

Rapport om kvalitetssikring: Bruk av kunstig intelligens i helse- og omsorgstjenesten

Først publisert: 10. januar 2025

Siste faglige endring: 10. januar 2025



Innhold

1. Forord	3
2. Sammendrag	4
3. Innledning	7
4. Fase 1: Beskrive utfordringen og vurdere om KI er løsningen .	10
5. Fase 2: Kartlegge KI-systemer på markedet	14
6. Fase 3: Vurdere egen virksomhets evne til å kunne ta i bruk et KI-system	28
7. Fase 4: Anskaffe et KI-system	34
8. Fase 5: Innføre og kvalitetssikre et KI-system	40
9. Fase 6: Drifte og forvalte et KI-system	45
10. KI-faktaark	51

Forord

Det nasjonale koordineringsprosjektet *Bedre bruk av kunstig intelligens* (KI) ble gjennomført i 2019-2023 som en del av arbeidet med Nasjonal helse- og sykehusplan 2019-2023. Gjennom arbeidet i koordineringsprosjektet ble det avdekket behov for tydeligere rammer for hvordan kvalitetssikring av KI-systemer kan utføres for å sikre trygg og effektiv bruk i helse- og omsorgssektoren.

Delprosjektet *Rammer for kvalitetssikring* ble igangsatt i august 2023, etter en idéfase i tett dialog med helsesektoren gjennom bl.a. workshoper våren 2023, og er nå del av Felles KI-plan for helse- og omsorgstjenesten 2024-2025 (heretter kalt *Felles KI-plan*). Arbeidet fra delprosjektet så langt presenteres i denne rapporten. Formidling, oppdateringer og videreutvikling av veiledningsmateriell vil skje gjennom prioriteringer i Felles KI-plan.

Representanter fra følgende helseetater har deltatt i arbeidet: Helsedirektoratet/Direktoratet for e-helse, Statens legemiddelverk/Direktoratet for medisinske produkter, Folkehelseinstituttet og Statens helsetilsyn, og etter hvert Direktoratet for strålevern og atomsikkerhet. Representanter fra Helse Sør-Øst RHF og Sykehuspartner HF har også deltatt. Helsedirektoratet har ledet arbeidet.

Det ble etablert et brukerpanel med eksterne, mulige brukere av støttematerialet, i tillegg til eksperter på området. Brukerpanelet har bestått av representanter fra Haukeland universitetssykehus, Akershus universitetssykehus, Vestre Viken HF, Bærum sykehus, LIVV Health, Senter for pasientnær kunstig intelligens ved Universitetssykehuset i Nord-Norge, Det norske Veritas, Norsk Regnesentral, Sykehusinnkjøp, Nasjonalt senter for e-helseforskning, Stavanger kommune, Drammen kommune, Trondheim kommune, Bergen kommune og KS (kommunesektorens organisasjon).

Kompetansenettverket Kunstig intelligens i norsk helsetjeneste (KIN) har løpende bidratt med verdifulle faglige innspill og diskusjoner.

Oslo, 18.desember 2024

Sammendrag

Denne rapporten er utarbeidet for å gi støtte til virksomhetene som planlegger å anskaffe et KI-system slik at de kan bruke det på en trygg og effektiv måte. Rapporten omhandler aspekter som virksomheter må være spesielt oppmerksom på for å kunne vurdere om et KI-system er tillitsverdig når de vurderer å anskaffe, innføre og ta i bruk et KI-system i helse- og omsorgstjenesten.

Det amerikanske National Institute for Standardization and Technology (NIST) karakteriserer tillitsverdig KI slik: *Gyldig og pålitelig, trygg, sikker og motstandsdyktig, ansvarlig og gjennomsiiktig, forklarbar og tolkbar, har godt personvern og er rettferdig med kontroll på skadelig skjevheter* [1]:

Europakommisjonen utarbeidet i 2019 retningslinjer med formål om å fremme tillitsverdig KI. I disse retningslinjene er det blant annet vist til tre komponenter som er sentrale for å sikre tillitsverdige KI-systemer gjennom hele livssyklusen til systemet:

"(1) det skal være lovlig, i samsvar med alle gjeldende lover og forskrifter, (2) det skal være etisk, og sikre etterlevelse av etiske prinsipper og verdier, og (3) det skal være robust, både fra et teknisk og sosialt perspektiv, da KI-systemer, selv med gode intensjoner, kan forårsake utilsiktet skade" [2].

Rapporten beskriver de juridiske rammene, hva som er pålagte krav etter relevant regelverk, og til dels hva som er lurt. KI er imidlertid et fagområde som er i rask bevegelse og med ny lovgivning på trappene. De mer konkrete grensegangene mellom EUs forordninger for medisinsk utstyr, personvern og kunstig intelligens må avklares. I tiden som følger vil det bli behov for oppdatert veiledningsmateriell etter hvert som kunnskapen og nytt regelverk formes.

Et KI-system kan ha ulike former: en selvstendig applikasjon på en PC eller mobilapplikasjon, en webtjeneste, en del av et fagsystem eller et journalsystem.

Med dette som bakteppe presenterer rapporten noen overordnede aspekter som virksomhetene må forholde seg til. Vi har delt inn disse i seks faser for anskaffelse, innføring og bruk av KI.

Juridiske rammer: Å ta i bruk KI-systemer innebærer vurderinger rundt blant annet helsefaglig forsvarlighet, personvern, informasjonssikkerhet, tilgang til helseopplysninger, datahåndtering og automatiserte beslutningsprosesser.

Helselovgivningen inneholder en rekke plikter og rettigheter for blant annet helse- og omsorgstjenesten og pasientene. Rett til helsehjelp, kravet til medisinsk forsvarlige tjenester og helsepersonells taushetsplikt er sentrale regler. Behandling av helseopplysninger er også omfattet av personvernforordningen, som stiller flere krav til dataansvarlig virksomhet. Likestillings- og diskrimineringsloven er også relevant [3]. Lov om medisinsk utstyr regulerer produsentenes ansvar for at medisinsk utstyr er sikkert og trygt, noe som også gjelder medisinsk utstyr som er eller består av KI [4]. Forordningen om kunstig intelligens skal sikre at produkter og systemer som anvender KI er trygge å bruke [5]. Regelverket stiller krav til både produsenter og brukere av KI (som kan være helse- og omsorgstjenesten). Det er også viktig å sikre at fagspesifikke regelverk blir identifisert og vurdert når KI-systemer skal anskaffes, valideres og tas i bruk

Risikohåndtering: KI kan bidra til forbedringer i helsevesenet, men introduserer også nye utfordringer. Uten gode risikovurderinger, kan KI-systemer påføre, forsterke eller videreføre urettferdige eller uønskede utfall for enkeltpersoner, helsevesenet og samfunnet for øvrig og utfordre pasientsikkerheten.

Validering: Før en KI-modell kan tas i bruk må den kvalitetssikres, noe som blant annet kan innebære klinisk validering. Klinisk validering betyr i denne rapporten å bekrefte at KI-modellen yter som den er ment, for et bestemt tiltenkt formål [6]. Dersom en KI-modell har tiltenkt formål å diagnostisere brudd på et røntgenbilde av et ben, er det akkurat dette KI-modellen skal valideres for.

Etiske prinsipper: KI-systemet må være trygt og etisk akseptabelt å bruke. Europakommisjonens ekspertgruppe foreslår syv prinsipper som også regjeringens KI-strategi legger til grunn for bruk av KI i Norge [7]. KI-strategien sier blant annet at bruk av KI skal bygge på etiske prinsipper, være gjennomskiktig og respektere menneskerettighetene og demokratiet [8]. Samtidig kan det være etisk betenkelig å ikke ta i bruk KI-systemer, hvis de kan gi bedre kvalitet, effektivisere tjenestene og dermed frigjøre ressurser som kan brukes på andre viktige områder.

Seks faser for anskaffelse, innføring og bruk av KI

Denne rapporten er inndelt i seks faser som en virksomhet kan ta utgangspunkt i når den skal vurdere å anskaffe og ta i bruk et KI-system. Fasene er kort beskrevet i listen nedenfor. Flere av fasene vil omhandle de samme risikoområdene eller kvalitetsaspektene, men vurderingene vil gjøres på økende detaljeringsnivå. For bruksområder for KI med lav risiko, trenger ikke virksomheten å adressere alle aspektene like detaljert. Underveis vil nye opplysninger kunne medføre behov for å gå tilbake til en tidligere fase og justere eller gjennomføre deler av denne på nytt.

1. **Beskrive utfordringen og vurdere om KI er løsningen:** Definere behovet og gjøre overordnet risiko- og kost-nyttevurdering.
2. **Kartlegge KI-systemer på markedet:** Vurdere KI-systemenes overenstemmelse med regelverk, samt KI-systemenes ytelse, kvalitet og nytte.
3. **Vurdere egen virksomhets evne til å ta i bruk et KI-system:** Vurdere både menneskelige og tekniske faktorer, samt gjøre risiko- og sikkerhetsvurderinger og oppdatere gevinstberegninger.
4. **Anskaffe et KI-system:** Planlegge anskaffelsesprosedyre, utarbeide kravspesifikasjon og vurdere tilbud.
5. **Innføre og kvalitetssikre et KI-system:** Gjøre risiko- og konsekvensvurderinger, validere KI-systemet på egne data, gjennomføre organisatoriske og tekniske tiltak og lage plan for drift og forvaltning.
6. **Drifte og forvalte et KI-system:** Overvåke, drifte og vedlikeholde KI-systemet, sørge for kontinuerlig kvalitetsforbedring, realisere gevinster og dele erfaringer.

KI-faktaark: Utdypende informasjon om utvalgte temaer som er behandlet i rapporten finnes i KI-faktaark som publiseres på den tverretatlige informasjonssiden om KI i helse- og omsorgstjenesten[9].

[1] [Artificial Intelligence Risk Management Framework \(AI RME 1.0\) \(nist.gov\)](#), side 12 og [Trustworthy and Responsible AI | NIST](#): Characteristics of trustworthy AI systems include: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed.

[2] European Commission: Directorate-General for Communications Networks, Content and Technology, Ethics guidelines for trustworthy AI, Publications Office, 2019, <https://data.europa.eu/doi/10.2759/346720>

[3] [Lov om likestilling og forbud mot diskriminering \(likestillings- og diskrimineringsloven\) - Lovdata](#)

[4] [Regelverket for medisinsk utstyr - Direktoratet for medisinske produkter \(dmp.no\)](#)

[5] EU vedtok en lov om kunstig intelligens (KI-forordningen) i 2024.
<https://www.regjeringen.no/contentassets/c499c3b6c93740bd989c43d886f65924/no/pdfs/nasjonal-digitalisering>

[6] <https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en>: I standarden ISO 9000:2015 "Quality Management System – Fundamentals and vocabulary" defineres "validation" slik: "Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled."

[7] [Ethics guidelines for trustworthy AI | Shaping Europe's digital future \(europa.eu\)](#) og <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/>

[8] <https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf>

[9] Faktaark om KI vil publiseres på [Kunstig intelligens](#)

Innledning

I Nasjonal helse- og samhandlingsplan for 2024-2027 understrekes det at KI vil kunne utgjøre et betydelig bidrag til en bærekraftig utvikling av vår felles helsetjeneste. Det forventes at kunstig intelligens og annen personellbesparende teknologi vil kunne bidra til at vi kan opprettholde kvaliteten i tjenesten i årene framover, i tillegg til at det kan bidra til å redusere ventetidene [10].

Rapporten reflekterer kunnskap ved årsskiftet 2024/2025. Siden KI er et område i rivende utvikling, vil ikke rapporten løpende kunne oppdateres med ny kunnskap.

Arbeidet fra delprosjektet *Rammer for kvalitetssikring* presenteres i denne rapporten og er status ved utgangen av 2024. KI er et fagområde i rask utvikling og det kommer nye lovgivninger fra EU som vil påvirke bruk av KI i Norge. I tiden som følger vil det bli behov for oppdatert veiledningsmateriell etter hvert som kunnskapen og nytt regelverk formes. Formidling, oppdateringer og hensiktsmessige måter å støtte helse- og omsorgstjenesten på, som videreutvikling av veiledningsmateriell vil skje gjennom prioriteringer i Felles KI-plan [11].

Rapporten beskriver de juridiske rammene, hva som er pålagte krav og til dels hva som er lurt for å sikre at KI-systemer som brukes i helse- og omsorgstjenesten er tillitsverdige. I retningslinjer om tillitsverdig KI peker Europakommisjonen blant annet på tre komponenter som er sentrale for å sikre tillitsverdige KI-systemer gjennom hele livssyklusen til systemet [12]:

- "1. Det skal være **lovlig**, i samsvar med alle gjeldende lover og forskrifter.*
- 2. Det skal være **etisk**, og sikre etterlevelse av etiske prinsipper og verdier.*
- 3. Det skal være **robust**, både fra et teknisk og sosialt perspektiv, da KI-systemer, selv med gode intensjoner, kan forårsake utilsiktet skade."*

Tillitsverdig KI karakteriseres av Det amerikanske National Institute for Standardization and Technology (NIST) slik: *Gyldig og pålitelig, trygg, sikker og motstandsdyktig, ansvarlig og gjennomiktig, forklarbar og tolkbar, har godt personvern og er rettferdig med kontroll på skadelig skjevheter* [13].

Definisjonen på et *KI-system* er etter KI-forordningen, og slik vi bruker betydningen i denne rapporten:

...et maskinbasert system som er designet for å operere med varierende nivåer av autonomi og som kan vise tilpasningsevne etter utplassering, og som, for eksplisitte eller implisitte mål, utleder, fra inndata det mottar, hvordan det skal generere utdata som prediksjoner, innhold, anbefalinger eller beslutninger som kan påvirke fysiske eller virtuelle miljøer [14].

Et KI-system kan ha ulike former: en selvstendig applikasjon på en PC eller mobilapplikasjon, en webtjeneste, en del av et fagsystem eller et journalsystem.

Det er noen iboende egenskaper ved KI som gjør det mer vanskelig å kvalitetssikre KI-systemer enn tradisjonelle IT-systemer. Når virksomheter skal anskaffe, innføre og ta i bruk KI-systemer er det viktig å ta hensyn til blant annet:

- **Sort-boksproblemet:** KI-modeller, spesielt de som bruker dype nevralt nettverk, kan være svært kompliserte å forstå, eller datagrunnlaget og logikken kan være utilgjengelig av kommersielle hensyn. Mangel på forklarbarhet og transparens kan gjøre det vanskeligere for helsepersonell å ha tillit til å bruke KI-systemet, samtidig som det i forlengelsen av dette kan utfordre pasienters rett til informasjon [15].
- **Bias:** er systematiske feil som oppstår i KI-systemer på grunn av skjevheter i treningsdataene. KI-modeller speiler dataene de trenes på, og må være representative for den befolkningen KI-modellen

skal brukes på. Dataene kan også representere diskriminering i samfunnet. Bias kan føre til at KI-modeller gir feilaktig, partisk eller diskriminerende innhold [16].

- **Hallusinasjon:** Generative KI-modeller skaper innhold basert på dataene modellen er trent på. Dette innholdet kan fremstå som sant og riktig, men kan være usant eller oppspinn. Dette vil kunne utgjøre en risiko ved KI-modellen som man bør være oppmerksom på ved bruk i virksomheten.

Mål og målgruppe

Denne rapporten er utarbeidet for å gi støtte til de som planlegger å anskaffe et KI-system slik at virksomhetene kan bruke det på en trygg og effektiv måte. Målsetningen er å gi veiledning for å identifisere, vurdere og håndtere risikoer knyttet til anskaffelse, innføring og bruk av ferdigutviklede KI-systemer i helse- og omsorgstjenesten. Rapporten omhandler aspekter som virksomheter må være spesielt oppmerksom på for å kunne vurdere om et KI-system er tillitsverdig.

Målgruppen for rapporten er personer som er involvert i vurderinger og beslutninger knyttet til å planlegge, anskaffe, ta i bruk og drifte KI-systemer i helse- og omsorgstjenesten. Rapporten kan også være relevant for ansatte i virksomheter som tar i bruk KI, leverandører av KI-systemer og helsemyndighetene.

Avgrensninger

Rapporten er avgrenset til anskaffelse av ferdigutviklede produkter. Den omfatter dermed ikke egenutvikling av KI-systemer. Problemstillinger knyttet til KI-systemer som lærer løpende er ikke særskilt behandlet. Dette er aktuelle temaer for videre arbeid i delprosjektet.

Generative KI-modeller er heller ikke særskilt behandlet, men et annet delprosjekt i Felles KI-plan vil omhandle vurdering av risiko ved store språkmodeller og hvordan legge til rette for at bruken av språkmodeller er tilpasset norske forhold [17].

Relevante kilder til innledningen

- [A buyer's guide to AI in health and care](#) – NHS Transformation Directorate (england.nhs.uk)
- [What is AI? Can you make a clear distinction between AI and non-AI systems? - OECD.AI](#)
- Digitaliseringsdirektoratet: [Veiledning for ansvarlig utvikling og bruk av kunstig intelligens i offentlig sektor](#)
- *Adoption of AI in healthcare* er en omfattende artikkel (white paper) på engelsk, utredet av DNV i et internasjonalt samarbeid, beskriver hva man bør vurdere i anskaffelsen av kunstig intelligens-baserte verktøy og kan leses her: <https://www.dnv.com/Publications/how-do-i-turn-this-on-what-to-consider-when-adopting-ai-based-tools-i>
- *The National Institute of Standards and Technology* (NIST) har utgitt et risikostyringsrammeverk for KI (NIST AI RMF 1.0), hvor det beskrives hvordan man kan jobbe systematisk med risikostyring av KI-systemer: [Artificial Intelligence Risk Management Framework \(AI RMF 1.0\)](#) NIST har i tillegg har en artikkel om forklarbar KI: Four Principles of Explainable Artificial Intelligence <https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf>
- I «[Etiske retningslinjer for pålitelig kunstig intelligens](#)» (digital-strategy.ec.europa.eu) utarbeidet av en ekspertgruppe oppnevnt av EU-kommisjonen, nevnes det tre hovedprinsipper for ansvarlig kunstig intelligens, nemlig lovlig, etisk og sikker kunstig intelligens. De samme prinsippene gjenspeiles i regjeringens [nasjonale strategi for kunstig intelligens \(regjeringen.no\)](#) fra 2020 og i [Fremtidens digitale Norge – nasjonal digitaliseringsstrategi 2024–2030 \(regjeringen.no\)](#)

- GENERATING EVIDENCE FOR ARTIFICIAL INTELLIGENCE-BASED MEDICAL DEVICES: A FRAMEWORK FOR TRAINING, VALIDATION AND EVALUATION (WHO , 2021)
<https://iris.who.int/bitstream/handle/10665/349093/9789240038462-eng.pdf>
- Norm for informasjonssikkerhet (Normen) har publisert en artikkel om sikkerhetsrisiko i systemer som benytter kunstig intelligens, og hvordan man kan stille sikkerhetskrav til leverandører av slike systemer:
<https://www.ehelse.no/normen/aktuelt-om-normen/sikkerhetsrisiko-i-systemer-som-benytter-kunstig-intelli>
- FDA: [U.S. Food and Drug Administration \(fda.gov\)](https://www.fda.gov/)
- [ISO/IEC 22989 Artificial intelligence concepts and terminology](#) etablerer en terminologi for KI og beskriver konsepter på KI-feltet og [Standard Norge | standard.no. ISO/IEC 22989:2022](https://www.standard.no/ISO/IEC-22989-2022)
- [Innebygd diskrimineringsvern: En veileder for å avdekke og forebygge diskriminering i utvikling og bruk av kunstig intelligens \(ldo.no\)](#)

[10] <https://www.regjeringen.no/no/dokumenter/meld.-st.-9-20232024/id3027594/>

[11] [Felles KI-plan for trygg og effektiv bruk av KI i helse og omsorgstjenesten 2024 - 2025](#)

[12] European Commission: Directorate-General for Communications Networks, Content and Technology, Ethics guidelines for trustworthy AI, Publications Office, 2019, <https://data.europa.eu/doi/10.2759/346720>

[13] [Artificial Intelligence Risk Management Framework \(AI RMF 1.0\) \(nist.gov\)](#), side 12 og [Trustworthy and Responsible AI | NIST](#): Characteristics of trustworthy AI systems include: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed.

[14] Artikkel 3 i KI-forordningen (oversatt):
https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689

[15] Mer om sort-boksproblemet:
<https://www.helse-nord.no/499f60/siteassets/dokumenter-og-blokker/forskning/strategi-for-kunstig-intelligens-i-h>

[16] Se også: <https://teknologiradet.no/ordliste-for-kunstig-intelligens/>

[17] [Felles KI-plan for trygg og effektiv bruk av KI i helse og omsorgstjenesten 2024 - 2025](#)

Fase 1: Beskrive utfordringen og vurdere om KI er løsningen

Hensikten i denne fasen er å:

- lage en god problembeskrivelse
- vurdere risikoer
- vurdere gevinster

Beskrive hvilket behov KI-systemet skal ivareta

Det å definere og tydeliggjøre virksomhetens behov er nødvendig for å kunne identifisere aktuelle KI-systemer som kan løse utfordringen. Dersom man for eksempel vil anskaffe et KI-system som er tiltenkt brukt for mennesker til et eller flere medisinsk formål, vil dette kunne anses som medisinsk utstyr (se fase 2)[18]. Hvem som skal bruke KI-systemet og hvem eller hva det skal brukes på, må klargjøres. Det samme gjelder med hvilken grad av autonomi KI-systemet skal fungere. Det må avklares om det er mulig og hensiktsmessig å løse problemet med et KI-system. Om problemet kan løses bedre, enklere og billigere ved en løsning uten KI, er det også viktig å kartlegge dette.

[18] Medisinsk utstyr defineres i MDR artikkel 2 (1), og omfatter programvare som ifølge produsenten er beregnet på å bli brukt på mennesker med henblikk på bl.a. følgende spesifikke medisinske formål: diagnostisering, forebygging, overvåking, prediksjon, prognostisering, behandling eller lindring av sykdom. <https://www.dmp.no/medisinsk-utstyr/utvikling-og-produksjon/kvalifisering-og-klassifisering/kvalifisering#Def>

Overordnet risikovurdering ved bruk av et KI-system

Produsenter av KI-systemer vil som utgangspunkt være pliktig til å foreta grundige risikovurderinger før de plasserer KI-systemer med høy risiko på markedet. Helsevirksomhetene som tar i bruk KI-systemer, må foreta egne vurderinger av risiko i egen virksomhet basert på risikoen som KI kan utgjøre i samvirke med virksomhetens andre systemer, prosedyrer og personell.

Selv om man tidlig i prosessen ikke har nok informasjon til å gjøre en grundig risikovurdering av de KI-systemene som kan være aktuelle, er det viktig å identifisere eventuelle betydelige negative effekter som bruk av et KI-system kan gi. De kan gjøre at ulempene er større enn fordelene. Aktuelle vurderingspunkter er:

- om den ivaretar kravet om forsvarlige tjenester
- hva det vil gjøre med kompetansen i virksomheten
- om det vil medføre overforbruk av helsetjenester
- om det vil påvirke pasientsikkerheten
- om det vil kunne medføre diskriminering [19]

- om det vil kunne medføre en risiko for personvernet

Anskaffelse av et KI-system som allerede er tatt i bruk i andre tilsvarende virksomheter med gode resultater, vil normalt kunne være mindre risikofyllt enn å anskaffe et helt ukjent KI-system.

[19] Det kan være vanskelig å påvise diskriminering dersom man ikke har de dataene som trengs, se [Hjerterom for etisk AI \(datatilsynet.no\)](#), et sandkasseprosjekt med Ahus. Se også [Innebygd diskrimineringsvern: En veileder for å avdekke og forebygge diskriminering i utvikling og bruk av kunstig intelligens \(ldo.no\)](#)

Overordnet gevinstvurdering

De tre prioriteringskriteriene (nytte, ressursbruk og alvorlighet) for helsetjenesten i Norge gjelder også for KI. En kost-nytte-vurdering bidrar til å avgjøre om et KI-system er den rette løsningen og om virksomheten skal investere i KI-systemet.

Et overslag kan inkludere kostnader forbundet med:

- Investeringer for å kunne bruke KI-systemet
- Test og validering
- Innføring
- Opplæring
- Bruk (personell, lisens og stykkpris)
- Forvaltning

Kostnadene settes opp mot antatte gevinster, knyttet til for eksempel økt pasientsikkerhet og/eller bedre ressursutnyttelse. Nedenfor listes det opp noen gevinster:

Kvalitet

- Økt diagnostisk treffsikkerhet eller raskere diagnostikk: Gir mer presise diagnoser og kortere tid fra undersøkelse til behandling.
- Optimalisering av behandling: Forbedret behandlingsregime, riktig dosering og økt grad av persontilpasset behandling.
- Forbedrede behandlingsutfall: Raskere rekonvalesens, færre bivirkninger, komplikasjoner, tilbakefall og reinnleggelser sammenlignet med dagens behandling.
- Bedre opplæring og kommunikasjon: Mer effektiv opplæring for pasienter, pårørende og helsepersonell.
- Tilpasset informasjon og beslutningsstøtte for pasienter og pårørende: Bedre tilrettelegging for egenomsorg.
- Bedre samhandling: Styrket samarbeid mellom aktører, både internt i virksomheten, mellom den kommunale helsetjenesten og spesialisthelsetjenesten, og på tvers av sektorer.

Økonomi

- Oppgave- eller personellbesparende: Reduksjon i behovet for manuell arbeidskraft gjennom automatisering av oppgaver.

- Administrativ effektivitet: Automatisering av administrative oppgaver som avtalehåndtering, logistikk og ressursallokering.
- Effektivisering av dokumentasjon: Bedre og raskere dokumentasjon av helsehjelp, inkludert medisinsk koding.
- Kortere ventetider: Raskere prosesser, som for eksempel svar på radiologiske undersøkelser og raskere hjemsendelse, som frigjør ressurser.

Utvikling

- Forskning og innovasjon: Analyse av store datasett for å oppdage nye sammenhenger og utvikle innovative løsninger.
- Bedre prognoser: Nye metoder for å forutsi behandlingsutfall og redusere risiko.
- Utvikling av teknologi og prosedyrer: Optimalisering av eksisterende systemer og innføring av nye løsninger for å styrke helsetjenesten.

Involvere rett fagkompetanse

Alle faser i prosessen med å anskaffe, innføre og bruke KI-systemer i helse- og omsorgstjenesten bør ha en tverrfaglig tilnærming.

For å tydeliggjøre ansvar, involvering og planlegging kan en ansvarsmatrise, ofte kalt en HUKI-matrise, for et anskaffelsesteam være nyttig [20][21]. En slik matrise benyttes for å beskrive forskjellige nivåer av involvering og ansvar som ulike roller bør eller kan ha, avhengig av oppgaven.

Tabell 1

Rolle	Fase 1 Beskrive utfordringen og vurdere om KI er løsningen	Fase 2 Kartlegge KI-systemer på markedet	Fase 3 Vurdere egen virksomhets evne til å kunne ta i bruk et KI-system	Fase 4 Anskaffe et KI-system	Fase 5 Innføre kvalitetskontroll av et KI-s
Prosjektledelse	I	HU	H	U	H
Bruker og/eller domenekompetanse (for eksempel kliniker)	HU	U	U	K	K
IT/virksomhetsarkitekt	I	U	U		
Fagspesifikk kompetanse (for eksempel strålevern)	K	K	U	U	U
Etisk kompetanse	K		K		
Innbygger/pasient/tjenestemottaker	I		I		I
Juridisk kompetanse	K	K	K	K	K
Personvern, informasjonssikkerhet		K	K	U	U
Informasjonsforvaltning			U	K	K
Analysekompetanse		U	I	I	U

Innkjøpskompetanse		K		U	
Løsningsarkitekt			K	U	K
Testleder		I	I	K	I
Drift og forvaltning			K		K
KI-superbruker (for eksempel en "KI-radiolog" eller "KI-medisinsk fysiker")					K
Systemansvarlig (IT-leverandør)	K	I	K	K	I
Systemeier (virksomhetsledelse)	K	I	K	H	I

Tabellen er et eksempel på en HUKI-matrise for en anskaffelse. Den viser hvilke roller som kan være med i et anskaffelsesteam og hvilket ansvar de har.

H: Hovedansvarlig (det endelige ansvaret, kan bare være én for hver fase), U: Utfører (utfører oppgaven, kan deles mellom flere roller/personer), K: Konsulteres (en som blir konsultert før endelig beslutning tas, krever to-veis kommunikasjon) og I: Informeres (en som bare informeres om beslutning, én-veis kommunikasjon)

For mer beskrivelse av rollene vises til [KI-faktaark 0](#) [22]

[20] HUKI står for hovedansvarlig, utførende, konsulteres og informeres. [Mal for ansvarsmatrise | Anskaffelser.no, eksempel_ansvarsmatrise_0.xlsx \(live.com\)](#)

[21] [Rolle- og ansvarsmatrise - KS](#)

[22] Faktaark om KI vil publiseres på www.helsedirektoratet.no/tema/kunstig-intelligens

Relevante kilder til fase 1

- [Kunstig intelligens - Helsedirektoratet](#)
- [Veiledning for ansvarlig bruk og utvikling av kunstig intelligens | Digdir](#)

Fase 2: Kartlegge KI-systemer på markedet

Hensikten med denne fasen er å kartlegge aktuelle KI-systemer som er tilgjengelige på markedet, for å få oversikt over om det finnes KI-systemer som:

- har tiltenkt formål i samsvar med behovet som er definert i fase 1
- er i tråd med kravene i gjeldende regelverk
- har en akseptabel ytelse og kvalitet

Identifisere aktuelle KI-systemer

Samlet informasjon om KI-systemer som er tilgjengelige på markedet finnes ikke per 1.11.2024. EU skal etablere en database over høy-risiko KI-systemer [23]. I tillegg er databasen EUDAMED for medisinsk utstyr under utvikling, men den vil først gi en fullstendig oversikt når den blir obligatorisk [24]. Dette skjer sannsynligvis i 2026. De amerikanske helsemyndighetene (FDA [26]) sin oversikt over medisinsk utstyr med KI kan være nyttig [27].

For radiologi finnes en nettside med oversikt over KI-systemer, som oppgir hvilken klasse de er CE-merket i og til hvilket regelverk [28] [29]. I tillegg oppgis annen dokumentasjon, som fagfellevurderte forskningsartikler og om de er godkjente av FDA.

Det kan være hensiktsmessig for anskaffelsesteamet å utføre en tidlig markedsdialog [30] for å få bedre kjennskap til KI-systemer på markedet [31]. For å kontrollere om KI-systemet samsvarer med regelverket for medisinsk utstyr er det relevant å etterspørre dokumentasjon som samsvarserklæring, EC-sertifikat og bruksanvisning fra produsenten til utstyret [32].

[23] EU-kommisjonen vil, i samarbeid med medlemslandene, etablere en database for KI-systemer med høy risiko i henhold til artikkel 71 i KI-forordningen. Databasen skal inneholde detaljert informasjon om de registrerte KI-systemene, inkludert blant annet tiltenkt formål, beskrivelse av data som er brukt og systemets operasjonslogikk (Vedlegg VIII i KI-forordningen).

[24] EUDAMED vil bestå av seks moduler knyttet til: aktørregistrering, unik enhetsidentifikasjon (UDI) og enhetsregistrering, meldte organer og sertifikater, kliniske undersøkelser og ytelsesstudier, årvåkenhet (*vigilance*) og markedsovervåking. <https://ec.europa.eu/tools/eudamed/>

[26] FDA: [U.S. Food and Drug Administration \(fda.gov\)](https://www.fda.gov/)

[27] [Artificial Intelligence and Machine Learning \(AI/ML\)-Enabled Medical Devices | FDA](#)

[28] AI for radiology har en database: [Radiology \(healthairegister.com\)](https://healthairegister.com/)

[29] [CE-merking \(standard.no\)](https://standard.no/) Merket betyr at produktet anses å oppfylle krav fra myndighetene som er gitt i direktiv eller forordning.

[30] Engelsk RFI; *request for information*

[31] Forespørsel om informasjon (request for information), en ikke-forpliktende markedsundersøkelse for å øke egen kompetanse om hva markedet tilbyr og hva vi ønsker å etterspørre.
[handbok-for-planlegging-og-gjennomforing-av-tidlig-markedsdialog-i-offentlige-anskaffelser-april-20201.pdf \(innovativeanskaffelser.no\)](#)

[32] En samsvarserklæring av medisinsk utstyr er en erklæring fra produsenten om at medisinsk utstyr oppfyller kravene i EU-regelverket for medisinsk utstyr og da får et CE-merke.

Regelverk for bruk av KI-systemer

Regelverkene for medisinsk utstyr er spesielt viktig for KI-systemer innen helse- og omsorgstjenesten. I tillegg definerer den kommende KI-forordningen viktige prinsipper og vil stille krav til både produsenten og brukeren (f.eks. helseinstitusjonen) av KI-systemer med høy risiko [33]. Det kan være hensiktsmessig å ta hensyn til disse kravene allerede nå, spesielt for høy-risikosystemer, selv om forordningen ikke er trådt i kraft per 2024 i Norge. Personopplysningsloven og den generelle helselovgivningen utgjør også viktige rammer for bruk av KI-systemer [34].

Tiltenkt formål

Hva KI-systemet er tiltenkt for, det vil si hvilket formål systemet skal brukes til, vil være bestemmende for hvilke(t) regelverk som gjelder [35]. KI-modeller for generelle formål [36] (eng: *general-purpose AI models (GPAI-models)*) har ikke et tiltenkt formål, men kan inngå i et utstyr som har et medisinsk formål. De faller inn under spesifikke reguleringer i EU-forordningen for KI.

KI-systemer er medisinsk utstyr etter lov om medisinsk utstyr dersom det er ment å skulle brukes på mennesker i den hensikt å, bidra til diagnostisering, forebygging, overvåking, prediksjon, prognostisering, behandling eller lindring av sykdom [37][38]. Andre KI-systemer som for eksempel logistikkssystemer og systemer for vaktplanlegging er i utgangspunktet ikke medisinsk utstyr. KI-systemer som for eksempel journalsystemer, kan imidlertid ha tilleggsfunksjoner som kan medføre at de klassifiseres som medisinsk utstyr. Koordineringsgruppen for medisinsk utstyr i EU (MDCG) [39] har publisert veiledende dokumenter som kan bidra til å vurdere når programvare faller inn under definisjonen av medisinsk utstyr [40].

- **Hvis det tiltenkte formålet er medisinsk, gjelder følgende:**
 - *KI-systemet betraktes som medisinsk utstyr* og virksomheten må bruke et CE-merket utstyr [41]. Utstyret må oppfylle dokumentasjonskravene i Lov om medisinsk utstyr.
 - KI-systemet vil i utgangspunktet betraktes som høyrisikosystem i henhold til KI-forordningen og må oppfylle kravene som stilles til disse systemene når den trer i kraft [42].
- **Hvis det tiltenkte formålet ikke er medisinsk, gjelder følgende:**
 - KI-systemet betraktes som høyrisikosystem i henhold til KI-forordningen dersom tiltenkt formål er listet i vedlegg III og må oppfylle kravene i denne, når den trer i kraft [43]. Det kan for eksempel være tilfelle dersom bruk av KI-systemet vil påvirke noens rett til å motta ytelser eller helsetjenester [44]. I dette tilfellet må offentlige tjenestetilbydere også utføre en risikovurdering av fundamentale menneskerettigheter (*Fundamental Rights Impact Assessment (FRIA)*) [45].
- **Hvis KI-systemet er ment å samhandle direkte med fysiske personer, gjelder følgende, med noen unntak:**
 - KI-forordningen krever transparens: KI-systemet må være designet og utviklet slik at berørte personer informeres om at de samhandler med et KI-system [46].

- I tillegg må virksomheten sjekke om KI-systemet kan falle innenfor annet produktregelverk [47]. I avsnittet *Alignment of product legislation* på denne websiden om *New legislative framework* gis en oversikt over EUs produktgrupper [48].

Lov om medisinsk utstyr

Lov om medisinsk utstyr [49] gjennomfører forordningene om medisinsk utstyr (MDR og IVDR). Dersom et KI-system skal benyttes til et medisinsk formål må det anskaffes et medisinsk utstyr. Det kan finnes lignende systemer som er satt på markedet under andre regelverk, men disse oppfyller ikke kravene til sikkerhet som stilles til medisinsk utstyr.

- **Medisinsk utstyr** kan kun markedsføres, distribueres eller tas i bruk når utstyret oppfyller de grunnleggende kravene til sikkerhet og ytelse etter regelverket, er samsvarsvurdert og dermed påført CE-merket, jf. håndteringsforskriften § 7. Denne forskriften pålegger at virksomheten anskaffer medisinsk utstyr som er egnet for det tiltenkte bruksområdet, og at utstyret brukes i samsvar med utstyrets vedlagte anvisninger.

Risikoklasser for medisinsk utstyr

Omfanget av samsvarsvurderingen etter lov om medisinsk utstyr avhenger blant annet av hvilken risikoklasse det medisinske utstyret tilhører [50].



Figur 1 Hentet fra Direktoratet for medisinske produkters nettsider:

<https://www.dmp.no/globalassets/documents/medisinsk-utstyr/figur-klassifisering---mdr---ny-versjon>.

Medisinsk utstyr klassifiseres i fire klasser (I, IIa, IIb og III) basert på risikoen forbundet med deres tiltenkte bruk. For de fleste typer utstyr i klasse I er det tilstrekkelig at produsenter selvdeklarerer utstyret. For utstyr i klasse IIa, IIb og III, (samt utstyr i klasse Ir, Is og Im), skal det foreligge sertifikat fra meldt organ som vurderer samsvar med kravene. Det finnes en oversikt over klassifiseringsregler på DMP's nettsider [51].

Medisinsk utstyr for in vitro-diagnostikk (IVD) har egne klasser (A, B, C og D) [52]. Krav om sertifikat fra meldt organ gjelder for IVD-utstyr i klasse B, C og D.

Produsentansvar

Hvis medisinsk utstyr brukes til annet tiltenkt formål enn beskrevet i samsvarserklæringen, eller virksomheten endrer utstyret slik at det kan påvirke samsvaret med gjeldende krav, medfører det at virksomheten tar på seg produsentansvaret, altså det juridiske ansvaret for oppfyllelse av alle krav i MDR slik at KI-systemet er sikkert og trygt [53].

Et eksempel på dette fra diagnostisk radiologi [54]:

- Eksempel: Et KI-system som er utviklet for bruk på voksne kan ikke uten videre tas i bruk på barn (0-18 år). Begrunnelsen for dette er at barn ikke kan behandles som små voksne. Det kreves særskilt kompetanse for å vurdere om et KI-system er trygt å bruke på barn og om KI-systemet gir god nok ytelse på barn sammenliknet med ytelsen på voksne. Dersom tiltenkt formål for KI-systemet er diagnostisk radiologi på voksne, men man i stedet vil bruke det på barn, må KI-systemet undersøkes for aldersgruppen gjennom en klinisk utprøving. Dersom virksomheten tar i bruk et slikt KI-system på barn, tar de på seg et produsentansvar.

Helseinstitusjoner har mulighet til å produsere medisinsk utstyr for intern bruk. DMP's nettsider beskriver hvilke krav som gjelder for dette [55].

Overgangsordninger

I tiden fremover vil det fortsatt være medisinsk utstyr på markedet som følger de gamle EU-direktivene [56]. Det er innført overgangsordninger mellom de gamle og de nye regelverkene for å forhindre mangler på det europeiske markedet, se DMP's nettsider [57].

Klinisk utprøving av et KI-system som ikke er CE-merket

Dersom det ikke finnes et alternativ som er CE-merket, eller hvor et KI-system uten CE-merking vurderes som egnet, vil det i tilfelle være nødvendig å søke om tillatelse for en klinisk utprøving hos DMP og Etikkomiteene for klinisk utprøving av legemidler og medisinsk utstyr (REK KULMU) [58].

FDA-godkjenning

Medisinsk utstyr som er godkjent av de amerikanske helsemyndighetene (US Food and Drug Administration (FDA)) [59], men som ikke er samsvarsvurdert i EU og CE-merket, kan ikke lovlig brukes i Norge eller i resten av EU/EØS, og bør derfor ikke anskaffes. Det hender at produsenter har ulike versjoner eller modeller for det amerikanske og det europeiske markedet. I slike tilfeller er det viktig å sørge for at virksomheten anskaffer det CE-merkede utstyret som er ment for det europeiske markedet.

KI-forordningen

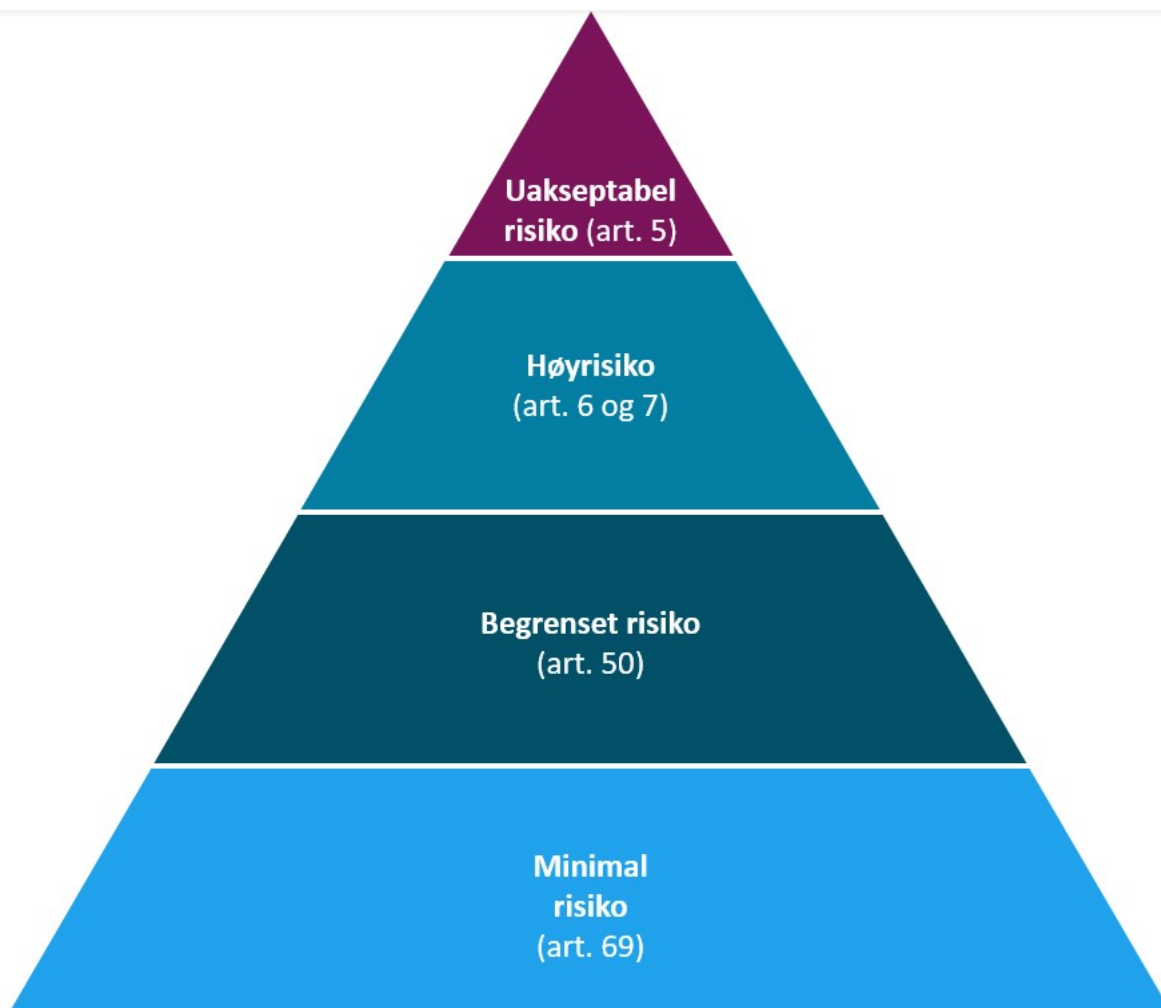
Formålet med [KI-forordningen \(eur-lex.europa.eu\)](https://eur-lex.europa.eu) er å fremme og øke bruk av menneskesentrerte og tillitsverdige KI-systemer, sikre et høyt nivå av beskyttelse av helse, sikkerhet og grunnleggende rettigheter, inkludert demokrati, rettssikkerhet og miljø, mot skadelige effekter som kan følge av KI-systemer og å støtte innovasjon.

KI-forordningen ble publisert 7. juli 2024 [60]. De ulike kravene i forordningen vil stegvis tre i kraft [61]. Regjeringen vil jobbe for at KI-forordningen blir innlemmet i EØS-avtalen så raskt som mulig [62].

EU-kommisjonen har bestilt europeiske harmoniserte standarder, som fungerer som et sett med tekniske spesifikasjoner som kan hjelpe med å oppfylle kravene i KI-forordningen [63].

Etter KI-forordningen er det den tiltenkte bruken av KI-systemet som bestemmer risikoklassen, ikke bare KI-systemet i seg selv. Det samme KI-systemet kan derfor ha lav risiko i én brukssituasjon, men høy risiko i en annen.

Pyramide som illustrerer forskjellige risikoklasser i KI-forordningen. Fra bunn til topp: minimal risiko (art. 95), begrenset risiko (art. 50), høyrisiko (art. 6 og 7.) og uakseptabel risiko (art.5).



Figur 2: Risikoklasser i KI-forordningen

Høyrisiko KI-systemer

Et høy-risiko KI-system kan etter KI-forordningen enten tilhøre kategorien regulerte produkter (som for eksempel medisinsk utstyr) eller kategorien regulerte bruksformål [64].

KI-forordningen beskriver to kategorier høyrisiko-KI-systemer

Regulerte produkter: Vedlegg I i KI-forordningen

Dersom det er krav om tredjepartssertifisering av et produkt eller sikkerhetskomponent i et produkt som allerede regulert av et av EU-regelverkene nevnt i vedlegg I, blant annet:

- Medical Device Regulation (MDR)
- In Vitro Diagnostic Device Regulation (IVDR)
- Maskindirektivet, Leketøysdirektivet m.fl.

Regulerte bruksformål: Vedlegg III i KI-forordningen

Frittstående systemer som:

- Biometrisk identifisering, følelsesgjenkjenning
- Kritisk infrastruktur
- Utdannelse og yrkestrening

- Ansettelse og arbeidsforhold
- Tilgang til grunnleggende velferdstjenester
- Rettshåndheverende myndigheter
- Migrasjon, asylsaker, grensekontroll
- Demokratiske prosesser og domstolenes virke

Kravene til KI-systemer som er eller inngår i medisinsk utstyr trer i kraft i EU 2. august 2027 [65]. Kravene til høyrisikosystemer brukt for visse formål, jf. vedlegg III, trer i kraft 2. august 2026.

KI-forordningen stiller blant annet følgende krav til høyrisikosystemer:

1. Risikostyringssystem (artikkel 9)
2. Data og datahåndtering (artikkel 10)
3. Teknisk dokumentasjon (artikkel 11)
4. Loggføring (artikkel 12)
5. Transparens og informasjon til brukere av systemet (artikkel 13)
6. Menneskelig overblikk (artikkel 14)
7. Nøyaktighet, robusthet og cybersikkerhet (eng: *Cyber Security*) (artikkel 15)

Artikkel 11 beskriver krav til teknisk dokumentasjon for et hvert høyrisiko KI-system og dette er ytterligere beskrevet i forordningens vedlegg IV [66]. Dette er dokumentasjon som leverandørene vil måtte utarbeide etter hvert som deler av KI-forordningen får anvendelse i EU, men vil allerede nå kunne være til hjelp i anskaffelsesprosessen.

Begrepene *transparens* og *menneskelig overblikk* (i artikkel 13 og 14) kan være nytt å forholde seg til i anskaffelsesprosessen. At KI-systemet er *transparent* betyr blant annet at det er utformet på en måte som gjør det mulig å forstå hvordan KI-modellen er utviklet. Videre at det er tydelig for brukerne at de samhandler med et KI-system. Når brukeren er klar over at systemet benytter KI, kan brukeren ta hensyn til dette når hen vurderer og benytter informasjon fra systemet. *Menneskelig overblikk* av et KI-system kan innebære en form for menneskelig tilsyn eller mulighet for menneskelig innblanding dersom det er nødvendig. Brukeren skal ikke være helt avhengig av systemet for kritiske beslutninger. Menneskelig overblikk forutsetter blant annet riktig kompetanse og opplæring [67].

KI-systemer med begrenset risiko

Systemer med begrenset risiko er ikke underlagt de samme strenge kravene som høy-risiko KI-systemer, men det er fortsatt krav om at KI-systemene må være transparente og at KI-generert innhold skal merkes i maskinlesbart format [68]. Virksomheter som tar i bruk KI-systemer har videre plikt til å sikre tilstrekkelig kompetanse hos personellet som vil bruke eller drifte KI-systemene.

KI-modeller for generelle formål

KI-modeller som er laget for generelle formål, også kalt GPAI-modeller (eng.: *General Purpose Artificial Intelligence* (GPAI) [69]), kan brukes enten direkte, som ChatGPT og Gemini, eller de kan tilpasses og/eller finjusteres av andre produsenter for et spesifikt formål. Produsenter som baserer seg på GPAI-modeller, vil trenge tilstrekkelig dokumentasjon om disse modellene for selv å kunne dokumentere kvaliteten på sitt produkt.

Selv om GPAI-modellene ikke har et spesifikt formål i seg selv, blir de likevel regulert i KI-forordningen dersom de oppfyller visse vilkår i artikkel 51 [70]. Leverandører av GPAI-modeller må bl.a. sørge for at de som bruker modellene deres forstår dem godt nok til å kunne bruke dem i sine produkter og oppfylle sine forpliktelser i henhold til loven [71]. Dersom slike modeller i tillegg har systemisk risiko, gjelder strengere krav [72]. Eksempler på systemisk risiko kan være at KI-systemet har evne til å påvirke valg, kritisk

infrastruktur og beredskap. Leverandører av GPAI-modeller med systemisk risiko må blant annet utføre nødvendige modellvurderinger for å identifisere, og kontinuerlig vurdere og redusere systemiske risikoeer gjennom risikostyring, overvåking etter at modellen er lansert, og samarbeide med relevante aktører. Alvorlige hendelser og mulige korrigerende tiltak skal umiddelbart rapporteres til EU-kommisjonen og nasjonale kompetente myndigheter.

I påvente av harmoniserte standarder har EU-kommisjonen lansert et første utkast til retningslinjer for KI-modeller for generelle formål [73].

Andre aktuelle regelverk

KI-ansvarsdirektiv

EU-kommisjonen har fremmet et forslag til et KI-ansvarsdirektiv gjelder sivilrettslig erstatningsansvar utenfor kontrakt når skaden forårsakes av et kunstig intelligent system. Forslaget fastsetter felles regler for fremleggelse av dokumentasjon for høyrisikosystemer med kunstig intelligens for å gjøre det mulig for saksøker å underbygge et eventuelt erstatningsansvar [74].

Automatiserte avgjørelser

Personvernforordningen (GDPR) har flere bestemmelser om automatiserte individuelle avgjørelser, særlig i artikkel 22, som handler om avgjørelser som fattes uten menneskelig innblanding. Reglene i personvernforordningen vil slå inn når KI-systemer foretar slike automatiserte avgjørelser eller når de brukes med helmaskinell behandling. Innbyggere har rett til ikke å være gjenstand for en avgjørelse som utelukkende er basert på automatisert behandling, herunder profilering, som har rettsvirkning for eller på tilsvarende måte i betydelig grad påvirker vedkommende [75]. Det gjelder flere unntak, men i så fall har innbyggerne og særskilte rettigheter, som å kreve at et menneske ser på avgjørelsen [76]. Krav om forklarbarhet og transparens angående informasjon om logikk som ligger til grunn for automatiserte avgjørelser er beskrevet i artiklene 13, 14 og 15 i personvernforordningen [77]. Bestemmelser om automatiserte avgjørelser finnes i [pasientjournalloven § 11](#) og [folketrygdloven § 21-11a](#), og åpner også opp for at det kan gis forskrift om at mer inngripende avgjørelser kan treffes automatisert.

Regelverk for tilgang og bruk av helseopplysninger

Helseopplysninger er etter norsk rett underlagt taushetsplikt og for all behandling av personopplysninger stilles det krav om behandlingsgrunnlag etter artikkel 6 og artikkel 9 i personvernforordningen for behandling av helseopplysninger [78].

I all behandling av person- og helseopplysninger er det krav om at det er en dataansvarlig [79]. Den dataansvarlige har ansvaret for at behandlingen av person- og helseopplysninger er i samsvar med gjeldende regelverk, blant annet å sørge for at helseopplysningene behandles lovlig og for tilfredsstillende informasjonssikkerhet og internkontroll.

Personvernforordningen gir rammer for dataansvarliges plikter og oppgaver, og legger opp til sanksjoner dersom pliktene ikke overholdes. Å sørge for at helseopplysningene behandles lovlig. Andre sentrale plikter er blant annet å sørge for tilfredsstillende informasjonssikkerhet og internkontroll.

Dataansvarlig kan inngå avtale med en databehandler om behandling av personopplysninger; en databehandleravtale. Dette er aktuelt når andre skal behandle opplysningene på deres vegne. Databehandleren kan bare behandle opplysningene etter avtalen med dataansvarlig.

Se mer utfyllende om taushetsplikt og tilgang til helseopplysninger for validering i fase 5.

Strålevernregelverket

Innføring av KI-systemer kan påvirke bruk av det strålegivende utstyret eller andre forhold relevant for strålevern, for eksempel beslutningsstøtte for vurdering av berettigelse av en radiologisk undersøkelse. Mange av de aktuelle KI-systemene er tatt i bruk innen fagområder som radiologi, og stråleterapi. Strålevernregelverket stiller også strenge krav til blant annet risikovurderinger, kvalitetskontroll av systemene og kompetanse til å bruke disse [80].

[33] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689

[34] [Lov om behandling av personopplysninger \(personopplysningsloven\) - Lovdata](#)

[35] Definisjon av "tiltenkt formål" i KI-forordningen art. 3 nr. 12

https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689#d1e2090-1-1 () og MDR art. 2 nr. 12 ([32017r0745.pdf \(lovdata.no\)](#))

[36] Det finnes ingen offisiell norsk oversettelse. I denne rapporten lener vi oss på den danske oversettelsen av KI-forordningen, som bruker de begrepet "AI-model til almen bruk".

[37] [Lov om medisinsk utstyr - Lovdata](#)

[38] Se for øvrig EØS-tillegget om medisinsk utstyr her. <https://lovdata.no/static/NLX3/32017r0745.pdf>

[39] [Medical Device Coordination Group Working Groups - European Commission](#)

[40]

https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2019_11_guidance_qualification_classification_

[41] Håndteringsforskriften: [Forskrift om håndtering av medisinsk utstyr - Lovdata](#)

[42] Systemer kategoriseres som høyrisiko-systemer hvis: 1) det inngår som en sikkerhetskomponent i et produkt eller systemet selv er et produkt som reguleres av regelverket opplistet i Annex I og 2) må undergå tredjepartssertifisering etter regelverket i Annex I i KI-forordningen. Artikkel 6.1 i KI-forordningen: [Classification rules for high-risk AI systems \(eur-lex.europa.eu\)](#)

[43] Artikkel 6.2 i KI-forordningen: [L_202401689EN.000101.fmx.xml \(europa.eu\)](#)

[44] Punkt 5 b og c i Vedlegg III i KI-forordningen

[45] Artikkel 27 i KI-forordningen

[46] Artikkel 50 i KI-forordningen

[47] [Krav til produkter i Norge | Direktoratet for samfunnssikkerhet og beredskap \(dsb.no\)](#)

[48] https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en

[49] Direktoratet for medisinske produkter (DMP) er fag- og tilsynsmyndighet for medisinsk utstyr i Norge og forvalter produktregelverket for medisinsk utstyr.[1] Utfyllende informasjon finnes på DMP's nettsider.

[50] [Klassifisering - Direktoratet for medisinske produkter \(dmp.no\)](#)

[51] [mdr--oversikt-klassifiseringsregler.pdf \(dmp.no\)](#)

[52]

<https://www.dmp.no/globalassets/documents/medisinsk-utstyr/figur-klassifisering---mdr---ny-versjon.png>

[53] Se MDR og IVDR artikkel 16

[54] KI-modeller utviklet for bruk på voksne kan ikke uten videre tas i bruk på barn. Se [Use of Artificial Intelligence in Radiology: Impact on Pediatric Patients, a White Paper From the ACR Pediatric AI Workgroup \(jacr.org\)](#)

[55]

<https://www.dmp.no/medisinsk-utstyr/for-helsetjenesten/egentilvirkning-av-medisinsk-utstyr-i-helseinstitusjon>

[56] Forordning om medisinsk utstyr (MDR) trådte i kraft 26. mai 2021 og opphevet direktiv 93/42/EØF om øvrig medisinsk utstyr og direktiv 90/385/EØF om aktivt implanterbart medisinsk utstyr med visse unntak. Forordning om in vitro-diagnostisk medisinsk utstyr (IVDR) trådte i kraft 26. mai 2022 og opphevet direktiv 98/79/EF om in vitro-diagnostisk medisinsk utstyr.

[57] Les mer om [overgangsordninger på DMPs nettsider](#).

[58] [Søknadspliktige kliniske utprøvinger - Direktoratet for medisinske produkter \(dmp.no\)](#)

[59] [U.S. Food and Drug Administration \(fda.gov\)](#)

[60] [Regulation - EU - 2024/1689 - EN - EUR-Lex \(europa.eu\)](#)

[61] Artikkel 113 i KI-forordningen

[62] [Fremtidens digitale Norge – nasjonal digitaliseringsstrategi 2024–2030 \(regjeringen.no\)](#) s. 67

[63] [https://ec.europa.eu/transparency/documents-register/detail?ref=C\(2023\)3215&lang=en](https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=en)

[64] Kategoriene beskrives i hhv. KI-forordningens vedlegg I og III. EU-kommisjonen har myndighet til å endre listen over regulerte bruksformål i vedlegg III etter nærmere angitte kriterier, som angitt i artikkel 7 i KI-forordningen.

[65] Artikkel 113 i KI-forordningen samt <https://artificialintelligenceact.eu/annex/3/>

[66] [L_202401689EN.000101.fmx.xml \(europa.eu\)](#)

[67] KI-forordningens artikkel 26 (2): [L_202401689EN.000101.fmx.xml \(eur-lex.europa.eu\)](#)

[68] KI-forordningens artikkel 50(1)(2): [L_202401689EN.000101.fmx.xml \(eur-lex.europa.eu\)](#)

[69] KI-forordningens artikkel 3 (3) definerer '*general-purpose AI model*' slik: *means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market;*

[70] KI-forordningens kapittel V: [L_202401689EN.000101.fmx.xml \(eur-lex.europa.eu\)](#)

[71] KI-forordningens artikkel 53: [L_202401689EN.000101.fmx.xml \(eur-lex.europa.eu\)](#)

[72] KI-forordningens artikkel 55: [L_202401689EN.000101.fmx.xml \(eur-lex.europa.eu\)](#)

[73]

<https://digital-strategy.ec.europa.eu/en/library/first-draft-general-purpose-ai-code-practice-published-written->

[74] <https://www.regjeringen.no/no/sub/eos-notatbasen/notatene/2022/des/ai-ansvarsdirektivet/id2950178/>

[75] https://lovdata.no/dokument/NL/lov/2018-06-15-38/KAPITTEL_gdpr-3-4#gdpr/a22

[76]

<https://www.datatilsynet.no/rettigheter-og-plikter/den-registrertes-rettigheter/rettar-ved-automatiserte-avgjor>

[77] Personvernforordningen:

https://lovdata.no/dokument/NL/lov/2018-06-15-38/KAPITTEL_gdpr-3-2#KAPITTEL_gdpr-3-2

[78] Les mer om dette på Helsedirektoratets nettsider om kunstig intelligens: [Regelverket for utvikling av kunstig intelligens - Helsedirektoratet](#)

[79] Dataansvarlig er definert i personvernforordningen som en fysisk eller juridisk person, en offentlig myndighet, en institusjon eller ethvert annet organ som alene eller sammen med andre bestemmer formålet med behandlingen av personopplysninger og hvilke midler som skal benyttes eller det er fastsatt i unionsretten eller nasjonal rett, jf. artikkel 4 nr. 7. Se også pasientjournalloven § 2 bokstav e som viser til definisjonen i personvernforordningen når det gjelder hvem som er ansvarlig for behandling av helseopplysninger i behandlingsrettede helseregistre, herunder i pasientjournaler.

[80] <https://lovdata.no/dokument/SF/forskrift/2016-12-16-1659>

Vurdering av kvaliteten til et KI-system

En viktig del av kartleggingen i fase 2 er å få oversikt over ytelse, sikkerhet og kostnadseffektivitet til KI-systemet for å kunne gjøre konkrete vurderinger og planlegging i neste fase.

Ytelsen til et KI-system

Ytelsen til et KI-system kan henge tett sammen med hvordan KI-modellen er utviklet, og er avhengig av hvilke data den er trent på. Dersom datasettene inneholder systematiske skjevheter, kan det føre til urettferdige eller upålitelige resultater. KI-modeller kan være svært fleksible, og kan ofte memorere treningsdataene, uten at dette sier noe om hvorvidt modellen kan generalisere til data den ikke har sett før. Det er derfor viktig at ytelsen av en KI-modell er testet på data som ikke har vært med i treningen.

Produsenten kan ha validert løsningen sin i varierende grad og det bør foreligge dokumentasjon om hvordan den tekniske valideringen er gjort, hvilke datasett som er brukt og at KI-modellen er testet på uavhengige data. [Forordningen om medisinsk utstyr \(lovdata.no\)](#) (MDR) krever at produsenten av CE-merket utstyr skal dokumentere ytelse. Det er ikke et generelt krav til klinisk validering for KI-modeller,

men for medisinsk utstyr i høyere klasse (spesielt IIb og III) kan det være krav om klinisk utprøving [81]. Denne rapporten inneholder spesifikke detaljer rundt klinisk evaluering av programvare som medisinsk utstyr [82]. Artikkel 61 i MDR omhandler klinisk evaluering av medisinsk utstyr, mens artikkel 62 omhandler generelle krav til kliniske utprøvinger.

For ytterligere beskrivelse, se senere faser og KI-faktaark 3 og 2 om henholdsvis validering og ytelse [83].

Sikkerhet til et KI-system

I denne fasen bør det vurderes, på overordnet nivå, om sikkerheten er ivarettatt i tilgjengelige KI-systemer på markedet.

Norm for informasjonssikkerhet og personvern (Normen) er et hjelpemiddel i arbeidet med å sikre informasjonssikkerhet og personvern i helse- og omsorgstjenesten. Normen er teknologinøytral, og har blant annet produsert artikler som omhandler KI [84].

Evaluering av metode og kunnskapsgrunnlag

Dersom det allerede er gjort en metodevurdering av KI-systemet før anskaffelse, vil dette være nyttig informasjon for virksomheten å sette seg inn. Fullstendige metodevurderinger kan tjene som fullverdig beslutningsgrunnlag fordi den vitenskapelige dokumentasjonen utarbeides basert på en grundig metodologisk fremgangsmåte, og fordi den kan inneholde en utfyllende konsekvensvurdering. KI er ny teknologi og så langt er det gjort få fullstendige metodevurderinger av ulike KI-systemer.

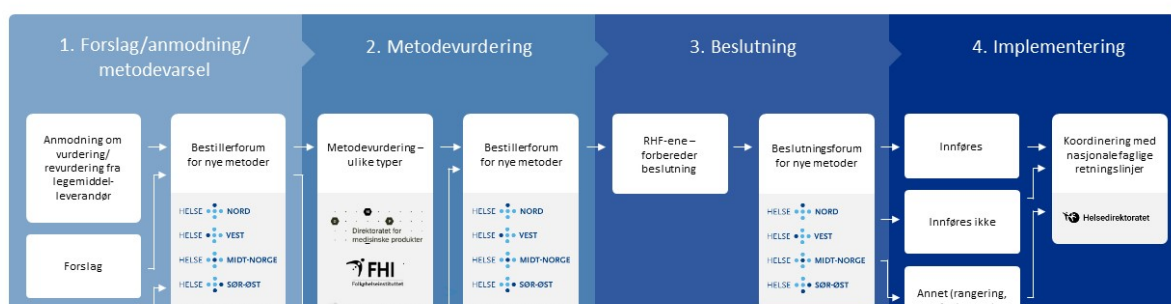
Medisinsk metodevurdering (HTA)

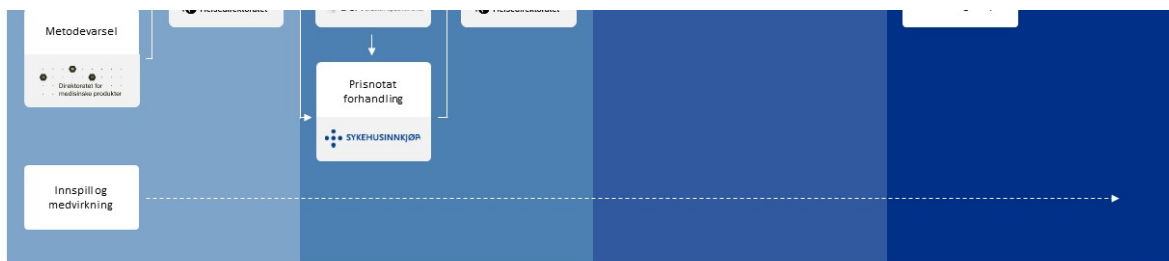
En medisinsk metodevurdering (*Health Technology Assessment*, HTA) er en tverrfaglig prosess med hensikt å belyse effekt, sikkerhet, etikk, organisatoriske forhold, kostnadseffektivitet og budsjettkonsekvenser knyttet til innføring eller utfasing av metoder i helsevesenet [85][86].

Nye metoder er et nasjonalt system for å beslutte hvilke behandlingsmetoder som kan tilbys i spesialisthelsetjenesten. Folkehelseinstituttet og Direktoratet for medisinske produkter gjennomfører metodevurderinger på oppdrag fra Bestillerforum for nye metoder [87]. Direktoratet for strålevern og atomsikkerhet (DSA) bistår inn i metodevurderingen dersom metoden anvender stråling eller kan få påvirkning på den medisinske strålebruken. Det pågår for tiden et arbeid med strategi for videreutvikling av Nye metoder [88]. I tillegg er det nedsatt en arbeidsgruppe som skal utarbeide et forslag til rammer, mål og kriterier for metodevurdering av medisinsk utstyr i Nye metoder.

Prosesskart – nasjonal prosess

NYE METODER





Figur 3: Prosesskart for Nye metoder [89]

Minimetodevurderinger

Mange beslutninger vil skje lokalt og da kan mini-metodevurderingene være aktuelt. En mini-metodevurdering er en nedskalert form av de nasjonale metodevurderingene, og er tilpasset for bruk både i sykehus og kommuner. FHI har ansvar for mini-metodevurderinger, og på deres nettsider finnes veiledning og nasjonal database over alle ferdigstilte og påbegynte mini-metodevurderinger [90]. Her kan man for eksempel finne kontaktperson for andre gjennomførte mini-metodevurderinger.

Strålevernforskriften stiller krav til generisk berettigelsesvurdering av nye metoder som involverer eller påvirker stråling og kan derfor omfatte anskaffelse av KI-systemer. Metodevurderinger der strålevernmessige aspekter er adressert ansees å være tilstrekkelig dokumentasjon.[91]

Erfaringer fra fagfelleverderte publikasjoner

Fagfelleverderte artikler gir uavhengig beskrivelse av teknologien og dokumenterer hvilken effekt den har, herunder vurderinger av klinisk nytteverdi. CE-merket er en dokumentasjon på sikkerhet og ytelse og samsvar med reguleringer.

Erfaringer fra brukere av tilsvarende KI-systemer

Systematisk erfaringsdeling er nyttig når ny teknologi skal tas i bruk. Virksomheten kan vurdere å kontakte andre virksomheter som har gjort seg erfaring med bruk av tilsvarende KI-systemer. Blant annet kan erfaring med brukergrensesnitt og gevinster være best å få fra en reell bruker ved andre helseinstitusjoner eller i andre land. Leverandøren kan oppgi referanser til andre brukere av deres KI-system.

[81] [MDR \(EU\) 2017/745.pdf \(lovdata.no\)](#)

[82] MDCG 2020-1 Guidance on Clinical Evaluation (MDR) / Performance Evaluation (IVDR) of Medical Device Software
https://health.ec.europa.eu/document/download/19d9e24f-2808-4e00-bfeb-75892047407d_en?filename=m

[83] Faktaark om KI vil publiseres på <https://www.helsedirektoratet.no/tema/kunstig-intelligens>

[84] [Sikkerhetsrisiko i systemer som benytter kunstig intelligens – hva vet vi, og hva kan vi gjøre? - ehelse](#)

[85] [Metodevurderinger - FHI](#)

[86] [Hva er metodevurdering? - FHI](#)

[87] [Metodevurdering - Nye metoder](#)

[88] <https://www.nyemetoder.no/om-systemet/strategi-for-nye-metoder-2023-2028/>

[89] [Bakgrunn: Hvorfor har vi Nye metoder? - Nye metoder](#)

[90] [Mini-metodevurdering \(minimetodevurdering.no\)](#)

[91] Se spesielt § 39 Berettigelse: <https://lovdata.no/dokument/SF/forskrift/2016-12-16-1659>

Relevante kilder til fase 2

- [Referansekatalogen for e-helse \(ehelse.no\)](#) gir oversikt over obligatoriske og anbefalte standarder for helse- og omsorgstjenesten, samt andre kravdokumenter som tekniske spesifikasjoner. Foreløpig er det ingen obligatoriske eller anbefalte standarder i referansekatalogen som omhandler KI spesielt. [EU-kommisjonen har imidlertid bedt de europeiske standardiseringsorganisasjonene CEN og CENELEC \(ec.europa.eu\)](#) om europeiske standarder som skal støtte opp under implementeringen av KI-forordningen. Etter hvert vil disse også harmoniseres for norske forhold og inngå i referansekatalogen.
- Den europeiske koordineringsgruppen for medisinsk utstyr (Medical Device Coordination Group: MDCG) er sammensatt av nasjonale eksperter og opprettelsen av gruppen er regulert i MDR (artikkel 103-108). De har bl.a. publisert veiledere om
 - [kvalifisering og klassifisering av medisinsk utstyr-programvare \(health.ec.europa.eu\)](#),
 - [om klinisk evaluering/ytelseevaluering av programvare \(health.ec.europa.eu\)](#),
 - [om kvalifisering av kombinasjoner av programvare og maskinvare \(health.ec.europa.eu\)](#) og
 - [om cybersikkerhet for medisinsk utstyr](#).
- Helsedirektoratet [vurderer for tiden \(1.10.2024\) normeringsgrad for en veileder og standard \(SN-CEN TS/ISO 82304-2:2021\) for kvalitetssikring av ikke-medisinske helseapper \(ehelse.no\)](#). Denne vil bli nyttig å bruke dersom virksomheten skal ta i bruk et KI-system for helse og velvære og som ikke skal brukes til medisinske formål.
- En rapport utgitt av UK National Cyber Security Centre (NCSC) og US Cybersecurity and Infrastructure Security Agency (CISA) med flere, gir retningslinjer for sikker utvikling av KI-systemer, og disse kan benyttes for å se hva man kan forvente at en leverandør gjør for å håndtere informasjonssikkerhetsrisiko i systemene de utvikler: [Guidelines for secure AI system development \(ncsc.gov.uk\)](#)
- The Open Worldwide Application Security Project (OWASP) har laget en oversikt over hvordan håndtere personvern under utvikling av et KI-system: [OWASP AI Security and Privacy Guide | OWASP Foundation](#)
- I tillegg er det publisert en [kronikk om kunstig intelligens i helsetjenesten \(tidsskriftet.no\)](#) i Tidsskriftet for den norske legeforening i oktober 2023 som gir nyttig veiledning.
- Nye metoder: [Forside - Nye metoder](#), [Strategi for Nye metoder 2023-2028 – Nye metoder](#).
- Folkehelseinstituttets side om minimetodevurderinger: [Mini-metodevurdering \(minimetodevurdering.no\)](#)
- Folkehelseinstituttets produktportefølje i Nye metoder: [Endringer i produktporteføljen \(fhi.no\)](#)
- [Erfaringer med utprøving av ISO 42001 ved AHUS](#)
- [Lov om likestilling og forbud mot diskriminering \(likestillings- og diskrimineringsloven\)](#)
- Rapport fra Likestillings- og diskrimineringsombudet: [Innebygd diskrimineringsvern](#)
- Ahus, sluttrapport: [Hjerterom for etisk AI \(datatilsynet.no\)](#). Rapport fra Datatilsynets regulatoriske sandkasse
- Helsedirektoratet: Rundskriv for regelverket for utvikling av kunstig intelligens: [Utvikling og bruk av klinisk beslutningsstøtteverktøy](#)
- [How to get ChatGPT regulatory approved as a medical device — Hardian Health](#)

- [Crafting an intended purpose in the context of software as a medical device \(SaMD\) - GOV.UK \(www.gov.uk\)](https://www.gov.uk/guidance/crafting-an-intended-purpose-in-the-context-of-software-as-a-medical-device-samd)
- [Medisinsk utstyr og kunstig intelligens - Direktoratet for medisinske produkter](#)
- [Veiledning og regelverk - Direktoratet for medisinske produkter](#)
- [Klinisk evaluering og ytelseevaluering - Direktoratet for medisinske produkter](#)
- [EU er i ferd med å etablere en database over CE-merkede produkter \(ec.europa.eu\)](https://ec.europa.eu/europa/)
- Relevant harmonisert standard: [ISO13485 - kvalitetssystem for medisinsk utstyr \(advisera.com\)](https://www.iso.org/standard/62411.html)

Fase 3: Vurdere egen virksomhets evne til å kunne ta i bruk et KI-system

Hensikten med denne fasen er å:

- kartlegge og vurdere virksomhetens menneskelige og tekniske faktorer
- utarbeide beslutningsgrunnlag for og planlegge anskaffelse av et KI-system

Menneskelige faktorer

Som ved innføring av annen teknologi, er menneskelige faktorer viktige for at KI-systemet vil fungere i praksis i virksomheten. KI-systemer er natur sosio-tekniske av natur, noe som betyr at de påvirkes av samspillet mellom medisinsk teknologi, klinisk praksis og menneskelige faktorer, inkludert pasienters og helsepersonells atferd og beslutninger [92]. Fordeler og risikoer knyttet til KI-systemet er avhengig av både konteksten og bruken av det [93].

Etiske vurderinger og retningslinjer

Virksomhetens normer, verdier og atferdsmønstre ligger til grunn for etisk ansvarlig bruk av KI. Helseforetakene har en klinisk etikk-komité som kan konsulteres ved behov. KS samarbeider med Senter for medisinsk etikk om å gi kommunene et best mulig tilbud innen relevant helsefaglig etikk. Virksomheten kan ha særlige retningslinjer som må tas hensyn til, for eksempel at KI ikke skal brukes til spesifiserte formål.

Noen spørsmål som kan melde seg er:

- Hvordan vil de som blir behandlet ved hjelp av et KI-system, akseptere at det ikke bare er et menneske som står bak en beslutning?
- Er systemet rettferdig? Sikrer man tilgang til likeverdige helse- og omsorgstjenester uavhengig av diagnose, bosted, personlig økonomi, kjønn, fødeland, etnisitet og den enkeltes livssituasjon? Tas det høyde for helse- og omsorgstjenester av like god kvalitet og resultatlikhet?

Virksomhetsledelse

Det kan være nyttig å vurdere å ta i bruk et ledelsessystem for KI, som for eksempel [ISO/IEC 42001 Ledelsessystem for kunstig intelligens \(standard.no\)](#). Den kan hjelpe virksomheter med å identifisere og redusere risiko knyttet til implementering av KI.

Virksomheten bør kartlegge hvilke endringer en innføring av et KI-system vil medføre i arbeidsprosessene, slik at ledelsen er godt informert om hvordan innføring av et KI-system vil påvirke de ansatte, og tjenesten som tilbys.

Kartleggingen kan omfatte:

- hvem som blir berørt av endringene og hvilke forventninger de har

- eventuelle endringer i arbeidsbelastning og/eller rutiner for helsepersonell under og etter innføring av systemet
- behov for ressurser og opplæring av de som skal både bruke og forvalte KI-systemet, inkludert å håndtere begrensningene til KI-systemet
 - hvilket behov det er for nye roller, ansvarsområder og samarbeid
 - hvilket behov det er for kompetanseheving
- hvordan virksomhetens skal håndtere uplanlagte situasjoner, og konsekvenser av disse, som for eksempel økte kostnader i en periode hvis
 - et eksisterende system må kjøres i parallell med innføringen av det nye KI-systemet
 - det er midlertidig nedetid for et KI-system som er i bruk

Når KI-forordningen trer i kraft vil virksomheter som skal innføre noen typer KI-systemer med høy-risiko være forpliktet til å gjennomføre en vurdering av konsekvenser for fundamentale menneskerettigheter (*Fundamental Rights Impact Assessment* (FRIA)) [94]. Dette er en vurdering som skal hjelpe virksomheten å kartlegge hvilke mulige negative konsekvenser et KI-system kan ha, og på bakgrunn av dette iverksette de tiltakene som er nødvendige for å ivareta fundamentale rettigheter og de øvrige kravene etter KI-forordningen.

Brukere av KI-systemet

Helsepersonelloven stiller krav om at helsepersonell yter forsvarlig helsehjelp, og det er viktig at de har de rette kvalifikasjoner og kompetanse til å bruke KI-systemet slik det er tiltenkt. De må være trygge på at KI-systemet bidrar til å understøtte deres arbeidsprosesser på en god måte og bør derfor involveres i anskaffelsesprosessen. Deres engasjement, eierskap og kunnskap er sentralt for å sikre aksept og god bruk av KI-systemet. Det er viktig å identifisere deres behov for opplæring og støtte i implementeringsprosessen, og eventuell motstand mot endring av arbeidsprosesser.

Hvordan innføring av KI som beslutningsstøtte påvirker kompetansen til de som skal bruke KI-systemet, kan være viktig å vurdere. Virksomheten må sikre at kompetansen til helsepersonell som bruker KI holder seg over tid, jf. forskrift om ledelse og kvalitetsforbedring i helse- og omsorgstjenesten [95]. Det kan vurderes om det skal opprettes ordninger for dette, som sikrer at helsepersonell unngår og reduserer risiko for automatiseringsbias, altså en overdreven tiltro til KI-verktøy som kan forringe faglig integritet og kvalitet.

KI-forordningen vil stille krav om transparens, som blant annet innebærer at det skal være mulig å forstå hvordan KI-systemet skal tolkes, og at det er tydelig for brukerne at de samhandler med et KI-system [96]. Etter KI-forordningen vil arbeidsgiver også ha en plikt til å informere arbeidstakere at KI-systemer tas i bruk på arbeidsplassen [97].

Personer som KI-systemet skal brukes på

KI-systemet vil ofte brukes for å få råd eller beslutningsstøtte som angår innbyggere, pasienter og andre tjenestemottakere. Virksomheten bør planlegge hvordan god informasjon skal gis til innbyggere, pasienter og tjenestemottakere og ta høyde for at bruk av KI kan skape både entusiasme og skepsis.

Bruk av et KI-system kan medføre endringer i hvordan den aktuelle tjenesten oppleves for innbyggere, helsepersonell, pasienter og tjenestemottakere. Åpenhet og god kommunikasjon om mål og hensikten med bruken, hvorfor og hvordan KI kan understøtte helsetjenesten, kan være nødvendig for å oppnå tillit i befolkningen til bruk av KI i helse- og omsorgstjenesten.

- En undersøkelse fra Storbritannia fant at ni av ti voksne briter mener bruk av KI for å oppdage kreft hovedsakelig er positivt, samtidig som over halvparten (56 %) er bekymret for å stole for mye på

denne teknologien fremfor på profesjonelle vurderinger, og 47 % er bekymret for vanskelighetene med å vite hvem som er ansvarlig for feil når man bruker denne teknologien [98].

Den nasjonale KI-strategien påpeker at gjennomsiktighet handler også om at datasystemer ikke skal utgis for å være mennesker – mennesker skal ha rett til å få vite om de samhandler med et KI-system [99].

Etter pasient- og brukerrettighetsloven har pasienten rett til å medvirke samt å få informasjon om innholdet i helsehjelpen som ytes, blant annet om mulige risikoer [100]. Personvernforordningen stiller krav til at den registrerte skal ha informasjon om hvordan deres personopplysninger brukes, og virksomheten bør vurdere om den skal informere om hvilke helseopplysninger som deles med KI-systemet og til hvilke formål [101].

Virksomheten bør vurdere hvilke eventuelle endringer i egne arbeidsprosesser som må gjøres for å ivareta pasientenes rettigheter ved innføring av KI-systemet. Et brukerutvalg kan konsulteres.

[92] "A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy". <https://dl.acm.org/doi/10.1145/3313831.3376718>

[93] [Artificial Intelligence Risk Management Framework \(AI RMF 1.0\)](#)

[94] Artikkel 27 i KI-forordningen. En FRIA skal bla inneholde en beskrivelse av prosessene der KI-systemet skal brukes, tidsperioden og hyppigheten for bruk, personer eller grupper som kan bli påvirket, spesifikke risikoer for skade, beskrivelse av menneskelige tilsyn og andre tiltak for å redusere risiko, samt håndtering av klager.

[95] [Forskrift om ledelse og kvalitetsforbedring i helse- og omsorgstjenesten - Lovdata](#)

[96] Artikkel 13 i KI-forordningen.

[97] Artikkel 26 i KI-forordningen

[98] Ada Lovelace Institute and The Alan Turing Institute, How do people feel about AI? A nationally representative survey of public attitudes to artificial intelligence in Britain (2023). Available at: <https://adalovelaceinstitute.org/report/public-attitudes-ai>

[99] <https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf>

[100] Pasient- og brukerrettighetsloven kapittel 3

[101] Personvernforordningen art. 13 og 14

Tekniske faktorer

Flere tekniske faktorer er viktige for at KI-systemet vil fungere i virksomhetens infrastruktur. Det må integreres godt i arbeidsflyten, passe inn i miljøet det skal brukes i og være enkelt å bruke. Det er viktig at man er åpen for å gjøre større grep og tenke nytt, som å prioritere etableringen av effektiv infrastruktur for å sikre god arbeidsflyt fremfor å forbedre eksisterende, men begrensede, systemer.

Dataforvaltning, -kvalitet og -kompatibilitet

De som har ansvaret for virksomhetens informasjonsforvaltning, er viktige i kartleggingsprosessen og må involveres. Virksomheten bør planlegge for helhetlig informasjonsforvaltning. Veilederen *Orden i eget hus* kan være nyttig i både kartlegging og planlegging [102]. Virksomheten bør ha vurdert hvilke krav som skal gjelde for behandling av data i KI-systemene og om virksomhetens data tilfredsstillende disse kravene. Dette gjelder for eksempel krav til datakvalitet, dataformat og -struktur, kryptering, komprimering, datamerking og lagringsmetode.

Integrasjon av KI-systemet

I vurderingen av om et KI-system passer for virksomheten, må det avklares hvilke grensesnitt virksomheten bruker for integrasjon mot lokale behandlingssystemer og journalsystemer, og hvilke åpne standarder disse grensesnittene følger. God håndtering av data i og samhandling mellom løsningene er viktig for å sikre effektive prosesser.

Lagring, prosessorkraft og infrastruktur

KI-systemet må vurderes opp mot eventuelle krav virksomheten har til teknisk infrastruktur, for eksempel til prosessering, lagring, nettverk, overvåkningsverktøy, sikkerhet og programvarelisenser. Det bør kartlegges om det er behov for endringer i teknisk infrastruktur, både til testing, etablering og i løpende drift.

Vurdere testmiljøer

Testmiljøer kan være relevant ved anskaffelse av KI-systemer slik at virksomheten har mulighet for å kvalitetssikre KI-systemene på en kontrollert måte. Krav til testmiljøer i de ulike fasene må avklares, inkludert hvilken teknisk infrastruktur som er nødvendig, om leverandøren skal stille med (deler av) testmiljøer og behov for testmiljøer for å teste integrasjon med eksisterende løsninger.

Kostnadene ved testmiljøene må estimeres, både de initiale kostnader og kostnader ved drift og forvaltning. Leverandøren må avklare om det vil påløpe ekstra kostnader for tilgang til deres egne testmiljøer eller ved bruk av KI-systemet i virksomhetens testmiljøer, for eksempel ekstra lisenskostnader.

[102] Veileder for orden i eget hus | Digdir

Planlegge anskaffelse

Kartleggingen som er gjort så langt har gitt innsikt i hvilke KI-systemer som er tilgjengelige, og hvilke menneskelige og tekniske faktorer virksomheten må ta hensyn til ved en anskaffelse. Denne kunnskapen bidrar til å kunne oppdatere risikovurderingen og gevinstberegningene slik at virksomheten har et godt beslutningsgrunnlag med tanke på om man vil gå videre med anskaffelse av et KI-system.

Oppdatere gevinstberegninger

Virksomheten bør så langt det lar seg gjøre vurdere hvilken nytte KI systemet vil ha, på samme måte som i en metodevurdering basert på tilgjengelig kunnskap. En helseøkonomisk evaluering, å sammenligne

kostnader og helseeffekter, er sentral i beslutningen om å anskaffe et KI-system. Elementer som bør inngå i vurderingen er:

- Hvilken nytte KI-systemet gir, klinisk, administrativt, i forskning eller annet
- Ressursbruk ved å innføre og bruke et KI-system
- Hvilken risiko det er forbundet med å innføre et KI-system

Den estimerte bruken av løsningen, for eksempel antall brukere eller antall undersøkelser kan påvirke kostnaden. Innføring av KI-systemer kan også medføre økt bruk av for eksempel diagnostiseringsmetoder. Derfor bør det gjennomføres en vurdering av økning i total kostnader ved en eventuell økt bruk av helsetjenester.

Ved å utføre og dokumentere relevante nullpunktsmålinger, for eksempel hvor mye tid helsepersonell bruker på en gitt arbeidsoperasjoner med dagens løsning, kan disse nullpunktsmålingene benyttes i det videre arbeidet med gevinster.

Oppdatere risikovurdering og planlegge sikkerhetstiltak

Informasjonssikkerhet og personvern knytter seg til hele systemet og dets bruk, og ikke kun KI-modellen. I virksomheten vil KI-systemet inngå i en helhet, hvor det driftes, forvaltes og brukes i en organisasjon som har flere andre systemer og sine måter å jobbe på, sin kultur og sin kompetanse. [Normen \(ehelse.no\)](#) og rammeverk for risikostyring av KI fra NIST kan være nyttig i vurderingen av om krav til informasjonssikkerhet og personvern er ivarettatt [103]. Basert på resultatet av risikovurderingen planlegger virksomheten hvilke tiltak som må iverksettes for å ivareta god risikohåndtering for en forsvarlig tjeneste, både ved innføring og når KI-systemet går over i drift [104].

Planlegge klinisk validering

Virksomheten bør vurdere om det er nødvendig med ytterligere klinisk validering før eller i forbindelse med implementeringen av et KI-system den planlegger å anskaffe. Kartleggingen av markedet gjennomført i fase 2 kan brukes som underlag for vurdering av grad og omfang av validering som virksomheten må gjøre selv. For ytterligere beskrivelse om validering i KI-faktaark 2 [105].

I noen tilfeller vil det være tilstrekkelig å teste løsningen på et lite utvalg av egne data, for deretter å sette produktet i drift og gjennomføre kvalitetskontroll som sikrer at man fanger opp eventuelle feil. I andre tilfeller kan det være fornuftig å gjennomføre prospektive studier, f.eks. randomiserte kontrollerte studier, for å validere KI-systemet i den tiltenkte arbeidsprosessen.

Det kan være hensiktsmessig å planlegge en pilot [106] før KI-systemet tas i bruk i andre deler av virksomheten. Ved å gjennomføre en pilot kan virksomheten redusere risiko, ved å kunne rette opp feil uten at konsekvensene blir for store, før en eventuell skalering. Det er viktig at helseinstitusjonen passer på at valideringen ikke er en klinisk utprøving uten at dette søkes godkjent i henhold til MDR.

[103] [Artificial Intelligence Risk Management Framework \(AI RMF 1.0\) \(nist.gov\)](#)

[104] [Forskrift om ledelse og kvalitetsforbedring i helse- og omsorgstjenesten - Lovdata](#) har som formål å bidra til faglig forsvarlige helse- og omsorgstjenester, kvalitetsforbedring og pasient- og brukersikkerhet, og at øvrige krav i helse- og omsorgslovgivningen etterleves.

[105] Faktaark om KI vil publiseres på [/helsedirektoratet/tema/kunstig-intelligens](#)

[106] En pilot er et begrenset forsøk eller testprosjekt i en virksomhet, så nært den faktiske hverdag som mulig. Formålet er å identifisere potensielle problemer, vurdere effektiviteten, og forstå konsekvensene av det nye systemet i en mindre og kontrollert setting.

Erfaringsdeling og forskning

Virksomheten kan vurdere om det er aktuelt med følgeforskning ved innføring av KI-systemet, eller annen forskning for å bidra til erfaringsdeling. I så fall må det etableres kontakt med relevante forskningsmiljøer, defineres mål og rammer for forskningen og sikres finansiering av arbeidet. Dette aspektet bør være vurdert før konkurransegrunnlaget for anskaffelsen kunngjøres. Klinisk forskning må søkes godkjent av de regionale etiske forskningskomiteene for medisinsk og helsefaglig forskning (REK), og utprøving av medisinsk utstyr må godkjennes av DMP.

Eksempel: Helse Nord RHF har finansiert et følgeforskningsprosjekt hos Nasjonalt senter for e-helseforskning (NSE) om gevinster og verdier ved innføring av kommersielle KI-løsninger på radiologiområdet i Helse Nord.

Relevante kilder til fase 3

- [Forskrift om ledelse og kvalitetsforbedring i helse- og omsorgstjenesten \(lovdata.no\)](#) har som formål å bidra til faglig forsvarlige helse- og omsorgstjenester, kvalitetsforbedring og pasient- og brukersikkerhet, og at øvrige krav i helse- og omsorgslovgivningen etterleves. Helsedirektoratet har laget en veileder til forskriften: [Veileder: Ledelse og kvalitetsforbedring i helse- og omsorgstjenesten](#)
- Artikkel fra Normen om sikkerhetsrisiko i systemer som benytter kunstig intelligens: <https://www.ehelse.no/normen/aktuelt-om-normen/sikkerhetsrisiko-i-systemer-som-benytter-kunstig-i>
- Normens veileder i risikostyring: <https://www.ehelse.no/normen/normen-dokumenter/Veileder-om-risikostyring-i-informasjessikkerhe>
- Rapport fra NSE om personvern fremmende teknologier: <https://ehealthresearch.no/rapporter/personvern-fremmende-teknologier-for-bruk-av-kunstig-intelligen>
- En rapport utgitt av UK National Cyber Security Centre (NCSC) og US Cybersecurity and Infrastructure Security Agency (CISA) med flere gir retningslinjer for sikker utvikling av KI-systemer: [Guidelines for secure AI system development](#)
- The Open Worldwide Application Security Project (OWASP) har laget en oversikt over hvordan håndtere personvern under utvikling av et KI-system: [OWASP AI Security and Privacy Guide | OWASP Foundation](#)
- Rapport fra Datatilsynet om kunstig intelligens og personvern, inkludert anbefalinger om godt personvern i utvikling og bruk av KI: <https://www.datatilsynet.no/globalassets/global/dokumenter-pdf/er-skjema-ol/rettigheter-og-plikter/rap>
- Helsedirektoratets veileder: [Ledelse og kvalitetsforbedring i helse- og omsorgstjenesten](#)
- Aktuell standard: [ISO SO/IEC FDIS 42001: Information technology — Artificial intelligence — Management system](#)

Fase 4: Anskaffe et KI-system

Hensikten i denne fasen er å anskaffe et KI-system og vil i hovedsak omfatte:

- Kravstille
- Lyse ut konkurranse
- Vurdere tilbud
- Inngå avtale med valgt leverandør

Vurdere samarbeid om anskaffelse av KI-systemer

Virksomheten bør vurdere i hvilken skala anskaffelsen skal gjennomføres. Det kan være ulike hensyn som gjør at det bør være en lokal prosess. Et eksempel er at et KI-system er så spesialisert at det bare vil gjøre nytte for noen få brukere i en spesifikk virksomhet. Andre ganger er det hensiktsmessig at virksomheten samarbeider om anskaffelsen med andre kommuner, eller andre sykehus, eventuelt kan anskaffelsen egne seg for en regional anskaffelse. Bildediagnostikk, for eksempel radiologi, er standardisert og gjøres ved mange virksomheter, og kan derfor egne seg for en anskaffelse i større skala.

I noen tilfeller kan det foreligge rammeavtaler som virksomheten kan gjøre avrop på. En rammeavtale gjør anskaffelsen enklere for de som kan gjøre avrop på denne. Det kan derfor være ønskelig å få på plass en rammeavtale dersom dette ikke er på plass i virksomheten

Regelverk for anskaffelser

Anskaffelser er regulert av lov og forskrift, se veiledning på anskaffelser.no. Her finnes aktuelle lover, forskrifter og veiledere for blant annet beregning av anskaffelsens verdi og hvilke terskelverdier som gjelder for å velge riktig anskaffelsesprosedyre. Også adgangen til å endre en allerede inngått kontrakt er regulert. Etter hvert som KI-forordningens regler trer i kraft, vil det bli viktig å avklare hvilken risikoklassifisering et KI-system har. Dette vil ha betydning for hvilke krav virksomheten skal stille til leverandøren og hvilke krav virksomheten må forholde seg til etter KI-forordningen, som bruker av KI-systemet.

Det vil være hovedsakelig tre typer anskaffelser av KI-systemer i helse- og omsorgssektoren: KI til klinisk bruk, KI til administrativ bruk og KI til bruk i forskning og innovasjon.

Egne bestemmelser for forskning og innovasjon

Dersom anskaffelsen gjennomføres som del av et forsknings- eller innovasjonsprosjekt, har anskaffelsesforskriften egne bestemmelser som kan benyttes som gir unntak fra plikt til å kunngjøre slike

kontrakter [107][108]. KI-forordningen vil ikke gjelde for KI-systemer og modeller, inkludert output, utviklet og tatt i bruk, dersom det eneste formålet er forskning og utvikling, jf. KI-forordningen art. 2 nr. 6. Reglene i KI-forordningen kommer imidlertid til anvendelse dersom KI-systemet skal settes ut på markedet.

[107] For gjennomføring av innovasjons- og utviklingsprosjekter, se:

<https://anskaffelser.no/innovasjon/finn-riktig-metode-gjennomforing-av-innovasjon-og-utviklingsprosjekter>

[108] For unntak for FoU-kontrakter, se:

<https://anskaffelser.no/avtaler-og-regelverk/anskaffelsesprosedyrer/forsknings-og-utviklingskontrakter/unntak>

Anskaffelsesprosedyre

Valg av anskaffelsesprosedyre påvirker tidsbruk, ressursbruk og muligheter for dialog med tilbydere i anskaffelsen. I et marked med rask teknologisk utvikling må behovet for kvalitetssikring veies mot risiko for å anskaffe noe som er utdatert dersom prosessene tar for lang tid. Hvor godt man kjenner markedet for det som skal anskaffes, vil påvirke valg av anskaffelsesform og danne grunnlag for konkurranse-strategien.

Konkurransepreget dialog kan være en relevant og fleksibel anskaffelsesprosedyre for KI-systemer dersom det er uklart fra starten hvilken løsning som er best egnet. I en prekvalifiseringsfase kan man da i første omgang begrense antallet leverandører som får delta, basert på deres kompetanse og erfaring, og deretter gjennomføre en dialog, der man sammen med leverandørene kan tilpasse kravspesifikasjonen [109].

- Som eksempel er Helse Sør-Øst RHF's anskaffelse av en rammeavtale for KI-plattformer en konkurransepreget dialog [110].

Ved mange virksomheter finnes det prosedyrer for fremgangsmåte ved anskaffelser. Som eksempel har Oslo universitetssykehus beskrevet en slik prosedyre i sitt elektroniske håndboksystem for anskaffelse av medisinsk teknisk utstyr, ehåndboka [111].

[109] [Konkurransepreget dialog | Anskaffelser.no](#)

[110] [Kvalifisering: Rammeavtale KI plattformer til Helse Sør-Øst | Doffin, Database for offentlige anskaffelser](#)

[111] [eHåndbok - Samhandlingsmodell for MTU-IKT i OUS \(ous-hf.no\)](#)

Prismodeller

KI-systemer har ulike prismodeller og de kan endre seg i større grad enn for mer modne produkter. Prismodellene kan inkludere engangskostnader, vedlikeholds- og oppgraderingskostnader og/eller kostnader basert på bruk. **Prismodellen kan også være skalerbar**, slik at kostnadene avhenger av bruken. Dersom KI-leverandøren har en fast grunnpris, er kartlegging av tiltenkt bruk og gevinstvurderingene som er gjort tidligere nyttige. Da kan virksomheten sikre seg at anskaffelsen er kostnadseffektiv.

Avtaler

[Det vises til Statens standardavtaler \(SSA\) | Anskaffelser.no](#) for standardavtaler for offentlige anskaffelser av for eksempel software og IT-løsninger.

I en anskaffelse er det en hovedavtale med en tjenestenivåavtale (*Service Level Agreement (SLA)*) og vedlegg som består av blant annet kravspesifikasjon.

En tjenestenivåavtale regulerer ansvar og respons ved drift og feil ved KI-systemet. Denne sier noe om hvem som har ansvar for hva og hvordan feil skal håndteres. Den bør også inneholde plan for versjonsoppdatering. Kontrakten kan åpne for videreutvikling av KI-modellen.

Dersom leverandøren skal behandle personopplysninger på vegne av virksomheten må det etableres en databehandleravtale [112]. Leverandøren kan ikke uten videre bruke virksomhetens data til egne formål, for eksempel til forskning eller forbedring av KI-systemet. Det er viktig å få avklart dette i tidlig leverandørdialog og få det inn i kontrakten.

[112] [Helsedirektoratets mal og veileder til standard databehandleravtale \(ehelse.no\)](#)

Kravspesifikasjon

Kartlegging i tidligere faser danner grunnlag for hvordan man skal utarbeide en kravspesifikasjon for anskaffelse. Dersom det i fase 2 ble gjennomført en tidlig markedsdialog, har dette gitt innsikt i hvilke produkter som er tilgjengelig og hvilke muligheter og begrensninger disse gir.

En kravspesifikasjon er et detaljert dokument som skisserer de spesifikke kravene systemet må oppfylle. Man må vurdere hvor detaljert man skal kravstille, da for detaljerte krav kan innskrenke mulighetsrommet. Det kan være nyttig å utforme brukerhistorier og overlate til leverandørene å komme med forslag om hvordan behovene kan løses, eller man kan beskrive detaljert hvordan man ønsker KI-systemet skal settes opp og fungere.

Virksomheten må kravstille at produsenten fremlegger dokumentasjon slik at KI-systemets ytelse, tillitsverdighet og sikkerhet kan vurderes grundig før kontraktsinngåelsen.

Den kommende KI-forordningen stiller krav til KI-systemer med begrenset og høy risiko, som det kan være lurt å ta høyde for i kravsettingen, i tillegg til andre regulatoriske krav [113]. Vi viser til fase 2. KI-forordningens artikkel 11, med tilhørende vedlegg IV beskriver mer detaljert hva dokumentasjonen skal inneholde for et høyrisiko KI-system [114][115].

Kravspesifikasjonen inneholder krav som deles i ulike grupper:

1. Generelle krav
2. Funksjonelle krav
3. Tekniske krav
4. Krav til sikkerhet

Generelle krav

Noen generelle krav kan være at leverandøren må redegjøre for etiske og samfunnsmessige problemstillinger knyttet til KI-systemet [116]. Krav om ivaretagelse av fundamentale menneskerettigheter,

som helse og sikkerhet, samt miljø og bærekraft kan inkluderes her. KI-systemer forbruker betydelig mengder energi, både til lagring og bruk av store datamengder til trening og bruk av KI-modeller. Det ble fra 1. januar 2024 gjort en endring i forskrift om offentlige anskaffelser. Ved alle anskaffelser, også av KI-systemer, må helsetjenesten vekte klima- og miljøhensyn med minimum tretti prosent når miljøkravet er relevant for anskaffelsen. Alle KI-systemer bør bidra til å oppfylle FNs bærekraftsmål.

Med KI-forordningen vil det også komme krav om at høyrisiko-KI-systemer må være utformet på en slik måte at transparens og menneskelig overblikk er sikret. Virksomheten kan be leverandøren beskrive hvilke trenings- og testmetoder som er benyttet. Krav om kvalitetsstyringssystem og logging som beskrevet i KI-forordningen, kan inngå her. For medisinsk utstyr er overvåking et krav til leverandør i henhold til for eksempel MDR.

Funksjonelle krav

Brukere av KI-systemet bør stille funksjonelle krav, for eksempel hvordan brukergrensesnittet skal fungere, og hvordan resultatet fra KI-systemet bør presenteres for brukeren av systemet. Det kan være behov for å stille krav om for eksempel forklarbarhet, tolkbarhet og transparens for de som skal bruke KI-systemet.

Datakvalitet er noe man spesielt må vurdere i en KI-anskaffelse, da dataene som brukes til å trene KI-modeller påvirker systemets ytelse, pålitelighet og upartiskhet. Det bør også stilles krav om at leverandøren kan redegjøre for hvilke data [117] de har brukt i utviklingen av KI-systemet, både i trening og valideringen, og hvilket behandlingsgrunnlag etter personvernforordningen. KI-forordningen stiller krav til datakvalitet. Artikkel 10 stiller krav til data og datahåndtering for KI-systemer med høy risiko. Data til trening, validering og testing skal tilfredsstillende kvalitetskriteriene og skal blant annet være relevante, tilstrekkelig representative, og i størst mulig grad være fri for feil og komplette med tanke på den tiltenkte bruken [118].

Krav om dokumentasjon på ytelse og kvalitet til KI-systemet i bruk må tas med. For eksempel hvordan klinisk evaluering eller utprøving er gjennomført, og på hvem, med hvilket utstyr.

Tekniske krav

KI-systemet skal passe i virksomhetens infrastruktur. Det stilles derfor krav til beskrivelser av dataflyt og dataformater som skal håndteres og utveksles mellom andre systemer og KI-systemet. Hvilken skalerbarhet KI-systemet har med tanke på antall brukere eller bruk, og eventuelle krav om testmiljøer kan inkluderes. Her vises det ofte til virksomhetens standarddokumenter, eller tilsvarende internasjonale eller norske standarder og tekniske godkjenninger.

KI-modeller kan benyttes både lokalt på egne servere og i sky. Benytter man seg av skyløsninger er det viktig å kravstille i henhold til [gjeldende anbefalinger \(ehelse.no\)](https://ehelse.no/gjeldende-anbefalinger).

Krav til sikkerhet

De fleste virksomheter i helse- og omsorgssektoren vil være forpliktet til å følge Normen ([Norm for informasjonssikkerhet og personvern i helse- og omsorgssektoren \(ehelse.no\)](https://ehelse.no/norm-for-informasjessikkerhet-og-personvern-i-helse-og-omsorgssektoren)). De tiltakene som er beskrevet i Normen og er viktige for å sikre digitale systemer generelt, er også viktige i beskyttelsen av KI-systemer.

Virksomheten må stille krav om at leverandøren har hensyntatt de spesielle utfordringene og truslene knyttet til KI-systemer. [Normen har publisert en artikkel om sikkerhet i KI-systemer \(ehelse.no\)](#) og [MITRE ATLAS™](#) er en oversikt over sikkerhetstrusler for KI-systemer.

[ENISA peker på følgende trusler i KI-livsløpet:](#)

- **Unnvikelse** (eng: evasion): KI-systemer kan ofte lures på nye måter, eksempel ved at angriper benytter spesielle inngangsverdier (eng: *adversarial examples, prompt injections*).
- **Forgiftning** (eng: *poisoning*): Angriper påvirker treningsdata eller KI-modellen slik at oppførselen til systemet blir påvirket i en bestemt retning.
- **Lekket informasjon om treningsdata eller KI-modell**: En angriper får hentet ut informasjon om KI-modellen, slik som konfigurasjon, parametre og treningsdata
- **Kompromitterte komponenter i KI-systemet**: Angriper kompromitterer en komponent, for eksempel ved å utnytte sårbarheter i open-souce-biblioteker som KI-systemet benytter.
- **KI-systemet feiler**: Hele KI-systemet feiler, for eksempel ved en feil eller ved at en angriper klarer å ta ned KI-systemet.

[113] [L_202401689EN.000101.fmx.xml \(europa.eu\)](#)

[114] [Art. 11 Technical Documentation - EU AI Act](#)

[115] [Annex IV: Technical Documentation Referred to in Article 11 \(1\) - EU AI Act](#)

[116] [Kunstig intelligens og etikk – SN-ISO/IEC TR 24368 \(standard.no\)](#)

[117] Hvordan data er selektert, opprinnelsen til dataene og prosessene når dataene oppsto. Hvilket utstyr som har blitt brukt for å innhente data. For eksempel hvilket kamerautstyr er benyttet i utvikling av en KI-modell for bildetolkning. Hvordan dataene er merket (eng: labelled), om de er korrekte, komplette og dekkende for formålet de skal brukes til.

[118] [Artikkel 10 i KI-forordningen \(eur-lex.europa.eu\)](#)

Vurdering av leverandørene

Ved gjennomføring av konkurransen gjør man vurdering av både leverandørene og de tilbudte KI-systemene basert på kravene i konkurransegrunnlaget.

Leverandørens verdikjede bør vurderes, det vil si alle ledd fra produsent til leverandør, og eventuelle underleverandører, og hvor robust denne verdikjeden er. Dette innebærer en vurdering av etikk, risiko og sårbarhet.

I et marked med rask teknologisk utvikling og mange oppstartsbedrifter, vil det det skje oppkjøp og konsolideringer som kan påvirke produktene. Noen bedrifter vil ikke lykkes og de kan beslutte å legge ned hele bedriften eller et eller flere av produktene sine. Virksomheter bør tenke igjennom hvordan dette vil kunne påvirke dem, hva som da eventuelt må gjøres, og om det er mulig å gjøre noen tiltak i forkant som kan gjøre det lettere å håndtere slike hendelser, eller om muligheten for å overføre kontrakten til annen leverandør bør reguleres i kontrakten.

Vurdering av tilbudte KI-systemer

Å teste KI-systemer med organisasjonens egne data kan være en måte å vurdere KI-systemets egnethet på i denne fasen. Det gir mulighet for å teste systemet rent teknisk med tanke på informasjonsflyt og vurdere KI-modellen med virksomhetens egne data.

Det kan også vurderes om en retrospektiv valideringsprosess skal gjøres i denne fasen. Dersom leverandørene åpner for dette, så er det viktig å samtidig ivareta krav til likebehandling av leverandørene [119]. Se fase 5 for lovlig behandlingsgrunnlag og unntak fra taushetsplikten for bruk av data fra virksomhetens journaler. Dersom validering er ønskelig, men ikke gjennomførbart før kontrakt signeres, kan virksomheten vurdere å kontraktsfeste muligheten for å gå ut av kontrakten dersom det viser seg at ytelsen ikke er som forventet når virksomheten får validert KI-systemet på egne data.

[119] Som også skrevet i fase 2, vil en situasjon der man har testet en løsning fullt ut (inkl. integrasjoner) fra en enkelt leverandør før man setter i gang en konkurransefase kunne skape utfordringer når det gjelder å gjennomføre en anskaffelse som ivaretar kravet til likebehandling i anskaffelsesforskriften. I slike situasjoner må man utjevne eventuelle fordeler en leverandør har fått, som gjerne i slike saker medfører at anskaffelsesprosessen kan ta ganske lang tid, fordi de øvrige leverandørene også må få det samme kunnskapsgrunnlaget og bør ha mulighet til å vise frem sine løsninger på tilsvarende måte, før det tildeles kontrakt.

Relevante kilder til fase 4

- [Standard databehandleravtale for bruk i helse- og omsorgssektoren, med veileder \(ehelse.no\)](#)
- [Veileder i bruk av skytjenester til behandling av helse- og personopplysninger \(ehelse.no\)](#)
- [Hvordan benytte Normens krav i anskaffelser \(ehelse.no\)](#)
- [Bruk av databehandler \(Normen faktaark 10\) \(ehelse.no\)](#)
- [Informasjonssikkerhet og personvern for leverandører til helse- og omsorgssektoren \(ehelse.no\)](#)
- [Veileder for rettigheter ved behandling av helse og personopplysninger \(ehelse.no\)](#)

Fase 5: Innføre og kvalitetssikre et KI-system

Hensikten med denne fasen er å kvalitetssikre innføringen ved å:

- Gjøre en ny risikovurdering
- Utføre nødvendige tiltak
- Teste og validere KI-systemet

Ny risikovurdering

I tillegg til vurderinger som gjøres ved innføring av vanlige IKT-systemer, som [personvernkonsekvensvurdering \(ehelse.no\)](https://www.ehelse.no/personvernkonsekvensvurdering) (eng. *Data Protection Impact Assessment*, DPIA) og risiko- og sårbarhetsanalyse (ROS), kan denne fasen innebære å gjøre en konsekvensvurdering av KI-systemet, for å fange opp andre og KI-spesifikke risikoer [120].

Krav om gjennomføring av en DPIA vil ofte kunne være tilfelle ved bruk av KI, siden det foreløpig er snakk om ny og innovativ teknologi (i henhold til personvernforordningen[121]). Når KI-forordningen trer i kraft, vil det bli et krav om at virksomheter gjennomfører en FRIA[122] for høyrisiko-KI-systemer brukt på enkelte områder som kan påvirke menneskerettigheter, dersom de leverer tjenester til allmenheten. Hensikten med en FRIA er å beskytte grunnleggende rettigheter, øke tilliten til KI og forebygge negative konsekvenser.

En konsekvensvurdering av grunnleggende rettigheter (FRIA) er en detaljert analyse av:

- Virksomhetens prosesser
 - Hvordan vil KI-systemet bli brukt i praksis?
- Tidsperiode og hyppighet
 - Hvor lenge og hvor ofte vil systemet bli brukt?
- Berørte personer
 - Hvilke kategorier av personer eller grupper vil sannsynligvis bli påvirket?
- Spesifikke risikoer
 - Hvilke risikoer for skade kan oppstå for de berørte personene?
- Tiltak for menneskelig tilsyn
 - Hvordan vil man sikre menneskelig kontroll og overvåking av systemet?
- Tiltak for å håndtere risiko
 - Hvilke tiltak vil bli iverksatt for å minimere eller eliminere de identifiserte risikoene?

En KI-konsekvensvurdering kan omfatte (men er ikke begrenset til) områder som påvirker pasientsikkerhet, rettferdighet, transparens, forklarbarhet, cybersikkerhet, informasjonssikkerhet, personvern, finansielle konsekvenser, tilgjengelighet og menneskerettigheter, og å oppdage eventuell ondsinnet bruk av systemet [123]. Det kan være noe overlapp mellom en slik KI-konsekvensvurdering, DPIA og FRIA. Det kan derfor være hensiktsmessig å gjennomføre disse vurderingene samtidig. Det finnes flere kilder til hvordan man kan gjøre konsekvensvurdering, og ulike maler [124][125][126]. Det danske Datatilsynet har publisert en mal tiltenkt en personvernkonsekvensvurdering ved bruk av personopplysninger i KI-systemer [127].

NIST har utgitt et risikorammeverk for kunstig intelligens ([NIST AI RMF 1.0](#)). Dette rammeverket kan benyttes til å bygge en struktur rundt håndtering av risiko knyttet til kunstig intelligens, og gir en oversikt over de ulike typene av risiko som bør vurderes.

Disse vurderingene ligger til grunn for virksomhetens plan for risikohåndtering og er bestemmende for hvilke tiltak som er nødvendige å implementere utover det som er beskrevet i KI-systemets bruksanvisning [128][129].

[120] Virksomhetene må følge kravene i personopplysningsloven og personvernforordningen, jf. https://lovdata.no/dokument/NL/lov/2018-06-15-38/gdpr/ARTIKKEL_25#gdpr/ARTIKKEL_25.

[121] GDPR artikkel 35: [L_2016119EN.01000101.xml \(eur-lex.europa.eu\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:2016119EN.01000101.xml)

[122] Fundamental rights impact assessment (FRIA), KI-forordningen artikkel 27

[123] Se også ISO/IEC 42001:2023: Ledelsessystem for kunstig intelligens.

[124] [AI and data protection risk toolkit | ICO](#)

[125] [Government of the Netherlands: AI Impact Assessment \(government.nl\)](#)

[126] [RAI Institute: Artificial Intelligence Impact Assessment \(AIIA\) - GOV.UK](#)

[127] [Nye skabeloner til gennemførelse af konsekvensanalyser \(datatilsynet.dk\)](#)

[128] [Forskrift om håndtering av medisinsk utstyr - Lovdata](#)

[129] Artikkel 26 i KI-forordningen

Iverksette tekniske og organisatoriske tiltak

Virksomheten må iverksette passende tekniske og organisatoriske tiltak med utgangspunkt i tidligere vurderinger. Noen tiltak som kan være relevante for implementering av KI-systemer er:

- **Tekniske tilpasninger og integrasjoner:** Eksisterende systemer må tilpasses det nye KI-systemet og nye arbeidsprosesser. Virksomheten kan vurdere å legge inn begrensninger i KI-systemet slik at det benyttes innenfor rammene det er utviklet og validert for [130]. Et eksempel kan være et KI-system som er utviklet og validert kun på voksne. Dersom en bruker av KI-systemet sender inn data for et barn, og alder inngår i inndataene, så kan KI-systemet gi tilbakemelding om at bruk av systemet på denne pasienten ikke er innenfor tiltenkt formål.
- **Opplæring:** Før systemet tas i bruk, må virksomheten sikre at relevante brukergrupper får opplæring og forstår KI-systemets funksjoner, begrensninger og etiske aspekter. Brukere må vite hvilke melderutiner, til hvilken instans eller hvilke meldesystemer som gjelder dersom det under bruk blir oppdaget feil og avvik forbundet med KI-systemet, for å sikre at feil oppdages og rettes opp i. Det bør tas utgangspunkt i allerede eksisterende avvikshåndteringssystemer i virksomheten [131]. Dette bør ivareta KI-forordningens krav til menneskelig overblikk.

[130] Se "tiltenkt formål" i fase 2

[131] [Forskrift om håndtering av medisinsk utstyr - Lovdata](#)

Testing av KI-systemet

Etter at tilpasninger og integrasjoner er utført, må de testes. Testingen omfatter både selve KI-systemet og helheten det skal inngå i, og at KI-systemet produserer utdata som forventet. Dette innebærer test av funksjonalitet, sikkerhet, robusthet, arbeidsflyt og test av integrasjoner i hele arbeidsflyten. CE-merket medisinsk utstyr har et definert tiltenkt formål og helsevirksomheten skal følge produsentens anvisninger før det kan tas i bruk.

Virksomheten har ofte bestemte rutiner for hvordan testing skal gjøres. Det finnes ulike begreper for ulike måter å teste på og noen av disse er overlappende. Testene bør være formelle og inngå i avtalen med leverandøren om hvilken part som gjør hva og hvilke feil virksomheten aksepterer, samt plan for å rette disse opp.

Klinisk validering

Enhver virksomhet innen helse- og omsorg er pålagt å tilby faglig forsvarlige helse- og omsorgstjenester og god pasientsikkerhet [132]. KI-systemet bør derfor testes med representative data fra virksomheten for å kvalitetssikre nøyaktighet, presisjon og pålitelighet. At et KI-system er CE-merket, betyr ikke nødvendigvis at KI-systemet vil fungere godt i praksis på virksomhetens pasienter, utstyr og undersøkelsesprotokoller.

Validering som begrep kan ha flere betydninger avhengig av kontekst. Validering av KI-modeller refererer til prosessen med å evaluere ytelsen til en modell, og gir et mål på hvor godt en modell generaliserer til ukjente data.

Begreper som ofte brukes om KI, er *intern validering*, som er gjort av produsent, *ekstern validering*, som er gjort av uavhengig part, *teknisk validering*, som gjøres i utviklingsprosessen og *klinisk validering*, som gjøres i en klinisk setting, etter at KI-modellen er ferdigutviklet. Valideringen kan enten gjøres *retrospektivt*, på data innsamlet i forkant, og muligens til et annet formål, eller *prospektivt*, på data fra en reell brukssituasjon.

Sort boks-problemet med KI refererer til utfordringen med at avanserte KI-modeller, spesielt de som er basert på dype nevrale nettverk, kan gi resultater uten å tydelig forklare hvordan de kom frem til dem. En slik manglende transparens og forklarbarhet bryter i utgangspunktet med de grunnleggende prinsipper for kunnskapsbasert medisinsk diagnostikk og behandling. I tillegg er logikken til modellen gjerne utilgjengelig for virksomheten av kommersielle hensyn [133]. Virksomheten må derfor sikre at KI-systemet er forsvarlig i bruk. Så lenge det er vanskelig å forklare hvordan en KI-modell virker, vil klinisk validering på virksomhetens egne data være en metode for å teste hvor godt KI-systemet er tilpasset den aktuelle befolkningen det skal brukes på.

Validering på representative datasett gjøres for å identifisere skjevheter i KI-modellen tidlig og for å kunne håndtere eventuelle svakheter når KI-systemet brukes. Hvis en KI-modell ikke gir tilfredsstillende ytelse for en befolkningsgruppe, må virksomheten planlegge hvordan helsetjenesten skal tilbys denne gruppen [134].

Hvor omfattende klinisk validering som er nødvendig, avhenger blant annet av hvordan KI-modellen er trent og på hvilke data, og hvordan det skal brukes. Dersom virksomheten skal ta i bruk et KI-system som allerede er innført og validert i en annen virksomhet med lignende pasientpopulasjon, med samme type utstyr og samme protokoller, vil omfanget av den kliniske valideringen kunne reduseres. Dersom virksomheten planlegger at KI-systemet skal fungere med høy grad av autonomi, vil dette kreve mer omfattende klinisk validering. Validering av KI-systemer bør gjøres risikobasert, og gjentas, helt eller delvis, hver gang leverandøren oppdaterer KI-systemet, for eksempel ved nye versjoner.

Innen radiologi har Kortenesniemi et al nylig publisert en artikkel om hvordan man i praksis kan teste et KI-system som skal tas i bruk innen radiologi [135]. Denne artikkelen beskriver fremgangsmåter som også kan være nyttig for andre fagområder.

Regelverk for tilgang til data til validering

Virksomheten har ansvaret for den kliniske valideringen, og vil som regel være dataansvarlig for dataene som skal brukes for denne typen validering. Helseopplysninger må behandles i samsvar med taushetsplikten. Dette kan skje etter samtykke fra den enkelte pasient, men det vil ofte ikke være praktisk mulig eller være for vanskelig å innhente samtykke når det gjelder store mengder data.

Virksomheten må som dataansvarlig ha lovlig behandlingsgrunnlag og et unntak fra taushetsplikten dersom det skal brukes pasientdata (til annet formål enn helsehjelp, pasientjournalloven § 20 [136]) fra virksomhetens journaler i virksomhetens kliniske valideringsprosess. Intern kvalitetssikring i virksomheten, hvor formålet er å sjekke at helsehjelpen som ytes opprettholder tilsvarende eller bedre kvalitet ved bruk av KI-systemet, kan gi unntak fra taushetsplikten, basert på lovbestemt unntak som helsepersonelloven §26 [137].

Denne hjemmelen kan også gjelde for samarbeidende virksomheter, med avtale som beskrevet i § 9 i pasientjournalloven [138]. Dersom flere virksomheter har gått sammen om å anskaffe et KI-system og det er behov for å dele opplysninger mellom virksomhetene, vil dispensasjonsadgangen i helsepersonelloven § 29 være aktuell som unntak fra taushetsplikten for kvalitetssikringsformål [139][140]. Et dispensasjonsvedtak gir både unntak fra taushetsplikten og et rettslig behandlingsgrunnlag for bruk av helseopplysninger fra pasientjournaler og andre behandlingsrettede helseregistre til blant annet formål om å utvikle, teste og bruke klinisk beslutningsstøtteverktøy, samt kvalitetssikring.

Det er som oftest nødvendig eller hensiktsmessig at leverandøren av KI-modellen også har en rolle i valideringen som gjøres i virksomheten. Som regel vil leverandøren være en databehandler for virksomheten i denne situasjonen, ved at den behandler data på vegne av virksomheten. Det vil i slike tilfeller være nødvendig etter personvernforordningen at det inngås en databehandleravtale mellom virksomheten og leverandøren for dette formålet [141]. Databehandleravtale inngås før virksomheten setter i gang med behandling av personopplysninger, se også fase 4 om avtaleform.

Vurdering av KI-systemets ytelse vil inngå i valideringsprosessen, se KI-faktaark 3. For ytterligere beskrivelse av validering se KI-faktaark 2 [142].

[132] [Forskrift om ledelse og kvalitetsforbedring i helse- og omsorgstjenesten - Lovdata](#)

[133] [Strategi for kunstig intelligens i Helse Nord for 2022-2025 \(PDF\)](#)

[134] [Algoritmer, kunstig intelligens og diskriminering : En analyse av likestillings- og diskrimineringslovens muligheter og begrensninger \(ldo.no\)](#)

[135] [Testing process for artificial intelligence applications in radiology practice - ScienceDirect](#)

[136] [Lov om behandling av helseopplysninger ved ytelse av helsehjelp \(pasientjournalloven\) - Lovdata](#)

[137] <https://lovdata.no/lov/1999-07-02-64/§26>

[138] [Lov om behandling av helseopplysninger ved ytelse av helsehjelp \(pasientjournalloven\) - Kapittel 2. Pasientjournaler og andre behandlingsrettede helseregistre - Lovdata](#)

[139] Mer omtale av § 29 i Helsedirektoratets rundskriv til helsepersonelloven: [Taushetsplikt og opplysningsrett - Helsedirektoratet](#)

[140] Les mer om hvordan man søker om dispensasjon: [Søknad om dispensasjon fra taushetsplikt - Helsedirektoratet](#)

[141] [Hvordan lage en databehandleravtale? | Datatilsynet](#)

[142] Faktaark om KI vil publiseres på [Kunstig intelligens](#)

Overlevering med plan for drift og forvaltning

Overlevering av KI-systemet innebærer å overlevere ansvaret til de som skal drifte og forvalte KI-systemet.

Det er viktig med avklaring av ansvarsforhold for både det tekniske og sikkerhetsmessige knyttet til KI-systemet, og for helsetjenesten. Bruk av KI-systemer kan kreve at virksomheten etablerer nye roller og ansvarsområder som ikke er kjent fra tidligere, knyttet til å ha menneskelig overblikk, overvåke ytelsen og håndtere datakvalitet gjennom hele livssyklusen til KI-systemet. Det vil være nødvendig å lage rutiner for hvordan feilfunksjoner og uventet ytelse fra KI-systemet skal registreres og håndteres videre.

Før overlevering av KI-systemet til driftsorganisasjonen, skal testing og opplæring være gjennomført, og alle roller og ansvarsområder ivarettatt.

Relevante kilder til fase 5

- [Normen beskriver organisatoriske og tekniske tiltak som anses egnet for å oppnå tilfredsstillende informasjonssikkerhet og personvern i sektoren \(ehelse.no\)](#)
- [Mal og veiledning for utfylling av personvernkonsekvensvurdering \(DPIA\) \(ehelse.no\)](#)
- [Veileder om risikostyring i informasjonssikkerhet og personvern \(ehelse.no\)](#)
- NSE har utgitt en kunnskapsoppsummering om bruk av personvernbevarende teknologier for kunstig intelligens i helse- og omsorgstjenesten:
<https://ehealthresearch.no/prosjekter/kunnskapsoppsummering-om-bruk-av-personvernbevarende-te>
- [ISO/IEC 42001 Ledelsessystem for kunstig intelligens](#)
- [EPIC har lansert et gratis verktøy \(open source\) for lokal validering av KI-modeller: AI validation software suite](#). Dette kan både brukes i virksomheter som benytter EPIC, og i virksomheter som benytter andre systemer.
 - GitHub hovedlenke: [GitHub - epic-open-source/seismometer: AI model evaluation with a focus on healthcare](#)

Fase 6: Drifte og forvalte et KI-system

Hensikten med denne fasen er å sikre at:

- Et kvalitetsstyringssystem er på plass for KI-systemet[143]
- KI-systemet overvåkes gjennom hele livsløpet med tanke på sikkerhet, stabilitet, ytelse og kvalitet på resultater

[143] I nødvendig omfang, tilpasset virksomhetens størrelse og risikoforhold.

Roller og ansvar

Virksomhetsledelsens ansvar er å etterleve blant annet krav om kvalitetssikring av helsetjenesten, jf. Forskrift om ledelse og kvalitetsforbedring i helse- og omsorgstjenesten [144]. KI-forordningen stiller krav til at leverandør og bruker har et kvalitetsstyringssystem på plass for høyrisiko-KI-systemer, jf. artikkel 17 [145]. KI-forordningen vil etter hvert tre i kraft, og virksomheten må forberede seg slik at de er i stand til å etterleve nytt regelverk. Eksisterende ledelsessystem bør derfor utvides til å omfatte kvalitetsstyring av KI-systemet. Ledelsessystem for KI (ISO/IEC 42001) er et verktøy som kan være nyttig for ledelsen [146].

Drift og bruk av KI-systemet kan kreve nye roller og ansvarsområder for å sikre at KI-systemet kan fungere trygt. Dersom virksomheten har anskaffet et høyrisiko-KI-system, vil det være flere områder som må ivaretas i henhold til KI-forordningen:

- KI-systemet skal, i likhet med medisinsk utstyr, brukes i henhold til leverandørens bruksanvisning [147]
- Menneskelig overblikk
- Transparens og informasjon til brukere av systemet (gjelder alle KI-systemer)
- Overvåkning av KI-systemet i drift, inkludert håndtere logg og avvik

Menneskelig overblikk (eng. *human oversight*) er beskrevet i KI-forordningens artikkel 14 under krav til høyrisiko-KI-systemer. Det stilles her krav til at KI-systemer skal utvikles slik at det kan gjøres menneskelige kontroller av systemene. Dette kan innebære å forstå hvilke muligheter og begrensninger KI-systemet har, slik at et menneske kan velge å overstyre resultatet, overvåke driften og oppdage og håndtere avvik, samt å forstå hvordan bruken kan gi automatiseringsskjevhet [148][149]. Det krever at et menneske med nødvendig kompetanse, støtte og autoritet har denne rollen.

- Ved Vestre Viken HF har de for eksempel navngitte KI-leger (KI-radiologer) som har særlig kompetanse på KI-systemet i bruk i klinikken.

I større virksomheter kan det være aktuelt å etablere en tverrfaglig gruppe med både operativt og strategisk ansvar for å sikre effektiv drift og forvaltning. Det kan være en nyttig organisering for å diskutere ytelse, forbedringsområder, og risikoer, og gi en klar struktur for hvordan beslutninger tas og kommuniseres. Et eksempel på en gruppe med personell med bred kompetanse samt tydelig ansvar:

- **Daglig drift:** Driftsansvarlig, systemansvarlig og IT-personell som spiller nøkkelroller for å sikre kontinuerlig og stabil drift, teknisk ansvar for ytelsen, infrastrukturen og sikkerheten. Vil koordinere tekniske oppdateringer, logging vedlikehold og integrasjon med andre systemer.

- **Domenekompetanse for relevans:** Klinisk ekspertise og/eller superbruker har stor betydning for å sikre at KI-systemet fungerer som forventet. Denne må være bindeledd mellom tekniske og kliniske team ved å oversette kliniske behov til tekniske krav. Vedkommende kan gi opplæring og støtte til andre brukere av KI-systemet og rapportere avvik eller begrensninger.
- **Samsvar og risiko:** Personvern, informasjonssikkerhet og juridisk støtte er mer sentrale i overordnet forvaltning enn i daglig drift. Skal sikre pasientenes tillit og overholdelse av lovverk og avtaler.
- **Informasjonsforvaltning:** Ansvar for orden i eget hus. God informasjonsforvaltning vil sørge for at data som brukes av KI-systemet, er korrekt, relevant og av høy kvalitet. Dette bidrar til å håndtere og unngå data bias og data drift.
- **Strategisk ledelse:** Systemeieren har det overordnede ansvar for å sikre at systemet gir verdi og følger virksomhetens mål og regelverk.

Det vises også til HUKI-matrisen i [fase 1](#) og [KI-faktaark 0](#).

[144] [Forskrift om ledelse og kvalitetsforbedring i helse- og omsorgstjenesten - Lovdata](#)

[145] KI-forordningens artikkel 17: [L_202401689EN.000101.fmx.xml \(eur-lex.europa.eu\)](#)

[146] [Ledelsessystem for kunstig intelligens – ISO/IEC 42001 \(standard.no\)](#)

[147] Dersom virksomheten for eksempel gjør betydelige endringer i KI-systemet eller ikke bruker det i henhold til bruksanvisningen, vil virksomheten gå over i rollen som produsent, og være forpliktet til å følge kravene som stilles til produsenter etter både forordningen for medisinsk utstyr og for KI-forordningen. Jmf MDR artikkel 16, og artikkel 25 i KI-forordningen.

[148] KI-forordningens artikkel 14: [L_202401689EN.000101.fmx.xml \(eur-lex.europa.eu\)](#)

[149] En tendens til automatisk å stole på eller stole for mye på resultatene fra et KI-system (automatiseringsskjevhet) som gir beslutningsstøtte.

Overvåkning

Overvåkning av KI-systemet krever en tverrfaglig tilnærming som inkluderer både tekniske, kliniske og sikkerhetsmessige aspekter. Virksomheten bør identifisere og iverksette tiltak for å gjøre virksomheten i stand til å oppdage og undersøke endringer i systemoppførsel [150].

KI-systemer er avhengige av robust teknisk infrastruktur som må overvåkes for å beskytte denne mot trusler som kan påvirke KI-systemets stabilitet og pålitelighet. Videre sikrer klinisk overvåkning at KI-systemet oppfyller sitt kliniske formål og gir verdi til helsepersonell og pasientene. Sikkerhetshendelser kan ha direkte kliniske konsekvenser, som for eksempel feil diagnostikk eller behandling, noe som setter pasientenes helse og sikkerhet i fare.

Cybersikkerhet, informasjonssikkerhet og personvern

KI-systemet kan representere ny eller mer kompleks teknologi sammenlignet med hva man tidligere er vant til i virksomheten. Siden kunnskapen om sikring av KI-systemer er ny, så må man forvente at kunnskapen

om sikkerhetsrisiko, samt god praksis for hvordan sikre disse systemene, vil endre seg over tid. Trusselbildet endrer seg stadig, og det kan derfor være behov for å ofte oppdatere risikovurderinger og gjennomføre tiltak. Koordineringsgruppen for medisinsk utstyr i EU (MDCG)[151] har publisert et veiledende dokument som kan bidra til å vurdere cybersikkerhet for medisinsk utstyr[152]. Andre ressurser på sikkerhet i KI-systemer er henvist til i [fase 4](#) [153][154][155].

Sikkerhetstesting må vurderes ved nye versjoner. Det må da ses på om det bør gjøres endringer i opplegget for sikkerhetstesting, for eksempel dersom en ny funksjonalitet i KI-systemet skal tas i bruk. Det er viktig å kjøre sikkerhetstester jevnlig, selv om det ikke er endringer i KI-systemet. Slik kan man oppdage nyoppstått risiko, for eksempel på grunn av at nye sårbarheter er blitt kjent, eller oppdage uautoriserte endringer av systemet. KI-systemer kan ha særegne sikkerhetsutfordringer ved at det foreligger spesielle behov knyttet til overvåking og logging av KI-systemet.

Risikoer kan håndteres ved å ha rutiner for jevnlige vurderinger og en plan for å holde seg oppdatert på sikkerhetsrisiko knyttet til den type KI-teknologi systemet benytter.

Overvåkning av ytelse

Et KI-system i drift må overvåkes som annet utstyr og behandling. KI-forordningen vil kreve overvåkning av høyrisiko KI-systemer, og også meldeplikt ved alvorlige hendelser [156].

Dersom virksomheten planlegger at KI-systemet skal fungere med høy grad av autonomi, vil dette kreve høy grad av overvåkning. KI-systemer og kontekstene de brukes i, er ofte komplekse, noe som gjør det vanskelig å oppdage og håndtere feil når de oppstår.

Modelldrift er spesielt viktig å følge med på og refererer til en forringelse av en KI-modells ytelse på grunn av endringer i data eller i sammenhengen mellom inndata og utdata. Dette innebærer blant annet å følge opp mulige endringer i inndataene, pasientpopulasjonen eller utdataene fra KI-systemet. Modelldrift kan ha en negativ innvirkning på modellens ytelse, noe som kan resultere i feilaktige beslutninger og dårlige prediksjoner [157]. Det er ulike årsaker til modelldrift, og KI-systemet må løpende overvåkes for disse.

For et klinisk beslutningsstøttesystem basert på KI kan overvåkning bestå av:

Tekniske tiltak for overvåkning av ytelse:

- Et KI-system er sensitivt til dataene det får inn og det bør derfor legges inn kontroller, for eksempel på at dataene er komplette og av riktig og forventet kvalitet og format.
- Overvåkning av teknisk stabilitet og robusthet.
- Validering av nye versjoner av modellen på et fast testdatasett. Oppdatere testdatasettet når det blir behov for det.
- Følge med på endringer i inndata og arbeidsprosesser og vurdere påvirkning på modellens ytelse.
- Implementere overvåkingsmekanismer for ytelse, som nøyaktighet og bias.

Kliniske tiltak for overvåkning av ytelse:

- Melde feil, avvik og begrensninger ved bruk av KI-systemet. Dette gjelder både feil i bruken av KI-systemet og resultatene det gir.
- Gjennomføre periodiske valideringer mot kliniske datasett for å avdekke potensielle feil som kan være vanskelig å oppdage i vanlig daglig bruk.

- Kvalitetskontroll av resultater ved å registrere avvik mellom KI-modellens prediksjoner og faktisk utfall i klinisk bruk. Evaluere om KI-systemets anbefalinger eller beslutninger stemmer overens med kliniske standarder og faglige retningslinjer.

[150] Normens dokumenter, inkludert [Sikkerhetsrisiko i systemer som benytter kunstig intelligens – hva vet vi, og hva kan vi gjøre? - ehelse](#)

[151] [Medical Device Coordination Group Working Groups - European Commission \(health.ec.europa.eu\)](#)

[152] [MDCG 2019-16 Rev.1 Guidance on Cybersecurity for medical devices \(health.ec.europa.eu\)](#)

[153] [Sikkerhetsrisiko i systemer som benytter kunstig intelligens – hva vet vi, og hva kan vi gjøre? - ehelse](#)

[154] [MITRE ATLAS™](#)

[155] [Multilayer Framework for Good Cybersecurity Practices for AI | ENISA](#)

[156] Krav til virksomheten i KI-forordningens artikkel 26: [L_202401689EN.000101.fmx.xml \(eur-lex.europa.eu\)](#)

[157] [What Is Model Drift? | IBM](#)

Logging og avvikshåndtering

For medisinsk utstyr er leverandør/produsent er forpliktet til å ta et ansvar for å overvåke KI-systemer som er CE-merket under MDR også etter at det er tatt i bruk (ettermarkedsovervåkning, eng. *post-market surveillance*). Dette inkluderer ansvar for å samle inn og evaluere erfaringer fra brukere av det medisinske utstyret. Formålet er å sikre fortsatt overholdelse av kravene, opprettholde sikkerhets- og ytelsesnivået, samt oppdage mulige behov for umiddelbare tiltak [158].

KI-forordningen vil stille lignende krav til høyrisiko-KI-systemer, men også brukerne skal bidra til å oppfylle dette kravet [159]. Virksomheten må påse at den automatiske loggen fra KI-systemet blir håndtert og tatt vare på [160]. Denne ettermarkedsovervåkningen er en systematisk dokumentasjon av ytelse og sikkerhetsproblemer i bruk som bidrar til å rette opp feil og kan forbedre systemene. Det kan være relevant med regelmessig rapportering til relevante myndigheter. Hvem som skal ha markedsovervåkningsmyndighet for KI-systemer i Norge er ikke endelig avklart [161].

Uønskede hendelser

Det er viktig å etablere rutiner for hvor og hvordan uønskede hendelser rapporteres når KI-systemet tas i bruk. Virksomheten bør sørge for at egne avvikssystemer er tilpasset for melding om avvik knyttet til KI-systemet, uavhengig om det er en klinisk eller en IT-hendelse, for å fange opp KI-relaterte uønskede hendelser.

Uønskede hendelser i helse- og omsorgstjenesten kan meldes på [melde.no](#). Denne kan også pasienter benytte til å melde uønskede hendelser, eller de kan henvende seg direkte på [HelseNorge.no](#).

Avvik som karakteriseres som alvorlig hendelse *skal* rapporteres. For KI-systemet er virksomheten pliktig til å melde om alvorlige hendelser som utgjør risiko til leverandøren umiddelbart, jf artikkel 26 (5) i KI-forordningen.

The Organisation for Economic Co-operation and Development (OECD) har etablert en global overvåkings-database for KI-hendelser som kan meldes her: [AI Incidents Monitor](#)

Kvalitetsforbedring

Ved å ha et kvalitetsstyringssystem som fungerer effektivt kan virksomheten overvåke KI-systemet slik at forverring i modellens ytelse, eller andre uønskede hendelser kan identifiseres i tide. Virksomheten har da mulighet til å håndtere avvik, lære av dem og iverksette nødvendige tiltak. Tiltak når det gjelder KI-systemet kan for eksempel dreie seg om å få leverandøren til å oppdatere KI-systemet, fase ut eksisterende eller anskaffe et nytt KI-system. Kontinuerlig kvalitetsforbedring av tjenesten kan ellers være å justere arbeidsprosesser, opplæring, informasjon og andre rutiner.

[158] Blant annet artiklene 83 i forordningen om medisinsk utstyr (MDR): <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0745> (eur-lex.europa.eu) og 78 i forordningen om in vitro-diagnostisk medisinsk utstyr: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R0746>

[159] KI-forordningens artikkel 72 om ettermarkedsovervåking: [L_202401689EN.000101.fmx.xml](#) (eur-lex.europa.eu)

[160] KI-forordningens artikkel 26 om krav til virksomheten: [L_202401689EN.000101.fmx.xml](#) (eur-lex.europa.eu)

[161] Direktoratet for forvaltning og økonomistyring (DFØ) anbefaler at Nkom skal være nasjonalt tilsynsorgan, markedsovervåkningsmyndighet og fungere som kontaktpunkt opp mot EU når det gjelder KI-forordningen. [Hvorfor skal Nkom få ansvaret for KI? \(cw.no\)](#)

Gvinstrealisering og erfaringsdeling

Gjennom anskaffelsen, implementering og testing vil man få bedre underlag for kost-nytteanalysen. Denne vil så ligge til grunn for gvinstrealiseringsplanen, med konkrete og målbare gevinster.

Ofte medfører innføring av et KI-system en ny måte å jobbe på. Virksomheten bør følge opp at KI-systemet blir brukt som tiltenkt og at nytten ved nye arbeidsprosesser blir tatt ut.

- Et eksempel på en ny arbeidsprosess: Pasienter som får avdekket "ikke brudd" ved en KI-analyse i Vestre Viken, blir sendt hjem i stedet for å måtte sitte og vente på legevakten til en radiolog har tid til å se på bildene. Radiologene har derfor ingen samtale med pasientene hvor det ikke avdekkes brudd, slik som de tidligere har hatt.

I forvaltningsfasen gjøres det målinger og utarbeides statistikk og rapporter av blant annet kvalitet og effekter av KI-systemet. Dette sammenstilles for å vise fremdriften i gvinstrealiseringen. Dersom virksomheten ikke får realisert de planlagte gevinstene, så bør den analysere årsakene til det, identifisere mulige tiltak og realisere de tiltakene som vurderes som mest hensiktsmessige og effektive. Mye av arbeidet med gvinstrealisering er ikke annerledes for KI-systemer enn for andre systemer [162].

Virksomhetens erfaringer med å innføre KI-systemet kan være nyttig kunnskap for andre, også på tvers av landegrenser. Resultater fra innføringen og eventuell følgeforskning kan være nyttig for andre og bør publiseres og kommuniseres. Å dele kvantitative erfaringstall om gevinster kan være god informasjon til andre som skal gjøre kost/nytte-vurderinger av KI-systemer. En effektiv kommunikasjonsmåte for kvalitative gevinster kan være å publisere intervjuer med helsepersonell som bruker KI-systemet, eller erfaringer fra innbyggere, pasienter og/eller tjenestemottakere som har fått raskere eller bedre helsehjelp. Likeledes bør negative læringspunkter deles, for å unngå at andre gjør samme investering uten å få den positive endringen de ønsker. Virksomheter som innhenter for eksempel pasientrapporterte utfallsmål og pasientrapporterte erfaringer kan bidra med nyttig kunnskap [163].

Siden kvalitetssikring av systemer som benytter KI er et tema som flere virksomheter i helse- og omsorgssektoren kommer til å møte framover, er dette et område der erfaringsutveksling har stort potensiale til å styrke arbeidet.

[162] Direktoratet for forvaltning og økonomistyring har laget en generell [Veileder til gevinstrealisering \(dfo.no\)](#) som kan være nyttig.

[163] PROM; Patient-Reported Outcome Measure og PREM; Patient-Reported Experience Measure

Relevante kilder til fase 6

- KI-forordningen: [L_202401689EN.000101.fmx.xml \(eur-lex.europa.eu\)](#)
- Forskrift for ledelse og kvalitetsforbedring med veileder, gir et godt grunnlag for vurdering av resultater for KI-systemer på samme måte som for andre helsetjenester: [Veileder: Ledelse og kvalitetsforbedring i helse- og omsorgstjenesten](#)
- Normen (Norm for informasjonssikkerhet og personvern i helse- og omsorgssektoren) beskriver organisatoriske og tekniske tiltak som anses egnet for å oppnå tilfredsstillende informasjonssikkerhet og personvern i sektoren: <https://www.ehelse.no/normen/normen-for-informasjonssikkerhet-og-personvern-i-helse-og-omsorgs>
- Det europeiske byrået for informasjonssikkerhet (The European Union Agency for Cybersecurity (ENISA)) foreslår et rammeverk for gode praksiser for informasjonssikkerhet (cybersecurity) for KI: <https://www.enisa.europa.eu/publications/multilayer-framework-for-good-cybersecurity-practices-for->
- Normen peker på følgende trusler med kunstig intelligens: <https://www.ehelse.no/normen/aktuelt-om-normen/sikkerhetsrisiko-i-systemer-som-benytt-kunstig-i>
- ISO/IEC 42001:2024: Ledelsessystem for kunstig intelligens: <https://standard.no/fagomrader/kunstig-intelligens/ledelsessystem-for-kunstig-intelligens--isoiec-42001-2024>
- ISO/IEC 22989:2023: Begreper og terminologi for kunstig intelligens (ISO/IEC 22989:2022): [Standard Norge | standard.no. NS-EN ISO/IEC 22989:2023](#)

KI-faktaark

Roller i anskaffelsen av et KI-system (KI-faktaark 0)

Om faktaarket

Formålet med faktaarket er å beskrive de ulike rollene som kan involveres og hvilket ansvar de har i en anskaffelsesprosess av KI-systemer i helse- og omsorgstjenesten.

Roller i anskaffelsen

Rapport om kvalitetssikring av KI-systemer beskriver et eksempel på en ansvarsmatrise, også kalt HUKI-matrise.

Følgende roller kan være de mest aktuelle i en ansvarsmatrise for anskaffelse av et KI-system.

Prosjektledelsen med prosjektgruppe for anskaffelsen består ofte av ressurser med kompetanse innen ulike områder, og kan dermed dekke mange av kompetansene for anskaffelsen.

Domenekompetanse refererer til ekspertisen som er spesifikk for et bestemt fagfelt eller bruksområde. Domenekompetanse i medisin innebærer forståelse av medisinske prosedyrer, sykdommer, behandlingsmetoder og helsevesenets dynamikk, som er viktig for å definere suksesskriterier og terskelverdier for et KI-system i bruk. Domenekompetanse er avgjørende i prosessen, ettersom det bidrar til mer effektive og informerte beslutninger.

Virksomhetsarkitektens rolle er å utvikle, vedlikeholde og forvalte den overordnede arkitekturen til en organisasjon og ha evnen til å se den store sammenhengen i hvordan ulike systemer og prosesser interagerer og bidrar til organisasjonens overordnede mål.

Fagspesifikk kompetanse er spesiell kompetanse innen ulike fagspesifikke områder som ofte går på tvers av domenekompetanse. Et typisk eksempel er strålevern. Medisinske fysikere har god kompetanse i klinisk strålevern av pasienter (og personell) og kan bidra til identifisering av nytteverdi og risikواسpekter forbundet med bruken av KI-systemet knyttet opp mot lovpålagte krav i strålevernregelverket.

Etisk kompetanse er ikke begrenset til en bestemt yrkesgruppe. Det refererer til evnen å forstå, reflektere over, og anvende etiske prinsipper og verdier i beslutningstaking.

Innbygger/pasient/tjenestemottaker er person som KI-systemet brukes på. Det innebærer blant annet at personen har persondata som blir brukt for å få et resultat fra et KI-system. Interessene til disse personene kan være representert via et sykehus' brukerutvalg eller en pasient-/brukerorganisasjon. Inkludering av disse kan bidra til å bygge tillit til teknologien, og sikrer deres perspektiv på etisk og trygg bruk av KI.

Juridisk kompetanse refererer til kunnskap og ferdigheter innen lovgivning, rettspraksis og juridiske prosedyrer. Det er mange ulike vurderinger som må foretas, blant annet forsvarlighet, pasientrettigheter, personvern og informasjonssikkerhet kontraktsrett og immaterielle rettigheter.

Kompetanse på **personvern og informasjonssikkerhet** innebærer å sikre at organisasjonens informasjonsteknologisystemer er sikre og at data behandles på en måte som er i samsvar med personvernlovgivningen. Dette inkluderer å identifisere og analysere potensielle risikoer forbundet med KI-systemet, inkludert datalekkasjer og cybertrusler. Kan bidra til å utvikle risikoreduserende strategier og sikkerhetstiltak. Virksomheten som anskaffer, vil som hovedregel være dataansvarlig og må ivareta ansvaret som følger med dette.

Analysekompetanse kan vurdere ytelsen til KI-systemet og designe robuste valideringsstudier for å teste ytelse. Dette inkluderer å bestemme passende datasett, velge relevante statistiske metoder, og gi anbefalinger for forbedring av modellens ytelse, for eksempel gjennom forslag til alternativ modellering eller dataforbedringer.

Innkjøpskompetanse sørger for at anskaffelsesprosessen følger lov og forskrift.

Løsningsarkitekt utformer og leder implementeringen av en spesifikk teknologisk løsning innenfor en virksomhet.

IKT drift og forvaltning overvåker og vedlikeholder KI-systemet når det er tatt i bruk.

En **KI-superbruker** kan være samme person som har domenekompetanse. Dennes rolle er å vite vet hvordan KI-systemet fungerer og sikrer at systemet blir brukt korrekt og effektivt. Kan for eksempel være ansvarlig for opplæring i virksomheten.

Systemansvarlig er den i virksomheten som har det tekniske og faglige ansvaret for KI-systemet.

Systemeier (virksomhetsledelse) er også risikoeieren, og er den i virksomheten som har det overordnede ansvaret for KI-systemet, dets funksjon og bruk inkludert risiko ved bruk av systemet.

Testleder forstår de tekniske og forretningsmessige kravene og kan utvikle en teststrategi som dekker blant annet krav til funksjonalitet, ytelse, sikkerhet, brukervennlighet og KI spesifikke forhold. Bidrar til å sikre løsningens kvalitet ved å identifisere og dokumentere feil.

KI-begreper (KI-faktaark 1)

Om faktaarket

Formålet med dette faktaarket er å forklare sentrale begreper knyttet til kunstig intelligens i helse- og omsorgstjenesten. Oversikten oppdateres løpende.

Relevante lenker

- Teknologirådet: [Ordliste for kunstig intelligens](#)

Beskrivelse av KI-begreper

Automatiseringsskjevhet: En tendens¹ til automatisk å stole på eller stole for mye på resultatene fra et KI-system som gir beslutningsstøtte.

Bias oppstår når datasettene som brukes for å trene kunstig intelligente systemer bærer preg av historiske skjevheter, er² ufullstendige eller inneholder feil. Det kan resultere i systematiske feil i modellens prediksjoner.

Finjustering (på engelsk fine-tuning) er en teknikk hvor en allerede forhåndstrent språkmodell trenes videre på et smalere eller oppgavespesifikt datasett for å løse en bestemt oppgave, eller for å tilpasse seg et miljø med spesifikt fagspråk. Dette kan for eksempel innebære å trene modellen på domsavsigelser for å oppsummere rettspraksis og gi juridiske råd, eller på helsedata for å avdekke mønstre i helsejournaler eller stille diagnoser.³

Generativ kunstig intelligens: Maskinlæringsmodeller som kan generere unikt innhold basert på informasjonen de har blitt trent på. Dette innholdet kan være tekst, bilder, lyd og video.⁴ Eksempler er ChatGPT og Bard som genererer tekst og Dall-E og Midjourney som genererer bilder.

KI for allmenn bruk (engelsk: *General Purpose AI - GPAI*) har ikke et tiltenkt formål, og faller således utenfor loven om medisinsk utstyr, men er regulert i EU-forordningen for KI. Reguleringen legger plikter på⁵ tilbydere av GPAI, og gir brukere og sluttbrukere rettigheter til en hel del informasjon om slike KI-modeller.

KI-modell: En algoritme er en oppskrift som forteller hvordan noe gjøres, og består enkelt forklart av et sett instruksjoner som forvandler inndata til utdata. En maskinlæringsalgoritme er oppskriften systemet bruker for å bygge en KI-modell av virkeligheten, basert på dataene som blir matet inn (treningsdataene). Denne modellen kan så brukes for å ta nye beslutninger om nye data som kommer inn.⁶

KI-plattform brukes ofte om en plattform som kan håndtere ulike KI-systemer til bruk i en virksomhet. Et eksempel er KI-plattformen anskaffet i Vestre Viken HF. På denne KI-plattformen håndterer virksomheten ulike KI-systemer, som for eksempel *Boneview*, som er et KI-system som bruker en KI-modell for å tolke røntgenbilder tatt på skadelegevakten for å avgjøre om det er *benbrudd* eller *ikke benbrudd*.

KI-system: Det finnes mange definisjoner av kunstig intelligens. Vi har valgt å lene oss på definisjonen til den europeiske KI-forordningen (oversatt fra engelsk): "Et KI-system er et maskinbasert system som er designet for å fungere med varierende nivåer av autonomi og som kan vise tilpasningsevne etter implementering. For eksplisitte eller implisitte mål, utleder systemet, fra inndata det mottar, hvordan det skal generere utdata som prediksjoner, innhold, anbefalinger eller beslutninger som kan påvirke fysiske eller virtuelle miljøer".⁷

Kvalitetssikring: Kvalitetssikring er planlagte og systematiske aktiviteter som gjøres for å oppnå at et produkt eller en tjeneste vil oppfylle kravene til kvalitet. Kvalitetssikring er en del av en organisasjons kvalitetsstyring.⁸

Løpende maskinlæring er en teknikk som ligner på menneskers måte å lære på: Hver gang vi opplever, ser eller lærer noe nytt, tilpasser vi modellene vi bruker til å tolke verden rundt oss. Det finnes flere teknikker for å sørge for at maskinlæringsmodellene kan lære av stadig nye data. Én måte er å inkludere de nye dataene i settet med treningsdata, kjøre hele læringsalgoritmen på nytt, og oppdatere modellen som anvendes. Dette er ikke et problem med små datasett, men med dagens datamengder kan det kreve stor regnekapasitet, og koste tid, penger og energi. Andre metoder handler om å bruke de nye dataene til å

justere på den modellen man allerede har. Dette er mye enklere, men har i noen tilfeller resultert i at de nye dataene får uforholdsmessig mye vekt, slik at modellen «glemmer» ting den egentlig hadde lært fra før.⁹

Modelldrift refererer til en forringelse av en KI-modells ytelse på grunn av endringer i data eller i sammenhengen mellom inndata og utdata. Modelldrift kan ha en negativ innvirkning på modellens ytelse, noe som kan resultere i feilaktige beslutninger og dårlige prediksjoner.¹⁰

Validering: Validering brukes om å bekrefte at KI-modellen yter som den er ment, for et bestemt tiltenkt formål.¹¹

[1] Oversettelse fra KI-forordningen, artikkel 14 (3a):

https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689#d1e38-124-1.

[2] Teknologirådet: <https://teknologiradet.no/ordliste-for-kunstig-intelligens/>

[3] Teknologirådet: <https://teknologiradet.no/ordliste-for-kunstig-intelligens/>

[4] Teknologirådet: <https://teknologiradet.no/ordliste-for-kunstig-intelligens/>

[5] <https://www.digi.no/artikler/debatt-alt-som-glimrer-er-ikke-ki-sjekk-om-ai-act-gjelder-deg/545862>

[6] Teknologirådet, side 16:

<https://media.wpd.digital/teknologiradet/uploads/2022/02/Kunstig-intelligens-i-klinikken.pdf>

[7] Artikkel 3.1 i KI-forordningen:

https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689#d1e38-124-1

[8] Store Norske Leksikon: <https://snl.no/kvalitetssikring>

[9] Teknologirådet, side 64:

<https://media.wpd.digital/teknologiradet/uploads/2022/02/Kunstig-intelligens-i-klinikken.pdf>

[10] [What Is Model Drift? | IBM](#)

[11] I standarden ISO 9000:2015 "Quality Management System – Fundamentals and vocabulary" defineres "validation" slik: "Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled:

<https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en>

Validering for bruk av KI i helse- og omsorgstjenesten (KI-faktaark 2)

Om faktaarket

Formålet med faktaarket er å forklare ulike måter å validere KI-systemer på i helse- og omsorgstjenesten.

Det er viktig å skille mellom produsenters og helsetjenestens ansvar ved innføring av medisinsk utstyr i klinikk. Produsenter vil ha ansvar for validering av ytelse i forbindelse med CE-merking, mens helsetjenesten skal ivareta det helsefaglige ansvaret, herunder validering i henhold til produsentens anvisninger.

For medisinsk utstyr er kravene til produsentens validering beskrevet i Guidance on Clinical Evaluation (MDR) / Performance Evaluation (IVDR) of Medical Device Software.¹ Produsenten skal også beskrive hvordan utstyret skal valideres ved ibruktaking i klinikk (der det sikres at utstyret er korrekt installert, fungerer korrekt og gir de svar som forventes). Kravene til helsetjenesten er omtalt i Forskrift om håndtering av medisinsk utstyr (Håndteringsforskriften).²

Informasjonen som videre følger i dette faktaarket er ikke spesifikt ment som veiledning om regelverket for medisinsk utstyr. Faktaarket forklarer ikke detaljerte krav til produsenter og virksomhet. Det vises til rapporten over, som beskriver spesifikke detaljer om produsentens ansvar ved klinisk evaluering av programvare som medisinsk utstyr.

Faktaarket omhandler følgende typer validering.

- Intern og ekstern validering
- Teknisk og klinisk validering
- Retrospektiv og prospektiv validering

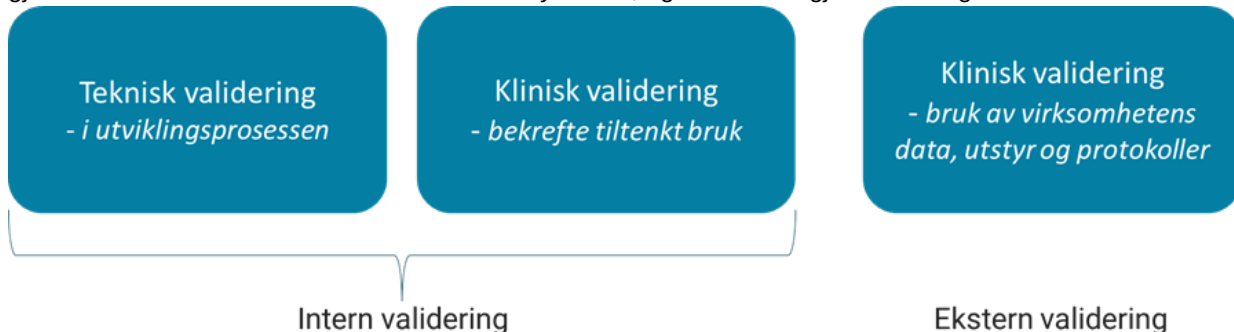
Valideringstypene beskrives først kort, deretter følger mer utdypende forklaring av dem.

Validering for bruk av KI i helse- og omsorgstjenesten

Validering av KI-modeller refererer til prosessen med å evaluere ytelsen og nøyaktigheten til en modell.

Validering betyr å bekrefte at KI-modellen yter som den er ment, for et bestemt tiltenkt formål.³ Dersom KI-modellen har tiltenkt formål å diagnostisere brudd på et røntgenbilde av en ben, er det akkurat dette KI-modellen skal valideres for. Ved validering av KI-modeller er det derfor viktig at det "tiltenkte formålet" er klart definert.

Det skilles mellom den valideringen som gjøres i utviklingen av et KI-system og den valideringen som gjøres av virksomheten som skal ta i bruk KI-systemet, og hvem som gjør valideringen.



Figur 1: Illustrasjon av intern og ekstern validering, og teknisk og klinisk validering

Intern og ekstern validering

Intern og ekstern validering viser til hvem som har gjort valideringen. Dersom det er produsenten i utviklingsprosessen kalles det intern validering. For eksempel at KI-modellens ytelse måles på en tilfeldig del av det opprinnelige datasettet, som er trukket ut på forhånd og holdt unna modelltreningen.

Hvis det er aktører som er eksterne og uavhengige av produsenten kalles det ekstern validering. Ytelsen måles da på et uavhengig datasett, som utvikleren ikke har hatt tilgang til. Ekstern validering kan utføres i en helsevirksomhet, som for eksempel vurderer å kjøpe KI-systemet.

Teknisk og klinisk validering

Teknisk validering gjøres i utviklingsprosessen og klinisk validering gjøres når modellen er ferdigutviklet.

- *Teknisk validering* gjøres ved hjelp av et datasett som ikke ble brukt under treningen av modellen. Hensikten med validering er å sikre at modellen generaliserer godt til ukjente data og ikke bare har lært seg mønstrene i treningsdatasettet.
- *Klinisk validering* refererer til prosessen med å evaluere en KI-modell i en klinisk setting, under reelle forhold. Dette innebærer å teste modellen med uavhengige kliniske data og pasienter for å bekrefte at den gir pålitelige og nyttige resultater som kan forbedre medisinsk behandling og beslutningstaking. Formålet med klinisk validering er å sikre at KI-modellen ikke bare er teoretisk solid, men også praktisk anvendbar og fordelaktig i virkelige medisinske situasjoner, slik at den kan bidra til forbedret helsevesen og pasientomsorg.⁴

Retrospektiv og prospektiv validering

Validering kan utføres enten retrospektivt og prospektivt.

I prospektive studier følges deltakerne fremover i tid. I retrospektive studier ser man tilbake.⁵

- *Retrospektiv validering* er å validere en KI-modell med et datasett som er innsamlet i forkant, og muligens til et annet formål. Slik validering kan være aktuelt i en anskaffelsesprosess der man i forkant av kontraktinngåelsen får testet KI-modellen på virksomhetens egne, allerede innsamlede data (for eksempel bilder i et arkiv).
- *Prospektiv validering* er å teste en KI-modell idet de oppstår i en reell brukssituasjon. Valideringen kan gjøres på virksomhetens egne data, for eksempel bilder som blir tatt under undersøkelse.

Ved prospektiv validering har man forhåndsbestemt hvilke data som skal brukes og hvordan man skal teste. Forskjellen fra retrospektiv validering ligger i at man har laget en protokoll for datainnsamlingen på forhånd, som dokumenterer hvilke data som skal samles inn, hvor lenge man skal samle data og eksakt hvilke tester man planlegger å gjøre. Når dataene foreligger, er prosessen med å måle ytelsen og dokumentere resultatene lik som for retrospektive studier.

I en retrospektiv validering analyserer man eksisterende data fra fortiden, og det trekkes konklusjoner basert på denne historiske informasjonen. Fordelen med retrospektiv validering er at det ofte er raskere og mindre kostbare enn prospektiv validering, siden dataene allerede eksisterer. Ulempen er at retrospektiv validering kan være begrenset av nøyaktigheten og fullstendigheten av de eksisterende dataene, og det er ofte vanskelig å kontrollere for forstyrrende variabler eller å etablere årsakssammenhenger.

For en retrospektiv studie kan det også være usikkerhet om datasettet er plukket ut for å passe spesielt godt med modellen (såkalt «cherry-picking») eller om eksklusjons- og inklusjonskriterier er valgt for å gi bedre resultater. Man kan også være bekymret for publikasjonsskjevhet, ved at studier med gode resultater har høyere sannsynlighet for å bli publisert mens dårlige resultater legges i en skuff.

Prospektiv validering er ofte mer nøyaktige, men er også mer tidkrevende og kostbare å gjennomføre sammenlignet med retrospektiv validering. Seriøse prospektive studier blir forhåndsregistrert med fast protokoll i internasjonale databaser, som reduserer de nevnte mistankene man kan ha til retrospektive studier.

Mer om teknisk validering

De viktigste elementene i teknisk validering inkluderer:

- *Datasplitting*: Datasettet deles vanligvis opp i tre sett: treningssett, valideringssett og testsett.
- *Treningssett* brukes til å trene modellen.
- *Valideringssett* brukes under treningen til å justere hyperparametere, gjøre modellutvalg og å evaluere modellen.
- *Testsett* brukes til å evaluere den endelige modellen etter treningen for å vurdere dens generaliseringsevne.
- *Justering av hyperparametre*⁴ gjøres under valideringsprosessen av modellen for å finne den beste kombinasjonen av hyperparametre som gir høyest ytelse på valideringssettet.
- *Måle nøyaktighet*: Under validering beregnes forskjellige ytelsesmetriker, som nøyaktighet, presisjon, tilbakekalling, F1-score, etc., avhengig av hvilken type problem (klassifisering, regresjon, etc.) modellen er designet for å løse.
- *Kryssvalidering*: er en teknikk for å evaluere modellens ytelse og spesielt om den er overtilpasset til treningsdataene. Teknikken innebærer å dele dataene i flere deler (typisk 5 eller 10). Modellen trenes på alle delene minus én, og valideres på den gjenværende delen. Denne prosessen gjentas slik at hver del brukes som valideringssett én gang. Ytelsesmålene til hver iterasjon sammenstilles og gir et samlet mål på ytelsen.
- *Over- og undertilpasning* (engelsk: *overfitting* og *underfitting*): Validering hjelper til med å identifisere en eventuell overtilpasning (modellen er for godt tilpasset treningsdataene) og undertilpasning (modellen er ikke godt nok tilpasset dataene).

Mer om klinisk validering

Klinisk validering av en modell med representative data er viktig for å vurdere om modellen vil fungere i bruk og på den populasjonen den er tenkt brukt på etter innkjøp. For eksempel, kan en modell for diagnostisering av hjerteinfarkt som kun baserer seg på data fra menn kanskje fungere dårlig på kvinner. Validering kan gjøres i flere omganger.

De viktigste elementene i klinisk validering inkluderer:

- Evaluering i klinisk miljø: Modellen testes på data samlet inn fra ekte pasienter i kliniske omgivelser for å se hvordan den presterer utenfor laboratoriet eller utviklingsmiljøet.
- Sammenligning med kliniske standarder: Modellens resultater sammenlignes med de etablerte standardene for medisinsk behandling for å vurdere om den gir sammenlignbare eller bedre resultater.
- Sensitivitet og spesifisitet: Måling av modellens evne til å korrekt identifisere positive tilfeller (sensitivitet) og korrekt utelukke negative tilfeller (spesifisitet).

- Prospektive studier: Studier hvor modellen brukes fremover i tid med pasienter som ennå ikke er diagnostisert, for å se hvordan modellen presterer i sanntid.
- Retrospektive studier: Studier hvor modellen testes på historiske data for å se hvordan den ville ha prestert på tidligere pasienter.
- Inter-rater reliability: Sammenligning av modellens resultater med vurderinger fra menneskelige eksperter for å sikre konsistens og pålitelighet.
- Klinisk nytte: Vurdering av om modellens bruk fører til bedre kliniske utfall, som forbedret diagnose, bedre behandlingsplaner, og økt effektivitet i pasientbehandling.
- Regulatorisk godkjenning: Klinisk validering kan også være en del av kravene for å oppnå godkjenning fra regulatoriske myndigheter, som FDA (Food and Drug Administration) i USA eller EMA (European Medicines Agency) i Europa.

Mer om intern validering

Enhver maskinlæringsmodell blir trent på ett (stort) datasett, som ofte deles opp i ulike sett før treningen: et treningssett, et valideringssett og et testsett.⁷ Treningssettet brukes til å trene KI-modellen. De fleste maskinlæringsmodeller har flere hyperparametere, som påvirker ytelsen. Dette kan for eksempel være antall skjulte lag i en nevraltmodell, dybden av et beslutningsstre, læringsraten man bruker under trening og hvor lenge modellen skal trene.

Under treningen av KI-modellen blir forskjellige verdier på hyperparameterne prøvd ut, og det gjøres løpende tester på valideringssettet. Den versjonen av modellen som gjør det best på valideringssettet velges etter hvert ut. Ytelsen på valideringssettet er imidlertid ikke en gyldig test av denne modellen fordi den ble plukket ut for å gjøre det bra på nettopp dette. Til slutt testes modellen derfor på testsettet, som består av usette data fra samme opprinnelige datasett. Dette gir et mer reelt estimat på ytelsen til modellen.

Mer om ekstern validering

I ekstern validering måles modellens ytelse på et datasett som produsenten (utvikleren) ikke har hatt tilgang til og som ikke har vært brukt i den interne valideringen. Den eksterne valideringen gjennomføres av en uavhengig person eller virksomhet, for eksempel en uavhengig forsker. Ideelt sett bør datasettet til ekstern validering inneholde datapunkter fra underrepresenterte demografiske grupper, slik at det er balansert og representativt. For å vurdere dette, må man etterspørre beskrivende detaljer om datasettet som er brukt i den eksterne valideringen.

Som beskrevet i «A Clinician's Guide to Artificial Intelligence»⁸, kan ekstern validering inkludere bruk av et datasett som er:

- Uavhengig av det originale datasettet, men lignende befolkning og setting.
- Uavhengig, men forskjellig enten i populasjon (for eksempel etnisitet, sosioøkonomisk status) eller setting (for eksempel screening, behandling i primær- eller spesialisthelsetjenesten, geografisk plassering).
- Representativ for de samme eller nye populasjoner over tid - dette vil bidra til å vurdere degradering av modellens ytelse ettersom populasjonen utvikler seg.
- Forskjellig av tekniske årsaker (for eksempel på grunn av bilder tatt av forskjellige apparater).

Skillet mellom intern og ekstern validering er ikke nødvendigvis helt skarpt. Man kan for eksempel tenke seg at en forskergruppe får tilgang til et egnet eksternt datasett og tester sin egen modell på dette fordi de ikke ønsker å dele modellen med andre forskere. Da mister man den garantien som ligger i å bruke en avhengig ekstern aktør, men ellers vil testen ha samme verdi som en ekstern validering.

Kliniske studier

Pragmatiske randomiserte kontrollerte forsøk (RCTer)

Intern validering, ekstern validering og prospektiv validering er observasjonelle, i den forstand at modellen ikke påvirker dataene. Dette kan medføre skjevheter sammenliknet med virkelig bruk, som skyldes årsakssammenhenger. Hvis bruken av en KI-modell ville ført til et annet hendelsesforløp, kan ikke dette fanges opp i en observasjonell studie. Et eksempel er mammografi-screening for brystkreft, hvor observasjonelle studier av sensitivitet og spesifisitet aldri blir helt balanserte, selv om krefttilfellene bekreftes med biopsi. Dette skyldes at enkelte krefttilfeller aldri vil manifestere seg som sykdom fordi pasienten dør av noe annet før kreften gir symptomer. Disse kan fremstå som falske negative hvis de oppdages av screening programmet, men ikke KI-modellen, eller som falsk positiv hvis KI-modellen oppdager dem, men ikke screening-programmet.

Hvis man ønsker å måle en slik KI-modells ytelse presist, må den derfor kunne påvirke hvilke tilfeller som plukkes ut til biopsi. I slike tilfeller kan man vurdere å gjennomføre pragmatiske kontrollerte forsøk («pragmatisk RCT»). Man bruker ordet pragmatisk fordi man tester ut en intervensjon som del av en kompleks virkelighet, framfor en ren «laboratorie-setting» hvor man har kontroll på alle faktorer. En ulempe med pragmatiske RCTer er at man ikke kan si med sikkerhet hvorfor og hvordan intervensjonen virker, men til gjengjeld fanger man opp indirekte effekter, som f.eks. at svarene fra den ene gruppen er mindre nøyaktige, eller kanskje bruker mer tid eller at antallet det tas biopsi av går opp. Pragmatiske RCTer gir gode estimater på effekten av en intervensjon i den aktuelle organisasjonen, men generalisering til andre enheter kan være problematisk, spesielt hvis arbeidsprosessene der er organisert annerledes.

[1] MDCG 2020-1 Guidance on Clinical Evaluation (MDR) / Performance Evaluation (IVDR) of Medical Device Software
https://health.ec.europa.eu/document/download/19d9e24f-2808-4e00-bfeb-75892047407d_en?filename=m

[2] <https://lovdata.no/dokument/SF/forskrift/2013-11-29-1373>

[3] I standarden ISO 9000:2015 "Quality Management System – Fundamentals and vocabulary" defineres "validation" slik: "Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled".
<https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en>

[4] Charlotte Haug, NEJM AI, Mission and vision: AI-tools must meet the same bar for clinical evidence that is expected from other clinical interventions.

[5] [Prospektiv og retrospektiv | Tidsskrift for Den norske legeforening \(tidsskriftet.no\)](https://tidsskriftet.no)

[6] Hyperparametere er konfigurasjonsinnstillinger som brukes til å strukturere maskinlæringsmodeller. Disse parameterne læres ikke fra data, men settes før treningsprosessen og forblir konstante under trening. De påvirker modellens ytelse og kapasitet til å lære fra data, og velges ofte gjennom hyperparameteroptimalisering eller -tuning.

[7] Begrepene «treningssett», «valideringssett» og «testsett» brukes ikke konsekvent i litteraturen.

[8] [A Clinician's Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies - PMC \(nih.gov\)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7346877/) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7346877/>

Ytelsesmål for maskinlæringsmodeller (KI-faktaark 3)

Om faktaarket

Formålet med faktaarket er å gi en kort innføring i ytelsesmål for noen hovedtyper av maskinlæringsmodeller.

Ytelsesmål for maskinlæringsmodeller

Faktaarket beskriver ytelsesmål for følgende hovedtyper maskinlæringsmodeller:

- Binære klassifikasjonsmodeller
- Klassifikasjon med flere klasser
- Regresjonsmodeller
- Segmentering og objekt-deteksjon i bilder
- Generative modeller

Binære klassifikasjonsmodeller

Den enkleste modellkategorien å evaluere er binære klassifikasjonsmodeller, hvor modellen velger mellom to kategorier, som "syk" eller "frisk".

Sensitivitet og spesifisitet

Sensitivitet og spesifisitet brukes for å kvantifisere ytelse av en modell eller en test. Under er dette eksemplifisert ved en diagnostisk test for en sykdom:

- *Sensitivitet* er modellens evne til å korrekt identifisere de med sykdommen (sanne positive) blant de som faktisk er syke. Med andre ord, det er sannsynligheten for at en syk person vil ha en positiv test.
- *Spesifisitet* er modellens evne til å korrekt identifisere de uten sykdommen (sanne negative) blant de som faktisk er friske. Med andre ord, det er sannsynligheten for at en frisk person vil ha en negativ test.

Positiv og negativ prediktiv verdi

Positiv prediktiv verdi (PPV) er sannsynligheten for at en person som fikk positivt testresultat virkelig er syk. Tilsvarende for negativ prediktiv verdi (NPV). Disse måler altså hvor ofte modellen har rett når den kommer med en positiv eller negativ prediksjon.

Disse prediktive verdiene påvirkes sterkt av sykdommens forekomst (prevalens) i utvalget som testes. En test med høy sensitivitet kan ha lav positiv prediktiv verdi hvis den gjøres hos grupper med lav sykdomsprevalens.¹ Hvis man skal tallfeste ytelsen til en maskinlæringsmodell i seg selv, er det derfor bedre å bruke sensitivitet og spesifisitet fordi PPV og NPV er avhengig av prevalensen i befolkningen. I

rabies-eksempelen er det for eksempel ikke rimelig å si at modellen er dårlig selv om PPV er nær 0. Hvis man skal bruke en maskinlæringsmodell i en klinisk situasjon for å ta beslutninger om f.eks. behandling, er derimot de prediktive verdiene sentrale, og sensitivitet og spesifisitet mindre relevant.

Krysstabell

Ytelsen til en klassifikasjonsmodell på et gitt datasett oppsummeres ofte med en matrise, også kalt krysstabell, se Figur 1. Krysstabellen viser antall faktiske tilfeller i kolonnene og antall predikerte tilfeller i radene. Relevante målinger kan deretter beregnes ut fra oppføring i tabellen. Krysstabeller er en nyttig måte å beskrive modellytelse på fordi de er intuitive å lese, og dekker alle mulige kombinasjoner av korrekte og feilaktige prediksjoner.

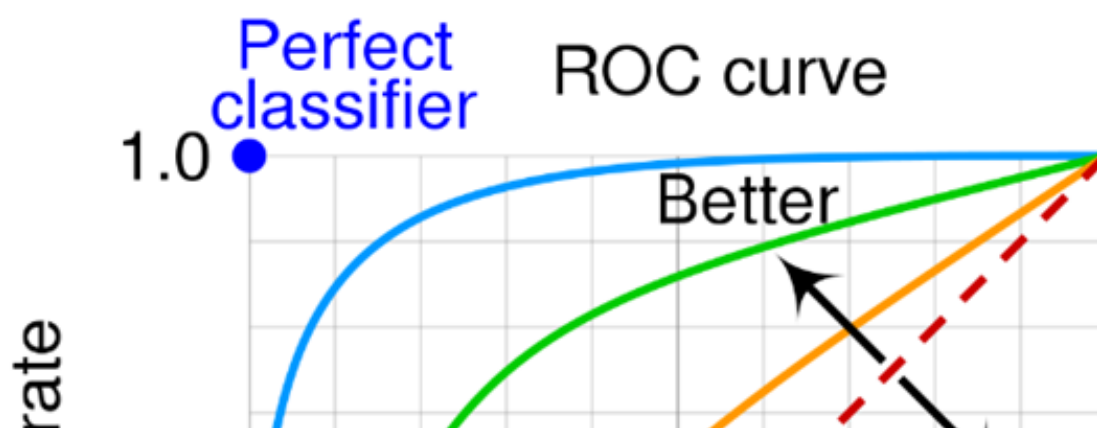
Figur 1: Innhold og utledede ytelsesmål i en krysstabell for en diagnostisk test

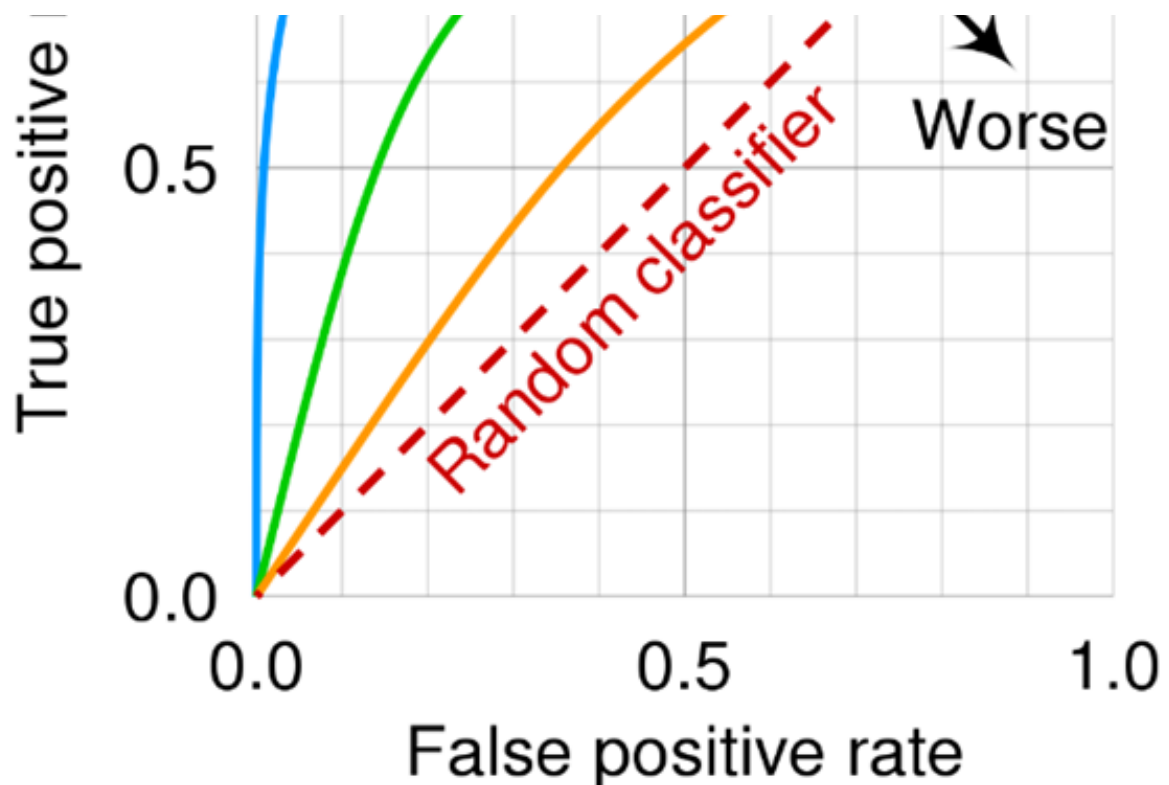
	Faktisk positiv (syk)	Faktisk negativ (frisk)	Ytelsesmål
Test positiv (syk)	Sanne positive (SP)	Falske negative (FN)	Positiv prediktiv verdi $=SP/(SP+FN)$
Test negativ (frisk)	Falske positive (FP)	Sanne negative (SN)	Negativ prediktiv verdi $=SN/(FP+SN)$
Ytelsesmål	Sensitivitet $=SP/(SP+FP)$	Spesifisitet $=SN/(SN+FN)$	

ROC-kurve og AUC

For binære klassifikasjonsmodeller vil det normalt ligge en kontinuerlig modell under, som tilordner en risikoskår til hvert inndata, og da vil en ROC-kurve (engelsk: *Receiver Operating Characteristics curve*) fange opp den relevante informasjonen². Punktene på denne kurven korresponderer med forskjellige terskelverdier for den underliggende risikoskåren. X-aksen er andelen falske positive ($1 - \text{spesifisitet}$) mens y-aksen er andel sanne positive (sensitivitet). Kurven representerer avveiningen mellom sensitivitet og spesifisitet, og hvis man f.eks. krever et bestemt nivå for sensitivitet (som 90%), kan man lese av hva den tilhørende spesifisiteten blir.

Arealet under ROC-kurven (AUC) er et vanlig generelt ytelsesmål, som kan benyttes hvis man ikke har noen klare føringer på sensitivitet eller spesifisitet. AUC er et tall mellom 0 og 1. En modell med 100% gale prediksjoner har en AUC på 0, mens en modell som har 100% riktige prediksjoner har en AUC på 1. Det er derfor et mål på hvor godt risikoskåren skiller mellom positive og negative tilfeller. AUC har en ekvivalent tolkning som sannsynligheten for at et tilfeldig positivt tilfelle får høyere risikoskår enn en tilfeldig negativ. Hvis man ikke har mulighet til å justere noen terskelverdi, vil sensitivitet og spesifisitet være faste verdier, som kan plottes som et punkt i det samme diagrammet.





Figur 2 Illustrasjon av en ROC-kurve (Receiver Operating Characteristics curve)³

ROC-kurven og AUC-målet er i prinsippet uavhengige av balansen i datamaterialet, og påvirkes for eksempel ikke hvis man doubler antallet negative tilfeller, men dette forutsetter at de man legger til trekkes fra samme underliggende fordeling som dem man har i utgangspunktet. AUC kan brukes til overordnet sammenlikning av maskinlæringsmodeller.

Klassifikasjonsmodeller gir ofte en sannsynlighet mellom 0 og 1 for at et tilfelle er positivt, i stedet for et binært resultat. Dette betyr at modellen gir en verdi som indikerer hvor sannsynlig det er at et tilfelle er positivt. For å klassifisere tilfellet som positivt eller negativt, setter man en terskel for sannsynligheten. Hvis verdien overstiger terskelen, klassifiserer modellen tilfellet som positivt; hvis verdien ikke overstiger terskelen, blir tilfellet klassifisert som negativt.

Terskelverdier

Før man gjør en anskaffelse bør man ta stilling til hvordan maskinlæringsmodellen skal brukes i praksis og foreta et valg om terskelverdi, og dermed prioritere mellom sensitivitet og spesifisitet. Denne avveiningen avhenger av å veie opp helsekonsekvensene ved å gå glipp av en diagnose med de eventuelle helseøkonomiske og andre implikasjoner overdiagnostisering medfører.

Et eksempel er en maskinlæringsmodell som skal brukes i tidlige brystkreftscreeningstester. Evnen til å fange opp positive tilfeller kan være spesielt viktig. For eksempel bør ikke noen som faktisk har en ondartet svulst overses. Dette innebærer at modellen krever høy sensitivitet - dvs. andelen *faktiske positive* tilfeller som blir korrekt identifisert. Samtidig, hvis sensitivitet prioriteres overdrevent, blir modellens spesifisitet - dvs. andelen faktiske negative tilfeller som blir korrekt identifisert - uakseptabelt lav. I praksis betyr dette at modellen kan utløse for mange unødvendige konsultasjoner og prøver av personer som ikke har brystkreft.

Hvis derimot maskinlæringsmodellen skal brukes til testing på et senere punkt i utredningen, kan evnen til å korrekt oppdage negative tilfeller bli viktigere, altså et høyere nivå av spesifisitet. For mange falske positive prøvesvar kan gi feilaktig diagnostisering, som kan gi unødig engstelse hos pasienten eller medføre overbehandling. Det å pålitelig fange opp positive tilfeller er fortsatt viktig, så avveiningen med sensitivitet må vurderes.

Klassifikasjon med flere klasser

Maskinlæringsmodeller kan også brukes i situasjoner hvor det finnes mer enn to klasser, for eksempel hvis modellen skal predikere forskjellige sykdommer. Flere av metodene fra binær klassifikasjon kan tilpasses denne situasjonen, og man kan for eksempel beregne sensitivitet, spesifisitet eller prediktive verdier for hver klasse. Det finnes også generiske mål som tar sikte på å vurdere hvor godt modellen fungerer alt i alt, som nøyaktighet (*accuracy* på engelsk) eller det mer komplekse målet F1-skår. Selv om disse kan være nyttige i en vitenskapelig prosess hvor man utvikler og sammenlikner modeller, er det viktig at disse ytelsesmålene vurderes tverrfaglig med analyse- og domenekompetanse for en anskaffelsesprosess.

Krysstabeller er ekstra nyttig når man har flere klasser, siden den illustrerer hvilke klasser modellen har problemer med å skille fra hverandre. Figur 3 er en illustrasjon på en krysstabell for en modell som predikerer hvilken legegruppe i en ortopedisk klinikk som skal behandle en gitt henvisning, basert på henvisningsteksten.

Figur 3 Krysstabell for en modell som fordeler henvisninger mellom legegrupper⁴

Korrekt\Valgt	Hånd	Rygg	Kne	Fot	Skulder	Bekken	Brudd
Hånd	838	49	14	10	4	5	36
Rygg	38	1319	10	11	7	9	11
Kne	10	4	1789	12	3	4	25
Fot	6	16	9	1560	5	19	79
Skulder	4	4	2	4	845	15	21
Bekken	2	4	4	14	10	1512	86
Brudd	36	18	54	93	40	78	2663

Regresjonsmodeller

Mens klassifikasjonsmodeller predikerer valg mellom diskrete klasser, predikerer regresjonsmodeller kontinuerlige størrelser. For eksempel ville regresjonsmodeller være nødvendig for å forutsi hvor mange mennesker som vil oppsøke legevakten på en gitt dag, eller hvor mange sykehjemssenger som vil være nødvendig for pasienter som skrives ut fra sykehuset neste uke.

Ytelsesmål for regresjonsmodeller er basert på statistikk over avviket mellom faktisk og predikert verdi. Et vanlig mål er 'Root mean square error' (RMSE). Den beregnes ved at man kvadrerer avvikene (som gjør dem positive) og tar gjennomsnitt av disse. Så tar man kvadratroten av dette, som gir et resultat med samme enhet som verdien man forsøker å predikere. Et beslektet ytelsesmål er 'Mean absolute error' (MAE), som er gjennomsnittet av absoluttverdien av feilen. MAE kan virke lettere å tolke, men RMSE er ofte et mer relevant mål fordi den straffer store avvik og er knyttet til estimering av gjennomsnittsverdier. Hvis man for eksempel har to caser med like prediktorer og forskjellige utdata-verdier (10 og 20), må modellen lage et kompromiss. For å få lavest mulig RMSE (5), må modellen predikere gjennomsnittet (15), mens alle modeller som gir verdi mellom 10 og 20 vil få minimal MAE (også 5). Derfor er RMSE også mer brukt som kriterium under trening av modeller.

RMSE og MAE måler størrelsen på feilen modellen gjør, men uten å ta hensyn til hvor vanskelig oppgaven er. Hvis verdien man vil predikere varierer lite, vil selv en «modell» som alltid gjetter på gjennomsnittet ha lave verdier på RMSE og MAE. Målet R^2 tar hensyn til dette ved å estimere hvor mye mindre variansen til

feilen er i forhold til variansen på verdien man predikerer. Jo, høyere R^2 , jo mer av variasjonen har modellen forklart, og en perfekt modell har $R^2 = 1$.

Feil med datainnsamling eller uvanlige hendelser kan resultere i "outliers", som er ekstreme observasjoner. Håndtering av slike er noe av det vanskeligste i praktisk medisinsk statistikk. Hvis vi for eksempel modellerer liggetiden i sykehus og en pasient er registrert med en liggetid på 100 år, er det klart en feil, som må fjernes fra datamaterialet. Hvis modellen estimerer gjennomsnittlig liggetid med RMSE som mål, vil denne observasjonen dominere beregningene fullstendig. Men hva så, hvis en liggetid er 1, 2 eller 3 år? Hvis man setter en grense for hva som er mulige liggetider, har det mye å si akkurat hvor denne settes. En vanlig løsning er å bruke MAE i stedet, som matematisk betyr å predikere median i stedet for gjennomsnitt. Men dette løser ikke problemet hvis modellen skal brukes til å planlegge belegg på et sykehus, for hvis liggetider på 2 år er reelle, vil slike pasienter faktisk legge beslag på sengeressursen så lenge. Derfor må man ta stilling til hvilke datapunkter som er feilregistreringer og hvilke som er reelle ekstremverdier.

Segmentering og objektdeteksjon i bilder

KI-modeller som vurderer bilder, vil sjelden være begrenset til ren klassifisering eller regresjon, men vil normalt også markere hvilke deler av bildet som bidrar til konklusjonen. En segmenteringsoppgave kan for eksempel være å tegne inn nyrene i et MR-bilde. Fasiten vil bestå i en eksakt markering av nyrenes areal og målet er at modellen markerer de samme pikslene. Et vanlig mål for samsvar er «Intersection over union», forkortet IoU, som er definert som arealet som er markert både av fasiten og modellen (intersection), delt på arealet som er markert av modellen og fasiten til sammen (union). Ytelsesmålet vil da være gjennomsnittlig IoU-verdi over en samling test-eksempler. IoU fungerer best hvis objektet man skal segmentere er omtrent like stort for de forskjellige eksemplene.

For objektdeteksjon er oppgaven å markere områder hvor en bestemt type objekter finnes med et kvadrat eller en annen geometrisk form, uten å markere hver enkelt piksel som inngår i objektene. Ytelsen til et slikt system gjøres ved at man teller hvor mange objekter som fanges opp og hvor mange falske positive modellen markerer. Man kan også beregne IoU som for segmentering ved å sammenlikne med annoterte markeringer.

Generative modeller

Dagens KI-modeller kan produsere tekst, lyd og bilder. For generering av tekst kan dette for eksempel dreie seg om å lage automatiske sammendrag av pasienthistorikk basert på lagrede journalnotater, referater fra kliniske konsultasjoner, eller oversetting av brev til pasientens morsmål. For lyd kan det for eksempel bli aktuelt med automatisk tolketjeneste, som løpende oversetter norsk tale til pasientens språk. Man kan også tenke seg prateroboter (*chatbots*), som for eksempel tester befolkningen for kognitiv svikt eller gir samtaleterapi for pasienter med angst og depresjon. Bilde- og video-generering for medisinske formål ligger kanskje lenger fram, men man kan tenke seg modeller som automatisk produserer informasjonsmateriell, skreddersydd til en bestemt pasient.

Det er neppe mulig å definere generelle ytelsesmål som automatisk evaluerer generativ KI, men noen anvendelser av denne teknologien vil også falle inn under de kvantitative kategoriene over. Screening for kognitiv svikt med chatbot passer for eksempel inn i rammeverket for klassifisering eller regresjon, avhengig av om utdataene er diskrete eller kontinuerlige.

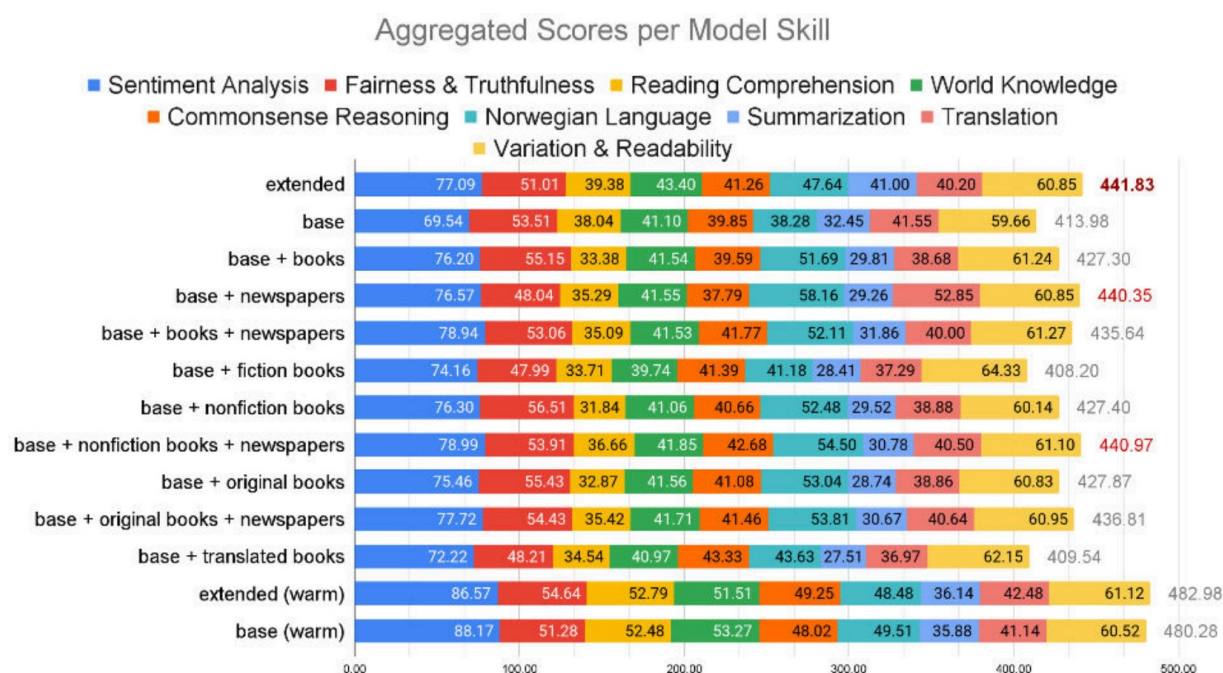
Eksempelet med chatbot som automatisert psykologtjeneste til behandling av angst og depresjon må vurderes på linje med medisinske intervensjoner generelt, fortrinnsvis med randomiserte kontrollerte studier. Selv om slike systemer kan være teknisk realiserbare i nær framtid, vil implementering i klinisk bruk

sannsynligvis ligge et stykke fram i tid fordi validering av sikkerhet og effekt er utfordrende. Dersom KI-systemet er medisinsk utstyr så skal det samsvarsvurderes jf. MDR og CE-merkes.

Ytelsen til KI-systemer som genererer tekst til bruk for mer oppgaver må normalt testes gjennom vurderinger av brukergruppen. Hvis man for eksempel vil innføre en KI-modell for å lage sammendrag av sykehistorier til støtte for klinikere, må nødvendigvis de aktuelle klinikerne vurdere hvor nyttig den er.

En måte å måle ytelsen til en språkmodell på er å utarbeide testsett rettet mot spesifikke oppgaver og sammenligne svarene fra modellen med testsettet. Det kan for eksempel være predefinerte spørsmål og svar (*ground truth*).

I Mimir-prosjektet har Nasjonalbiblioteket sammenlignet ytelsen til noen språkmodeller på tvers av en rekke språklige og naturlige oppgaver for språk, som tekstgenerering, oversettelse, spørsmål/svar og sentimentanalyse.⁵ Figuren under viser en sammenligning av hvordan ulike generative språkmodeller skårer på ulike ferdigheter som rettferdighet, sannferdighet og norsk språk.



Figur 4: Sammenligning av ulike generative språkmodeller

[1] [Hva er sannsynligheten for riktig resultat av en diagnostisk test? | Tidsskrift for Den norske legeforening \(tidsskriftet.no\)](https://tidsskriftet.no/2018/09/medisin-og-tall/roc-kurver-og-diagnostiske-tester)

[2] ROC står for Receiver operating characteristic curve. ROC-kurven er et grafisk verktøy som brukes for å vurdere hvor godt en binær klassifikasjonsmodell skiller mellom to klasser. Uttrykket kommer fra andre verdenskrig, der det ble brukt for å måle radaroperatørenes evne til å skille mellom fiendtlige objekter og støy. <https://tidsskriftet.no/2018/09/medisin-og-tall/roc-kurver-og-diagnostiske-tester>

[3] <https://spotintelligence.com/2024/06/17/roc-auc-curve-in-machine-learning/>

[4] Tabellen er produsert ved hjelp av ChatGPT 4o for illustrasjonsformål.

[5] https://www.nb.no/content/uploads/2024/08/Mimirprosjektet_teknisk-rapport.pdf

