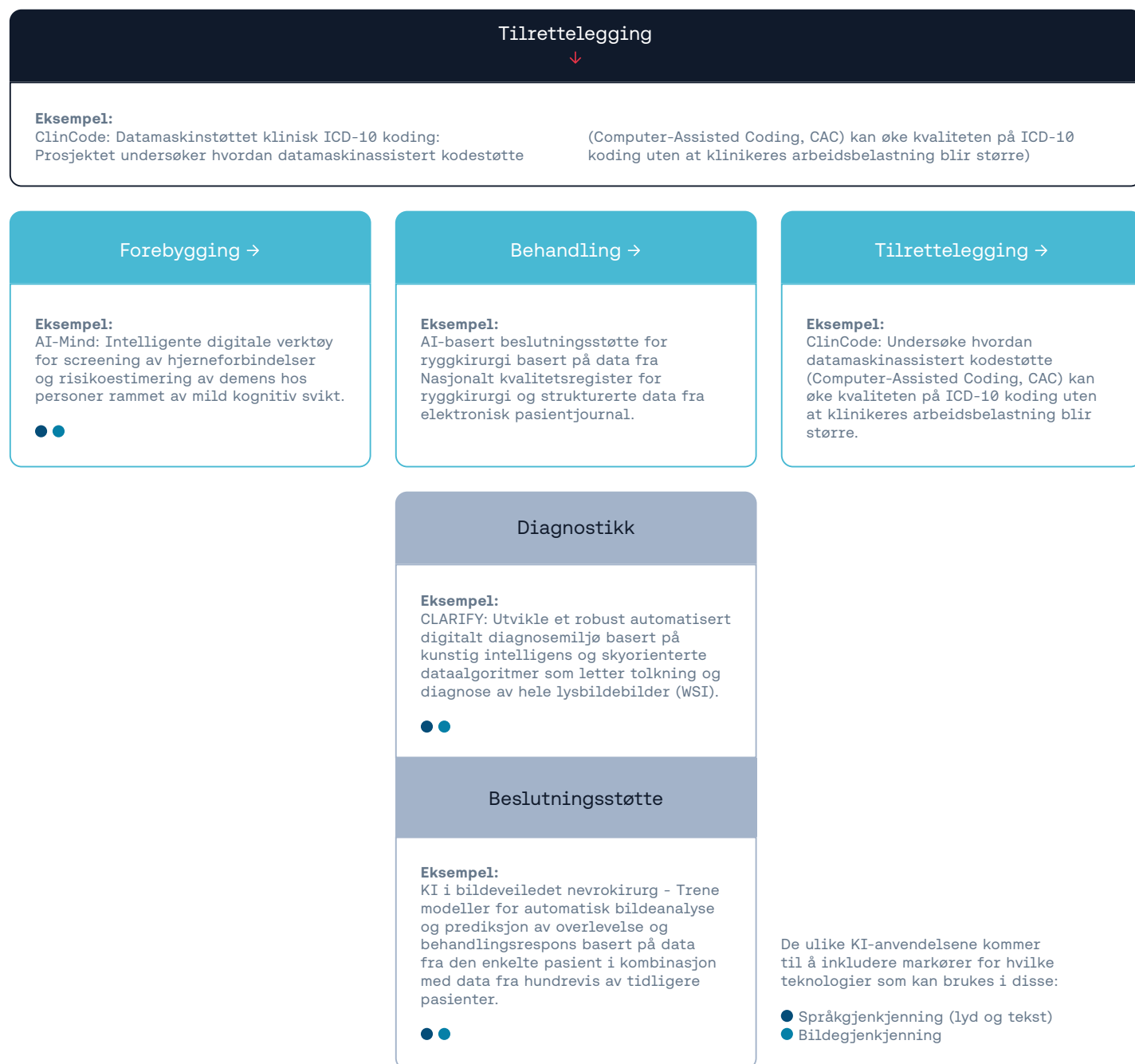
Muligheter for KI i norsk helsetjeneste

Norge har gode forutsetninger for å lykkes med KI i helsetjenesten på grunn av sterke forskningsmiljøer, økende KI-kompetanse og store mengder helsedata (selv om denne ikke er gjort tilgjengelig i dag). KI kommer til å ha mange og varierte bruksområder i helsetjenesten fremover. Helt fra forebyggingsstadiet vil KI kunne bidra med å analysere og overvåke trender, og oppdatere risiko for ulike sykdommer. Videre vil KI kunne spille en nøkkelrolle gjennom behandlingsfasen med diagnostisering og beslutningsstøtte ved bildeanalyse, tekstanalyse og behandling av store data. Også under rehabilitering vil det være muligheter for KI-løsninger til å følge opp pasienters utvikling. Tilrettelegging utgjør også et stort mulighetsrom for KI, selv om denne kategorien ikke virker direkte inn på pasientforløpet. KI vil kunne tilrettelegge for bedre ressursutnyttelse og automatisering av administrative prosesser gjennom aktiviteter knyttet til databruk, IKT-infrastruktur, lovverk og kunnskapsheving⁵⁵.

Flere eksperter og klinikere som er intervjuet peker på at de viktigste formålene med KI-løsninger vil være å bedre helse- og omsorgstjenester gjennom mer presis behandling, reduksjon av overbehandling og forbedret livskvalitet for pasienter. Dette gir økt pasientsikkerhet, unngåtte kostnader og frigjøring av tid til kvalifisert personell som kan fokusere mer på pasientbehandling. Det forventes også at bruk av KI sammen med store mengder helsedata kan gi bedre grunnlag for mer persontilpasset behandling fremover⁵⁶.

I dag er det primært snakk om å bruke KI som klinisk beslutningsstøtte og ikke til å ta autonome beslutninger. Det vil med andre ord ikke være KI-løsninger som erstatter mennesker, men heller jobber sammen med og støtter helsepersonell gjennom utføring av spesifikke forhåndsbestemte oppgaver. Dette kan for eksempel være et system som kan analysere og skille ut hvilke røntgenbilder som krever selvstendig vurdering av en radiolog. Dette fordi ansvaret for de medisinske beslutningene, selv med bruk av KI-løsninger, ligger hos helsepersonellet.

Figur 2: Relevante bruksområder for KI i helsesektoren med eksempelprosjekter



Bruksområdene representerer stadier i pasientforløpet gjennom kontakt med helsetjenesten, eller hvilket stadium i sykdomsutviklingen KI-prosjektet er ment å bistå i. Tabellen er inspirert av Nasjonalt senter for e-helseforsknings kategorisering av KI-prosjekter <https://ehealthresearch.no/kin/prosjekter>.

Eksempler på sentrale KI-anvendelser som det jobbes med i den norske helsesektoren i dag

Tabell 1: Oversikt over utvalgte eksempler på typer av KI-løsninger i dagens helsesektor^{57,58,59}

KI-anvendelse	Beskrivelse	Risikoen for bias	Eksempel-prosjekter
Bildeanalyse	Bildeanalyse som kan brukes til mønstergjenkjenning og tolkning av medisinske bilder innen radiologi anses som det mest lovende på kort sikt.	Studier viser at ubalanserte datasett som brukes for å trene opp dyplæringssystemer i medisinsk bildebehandling gir opphav til bias ved at det presterer dårligere for pasientgrupper som er underrepresentert i dataen. Når det skiftes mellom bilder av menn og kvinner skjer det små strukturelle endringer i bildene som påvirker prestasjonen til systemet.	BoneXpert: Beskrivelse: Medisinsk bildeanalyse for å detektere benalder og deretter gi riktig hormonbehandling. Status: BoneXpert er i dag CE-merket, og brukes av omtrent 200 sykehus globalt – inkludert seks helseforetak i Helse Sør-Øst. BoneXpert i dag den eneste KI-anvendelsen som er i bruk i behandling i Norge.
	Mange bildeanalyse-prosjekter er fortsatt under forskning. Likevel finnes det KI-løsninger som allerede er tatt i bruk innenfor radiologi i flere av landets helseregioner.		DoMore Diagnostics: Beskrivelse: Produktet Histotype Px Colorectal gir klinisk beslutningsstøtte som er i stand til å forutsi utfallet av en pasients kreft og kan veilede en mer presis behandling for hver pasient, og dermed unngå over- og underbehandling. Status: Produktet er CE-merket, og i samarbeidsprosesser med et universitetssykehus og en kunde

KI-anvendelse	Beskrivelse	Risikoen for bias	Eksempel-prosjekter
Språkanalyse (Natural language processing (NLP))	NLP handler om å analysere ustrukturerte data ved å gjøre tekst eller tale om til kodet, strukturert informasjon. NLP-systemer for bruk i helse er kun på forskningsstadiet.	Kjønnsdimensjonen er viktig for å utvikle effektive NLP løsninger i helse. Dette er fordi vi registrerer forskjeller i hvordan kvinner og menn snakker og skriver.	IKKS (informasjons-system for klinisk konseptbasert søk): Beskrivelse: datasystem som ved hjelp av språkanalyse søker etter kritisk informasjon om pasienters allergier fra deres journaler som beslutningsstøtte til å gjennomføre behandlinger eller operasjoner. Datasystemet har vist seg å oppfange allergier, samt se sammenhenger i pasientjournalene nesten like presist som erfarne leger, men mye hurtigere. Status: prosjektet har vært avsluttet grunnet de norske reglene for personvern og håndtering av helsedata, men det er mulig å søke om unntak for å få tilgang til dette.
			KI som beslutningsstøtte i medisinske nødansrop: Beskrivelse: Bruk av KI til å identifisere / bistå operatørene på AMK-sentralen til å identifisere hjertestans ved å sammenstille strukturert informasjon som finnes i systemet og forsøke å oppdage mønstre. Det skal ikke være et ekspertsystem som sier at denne pasienten har hjerneslag, men kunne gi signal til AMK operatøren om at pasienten har en historikk som sier at denne pasienten kan ha hjerneslag diagnose som en av flere ting. Status: Tidlig i forskningsstadiet

Bildeanalyse

Analyse og tolkning av medisinske bilder er i dag det området innen KI hvor utviklingen har kommet lengst, og hvor det er størst potensiale for bruk på kort sikt⁶⁰. Helsedirektoratet peker på at KI innen radiologi kan være særlig nyttig til: 1) utføring av tidkrevende og repeterende oppgaver ved prosessering av bilder, 2) beslutningsstøtte, 3) administrasjon og kvalitetssikring og 4) finne nye sammenhenger, også med nye datakilder⁶¹. Helsedirektoratet beskriver at KI innen radiologi kan hjelpe *“helsepersonell med å øke kvaliteten på arbeidet, som å planlegge undersøkelser basert på tidligere undersøkelser, få riktig posisjon på kroppsdelene som skal undersøkes, merke undersøkelsen rett og riktig eksponering av røntgenstråler”*⁶².

Det finnes flere grunner til at KI er egnet til bildeanalyse og mønstergjenkjenning. Store deler av utviklingen innen dyplæring har vært knyttet til bildeanalyse, og man har dermed utviklet avanserte algoritmer. I tillegg finnes det internasjonale standarder for medisinske bilder (DICOM®) som muliggjør samarbeid. Tilgjengeligheten av bilder til å trene opp algoritmene er også god⁶³.

En utfordring i norsk helsetjeneste er det lave antallet radiologer, og dermed behandlingstiden per pasient. Det er estimert at det vil bli en årlig økning på **12,5%** i antall bilder som må analyseres av radiologer, og raske bildeanalyseverktøy vil dermed spille en viktig rolle i å effektivisere dette arbeidet og håndtere økningen⁶⁴.

Det finnes både velkjente fysiologiske forskjeller, men også mulige forskjeller som man ikke kjenner godt nok til i dag mellom menn og kvinner. Dermed utgjør dette en viss risiko for kjønnsbias innen bildeanalyse dersom det ikke tas godt nok høyde for forskjellene.

Språkanalyse/Natural language processing (NLP)

Språkanalyse, eller natural language processing (NLP) på engelsk, er en betegnelse innenfor KI som handler om algoritmers evne til å forstå og effektivt analysere kompleks menneskelig kommunikasjon som tekst og tale⁶⁵.

Bruk av språkanalyse i helsesektoren har et enormt potensial fordi det raskt kan søke, analysere og tolke store mengder med informasjon om pasienter i ulike journaler. Dette gjør det mulig å raskt trekke ut kritisk informasjon. Slik kan rett anvendelse gi presis informasjon fra ustrukturert data, og dermed forbedre behandling og resultater for pasienter⁶⁶.

Imidlertid er gjennomføring av tekstanalyse en svært krevende øvelse fordi dataene ikke er standardisert. Det finnes også store språklige variasjoner mellom enkeltpersoner knyttet til variabler som geografi (dialekt), kjønn, etnisitet og andre sosiale faktorer. Det er et omfattende lovverk knyttet til personvern og helsedata. Dette førte til at samarbeidsprosjektet mellom Sørlandet sykehus og Universitetet i Agder, IKKS (Informasjonssystem for klinisk konseptbasert søk), i en lengre periode ble stanset fordi bruken av helsedata var lovstridig⁶⁷.

For å utvikle effektive verktøy for språkanalyse innen helse er det viktig at kjønnsdimensjonen er med. Dette skyldes at naturlig språkanalyse er en KI-kategori som er særlig utsatt for kjønnsbias, fordi man vet at språk inneholder mange markører for underliggende faktor som kjønn. Dette gjelder f.eks. ved bruk av maskinlæring på pasientjournal-data, eller for hvordan mannlige og kvinnelige pasienter selv beskriver sine symptomer (muntlig eller skriftlig)⁶⁸.

Sentrale utfordringer for innføring av KI i helsesektoren

Det er en rekke årsaker til at KI-løsninger ikke er utbredt i helsesektoren i Norge i dag, og at de fleste prosjekter befinner seg på forskningsstadiet. Direktoratet for e-helse har beskrevet flere av dagens sentrale utfordringer for innføring av KI i den norske helsesektoren⁶⁹, og denne rapporten trekker frem noen eksempler:

Definisjon av medisinsk utstyr

Medisinsk utstyr er ethvert instrument, apparat, utstyr, programvare, implantat, reagens, materiale eller annen gjenstand som ifølge produsenten er beregnet på å bli brukt, alene eller i kombinasjon, på mennesker med henblikk på ett eller flere av følgende spesifikke medisinske formål:

- diagnostisering, forebygging, overvåking, prediksjon, prognostisering, behandling eller lindring av sykdom
- diagnostisering, overvåking, behandling, lindring av eller kompensasjon for skade eller funksjonshemning
- undersøkelse, utskifting eller endring av anatomien eller av en fysiologisk eller patologisk prosess eller tilstand
- for å frambringe informasjon ved hjelp av in vitro-undersøkelse av prøvemateriale fra menneskekroppen, herunder organ-, blod- og vevdonasjoner

Forskrift om medisinsk utstyr⁷³

Regelverk

CE-merking

I dag blir mange forskningsorienterte KI-prosjekter ikke satt inn i operativ bruk, til tross for stort potensial på spesifikke utfordringer. Dette skyldes ofte at de ikke oppnår CE-merking. All teknologi som har et tiltenkt medisinsk formål, eller som fra produsentens side skal brukes på mennesker til å diagnostisere, forebygge, overvåke, behandle eller lindre sykdom, skade, eller handicap, er å betrakte som medisinsk utstyr under EUs reguleringer for medisinsk utstyr. Alt av slik teknologi og utstyr som skal brukes klinisk i Norge må CE-merkes og klassifiseres før det kan tas i bruk⁷⁰.

Veien til CE-merking er en krevende prosess for å gjøre produktene markedsklare, og er en forutsetning for utbredelse til flere HF/sykehus⁷¹. Helsedirektoratet viser til at det er et stort behov for økt kunnskap, forståelse og veiledning knyttet til regelverket⁷².

Proessen for å oppnå CE-merking kan være vanskelig for leverandører å navigere i. Den består også av flere steg, ifølge Nasjonalt senter for e-helseforskning⁷⁴. Først må programvaren sjekkes opp mot reguleringen for medisinsk utstyr for å se om den faller innenfor definisjonen som er satt av fag- og tilsynsmyndigheten - Statens Legemiddelverk. Risikoklassifisering må etableres, systemdokumentasjon må gjennomgås, og et ISO-godkjent kvalitetskontrollsystem må være på plass. Utstyret må igjennom en klinisk evaluering og resultatene må til fagfelleevaluering. Autorisert organ vurderer evalueringsrapporten, og hvis den blir godkjent, gis utstyret CE-merking og et ISO13485-sertifikat⁷⁵.

For programvare som faller innenfor definisjonen for medisinsk utstyr, er det vanlig at Statens Legemiddelverk bidrar opp mot prosjektet hovedsakelig i de tidlige fasene. En ny forskrift om medisinsk utstyr trådte i kraft 26. mai 2021, og har blitt oppdatert senest 26. mai 2022. På dette området skjer det kontinuerlig utvikling.

Det er viktig å forstå at selv om en CE-merking indikerer at programvaren oppfyller krav om medisinsk utstyrs prestasjon, kvalitet, effektivitet og sikkerhet, er det ikke garantert at algoritmene er bygget på data som er innhentet på etisk vis, eller at algoritmen vil gi like gode resultater til ulike pasientgrupper.

Personvern og GDPR

Flere norske utredninger peker på at manglende tilgang til data er en av de sentrale utfordringene som må løses for å oppnå effektiv bruk av KI i helsetjenesten i Norge⁷⁶. Tilgang på data av høy kvalitet som er relevant for pasientgruppen er helt nødvendig for å utvikle gode KI-systemer. Den juridiske utfordringen er at KI basert på maskinlæringsalgoritmer kan bruke personlige helseopplysninger. Dette er noe som skaper utfordringer mot GDPR sine krav til innsamling, bruk og oppbevaring av personopplysninger.

Dagens regelverk hjemler i liten grad å bruke helseopplysninger om én pasient til å gi helsehjelp til neste pasient, med mindre man har samtykke. Samtidig kan det gis dispensasjon fra taushetsplikten for å bruke pasientdata til forskning gjennom Regionale komitéer for medisinsk og helsefaglig forskningsetikk (REK). KI utfordrer skillet mellom forskning og pasientbehandling. Dette fordi det ofte vil være behov for å fremsette pasientopplysningene fra forskningen når verktøy basert på KI utviklet i et forskningsprosjekt skal tas i bruk i pasientbehandling. Dispensasjonen fra taushetsplikten vil da ikke lenger gjelde, og bruken av personopplysningene er ikke lenger lovlig.

Etiske utfordringer

Det er også en rekke utfordringer og risikoer knyttet til etiske vurderinger om ansvar, tillit og ikke-diskriminering som bør vurderes nøye ved innføringen av KI-løsninger. Den økte bevisstheten knyttet til blant annet bias kan være en forklarende faktor til hvorfor norsk helsetjeneste har en forsiktig tilnærming til implementering av KI.

En annen problemstilling relatert til etiske vurderinger er å forstå hvordan algoritmen lærer, gjennomfører beslutninger og utvikler seg over tid. Dette er spesielt viktig i helsesektoren fordi uten innsyn kan det være vanskelig å avdekke feil eller å forstå algoritmens beslutningsgrunnlag. Dette er noe som er viktig når det skal tas kliniske beslutninger. Eksempler på årsaker til dette er upresise metodebeskrivelser, lukket kildekode fra leverandør, private helsedata og ufullstendige evalueringer. Dermed er påliteligheten til algoritmene lavere for helsepersonell og pasienter⁷⁷. Manglende transparens og forklarbarhet beskrives ofte som “svart boks-” problematikk. Det vil finnes algoritmer som selskaper bevisst ikke gir innsyn i fordi dette er forretningshemmeligheter. Eller at oppbyggingen til algoritmen er for komplisert til å kunne forklares uten dyp ekspertkompetanse. Sistnevnte gjelder særlig ettersom KI-systemene blir mer datadrevet og baseres på nevralt nett⁷⁸. Dette gjør det krevende å forstå resultatene som systemet produserer. Også Helsedirektoratet understreker viktigheten av transparens i KI-løsninger. De beskriver at “når det gjelder etiske problemstillinger er det behov for å undersøke hvordan man kan sørge for transparente verdikjeder for å sikre etisk datainnsamling, trening og utvikling av kunstig intelligens⁷⁹.”

I dag finnes det derimot flere norske initiativer for å møte både de juridiske og etiske utfordringene. Helsedirektoratet har i samarbeid med Direktoratet for e-helse, Statens legemiddelverk og Statens helsetilsyn utarbeidet en landingsside som fungerer som en “informasjonsstartpakke” for forskere, utviklere og innkjøpere av KI innen helse⁸⁰. Denne gir både en oversikt over relevant regelverk, men også en oversikt over flere av de internasjonale etiske rammeverkene. I tillegg er Datatilsynets sandkasse et populært initiativ for å stimulere til innovasjon av etiske og ansvarlige KI-systemer fra et personvernsperspektiv. Sandkassen tilbyr gratis veiledning til utvalgte private og offentlige prosjekter. Et av områdene det legges til rette for gjennom sandkassen er å unngå diskriminering og bias ved utvikling av KI-systemer. De legger til grunn de samme etiske prinsippene som EUs etiske retningslinjer for pålitelig KI. Målet til sandkassen er at det skal utvikles mer etiske og ansvarlige KI-systemer. Sandkassen vil promotere en bedre forståelse av de regulatoriske kravene fra KI-baserte produkter og tjenester⁸¹. EU har også utarbeidet et initiativ kalt AI-WATCH for å følge opp sosioøkonomiske, juridiske og etiske konsekvenser av KI. Dette er også Norge en del av. I initiativet følger de også med på sektorielle områder, som blant annet helse⁸².

Anskaffelser av KI-løsninger til klinisk bruk

Nasjonalt senter for e-helseforskning beskriver at det overordnet finnes tre ulike måter å anskaffe KI-løsninger i norsk helsetjeneste i dag⁸³. Det første alternativet er å gjennomføre et forskningsprosjekt i en helseinstitusjon sammen med klinikere og teknologikyndige. Det andre er å danne et samarbeid mellom en helseinstitusjon og et privat teknologiselskap. Det tredje alternativet innebærer å direkte anskaffe en ferdig utviklet og CE-merket KI-løsning. Disse anskaffelsesalternativene byr på særegne fordeler og ulemper som beskrives i detalj i *“Implementation of artificial intelligence in Norwegian healthcare: The road to broad adoption”* (2022)⁸⁴.

I hovedtrekk eksisterer det noen sentrale avveininger knyttet til valg av anskaffelsesform innen KI. Viktige vurderinger er knyttet til hvor lang tid det går fra forskningsprosjekt settes i gang til det tas i bruk klinisk, og ressurs- og kompetansebehov opp mot tilgjengelig kapasitet ved institusjonen. Ved å utvikle egne løsninger oppnår man fordelene av å ha større nærhet til å kunne justere og forbedre løsningen. I tillegg til å kunne skreddersy løsningen til arbeidsflyten ved helseinstitusjonen. Dette fordi det er bygget på intern kompetanse. Derimot vil innkjøp av en ferdig utviklet og CE-merket KI-løsning kreve mindre intern kompetanse, og kunne føre til at man raskere kan ta løsningen i bruk. En vesentlig ulempe ved innkjøp av eksterne løsninger er at det ikke nødvendigvis er tilstrekkelig med transparens knyttet til datasettene som algoritmene er trent på, eller beslutningsprosessen til KI-løsningen. Dette kan føre til manglende tillit blant helsepersonell som skal ta løsningen i bruk, men øker også risikoen for uønskede bias.



4

Utfordringer knyttet til kjønnsbias i KI i norsk helsetjeneste

Risikofaktorer for kjønnsbias

Det er i dag risiko for negative konsekvenser knyttet til bruk av KI i helsesektoren. For eksempel er det avdekket at en rekke ulike bias kan oppstå gjennom stegene i prosessen – helt fra KI-utviklingen til den endelige kliniske bruken⁸⁵. Ved å ikke gjøre godt nok rede for biologiske og sosiokulturelle forskjeller mellom kjønnene, vil man oppnå suboptimale resultater, og risikere diskriminerende behandling⁸⁶.

Til tross for økende kjennskap til problemstillingene rundt bias, er ikke nødvendigvis kapasiteten for å teste dette på plass. Helsearbeidere er knappe ressurser som bruker store deler av sin tid på pasientbehandling. Som beskrevet tidligere er det komplekst å kontrollere resultatet av en KI-algoritme, noe som totalt sett gjør det krevende for helsepersonell å oppdage og håndtere kjønnsbias⁸⁷.

I denne studien har vi identifisert seks risikofaktorer for kjønnsbias i norsk helsetjeneste.



Det er en manglende bevissthet og forståelse av kjønnsbias blant leverandører av KI-løsninger til norsk helsesektor

En hovedutfordring for god bruk av KI innen helse er den manglende innsikten og forståelsen av de etiske dimensjonene til KI-løsningene blant leverandører. Dette inkluderer kunnskap om hvilke kjønnsbias som kan oppstå, og fremgangsmåter for å teste disse. Flere vi har intervjuet som jobber med å utvikle KI forteller at de ikke har gode rutiner for å undersøke dette, og at ingen har spesifikt ansvar for håndtering av bias.

Den lave bevisstheten og kunnskapen rundt bias er problematisk fordi den menneskelige dimensjon står så sentralt i KI-utvikling, og man bevisst må velge hvilke kriterier og bias man kan akseptere. Innen helse kan det være ekstra krevende fordi man ofte må behandle kvinner og menn forskjellig, og ta stilling til hvilke forskjeller som bør være der og ikke.

Disse funnene underbygger også andre lignende studier. Dette inkluderer blant annet den norske studien til Cathrine Bui som fant at mange KI-utviklere mangler kunnskap og bevissthet om hvordan kjønnsdimensjonen spiller inn⁸⁸. Flere utviklere i denne studien antok at deres egne algoritmer er nøytrale og dermed "rettferdige". KI-utviklerne som ble intervjuet i studiene "kunne lite om etiske problemer", og antok at dette er uproblematisk ettersom deres teknologi ble oppfattet som "nøytral". Stemmegjenkjenning er blant KI-anvendelsene som ikke ble testet på begge kjønn, noe som kan være problematisk ettersom kvinner og menn kommuniserer ulikt. Flere manglet kunnskap om hva som kunne føre til diskriminering og hva de burde teste algoritmene for⁸⁹.

Studien inkluderte også personer som jobber direkte eller indirekte med helse, og flere manglet kunnskap om hva som kunne føre til diskriminering, og hva de burde teste algoritmene for. I tillegg bekrefter internasjonal faglitteratur at, til tross for de store fremskrittene innen maskinlæring og KI, har mye av dagens teknologi ikke tilstrekkelig fokus på å oppdage bias. I utviklingen av algoritmer blir ofte kjønnsdimensjonen, og påvirkningen dette har på helse- og sykdomsforskjeller mellom individer, ignorert⁹⁰.

Mangelfull bevissthet knyttet til kjønnsbias kan skyldes en rekke forhold. Sannsynlige forklaringer inkluderer blant annet at flere menn enn kvinner er utviklere, at vi har utilstrekkelig forskning på kjønnsforskjeller, samt et generelt lavt kompetansenivå om bias blant beslutningstakere. Videre hevder Bui at underliggende problemer for dette inkluderer feil data, vurderingene til de som velger ut dataen, samt troen på at det finnes en gullstandard for hva som er nøytralt (uten å ta innover seg alle bias).

Samtidig beskriver våre intervjuobjekter at de opplever at det stadig blir en større forståelse for bias innen KI og helse, at man er bevisste på problemstillingene, men hevder at ingen ennå har funnet en god måte å balansere det på. De erkjenner at det ikke er lett å unngå at algoritmer kan ha uønsket bias, men peker på at det er viktig å bygge opp en forståelse for utfordringene det gir.



For leverandører av KI-løsninger er det manglende insentiver til å prioritere biastesting (inkl. kjønnsbias), utover krav satt av CE-merking

Dagens insentivsystem tilrettelegger ikke for å fokusere på bias fra starten av utviklingen. Våre funn tilsier at flere leverandører primært ønsker å utvikle KI-løsninger som presterer godt på de tiltenkte oppgavene, og som tilfredsstiller kravene som stilles for å bli CE-merket. Dette er problematisk fordi bevissthet om bias helt fra starten av utviklingsfasen kan gjøre det enklere å forstå hvor bias oppstår.

I dag prioriteres det å få tak i tilstrekkelig helsedata som datagrunnlag, deretter å trene opp algoritmen på dataene til å utføre en spesifikk oppgave, og til slutt å få produktet CE-merket.

I intervjuer kommer det frem at de som leverer KI-løsninger ofte må nøye seg med datasettene de har tilgjengelig, og at en sentral utfordring er at man ikke i større grad kan velge utformingen av datasett. Deretter er neste prioritet å produsere et produkt som løser det aktuelle problemet, og som kan CE-merkes. Etablert praksis er å teste for bias knyttet til for eksempel kjønn først som en del av innkjøpsprosessen.

Videre beskriver KI-leverandører selv at det å undersøke for bias er en krevende oppgave. Et norsk KI-selskap hevder at å legge for mye vekt på å oppdage kjønnsbias i en tidlig fase av produktutviklingen kan hindre fremdriften mot å lansere et markedsklart produkt. Selskapet mener at det kan føre til svekket konkurransefortrinn hvis man ikke utfører kvalitetskontroller på andre områder. Hvis konkurrerende selskaper lar være å teste, gir det trolig mange aktører dempet insentiv til å gjennomføre slike aktiviteter.



Å undersøke for bias i en tidlig fase kan begrense vår vei mot et markedsklart produkt.

Norsk KI-leverandør



Få bruker etiske rammeverk i utviklingen av KI-løsninger

På spørsmål om bruk av etiske prinsipper eller retningslinjer som EUs etiske rammeverk, finner vi at det er få som har egne rutiner for, eller som følger disse rammeverkene. I dag finnes det mange ulike rammeverk, retningslinjer og prinsipper for KI. I tillegg til at disse utarbeides for ulike fagområder, har de også varierende overlapp i innhold, og det er dermed ikke lett å velge hvilket man skal bruke. I tillegg er de generelt beskrevet på et overordnet nivå og i generelle termer. Det er dermed ikke nødvendigvis enkelt å vite hvordan man skal bruke disse i praksis. Eksempelvis kan et etisk rammeverk beskrive detaljert hva 'rettferdig KI' er, men ikke presentere spesifikke tiltakene for å teste for dette. Det er med andre ord behov for en sektortilpasset konkretisering som KI-utviklere kan benytte. Det er i dag stor interesse for dette området, og det diskuteres jevnlig hvilken tilnærming som er mest hensiktsmessig.

Spørsmålet om hvem som bør ha ansvar for å utarbeide en slik guide til bruk av etiske rammeverk er krevende fordi det stiller høye krav til helsefaglig og medisinsk kompetanse i tillegg til en dyp forståelse av kunstig intelligens. Dette bør også kombineres med god forståelse for de sosiale og etiske aspektene. Intervjuobjekter er heller ikke sikre på om én enkelt instans bør ha dette ansvaret.



Innkjøpere av KI-løsninger tar i en viss grad for gitt at bias har blitt håndtert gjennom CE-merkingen

Gjennom våre intervjuer har vi funnet at innkjøpere av KI-løsninger i en viss grad stoler på at dataene er testet og er pålitelige gjennom CE-merkingen. En innkjøper av KI-produkter for et stort helseforetak forklarer at de stoler på at det allerede er godkjent, korrigert for bias, og vil fungere godt. Flere intervjuobjekter forteller også at de ikke har noen med spesifikt ansvar for etiske vurderinger knyttet til teknologien.

Derimot finnes det også et eksempel på en innkjøper som har tatt en mer aktiv tilnærming og forespurt spesifikke beskrivelser av hvordan produktene tilfredsstiller anbefalingene til EUs etiske rammeverk. Dette eksempelet er Vestre Viken HF som har anskaffet en KI-løsning fra Phillips til beslutningsstøtte i tolkningen av radiologisk billedmateriell. I denne prosessen fikk de leverandøren til å svare på hvordan EUs etiske rammeverk var hensyntatt. Deretter vil Vestre Viken gjennomføre en medisinsk metodevurdering (HTA) som beslutningsgrunnlag for å oppsummere dagens forskning, og for å prøve å finne svar på etikk. Videre vil det gjøres en retrospektiv validering og en MS-kontroll for å sjekke om KI-plattformen fungerer som ønsket før den kan tas i bruk fremover.



Det er uklart hvor ansvaret for validering ligger i dag, og hvordan ansvarsfordelingen ideelt sett bør være

Gjennom flere intervjuer beskrives det at det trolig ikke er gode nok og standardiserte prosesser for validering blant innkjøpere av KI-løsninger. Med validering menes her Helsedirektoratets definisjon om “helseinstitusjonens evne til å finne ut om utstyret er egnet for den aktuelle pasientgruppen eller for bruk i den aktuelle helseinstitusjonen”⁹¹.

Flere aktører i helsesektoren mener også at det er uklart hvem som bør ha ansvar for å oppdage og håndtere bias, og hvor ansvaret for validering bør ligge. Noen mener at ansvaret for validering trolig bør være delt mellom helseforetakene og nasjonalt plan.



“Ansvaret for validering bør ligge dels hos oss som tar det i bruk og på nasjonalt nivå. Leverandørene kommer bare til å prøve å få CE-merkingen (...), men vil ikke ta ansvar for hvordan det fungerer videre. Vi må selv ta ansvar for validering på hvert sykehus. Det bør finnes nasjonale føringer for hva som skal kreves for å kunne ta det i bruk (krav til antall datasett og representasjon for pasientgruppen det skal brukes på osv.) Det bør være samme testing rundt i Norge, så det gjelder å finne normer på dette.”

Innkjøper av teknologi i et stort norsk helseforetak



I dag har algoritmene til flere KI-løsninger manglende forklarbarhet og tolkbarhet. Denne manglende transparens kan hindre at man oppdager hvor kjønnsbias oppstår

Flere intervjuobjekter uttrykker at bedre forklarbarhet og tolkbarhet til KI-løsningene er viktige faktorer for å forstå beslutningsprosessen, hva dataen inneholder, og om det kan finnes bias. Dette er viktig fordi det å lykkes med å ta i bruk KI-verktøy i norsk helsetjeneste, krever at klinikerne har god tillit til verktøyene og resultatene disse produserer.

Denne rapporten retter søkelyset mot utfordringen knyttet til at klinikere muligens ikke har god nok forståelse av KI-verktøyenes virkemåte og resultater. Dette utgjør en risiko for at de ikke får nok tillit til dette, og kan ha vanskeligheter med å forklare dette til pasienter. Det vil ikke alltid være mulig for helsepersonell å ha en dyp forståelse av hvordan KI-systemene fungerer. I dag varierer de ulike KI-algoritmene som det forskes på i grad av kompleksitet, men særlig nevralt nettverk er komplekse og krever en dypere forståelse for helsepersonell⁹². På den ene siden kan slike KI-løsninger som er basert på nevralt nettverk og mindre forklarbare algoritmer prestere bedre og mer presist enn andre KI-løsninger. På den andre siden eksisterer det en avveining hvis ikke beslutningsprosessene eller resultatene kan forklares like godt, og de dermed oppleves som mindre pålitelige. For at KI-løsningene skal tas i bruk er det nødvendig både med tillit og en viss evne til å forklare resultater for pasienter. Dette er viktig fordi pasienter har rett til å få forklaring på hvorfor en viss behandling velges - også når et KI-verktøy er involvert som beslutningsstøtte. I en større undersøkelse fra Helsedirektoratet basert på dialog med de regionale helseforetakenes brukerutvalg og Kreftforeningen, kommer det frem at "pasientene ønsker å vite hvordan produktene virker og hvor presise de er i forhold til dagens metoder".⁹³

Manglende transparens og forklarbarhet beskrives ofte som "svart boks"-problematikk. En KI-ekspert som er blitt intervjuet hevder derimot at denne problematikken i stor grad er misforstått – særlig kategoriseringen av hva som betegnes som "svart boks" og ikke. KI-løsningene som brukes i helsetjenesten i Norge i dag er ikke nødvendigvis en "svart boks" fordi det er mulig å forstå hva algoritmene gjør, men det vil være ulike nivåer av forståelse mellom en utvikler og en helsearbeider.

I fravær av tilgang på norsk helsedata, blir dessuten ofte utenlandsk data brukt, og det knytter seg dermed større usikkerhet til hvor godt den vil fungere for en norsk populasjon, og hvilke bias som kan finnes dersom algoritmene ikke er trent på en like relevant populasjon.

Det finnes eksempler på bruk av KI innen helsetjenesten som understreker viktigheten av forklarbarhet. At det blir stadig viktigere å forstå beslutningsprosessen for helsepersonell som tar i bruk KI som beslutningsstøtteverktøy, ble tydelig ved bruk av IBM sitt KI-system, IBM Watson for Oncology. Dette verktøyet har flere ganger anbefalt feil behandling av kreftpasienter, og dette kom blant annet frem ved forsøk gjennomført på Rigshospitalet i Danmark. Der hadde IBM Watson en svært høy feilrate på anbefalingene, og ifølge en direktør ved Rigshospitalet foreslo det “en behandling med stoffer som pasienten aldri mer skulle ta i sitt liv, fordi det var livsfarlig”⁹⁴. I slike tilfeller ville det vært nyttig for helsepersonell å kunne forstå og undersøke hvorfor og hvordan KI-løsningen kom frem til akkurat den avgjørelsen. Med en innsikt i dette, ville legen kunnet gå inn og korrigere der det var nødvendig. Derimot var IBM Watson-løsningen ikke en transparent løsning som kunne forklares, fordi algoritmene var en forretningshemmelighet. Dette førte til at hele avtalen med IBM Watson ble avsluttet.

Et forskningssenter som ble intervjuet understreker viktigheten av forklarbarhet, og hevder at særlig tolkbarhet har blitt løftet opp på agendaen gjennom en oppvåkning rundt disse problemstillingene. Senteret peker videre på at forklarbare modeller kan gjøre det enklere å forstå om det er diskriminering knyttet til kjønn.





5

Hvordan bedre forstå og motvirke uønskede kjønnsbias i KI

Anbefalinger til håndtering av kjønnsbias

Basert på dokumentgjennomgang og gjennomførte intervjuer, presenterer vi seks anbefalte tiltak for å fremme trygg, rettferdig og ikke-diskriminerende bruk av KI i den norske helsetjenesten. I stor grad er disse tiltakene rettet mot bevisstgjøring, kunnskapsbygging, og tydeliggjøring av ansvar for de forskjellige aktørene og interessentene. Disse tiltakene må iverksettes på flere nivåer: Fra nasjonalt hold, og helt ned til det enkelte utviklerteam. Samarbeid mellom interessentene i helsesektoren er også en nøkkelforutsetning for å lykkes.



Mangfold og tverrfaglig kompetanse i team som utvikler og leverer KI-løsninger

Mangfoldige team med komplementær kompetanse har høyere sannsynlighet for å avdekke fordommer⁹⁵. Å øke andelen av kvinner som deltar i utviklingen av KI vil bidra til å oppdage bias, og bedre kjønnsbalanse i utviklingsprosjekter vil føre til økt bevissthet rundt problemstillingene.

Bruk av KI i klinisk sammenheng er svært komplekst, og det kreves derfor også faglig komplementær kompetanse. Utvikling av gode KI-løsninger til bruk i helsesektoren krever tverrfaglig kompetanse innen medisin, teknologi, samt samfunnsvitenskapelige disipliner, juridiske fag og etikk. I dag gjøres dette blant annet gjennom samarbeid mellom KI-forskningsmiljøer og sykehus (f.eks. Sørlandet Sykehus og Centre for Artificial Intelligence Research (CAIR) ved UiA).

Mangfoldige team er også noe som løftes frem som et viktig tiltak i rapporten til Nasjonalt senter for e-helseforskning *“Implementation og artificial intelligence in Norwegian healthcare: The broad adoption”*, om hvordan KI skal benyttes i norsk helsesektor. Rapporten beskriver viktigheten av at IT-eksperter og helsepersonell samarbeider i utviklingen av KI-systemer for helse, og at det legges til rette for en interaktiv utviklings- og kommunikasjonsprosess. Dette gjelder også behovet for tverrfaglig kompetanse i benyttelsen av KI i helsetjenesten.



Kunnskapsdeling og bevisstgjøring om kjønnsbias

Økt kunnskapsdeling og bevisstgjøring om kjønnsbias er viktig for å håndtere risikoen. Det er i dag en manglende forståelse av problemstillingene knyttet til bias blant leverandører av KI-løsninger. Det er derfor viktig å heve kompetansen blant utviklere av KI-løsninger. Det bør tilbys opplæring innen etiske, sosiale og helsespesifikke aspekter ved KI, i studier innen KI og maskinlæring.

Samtidig er det også sentralt at klinikere som tar i bruk verktøyene opparbeider seg kompetanse om utfordringene knyttet til KI og bias. Det er krevende å oppdage og å håndtere kjønnsbias da dette stiller krav både til tid og kompetanse⁹⁶. Erfaring og kunnskap om hvilke bias som er oppdaget og hvordan disse har blitt håndtert, bør tilgjengeliggjøres slik at andre kan lære av det. Det er også viktig med gode samarbeidsfora og strukturer slik at erfaringene som høstes kan deles. Som påpekt i Nasjonal strategi for KI, bør Norge fortsette å aktivt delta i internasjonale fora som tilrettelegger for samarbeid om ansvarlig og pålitelig KI-anvendelse.

Denne kunnskapsdelingen og bevisstgjøringen bør skje blant etablerte nettverk for KI i den norske helsetjenesten, og vår anbefaling er at sekretariatet i Kunstig intelligens i norsk helsetjeneste (KIN) inkluderer dette tiltaket i nettverkets formål.



Økt grad av forklarbarhet, sporbarhet og transparens i KI-løsningene som et konkurransefortrinn

Det er behov for økt grad av forklarbarhet, sporbarhet og transparens i KI-løsningene. For klinikere er det viktig å forstå beslutningsprosessen, relevansen til treningsdataene, og om det fungerer godt for deres pasientgruppe. Dette er viktig fordi det å lykkes med å ta i bruk KI-verktøy krever at klinikerne har tillit til verktøyene og resultatene av dem. I tillegg har pasienter rett til å være aktive deltagere i sitt behandlingsforløp og i valgene som gjøres som en del av forløpet.

Utviklere må legge til rette for dette gjennom transparens om produktene sine og hvilken data som benyttes, men også ved å utvikle egnede visualiserings- og forklarbarhetsverktøy. Ifølge et intervjuobjekt vil det å gjøre et produkt mer forklarbart komplisere utviklingen, og det er derfor ikke insentiver til å gjøre dette med mindre kunder spesifiserer det eksplisitt. Samtidig er det viktig at innkjøpere av KI-løsninger stiller krav om forklarbarhet ved anskaffelser.

EUs ekspertgruppe for KI har fremmet viktigheten av at europeiske KI-leverandører utvikler pålitelige og trygge KI-løsninger for å kunne oppnå et konkurransefortrinn vis-a-vis konkurrenter med lavere grad av transparens. Økt grad av transparens vil gi bedre produkter og dermed være et konkurransefortrinn.

Vår anbefaling er at leverandører av KI-løsninger gjør sporbarhet og forklarbarhet til en del av produktenes funksjonalitet.



Tilgjengeliggjøring av etiske rammeverk gjennom praktiske guider

Det er få som bruker etiske prinsipper eller retningslinjer (slik som EUs etiske rammeverk), og det er behov for veiledning til praktisk bruk. Mange av dagens rammeverk, retningslinjer og prinsipper for KI er skrevet på høynivå og for tverrsektoriell bruk. Noe som gjør at det er krevende for utviklere å ta disse i bruk. For at KI-utviklerne skal få en konkret forståelse av de etiske vurderingene, er det derfor behov for konkretisering. Dette vil kunne bidra til å redusere “avstanden” mellom f.eks. EUs etiske rammeverk og den praktisk KI-utviklingen i Norge.

EU utviklet i 2020 en selvevalueringsliste for pålitelig KI (Assessment List for Trustworthy Artificial Intelligence (ALTAI)). Dette er et godt steg i riktig retning⁹⁷. Denne type verktøy bør tilpasses helsesektoren. Det er ikke åpenbart hvor dette ansvaret bør ligge. Samtidig som arbeidet er krevende fordi det stiller høye krav til helsefaglig kompetanse og kunstig intelligens – i kombinasjon med god forståelse for de sosiale og etiske aspektene.

Vår anbefaling er at nasjonale helsemyndigheter i samarbeid med likestillingsmyndigheter utarbeider praktiske guider til bruk for leverandører og innkjøpere. Dette bør koordineres og eies av det nasjonale koordineringsprosjektet for kunstig intelligens i norsk helsetjeneste “Bedre bruk av kunstig intelligens”.



Bias-testing som del av risikovurderingen og utarbeidelse av retningslinjer

For å sikre at bias blir håndtert grundig, bør biastesting inngå som en egen del av risikovurderingen. Dette for å oppdage bias knyttet til bl.a. kjønn, etnisitet og geografi. I dag er utviklere av KI pålagt å gjennomføre risikovurderingen knyttet til personvern og informasjonssikkerhet. Et godt grep kan være å inkludere bias-testing i risikovurderingen. Dette forskes det nå på ved universitetet i Tromsø.

Vår anbefaling er at nasjonale helsemyndigheter utarbeider retningslinjer til bruk for leverandører og innkjøpere. Dette bør koordineres og eies av det nasjonale koordineringsprosjektet for kunstig intelligens i norsk helsetjeneste "Bedre bruk av kunstig intelligens". Her kan koordineringsprosjektet for eksempel ta i bruk erfaring fra implementeringsarbeidet til arbeidet med Normen for informasjonssikkerhet i helsesektoren.



Tydeligere forklaring av hva et CE-merket produkt innebærer, og hvilke krav som stilles til den som skal ta det i bruk

Til tross for at CE-merkingen er en etablert standard for at KI-løsninger får godkjent markedstilgang, betyr ikke denne merkingen at produktet har tatt høyde for bias som beskrevet i denne rapporten. Flere personer som har blitt intervjuet stiller spørsmål ved hvor godt ordningen med CE-merking fungerer i dag, og peker på at CE-merkingen ikke er god nok for å håndtere kjønnsbias.

Det er en manglende forståelse for hva CE-merkingen dekker, og for hvilke krav som stilles til den som tar dette i bruk. Dette kan føre til at innkjøpte KI-løsninger som er CE-merkede blir tatt i bruk som om de hadde blitt testet godt nok for bias. Vår anbefaling er at nasjonale helsemyndigheter utarbeider informasjonsmateriell til bruk for leverandører og innkjøpere. Dette bør koordineres og eies av det nasjonale koordineringsprosjektet for kunstig intelligens i norsk helsetjeneste "Bedre bruk av kunstig intelligens".

APPENDIKS 1 →

Bidragstere til rapporten

PA Consulting og CARE Norge ønsker å takke alle som har bidratt til denne rapporten:

Nasjonale eksperter:

- Direktorat for e-helse
- Helsedirektoratet
- Nasjonalt senter for e-helseforskning
- Norsk radiologisk forening
- Senter for pasientnær kunstig intelligens (SPKI)
- Statens legemiddelverk (SLV)

Akademia:

- OsloMet
- Universitet i Tromsø
- Universitetet i Agder
- Universitetet i Oslo

Helseforetak:

- Aker sykehus
- Helse Sør-Øst
- Oslo Universitetssykehus
- Sykehuspartner
- Sørlandet Sykehus
- Vestre Viken HF
- Ålesund sykehus

KI-leverandører, KI-prosjekter og KI-miljøer:

- AI Mind (OUS)
- BoneXpert
- DoMore Diagnostics
- EKG AI (Akershus universitetssykehus)
- Medsensio
- Prosjekt "KI som beslutningsstøtte i medisinske nødannop"
- Computational Radiology & Artificial Intelligence (CRAI)
- LIFELOTS Exchange Extended (LEE)
- Norwegian Artificial Intelligence Research Consortium (NORA)
- Norsk råd for digital etikk (NORDE)
- Senter for forskning på kunstig intelligens (CAIR)

APPENDIKS 2 →

Bias typer

Bias i datagrunnlaget	
Historisk bias	Historiske data eksisterer i den kontekst de ble samlet inn. Eldre data kan inneholde grupperinger som har hatt mindre tilgang til ressurser eller hatt lavere prioritering historisk sett, slik at modeller utviklet på denne dataen propagerer uønsket praksis.
Registreringsbias	Registrerte data skal ideelt sett gi en riktig fremstilling av en tilstand på registreringstidspunktet. Allikevellikevel kan den sosiale konteksten for registreringen påvirke hvilken og hvordan informasjon blir registrert, derav oppstår en registreringsbias.
Målebias	Målebias oppstår når det er en tilfeldig eller systematisk målefeil under datainnsamling, eller ved bruk av en unøyaktig nærliggende variabel (variabler som ikke måler direkte observerbar data). Dette kan da føre til at dataen blir en tilnærming eller overforenkling av virkeligheten.
Representasjonsbias	Representasjonsbias oppstår når en modell bygger på data som ikke representerer populasjonen den skal brukes på. Dette kan oppstå der dataen som brukes er mangelfull, utvelgelsen ikke er tilfeldig, eller at utvalgspopulasjonen ikke samsvarer med populasjonen modellen skal brukes på.
Utvalgsbias	Utvalgsbias er et spesialtilfelle av representasjonsbias, der utvelgelsen av observasjoner ikke er tilfeldige, og dermed er datasettet ikke representativt for dataen som skal brukes

Modellvalg bias	
Læringsbias	Læringsbias oppstår når man velger tapsfunksjon for maskinlæringsmodellen. Når én tapsfunksjon er optimalisert, går det på bekostning av de andre tapskriteriene. Dette skjer når modellen er bygget på et helt datasett, og går på bekostning av de underrepresenterte gruppene i dataen.
Referansebias	Referansebias oppstår når ferdige referansedata brukes til å evaluere maskinlæringsmodellens resultater. Dette skjer ved at modellen som produserer de beste resultatene på referansedataen foretrekkes, uten at denne modellen generaliserer best til dataen modellen brukes på etter produksjon-settingen.
Aggregeringsbias	Aggregeringsbias oppstår når en modell er bygget på en ekstrapolering fra en majoritetsbefolkning, derav antar at undergrupper i dataen deler egenskaper og attributter med majoriteten.
Konseptdrift	Konseptdrift er et fenomen som oppstår som følge av endringer over tid. Der relasjonen mellom forklaringsvariablene og den forklarte variabelen endres. Hvis en maskinlæringsmodell ikke tar høyde til denne endringen vil ytelsen gradvis bli verre når endringene blir større. Konseptdrift er vanskelig å skille fra vanlig støy i datasettet, eller utliggere i dataen.

APPENDIKS 3 →

Sentrale etiske rammeverk

WHO: Veiledning for bruk av etiske prinsipper for bruk av KI i helse (oversatt og kort forklart)⁹⁸.

Prinsipp	Hva
Beskytte menneskelig autonomi	Mennesket bør ha kontroll over de medisinske beslutningene som tas ved å ha tilstrekkelig med informasjon om systemet, samt sikre at innhenting av informasjon foregår på en måte som er i tråd med personvern og konfidensialitet.
Fremme menneskers velvære og sikkerhet	KI skal ikke skade mennesker, og de som utvikler systemene må utvikle dem i tråd med regulatoriske krav til sikkerhet og kvalitetskontroller.
Åpenhet, forklarbarhet og forståelighet	Dette innebærer at KI-systemet bør være forståelig for utviklere og brukere av systemet. Dette krever dokumentert og tilgjengelig informasjon om hvordan teknologien er utviklet samt hvordan den skal brukes og ikke.
Ansvarlighet	Selv om KI-systemet tar en beslutning, er ansvaret for beslutningen hos ansvarlige personer og ikke hos teknologien.
Sikre inkludering og rettferdighet	Systemene må utvikles for bredest mulig bruk og tilgang uavhengig av alder, kjønn, inntekt, etnisitet, seksuell legning eller andre faktorer. Det må ikke utvikles skjevheter i systemet som gir ulemper for noen grupper eller personer.
Fremmer bærekraft	Minimere miljøkonsekvenser og øke energieffektivitet.

EUs etiske retningslinjer for KI

EU har utviklet etiske retningslinjer for utvikling av troverdige KI-systemer, hvor disse består av tre punkter som bør etterkommes i hele KI-systemets livssyklus:

- systemet bør følge alle gjeldende lover og regler
- det bør være etisk og i tråd med etiske prinsipper og verdier
- systemene bør være robuste både fra et teknisk og sosialt perspektiv med gode intensjoner⁹⁹.

Basert på dette har EU-kommisjonen utviklet syv krav som KI-systemer bør etterkomme for å være troverdige. Disse innebærer ansvarlig og etiske utvikling av kunstig intelligens. Det er disse prinsippene regjeringen baserer seg på ved utvikling av kunstig intelligens i Norge¹⁰⁰.

Krav	Hva
Menneskelig handlekraft og tilsyn	KI kan gi mennesker bedre tilgang til personopplysninger eller øke tilgjengeligheten til utdanning. Samtidig kan KI ha en negativ effekt på grunnleggende rettigheter dersom dataen ikke behandles godt nok. Mennesker skal også kunne bidra i beslutningsprosessene og gi tilbakemeldinger gjennom styringsmekanismen "human-in-the-loop".
KI-systemet skal være teknisk robust og utvikles forebyggende mot risiko	Systemene skal utvikles til å være nøyaktige, pålitelige og etterprøvbare.
KI-systemer skal prosessere data på en måte som tar hensyn til personvern	Data som samles inn og brukes i KI-systemene skal ikke brukes på en diskriminerende eller ulovlig måte. Kvaliteten på dataen som sames inn må også undersøkes for om den inneholder sosialt konstruerte skjevheter, unøyaktigheter eller feil.
KI-systemer skal være gjennomsiktige	Alle beslutninger som systemet tar skal være sporbare og forklarbare. Det vil si at beslutningen som KI-systemet gjennomfører skal være forståelig og kunne etterprøves. Det må også kommuniseres til bruker av systemet at det kommuniseres med et KI-system.
KI-systemer skal legge til rette for mangfold og rettferdig behandling	KI-systemer kan inneholde ukomplette datasett som for eksempel at de inneholder historiske bias. Dette kan føre til diskriminering mot enkeltpersoner eller grupper. Derfor bør bias fjernes fra datainnsamlingsfasen. Ved utvikling av KI-systemer kan algoritmene også inneholde skjevheter. Dette kan motvirkes ved å utvikle tilsynsprosesser som analyserer systemets formål, begrensninger, krav og beslutninger på en transparent måte.
KI-systemer skal utvikles for å være til nytte for samfunnet og miljøet	Systemene bør derfor utvikles i tråd med bærekraftsmålene.
Ansvarlighet	Det skal utvikles mekanismer som sikrer ansvarlighet for KI-systemet gjennom hele livssyklusen til systemet.

Kilder

- 1 United Nations. (2022, 11 28). United Nations - Department of Economic and Social Affairs. Hentet fra Sustainable Development Goals: <https://sdgs.un.org/goals>
- 2 Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., . . . Valencia, A. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. npj digital medicine.
- 3 Helse og omsorgsdepartementet. (2020). Nasjonal helse og sykehusplan 2020-2023.
- 4 Helse-vest. (2022). Fremtidens helse og omsorg 2030.
- 5 Direktoratet for e-helse. (2019). Utredning om bruk av kunstig intelligens i helsesektoren.
- 6 Direktoratet for e-helse. (2022). Innsiktsrapport: Behov for data til kunstig intelligens i helsetjenesten. Oslo: Direktoratet for e-helse.
- 7 Direktoratet for e-helse. (2019). Utredning om bruk av kunstig intelligens i helsesektoren.
- 8 Regjeringen. (2020). Nasjonal strategi for kunstig intelligens.
- 9 EU. (2019). Ethics Guidelines For Trustworthy AI.
- 10 Ibid.
- 11 Regjeringen. (2020). Nasjonal strategi for kunstig intelligens.
- 12 Direktoratet for e-helse. (2019). Utredning om bruk av kunstig intelligens i helsesektoren. Direktoratet for e-helse. Retrieved from Utredning om bruk av kunstig intelligens i helsesektoren: <https://www.ehelse.no/publikasjoner/utredning-om-bruk-av-kunstig-intelligens-i-helsesektoren>
- 13 Bakkevoll P.A., M. A. (2018). Health analytics: Kunstig intelligens –nye muligheter for helsetjenesten. Oslo: Nasjonalt senter for e-helseforskning.
- 14 Teknologirådet. (2018). teknologiradet.no. Retrieved from Kunstig intelligens - muligheter utfordringer og en plan for Norge: <https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf>
- 15 SNL. (2021). SNL.no. Retrieved from ekspertsystemer: <https://snl.no/ekspertsystem>
- 16 Bakkevoll P.A., M. A. (2018). Health analytics: Kunstig intelligens –nye muligheter for helsetjenesten. Oslo: Nasjonalt senter for e-helseforskning.
- 17 IBM. (n.d.). ibm.no. Retrieved from Supervised Learning: <https://www.ibm.com/cloud/learn/supervised-learning>
- 18 Teknologirådet. (2018). teknologiradet.no. Retrieved from Kunstig intelligens - muligheter utfordringer og en plan for Norge: <https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf>
- 19 Ibid.
- 20 Kleppe, A., Skrede, O.-J., Raedt, S., Liestøl, K., Kerr, D., & Danielsen, H. (2021). Designing deep learning studies in cancer diagnostics
- 21 IBM. (2020). ibm.com. Hentet fra Deep Learning: <https://www.ibm.com/cloud/learn/deep-learning>
- 22 Ibid.
- 23 Oxford Languages. (2022, 25 11). Oxford Languages and Google. Hentet fra Definition Bias: <https://languages.oup.com/google-dictionary-en/>

- 24 Gerke, S., Minssen, T., & Glenn, C. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. National Library of Medicine.
- 25 Direktoratet for e-helse. (2019). Utredning om bruk av kunstig intelligens i helsesektoren. Direktoratet for e-helse. Retrieved from Utredning om bruk av kunstig intelligens i helsesektoren: <https://www.ehelse.no/publikasjoner/utredning-om-bruk-av-kunstig-intelligens-i-helsesektoren>
- 26 Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., . . . Valencia, A. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. npj digital medicine.
- 27 Gerke, S., Minssen, T., & Glenn, C. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. National Library of Medicine.
- 28 Cirillo, D., Catuara-Solarz, S., Morey, C., Guney, E., Subirats, L., Mellino, S., . . . Valencia, A. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. npj digital medicine.
- 29 Reuters. (2018). reuters.com. Hentet fra Amazon scraps secret AI recruiting tool that showed bias against women: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCNIMK08G>
- 30 Språkrådet. (2017). termwiki.sprakradet.no. Hentet fra Kjønnssfordom: <http://www.termwiki.sprakradet.no/wiki/Kjønnsfordom>
- 31 Straw, I., & Wu, H. (2022). Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction. National Library of Medicine. KI-systemet var trent på datasett fra indiske pasienter (Indian Liver Patient Dataset). Studien viste i første omgang at algoritmene hadde en suksessrate på 70% for å identifisere leversykdom. Videre testet de for om algoritmen identifiserte like godt for menn og kvinner, og det viste seg da at algoritmen overså 44% av tilfellene for kvinner og 23% av tilfellene for menn. Studien var basert på fire algoritmer, hvor de to algoritmene som var best til å diagnostisere sykdom totalt sett viste seg å ha størst kjønnsforskjeller og presterer dårligere for kvinner enn menn. Det betyr at selv om algoritmer viser seg å være mer presise må det undersøkes hvem algoritmene er presise for.
- 32 Cirillo, D. C.-S. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ digital medicine, 1
- 33 Ibid.
- 34 Teknologirådet. (2021). KUNSTIG INTELLIGENS I KLINIKKEN Seks trender for fremtidens helsetjeneste.
- 35 Direktoratet for e-helse. (2022). Behov for data til kunstig intelligens i helsetjenesten.
- 36 Den nasjonale forskningsetiske komiteene. (2019). forskningsetikk.no. Hentet fra Retningslinjer for inklusjon av kvinner i medisinsk forskning: <https://www.forskningsetikk.no/retningslinjer/med-helse/inklusion-av-kvinner/>
- 37 Cirillo, D. C.-S. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ digital medicine,
- 38 Sundby, J. (2020). Universitetet I Oslo. Retrieved from Ulempe for kvinner når medisiner testes på menn: <https://www.med.uio.no/om/aktuelt/blogg/2020/ulempe-for-kvinner-nar-med-testes-pa-menn.html>
- 39 Cirillo, D. C.-S. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ digital medicine,
- 40 Rustagi, G. S. (2021). When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. Retrieved from ssir.org: https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_advance_ai_gender_equity
- 41 De nasjonale forskningsetiske komiteene. (2019). Retningslinjer for inklusjon av kvinner i medisinsk forskning.
- 42 Bergstrøm, I. I. (2014). forskning.no. Hentet fra Derfor forsker vi mest på menn: <https://forskning.no/kjonn-og-samfunn-menneskekroppen-psykologi/derfor-forsker-vi-mest-pa-menn/566222>
- 43 Cirillo, D. C.-S. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ digital medicine,
- 44 Norori, N., Hu, Q., Marcelle Aellen, F., Dalia Faraci, F., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. National Library of Medicine.
- 45 Smith, G., & Rustagi, I. (2020). Mitigating Bias in Artificial Intelligence.
- 46 SSB. (2021). Kraftig vekst i antallet med IT-yrker. Retrieved from ssb.no: <https://www.ssb.no/arbeid-og-lonn/sysselsetting/artikler/kraftig-vekst-i-antallet-med-it-yrker>

- 47 WHO. (2021). Ethics And Governance Of Artificial Intelligence For Health.
- 48 EU. (2019). Ethics Guidelines For Trustworthy AI
- 49 Ibid.
- 50 Teknologirådet. (2021). KUNSTIG INTELLIGENS I KLINIKKEN Seks trender for fremtidens helsetjeneste.
- 51 Gerke, S., Minssen, T., & Glenn, C. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. National Library of Medicine.
- 52 Direktoratet for e-helse. (2019). Utredning om bruk av kunstig intelligens i helsesektoren.
Direktoratet for e-helse. Retrieved from Utredning om bruk av kunstig intelligens i helsesektoren:
<https://www.ehelse.no/publikasjoner/utredning-om-bruk-av-kunstig-intelligens-i-helsesektoren>
- 53 Smith, G. (2020). What does “fairness” mean for machine learning systems?
- 54 Kommunal- og moderniseringsdepartementet. (2020). Nasjonal strategi for kunstig intelligens. Oslo: Kommunal- og moderniseringsdepartementet. S. 6
- 55 E-Health Research. (2022). E-Health Research. Hentet fra Kunstig intelligens i norsk helsetjeneste:
<https://ehealthresearch.no/kin/prosjekter>
- 56 Helse og omsorgsdepartementet. (2020). Nasjonal helse og sykehusplan 2020-2023.
- 57 Helsedirektoratet. (2021). Tilrettelegging for bruk av kunstig intelligens i helsetjenesten: Med utgangspunkt i det radiologiske fagområdet. Oslo: Helsedirektoratet.
- 58 Larrazaval, A., Nieto, N., Peterson, V., & Enzo, F. (2020). Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. PNAS.
- 59 Cirillo, D. C.-S. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ digital medicine,
- 60 Bakkevoll P.A., M. A. (2018). Health analytics: Kunstig intelligens –nye muligheter for helsetjenesten. Oslo: Nasjonalt senter for e-helseforskning.
- 61 Helsedirektoratet. (2021). Tilrettelegging for bruk av kunstig intelligens i helsetjenesten: Med utgangspunkt i det radiologiske fagområdet. Oslo: Helsedirektoratet.
- 62 Ibid.
- 63 Makhlysheva A., M.-R. L. (2022). Implementation of artificial intelligence in Norwegian healthcare: The road to broad adoption. Oslo: Norwegian Centre for E-health Research
- 64 Intervju med en representant for et norsk helseforetak
- 65 IBM. (2020). [IBM.com](https://www.ibm.com/cloud/learn/natural-language-processing). Retrieved from Natural Language Processing (NLP):
<https://www.ibm.com/cloud/learn/natural-language-processing>
- 66 Foresee Medical. (2022, 11 10). Foresee Medical. Hentet fra Natural language processing in healthcare :
<https://www.foreseemed.com/natural-language-processing-in-healthcare>
- 67 Bjørkeng, P. K. (2019, 11 16). Aftenposten. Hentet fra Tror de kan redde liv med kunstig intelligens. Men trolig er det ulovlig.: <https://www.aftenposten.no/norge/i/dOadGo/tror-de-kan-redde-liv-med-kunstig-intelligens-men-trolig-er-det-ulovlig>
- 68 Garimella, A., Banea, C., Hovy, D., & Mihalcea, R. (2019). *Women’s Syntactic Resilience and Men’s Grammatical Luck: Gender-Bias in Part-of-Speech Tagging and Dependency Parsing*.
- 69 Direktoratet for e-helse. (2019). Utredning om bruk av kunstig intelligens i helsesektoren.
- 70 Helse- og omsorgsdepartementet. (2021). Lovdata . Forskrift om medisinsk utstyr:
[Forskrift om medisinsk utstyr - Lovdata](#)
- 71 Direktoratet for e-helse. (2019). Utredning om bruk av kunstig intelligens i helsesektoren.
- 72 Helsedirektoratet. (2021). Tilrettelegging for bruk av kunstig intelligens i helsetjenesten: Med utgangspunkt i det radiologiske fagområdet. Oslo: Helsedirektoratet. (Side 5)
- 73 Helse- og omsorgsdepartementet. (2021). Lovdata . Forskrift om medisinsk utstyr:
[Forskrift om medisinsk utstyr - Lovdata](#)
- 74 Makhlysheva A., M.-R. L. (2022). Implementation of artificial intelligence in Norwegian healthcare: The road to broad adoption. Oslo: Norwegian Centre for E-health Research
- 75 Ibid.
- 76 Direktoratet for e-helse. (2022). Innsiktsrapport: Behov for data til kunstig intelligens i helsetjenesten. Oslo: Direktoratet for e-helse.

- 77 Øye, O.-J., & Normann, M. (2022, 10 19). OsloMet.no. Hentet fra Kunstig intelligens kan gi oss riktig diagnose. Men vi vet for lite om hvordan den gjør det.: <https://www.oslomet.no/forskning/forskningsnyheter/kunstig-intelligens-kan-gi-riktig-diagnose-men-vi-vet-for-lite-om-hvordan>
- 78 Teknologirådet. (2018). teknologiradet.no. Retrieved from Kunstig intelligens - muligheter utfordringer og en plan for Norge: <https://teknologiradet.no/wp-content/uploads/sites/105/2018/09/Rapport-Kunstig-intelligens-og-maskinlaering-til-nett.pdf>
- 79 Helsedirektoratet. (2021). Tilrettelegging for bruk av kunstig intelligens i helsetjenesten: Med utgangspunkt i det radiologiske fagområdet. Oslo: Helsedirektoratet. (Side 5)
- 80 Helsedirektoratet. (,). Helsedirektoratet.no. Hentet fra Kunstig intelligens: <https://www.helsedirektoratet.no/tema/kunstig-intelligens>
- 81 Datatilsynet. (2021). datatilsynet.no. Hentet fra Framework for the Regulatory Sandbox: <https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/framework-for-the-regulatory-sandbox/>
- 82 European Commission. (u.d.). ai-watch.ec.europa.eu. Hentet fra AI Watch: https://ai-watch.ec.europa.eu/index_en
- 83 Makhlysheva A., M.-R. L. (2022). Implementation of artificial intelligence in Norwegian healthcare: The road to broad adoption. Oslo: Norwegian Centre for E-health Research (Side 16-18)
- 84 Ibid. (Side 16-18)
- 85 OECD. (2020). Trustworthy AI in health (Background paper for the G20 AI Dialogue, Digital Economy Task Force). Paris: OECD.
- 86 Cirillo, D. C.-S. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ digital medicine
- 87 OECD. (2020). Trustworthy AI in health (Background paper for the G20 AI Dialogue, Digital Economy Task Force). Paris: OECD.
- 88 Bui, C. K. (2021, October 12). Exploring Bias Against Women in Artificial Intelligence. Hentet fra University of Oslo: https://www.duo.uio.no/bitstream/handle/10852/88551/All_Chapters_Master-s_thesis-Gender_Bias_in_AI-ver--9.pdf?sequence=1&isAllowed=y
- 89 Bui, C. (2022). Må kunstig intelligens være kjønnsdiskriminerende? (Norsk råd for digital etikk (NORDE)), Interviewer) I podcast-episoden "Må kunstig intelligens være kjønnsdiskriminerende?" av NORDE beskriver Bui sitt prosjekt for å finne ut hvor mye norske utviklere vet om hvordan kunstig intelligens kan diskriminere kvinner og hvilke rutiner de har for å forebygge diskriminering. I den kvalitative studien ble 13 personer som jobber og forsker på KI i Norge ble intervjuet for å undersøke de etiske sidene. Blant disse jobbet flere direkte mot helsesektoren.
- 90 Cirillo, D. C.-S. (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ digital medicine
- 91 Helsedirektoratet. (2022, 11 08). Helsedirektoratet.no. Retrieved from Kunstig intelligens - Regelverk - Medisinsk utstyr og CE-merking: <https://www.helsedirektoratet.no/tema/kunstig-intelligens/regelverk/medisinsk-utstyr-og-ce-merking>
- 92 Makhlysheva A., M.-R. L. (2022). Implementation of artificial intelligence in Norwegian healthcare: The road to broad adoption. Oslo: Norwegian Centre for E-health Research
- 93 Helsedirektoratet. (2021). Tilrettelegging for bruk av kunstig intelligens i helsetjenesten: Med utgangspunkt i det radiologiske fagområdet. Oslo: Helsedirektoratet (Side 5)
- 94 Brombach, H. (2018, 07 28). Digi.no. Hentet fra Rapport: IBM Watson har flere ganger anbefalt fullstendig feil behandling av kreftpasienter: <https://www.digi.no/artikler/rapport-ibm-watson-har-flere-ganger-anbefalt-fullstendig-feil-behandling-av-kreftpasienter/442634>
- 95 Cathrine Bui (2021) avdekket at det var høyere sannsynlighet for bevissthet rundt bias i team som bestod av flere minoriteter <https://www.duo.uio.no/handle/10852/88551>
- 96 OECD. (2020). Trustworthy AI in health (Background paper for the G20 AI Dialogue, Digital Economy Task Force). Paris: OECD.
- 97 European Commission. (2020, 07 17). Shaping Europe's digital future. Hentet fra Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment: <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>
- 98 WHO. (2021). Ethics And Governance Of Artificial Intelligence For Health
- 99 EU. (2019). Ethics Guidelines For Trustworthy AI
- 100 Regjeringen. (2020). Nasjonal strategi for kunstig intelligens
- 101 <https://www.dsb.no/lover/produkter-og-forbrukertjenester/artikler/fakta-om-ce-merking/>

**Bringing
Ingenuity
to Life.**



About PA

We believe in the power of ingenuity to build a positive human future.

As strategies, technologies and innovation collide we create opportunity from complexity.

Our diverse teams of experts combine innovative thinking and breakthrough use of technologies to progress further, faster, together. Our clients adapt and transform, and together we achieve enduring results.

We are over 4,000 strategists, innovators, designers, consultants, digital experts, scientists, engineers and technologists. And we have deep expertise in consumer and manufacturing, defence and security, energy and utilities, financial services, government and public services, health and life sciences, and transport.

Our teams operate globally from offices across the UK, Ireland, US, Nordics and Netherlands.

Discover more at paconsulting.com and connect with PA on [LinkedIn](#) and [Twitter](#).

PA. Bringing Ingenuity to Life.

Corporate Headquarters

PA Consulting
10 Bressenden Place
London SW1E 5DN
United Kingdom

+44 20 7730 9000

Oslo Office

Verkstedveien 1
PO Box 150, Skøyen
0212 Oslo
Norway

+47 67 58 67 58

paconsulting.com

This document has been prepared by PA.
The contents of this document do not constitute any form of commitment or recommendation on the part of PA at the date of their preparation.

© PA Knowledge Limited 2023.
All rights reserved.

No part of this documentation may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the written permission of PA Consulting.

Produced by Design Studio at PA
7_300451 May 2023.