

PETROLEUMSTILSYNET

## ET MENNESKESENTRETT PERSPEKTIV PÅ KOGNITIV TEKNOLOGI I PETROLEUMSINDUSTRIEN

HOVEDRAPPORT

ST-16407-2

Type dokument:

Hovedrapport

Rapport tittel:

Et menneskesentrert perspektiv på kognitiv teknologi i petroleumsindustrien

Kunde:

Petroleumstilsynet

Dokument nr. ST-16407-2				
Forfattere J. Ernstsens, E. G. Aulie, G. B. Sætren, H. C. Stenhammer, R. Phillips				
<i>Referanse til deler/utdrag av dette dokumentet som kan føre til feiltolkning, er ikke tillatt.</i>				
Rev.	Dato	Grunn for rev.	Kontrollert	Godkjent
1.0	06.10.2021	Utkast	G. Hauland	J. C. Rolfsen
2.0	15.10.2021	Endelig	G. Hauland	J. C. Rolfsen
3.0	18.11.2021	Endelig	G. Hauland	J. C. Rolfsen

## OPPSUMMERING

I dette prosjektet er det innhentet og sammenstilt kunnskap om utvikling og implementering av kognitiv teknologi i sikkerhetskritiske operasjoner innen høyrisikoindustri, primært petroleumsindustrien. Kunnskapen er hentet fra forskningslitteratur, næring og myndigheter.

Problemstillingene som vi har undersøkt i dette prosjektet er:

- Hva er kognitiv teknologi i petroleumsnæringen i Norge i dag?
- Hvordan samhandler mennesket og kognitiv teknologi?
- Hva er sluttbrukers rolle i utviklingen av kognitiv teknologi?
- Hvilke premisser gjelder for arbeid med kognitiv teknologi i dag?

For å undersøke problemstillingene ovenfor gjorde vi først en systematisk litteraturgjennomgang. Etterfulgt av semistrukturerte intervjuer med nøkkelaktører fra selskaper i petroleumsindustrien samt myndigheter i industrier som maritim industri og bilindustri. Rapporten gir basert på sammenfattet kunnskap og drøfting av denne opp mot problemstillingene.

Kognitiv teknologi i petroleumsnæringen i dag:

1. Kognitiv teknologi benytter i dag enkle tolkbare modeller. Tilnærminger og metoder for å jobbe med samhandlingsproblematikk, som f.eks. funksjonsallokering, må gjøres i sammenheng med den faktiske bruken av kognitiv teknologi.
2. Kognitiv teknologi bør utvikles for å løse definerte problem, ikke utvikles som løsninger som kan være nyttige for å løse problemer man ikke har definert. Behovet for teknologien må være drevet av problemløsning for å sikre god forankring i hele utviklingsprosessen.
3. Sluttbrukers kriterier inkluderes for lite ved valg av datamodell. Praxis og erfaring viser at det er vanskelig å inkludere sluttbrukers kriterier i valg av datamodell på grunn av usikkerhet knyttet til hvordan operasjoner vil bli utført etter introduksjonen av kognitiv teknologi.

Utvikling og implementering av kognitiv teknologi:

4. En tilnærming basert på et sosioteknisk rammeverk bør ytterligere nyanseres slik at cyberdelen i systemet, som inkluderer utvikling av datamodell og software, får økt fokus i arbeid med kognitive teknologier. HCAI (Human Centered Artificial Intelligence)-rammeverket kan legges til grunn for å fremme brukersentrert utvikling av kognitive teknologier. Særlig i systemer der det er både høy grad av automatisering og menneskelig kontroll.
5. Det er behov for metoder og tilnærminger for involvering av sluttbruker i tidlig utviklingsfase av datamodeller til bruk i sikkerhetskritiske operasjoner. AutoML og interaktiv maskinlæring (IML) er to tilnærminger som involverer sluttbruker og som kan fungere som et utgangspunkt til videre arbeid.
6. Involvering av sluttbruker fungerer i sen utviklingsfase. Det ble understreket at sluttbrukere, som en knapp ressurs, prioriteres i sen utviklingsfase av selskapene. Blant annet grunnet mer konkrete krav satt av regelverk og standarder og usikkerheter knyttet til metoder for brukerinvolvering i tidlig fase.
7. Tilbakemeldinger fra sluttbrukere i sen utviklingsfase er viktig for videre oppdatering og utvikling av modellen. Tilbakemeldingene baseres på trening med kognitiv teknologi i simulatorer og bruk i drift etter at teknologien er implementert offshore.

8. Sluttbrukers forståelse av hvordan kognitiv teknologi fungerer er viktig for samhandling. Viktige faktorer for å danne seg en god mental modell er trening og bruk, i tillegg til at datamodellen er gjennomsiktig.

Samhandling menneske og kognitiv teknologi:

9. Datamodeller som er tolkbare for sluttbruker er å foretrekke. Dersom black box-modeller må brukes bør det benyttes verktøy for å øke modellens tolkbarhet.
10. Å balansere riktig grad av tillit i sikkerhetskritiske operasjoner er viktig og krevende. En del av denne balansen handler om at den menneskelige aktøren må forstå sammenhengen mellom inputvariablene og output, som igjen setter føringer for datamodellens gjennomsiktighet.
11. Metoder og tilnærminger for funksjonsallokering må passe til den faktiske bruken av kognitiv teknologi. Det finnes mange metoder og tilnærminger til temaet som antar mer avansert teknologi enn det som brukes i praksis.

Premisser for arbeid med kognitiv teknologi:

12. Det er usikkerhet i næringen knyttet til sluttbrukers fremtidige kompetansebehov dersom kognitiv teknologi blir mer avansert eller utbredt.
13. Det er behov for å avklare en datadelingsfilosofi for å sikre/få til en åpen og rettferdig deling av data innen næringen. Hensikten med datadelingsfilosofien vil være å utvikle sikre og effektive datamodeller.
14. Maskinlæring er et begrep som benyttes vidt i markedsføringsøyemed. Det kan være vanskelig å få et riktig bilde av den kognitive teknologien ulike selskaper faktisk anvender, basert bare på hva som beskrives i markedsføring.
15. Innovasjonsgraden til kognitive teknologier kan påvirkes av samfunnets interesser i O&G. Det er derfor en usikkerhet knyttet til om utfordringene med fremtidens kognitive teknologi blir gjeldende.
16. Det er hensiktsmessig med samarbeid for å løse felles problemer knyttet til bruk av kognitiv teknologi på tvers av industrier og myndigheter.

## Innhold

1	INTRODUKSJON .....	7
1.1	Kognitiv teknologi.....	7
1.2	Hvorfor må vi forstå mer om kognitiv teknologi i høyrisikoindustri? .....	8
1.3	Rask utvikling innen kognitiv teknologi.....	8
1.4	Fremme menneskets samhandling med kognitiv teknologi .....	9
1.5	Omfang og begrensninger .....	9
1.6	Metode.....	10
1.7	Hovedmål, problemstilling og videre arbeid .....	11
2	KOGNITIV TEKNOLOGI I PETROLEUMSNÆRINGEN I DAG.....	12
2.1	Kognitiv teknologi i høyrisikoindustri.....	12
2.2	Bruk av kognitiv teknologi i norsk petroleumsindustri .....	12
2.3	Vurdering av kognitiv teknologiens merverdi .....	13
2.4	Vurdering av kriterier for både tekniske systemer og sluttbruker.....	14
3	UTVIKLING OG IMPLEMENTERING AV KOGNITIV TEKNOLOGI.....	15
3.1	Rammeverk for arbeid med kognitiv teknologi.....	15
3.2	Involvering av sluttbruker .....	16
4	SAMHANDLING MENNESKE OG KOGNITIV TEKNOLOGI.....	22
4.1	Mentale modeller viktig for samhandling.....	22
4.2	Gjennomsiktige datamodeller .....	22
4.3	Tillit til kognitiv teknologi .....	24
4.4	Funksjonsallokering.....	25
5	PREMISSER FOR ARBEID MED KOGNITIV TEKNOLOGI.....	28
5.1	Usikkerhet knyttet til sluttbrukers fremtidige kompetansekrav.....	28
5.2	Behov for å avklare datadelingsfilosofi .....	29
5.3	Maskinlæring som begrep er utsatt for effektiv markedsføring.....	29
5.4	Innovasjon påvirkes av samfunnets interesser i O&G.....	30
5.5	Koordinering med andre myndigheter.....	30
6	KONKLUSJON .....	31
7	MULIG VIDERE ARBEID MED KOGNITIV TEKNOLOGI.....	33
7.1	Standarder og metoder for brukersentrert utvikling av kognitiv teknologi.....	33
7.2	Strategi for autonome systemer .....	33
8	REFERANSER.....	34

Vedlegg A: Artikler fra litteraturstudien

Vedlegg B: Detaljert metode

Begrep	Forklaring
Datadrevet automatisering	En modell som er utviklet basert på data, gjerne med hjelp av ML-algoritmer.
Datamodell	En modell av virkelighet basert på data. I kognitive teknologi brukes datamodell om det som prosesserer inputdata og kommer med output.
Fysikkbasert modell	Samlebetegnelse for datamodeller basert på naturlover og fysikk.
Maskinlæring	Læringsprosessen der blant annet ML-algoritmer brukes til å utvikle ML-modeller.
Menneskesentrert design	En tilnærming til interaktiv utvikling av systemer som har et mål om å gjøre systemet nyttig og brukervennlig for sluttbruker.
ML-algoritme	Samlebetegnelse for algoritmer som brukes til å utvikle en ML-modell. Eksempler er lineær- og logistisk regresjon, beslutningstrær, eller nevrale nettverk.
ML-modell	Samlebetegnelse for datamodeller utviklet med ML-algoritmer.
Nevrale nettverk	Samlebetegnelse for datastrukturer og algoritmer som er inspirert av måten nervecellene i en hjerne er organisert på.
Kognitivt system	Et system bestående av to eller flere både menneskelige og tekniske kognitive aktører som jobber sammen mot et felles mål, og kan tilpasse seg endringer.
Kognitiv teknologi	Et teknisk system som a) oppfatter (input), b) beregner/prosesserer (datamodellen), og c) foreslår eller iverksetter aksjoner (output), der beregningen er basert på ML-modeller.
Kognitiv aktør	Et menneske eller en ML-modell.
Kompleksitet	Et begrep om noe som består av flere faktorer (dimensjonalitet) og gjensidig avhengighet mellom disse (mellomavhengighet).
Sikkerhetskritisk operasjon	En operasjon med storulykkespotensiale, f.eks. flere personer mister livet, betydelig skade på miljø eller betydelige kostnader.
Situasjonsforståelse / situasjonsbevissthet	Forståelse av hva som «foregår rundt oss» (forenklet definisjon av et komplekst begrep).
Sluttbruker	En menneskelig- eller teknisk aktør som interagerer med det kognitive systemet for utføre sin funksjon. Sluttbruker står i kontrast til roller som vedlikeholder, støtter, utvikler og designer det kognitive systemet. Sluttbruker har erfaring og kompetanse på drift av systemet og operasjon.

Forkortelse	Engelsk	Norsk
AAL	Automation Allocation Logic	Automatisert fordelingslogikk
AI	Artificial Intelligence	Kunstig intelligens
CoCyS	Cognitive-Cyber-Symbiosis	Kognitiv-cybersymbiose
HCAI	Human-Centered Artificial Intelligence	Menneskesentrert kunstig intelligens
HF	Human Factors	Menneskelig faktorer
HMI	Human Machine Interaction/interface	Menneske-maskin interaksjon
IML	Interactive Machine Learning	Interaktiv maskinlæring
ML	Machine Learning	Maskinlæring
MTO	Man Technology Organisation	Menneske-teknologi-organisasjon
Ptil/PSA	Petroleum Safety Authority Norway	Petroleumstilsynet
SLR	Systematic Literature Review	Systematisk litteraturgjennomgang

## 1 INTRODUKSJON

Denne studien er en del-leveranse i Petroleumstilsynets (Ptil) satsning innen IKT-sikkerhet. Bakgrunnen for dette oppdraget er behovet for innhenting og sammenfatning av kunnskap om hvordan vi kan fremme kognitive systemers pålitelighet ved å inkludere kunnskap om menneskelige faktorer.

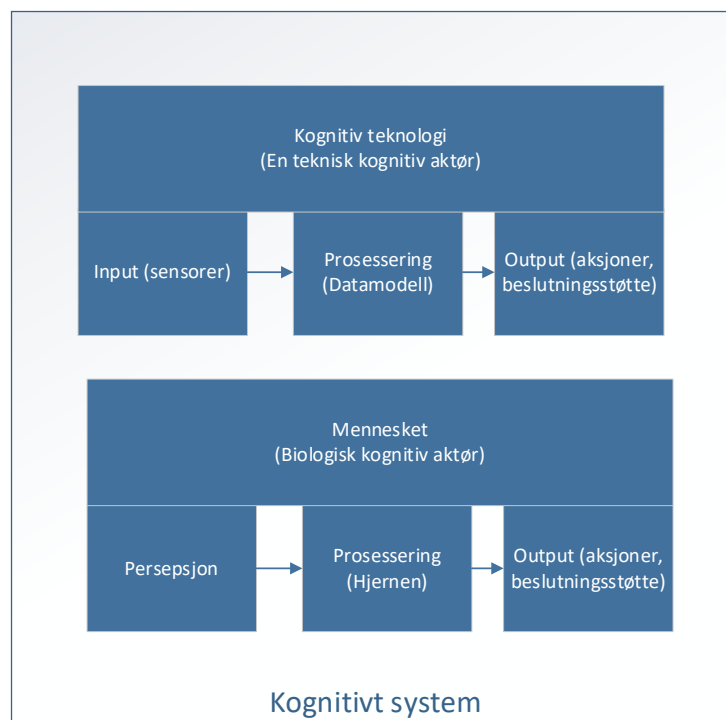
Mange bedrifter i petroleumsindustrien viser en økende interesse for kognitiv teknologi, fordi slik teknologi kan bidra til at sikkerhetskritiske oppgaver utføres på en mer effektiv, sikker eller bærekraftig måte. Dette medfører også at en del tekniske, organisatoriske og menneskelige utfordringer må løses for å realisere potensialet som kognitiv teknologi gir.

### 1.1 Kognitiv teknologi

Teknologisk optimalisering, og automatisering av drift og vedlikehold, kan øke produktiviteten og sikkerheten i petroleumsindustrien. I et globalt konkurransepreget marked jaktes det kontinuerlig på nye kostnadseffektive måter å øke utvinning og salg av petroleumsprodukter. Et ledd i dette er å optimalisere og automatisere drift og vedlikehold, ved å utvikle og implementere kognitiv teknologi. Kognitiv teknolog kan forstås som en videreutvikling av kybernetikk.

En viktig del av kybernetikk er å innhente data via sensorer, prosessere informasjonen, beslutte og så aksjonere, samt læring/tilbakemelding som en respons til aksjonen. Denne prosessen er ofte referert til som «kybernetikk-løkken». Det finnes flere forslag på definisjoner av kybernetikk i faglitteraturen. Det viktigste for dette prosjektet er forståelsen av at kybernetikk handler om kommunikasjon og kontroll av og mellom kognitive aktører. Kognitive aktører inkluderer både tekniske- og menneskelige aktører, men det er kun tekniske kognitive aktører som utgjør kognitiv teknologi.

Mennesket er også en kognitiv aktør. Når flere kognitive aktører, enten de er tekniske- eller menneskelige, jobber sammen mot et felles mål, kalles det et kognitivt system (se Figur 1-1). Mulighet for å tilpasse seg endringer er sentralt i et kognitivt system.



Figur 1-1: Et kognitivt system bestående av en menneskelig kognitiv aktør og en teknisk kognitiv aktør.

Datamodellen er «hjernen» i den kognitive teknologien. Det finnes flere typer tilnærminger for å programmere datamodellen. De kan være enkle, som eksempelvis «hvis temperatur over 90 grader, gi alarm til operatør», til å kreve tusenvis av linjer med programmering for å utføre (som naturligvis igjen krever enda mer av teknologien/hardware). Slike datamodeller kan bidra til økt sikkerhet og effektivitet, f.eks. ved bruk av en datamodell til å tolke hvorvidt signalene fra brønnen kan være tidlige tegn på en ukontrollert inn/utstrømning i brønnen eller ikke.

Kognitiv teknologi utfører ofte oppgaver sammen med menneskelige aktører. Et slikt samarbeid kan bestå i at kognitiv teknologi observerer operasjonen i felt (med sensorer) og presenterer et bilde til en menneskelig aktør som tolker og beslutter basert på informasjonen som kognitiv teknologi gir til mennesket. Etter hvert som kognitiv teknologi videreutvikles og blir mer sofistikert, kan informasjonen som presenteres til den menneskelige aktøren bli stadig mer gjennomarbeidet og kompleks. Dette kan gjøre det enklere for menneskelig aktør å forstå hva som skjer i felt, men samtidig også vanskeligere å forstå hvorfor og hvordan den kognitive teknologien kommer frem til den informasjonen og beslutninger som den gjør. Det viser seg også å være mer effektivt at tekniske aktører tar over flere menneskelige funksjoner som å forstå og beslutte hva som skjer i felt, slik at menneskets oppgave flyttes i økende grad til å overvåke automatiserte tekniske prosesser.

## **1.2 Hvorfor må vi forstå mer om kognitiv teknologi i høyrisikoindustri?**

Petroleumsindustrien er en høyrisikoindustri som i økende grad tar i bruk kognitiv teknologi for å utfylle og erstatte menneskelige aktører, noe som gir et økt behov for å sikre bedre forståelse av denne teknologien og hvordan samhandlingen er med mennesket.

Høyrisikoindustri utfører flere sikkerhetskritiske operasjoner. Kognitiv teknologi kan gjøre det mulig å utføre slike operasjoner på en sikrere og mer effektiv måte. I dag, må slik teknologi i stor grad samhandle med mennesker, spesielt sluttbrukere, for å sikre at teknologien fungerer. For å realisere potensialet i kognitiv teknologi må vi forstå mer om denne samhandlingen.

Det er flere utfordringer for samhandling som presenteres etter hvert som kognitiv teknologi introduseres i sikkerhetskritiske operasjoner. En sentral og generell utfordring går på behovet for at sluttbruker forstår teknologien, hva den baserer forslag og aksjoner på, og hvordan den prosesserer innkommende variabler /1/. Så lenge det er mennesket som sitter med det øverste ansvaret i operasjonen er det problematisk dersom den kognitive teknologien blir så kompleks at det er vanskelig for sluttbruker samhandle sikkert og effektivt med den /2/.

En slik utfordring introduserer flere problemstillinger; fra hvordan kognitive teknologi utvikles og implementeres, til hvilke kompetansekrav som kan og bør stilles til sluttbruker.

## **1.3 Rask utvikling innen kognitiv teknologi**

Bruk av kognitive teknologier er ikke nytt. Flere forskere har i lang tid sett på hvordan mennesket og teknologi kan fungere sammen for å oppnå den «kognitive flyten», som er nødvendig for optimal styring av prosesser i både rutine og kritiske situasjoner /3/. Det har blitt pekt på at kognitiv teknologi må støtte menneskelige evner, begrensinger og forutsetninger. Dette er viktig fordi mennesket alltid forstå situasjonen slik at raske beslutninger kan tas for å unngå uønskede- og kritiske hendelser /4/.

Etter hvert som kognitiv teknologi utvikles slik at den kan overta mer av det operasjonelle- og taktiske ansvaret enn i dag, stilles nye krav til samhandlingen mellom mennesket og kognitiv teknologi. Operasjonelt- og taktisk ansvar handler om å utføre operasjonen i tillegg til å overvåke og tilpasse seg umiddelbare endringer i og rundt operasjonen. De nye kravene som må stilles er både tekniske; som



behov for dataoverføringskapasitet, organisatorisk; som presise industristandarder, og samhandlingen mellom mennesket og kognitiv teknologi /5/. Et eksempel på nye krav kan være å oppdatere designstandarder til at kognitive systemer har tekniske aktører med ansvar for overvåking og styring av prosess. Dette er avklaringer som må tas for å fremme god samhandling mellom mennesker og kognitiv teknologi /4/.

#### **1.4 Fremme menneskets samhandling med kognitiv teknologi**

For å løse utfordringene knyttet til hvordan kognitiv teknologi utvikles og implementeres er det viktig at sluttbrukere får medvirke tidlig i utvikling og implementering av kognitiv teknologi. Denne påstanden er støttet av omfattende forskning i ulike fagfelt som menneskelige faktorer /4/, sosiotekniske systemer /5/, kognitive systemer /6/, human-maskin interaksjon (HMI) /7/ og meningsfull menneskelig kontroll /4/. Likevel utvikler dataingeniører og systemutviklere i ulike industrier, design som fortsatt er mer teknologisentrert enn brukersentrert /8; 9/. Ved utvikling og implementering av for eksempel teknologi for automatisert boring tar utviklere og utviklingsprosesser for lite hensyn til menneskelige faktorer og sluttbrukerens behov, kognitive arbeidsmåte eller arbeidskontekst 10<sup>[OB]</sup>/.

I implementering av kognitiv teknologi vil det være viktig å ta hensyn til det felles kognitive systemet, ved optimalisering av både sosiale og tekniske komponenter enkeltvis og samlet i systemet. Utvikleren må forstå hvordan sluttbrukeren samhandler med kognitiv teknologi slik at sluttbrukeren kan oppnå en god situasjonsforståelse for å styre risikoene når kritiske situasjoner oppstår /11/. De må også forstå hvordan kognitiv teknologi kan produsere data som tolkes forskjellig avhengig av bruker og situasjon /5/.

#### **1.5 Omfang og begrensninger**

Det er et begrenset antall kognitive teknologier med maskinlærte datamodeller (ML-modeller) i bruk i olje- og gassindustrien i dag. De aller fleste er i konseptutviklingsfasen, og operasjonell erfaring med bruk av ML-modeller er derfor begrenset. Siden kognitive teknologier testes og implementeres også i andre industrier, har det vært ønskelig å lære om hvordan andre bransjer håndterer utfordringer fra samspill mennesket og kognitiv teknologi. Basert på følgende kriterier ble "andre bransjer" begrenset til ulike organisasjoner knyttet til maritim industri og bilindustri:

- høy grad av teknologisk utvikling.
- funnene har relevans for petroleumsbransjen.
- tilgjengelig for prosjektteamet.

Vi har ikke sett på petroleumsindustrier utenfor Norge. Vi snakker om kognitive teknologier der de brukes, eller er tenkt og brukt, i sikkerhetskritiske operasjoner. Dette ekskluderer ML-modeller som f.eks. brukes i analyser utført onshore (der det er god tid og tilgjengelige ressurser til å kvalitetssikre output).

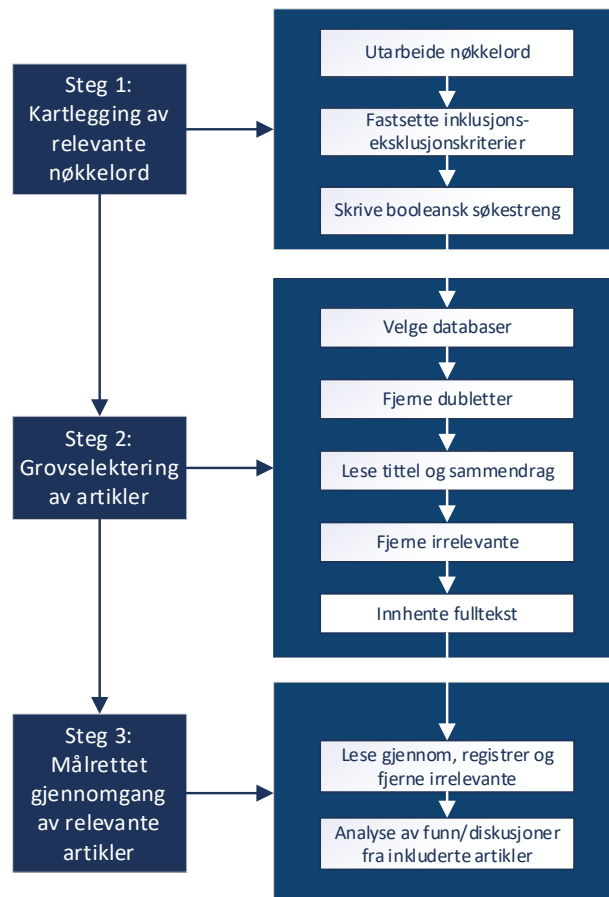
I denne rapporten beskriver vi hvordan vi har innhentet informasjon og sammenfattet dette til kunnskap som kan brukes i møte med de utfordringene som oppstår, og vil oppstå, i samspill mellom mennesket og kognitive teknologier.

## 1.6 Metode

Innhenting av kunnskap ble gjennomført med en systematisk litteraturgjennomgang og semistrukturerte intervjuer. Kunnskap fra litteraturgjennomgangen ble brukt til å utvikle intervjuguide, i tillegg til å bidra inn i diskusjonen av problemstillingene.

### Litteraturstudie

Vi benyttet systematisk litteraturgjennomgang (Systematic Literature Review, SLR) /12; 13; 14/ som metode til dette litteraturstudiet.



Figur 1-2 Prosess for litteraturstudie, tilpasset fra /15/.

I den systematiske litteraturgjennomgangen gjennomførte vi en akademisk komponent (steg 1 til 3) og en pragmatisk målrettet gjennomgang av relevante artikler. Dette sikrer at prosjektet både vurderer underlag som er definert relevant til arbeidsomfanget, men også bidrar med ny kunnskap innhentet med formål om å redusere subjektiv innvirkning på selektering av artikler samt være transparente på hvordan disse er selektert.

Det første steget var en kartlegging av relevante nøkkelord som skulle brukes i litteratursøket. Deretter ble det i det andre steget gjennomført en grovselektering av artikler returnert fra nøkkelordene. I det tredje steget ble de selekterte artiklene gjennomgått, og det fjerde steget innebar en målrettet gjennomgang av relevante artikler.

## **Intervjustudie**

Det ble benyttet semistrukturerte gruppeintervjuer og individuelle intervjuer for å innhente kunnskap fra næring og andre myndigheter. Analysemetoden tok utgangspunkt i tematisk analyse /16/. Intervjuene og analysen ble utført av et tverrfaglig forskerteam med bakgrunn fra ulike ingeniørfag, ML, HF, bioteknologi og sikkerhet, inkludert arbeid innenfor prosess- og boreteknologi.

31 personer fra 10 ulike selskaper ble intervjuet, i totalt 15 intervjuer. Blant intervjuobjektene var det både store og mindre operatørselskaper, leverandørselskaper og boreselskaper. Intervjuene varte i omtrent 45-60 minutter og 2-4 forskere deltok. I tillegg ble det gjennomført intervjuer med andre myndigheter innenfor sjøfart og bilindustri. Ptil ble ikke intervjuet i dette prosjektet.

Analysen tok utgangspunkt i tematisk analyse /16/. Umiddelbart etter gjennomføring av hvert intervju diskuterte forskerne intervjuene, delte notater og oppsummerte hva som var de viktigste funnene fra intervjuet. På bakgrunn av disse samtalene ble intervjuguiden og intervjuene videreutviklet. Etter at alle intervjuene var gjennomført, samlet forskerne seg i ukentlige analysemøter.

Temaet for prosjektet ble ikke vurdert å være sensitivt eller omhandle informasjon som var egnet til å kunne skade informantene. Innsamlede data er anonymisert i henhold til GDPR. Ingen informanter kan gjenkjennes i publisert materiale basert på utsagn fra intervju.

## **1.7 Hovedmål, problemstilling og videre arbeid**

Hovedmål: Øke forståelsen av samhandlingen mellom mennesker og kognitiv teknologi ved:

- Sammenstilling av kunnskap fra forskning, myndigheter og næringsliv.
- Utføre en kartlegging av relevante kognitive teknologi-prosjekter.
- Kartlegge bruk av kognitive teknologier i dag.

For å forstå menneskets samhandling med kognitive teknologier er det en rekke kartlegginger og spørsmål som må stilles. I dette prosjektet problematiserer vi følgende tema:

- Hva er kognitiv teknologi i petroleumsnæringen i Norge i dag?
- Hvordan samhandler mennesket og kognitiv teknologi?
- Hva er sluttbrukers rolle i utviklingen av kognitiv teknologi?
- Hvilke premisser gjelder for arbeid med kognitiv teknologi i dag?

Vi gir deretter en refleksjon om mulig videre arbeid med disse problemstillingene.

## **2 KOGNITIV TEKNOLOGI I PETROLEUMSNÆRINGEN I DAG**

For å spisse arbeid og forståelse av kognitiv teknologi må det gjøres en kartlegging av hvordan teknologien brukes i praksis. Først presenteres en generell bruk av kognitiv teknologi i høyrisikoindustri. Deretter, basert på samtaler med næringen om dette temaet, gir vi en realitetsorientering om bruk av teknologien i dag. Vi legger også frem diskusjonen fra intervjuene om hvilke kriterier som legges til grunn i valget om bruk av kognitiv teknologi og om hvordan relevante og potensielle prosjekter velges for å ende på gode tekniske løsninger.

### **2.1 Kognitiv teknologi i høyrisikoindustri**

Operasjoner i høyrisikoindustrier er ofte komplekse og derfor vanskelig å fullstendig automatisere. En årsak til dette er utfordringer med å beskrive og predikere slike operasjoner, tradisjonelt gjort med deterministiske fysikkbaserte datamodeller. I mange deler av operasjonen kjenner man ofte ikke til hele virkeligheten, noe som gjør det vanskelig å bygge fysikkbaserte modeller av virkeligheten. ML, derimot, viser seg å være et nyttig verktøy for å lage datamodeller av mange operasjoner der deterministiske fysikkbaserte modeller ikke strekker til /17/.

Det finnes flere typer ML som kan brukes til å utvikle slike datamodeller. En sentral styrke i de fleste ML-algoritmene er at de har evnen til å raskt lese og forstå store datasett. Denne læringen kan gjøres på flere måter. En mye brukt tilnærming er å repetere algoritmen mange ganger (iterativ prosess), og dermed lære modellen mer og mer om datasettet. Denne repeteringen kan gjøres millioner av ganger for å trene modellen.

Før ML-algoritmen kan repeteres må den først vite hvilke egenskaper i datasettet den skal se etter. Dette kan gjøres ved at ML algoritmen bruker statistiske metoder til å finne sammenhenger i datasettet uten at mennesket forteller algoritmen hvor og hva den skal lete etter. Dette kan for eksempel være faktoranalyse, gjenkjenne avvik i trender eller finne assosiasjoner i datasettet. Denne tilnærmingen til ML er ofte referert til som ikke-assistert maskinlæring.

En annen tilnærming til ML er assistert maskinlæring /18/. I denne tilnærmingen navngir mennesket hvilke egenskaper i datamaterialet datamaskinen skal se etter. Dette kan f.eks. være å fortelle datamaskinen at når et sett av trykkmålinger danner et bestemt mønster, så kan dette mønsteret forstås som f.eks. et brønnsplask (kick). Når dette er repetert mange nok ganger, kan modellen brukes til å kjenne igjen brønnsplask basert på trykkmålinger i sanntid (merk at eksempelet brønnsplask er et mer komplisert fenomen enn fremstilt her).

Dersom operasjonen er så kompleks at det ikke er mulig å formulere klare matematiske sammenhenger, kan det være behov for dypere læring av datasettet. En tilnærming til dyp læring er å bruke nevralt nettverk til å lære datasettet. Et nevralt nettverk er en datastruktur med et sett av algoritmer organisert på en måte som er inspirert av måten nervecellene i en hjerne er organisert på. For å bruke det samme eksempelet: Dersom brønnsplask består av trykk, temperatur og pumpeslag, så kan man trene en datamodell med et nevralt nettverk til å finne dypere sammenhenger mellom trykk, temperatur og pumpeslag som tidlig indikasjon på brønnsplasking. Nevrale nettverk må trenes opp på store datasett for å få god treffsikkerhet.

### **2.2 Bruk av kognitiv teknologi i norsk petroleumsindustri**

I forskningslitteraturen, i media, i diskusjoner på relevante fagkonferanser og i markedsføring er det lett å få en entusiastisk holdning til kognitiv teknologi. Det kan fort mislede interesserte til å tenke på avanserte tekniske løsninger der komplekse black box-modeller brukes til å utføre operasjoner.

I den empiriske kartleggingen ble det gitt en realitetsorientering omkring faktisk bruk av kognitiv teknologi. Det kom frem at det er viktig å skille mellom maskinlærte datamodeller som brukes i «onshoreanalyser», hvor det er nok tid og ressurser til å kvalitetssikre output, og de som brukes i sikkerhetskritiske operasjoner. Der avanserte ML-modeller, som nevrale nettverk, kan brukes i onshoreanalyser, er det kun identifisert eksempler på enkle tolkbare modeller til bruk operasjonelt. Det ble heller ikke identifisert konseptprosjekter med bruk av black box-modeller i sikkerhetskritiske operasjoner.

Samtidig ble det presisert i intervjuene at problemstillinger knyttet til samhandling med kognitiv teknologi oppleves som relevante, til tross for at bruken av kognitiv teknologi i olje og gass næringen i dag er begrenset til enklere tolkbare datamodeller. Det ble understreket at samhandlingsproblematikk må vurderes i sammenheng med den faktiske kognitive teknologien som benyttes, og at forskningslitteraturens komplekse metoder for å tilnærme seg dette temaet må vurderes i lys av faktisk bruk.

*Funn 1: Kognitiv teknologi benytter i dag enkle tolkbare modeller. Tilnærminger og metoder for å jobbe med samhandlingsproblematikk, som f.eks. funksjonsallokering, må gjøres i sammenheng med den faktiske bruken av kognitiv teknologi.*

### **2.3 Vurdering av kognitiv teknologiens merverdi**

Funn fra intervjustudien viser at kognitiv teknologi må kunne gi merverdi til operasjonen. Selv om det er mulig å lage en avansert datamodell, så er det ikke gitt at det gir merverdi. For eksempel, dersom en deterministisk fysikkbasert datamodell gjør jobben godt nok, gir det ikke merverdi å implementere en avansert ML-basert datamodell. Det ble i flere intervjuer understreket at datamodellene må løse et definert problem eller en oppgave, heller enn å finne et problem som en avansert datamodell kan løse. En utviklingsprosess drevet av problemløsning gir en sterkere forankring i hele prosessen, som er viktig for et godt og sikkert sluttprodukt.

Det ble videre diskutert viktigheten av å skille mellom datamodeller for kostnadseffektivisering og datamodeller for å redusere risiko og for å forbedre fysisk arbeidsmiljø. Datamodeller som kun implementeres for kostnadseffektivisering gjør ikke nødvendigvis systemet bedre. De 4 D'ene «dirty, dull, dangerous and distance» er en tilnærming som kan brukes for å synliggjøre hvilke problemer en datamodell kan løse. Målet er da å redusere D'ene og ikke nødvendigvis redusere kostnader (selv om det ofte er en bonus).

Kort fortalt omhandler kriteriet *dirty* å redusere eksponering, eksempelvis for farlige kjemikalier. *Dull* går ut på å redusere repetitivt og kjedelig arbeid. *Dangerous* går ut på å redusere eksponering for farlig arbeid. Siste kriteriet er *distance*, som omhandler fjernopererte systemer, der formålet er å redusere lange reiseveier. En slik vurdering av datamodeller bidrar til at selskaper og leverandører kan synliggjøre at implementeringen gir merverdi i det totale kognitive systemet. At datamodellen har en nytteverdi utover «bare» kostnadseffektivisering, er viktig for sluttbruker.

*Funn 2: Kognitiv teknologi bør utvikles for å løse definerte problem, ikke utvikles som løsninger som kan være nyttige for å løse problemer man ikke har definert. Behovet for teknologien må være drevet av problemløsning for å sikre god forankring i hele utviklingsprosessen.*

## 2.4 Vurdering av kriterier for både tekniske systemer og sluttbruker

Ved valg av data modeller er det kriterier for det tekniske systemet som primært blir vurdert først. Deretter gjøres en vurdering av sluttbrukers kriterier. Det ble diskutert i intervjuene at det tekniske systemets kriterier omhandler blant annet:

- Tilgang på kunnskap og empiriske data om operasjonen.
- Tilgang til og god kvalitet på statistikken som kan brukes til å bygge og trene datamodellen.
- Overførbarhet av datamodellen. Dette går ut på å undersøke behovet for, og mulighetene for, å bruke datamodellen også i andre operasjoner. For eksempel i andre brønner.

Videre diskuteres det i intervjuene at først etter at kriteriene for det tekniske systemet er vurdert og bestemt, blir sluttbrukers behov til valg av datamodell vurdert. Men det er heller ikke alltid at denne vurderingen gjøres. Sluttbrukers kriterier, ifølge informantene, er blant annet:

- Sluttbrukers behov for at datamodellen er forklarbar.
- Sluttbrukers behov for integrering av datamodell inn i HMI.
- Sluttbrukers behov for tillit til datamodellen.
- Hvordan datamodell og sluttbruker skal allokere funksjoner mellom seg.
- Hvilke kompetansekrav må sluttbruker ha for å utnytte datamodellen.

Selv om listene med kriterier ikke er fullstendig, så er poenget likevel at sluttbrukers kriterier må inkluderes og tydeliggjøres når datamodeller som skal brukes i sikkerhetskritiske operasjoner skal velges, utvikles og implementeres. Det kom frem i intervjuene at sluttbrukers kriterier var vanskelig å vurdere, fordi sluttbruker ikke alltid kunne forutsi, eller kjente til helt konkret, hvordan operasjonen ble utført med en ny og innovativ datamodell.

I dag tyder det på at kriterier om krav, tilgjengelighet og sluttbrukerkompetanse ikke vurderes ved valg av datamodeller. Dette kan mulig føre til en «vi tar det vi får»-holdning etter at design og utvikling er igangsatt. Tidlig spesifisering av sluttbrukers kriterier kan også bidra til at mindre selskaper/underleverandører lettere kan legge press på kunde om å stille med riktig kompetanse tidlig i utvikling og implementering av kognitive teknologi.

*Funn 3: Sluttbrukers kriterier inkluderes for lite ved valg av datamodell. Praksis og erfaring viser at det er vanskelig å inkludere sluttbrukers kriterier i valg av datamodell på grunn av usikkerhet knyttet til hvordan operasjoner vil bli utført etter introduksjonen av kognitiv teknologi.*

### **3 UTVIKLING OG IMPLEMENTERING AV KOGNITIV TEKNOLOGI**

I dette avsnittet har vi sett på kognitiv teknologi fra ulike perspektiv. Først presenteres en nyansering av det sosiotekniske perspektiv som rammeverk for å skille mellom de ulike typene aktører i systemet. Deretter diskuteres funn fra litteratur- og intervjustudien knyttet til utvikling og implementering av kognitiv teknologi.

#### **3.1 Rammeverk for arbeid med kognitiv teknologi**

En sikkerhetskritisk operasjon består av et system med flere aktører, der konsekvensen av at systemet feiler kan være kritisk. Et sosioteknisk rammeverk kan benyttes for å navngi ulike typer aktører i kognitive systemer. At et kognitiv system faller innunder et sosioteknisk rammeverk betyr at systemet består av både tekniske- og av menneskelige aktører som er gjensidig avhengig av hverandre. I praksis har det vært et tydelig skille mellom hvilke funksjoner som hører til de tekniske aktørene og hvilke som hører til mennesket. Dette skillet er på vei til å viskes ut etter hvert som de tekniske aktørene blir mer kognitive, og behovet for å ytterligere nyansere det sosiotekniske rammeverket økes.

Den økende bruken av ML-modeller gjør at funksjoner som tidligere kun tilfalt enten tekniske aktører eller mennesket, nå dynamisk flyter mellom aktørene i systemet. Et eksempel kan være under retningsboringsoperasjoner, der ML-modeller bidrar til å bestemme hvordan borekronen skal nå målet, litt som en autopilot. Dette krever en kontinuerlig veksling mellom datamaskinen som styrer borestrengen (teknisk kognitiv aktør) og retningsboreren (menneskelig kognitiv aktør) som overvåker og griper inn dersom det er nødvendig. Ved å selvstendiggjøre «cyber» i det sosiotekniske systemperspektivet, altså datamodellen, vil diskusjonen om hvordan og hvorfor en datamodell skal utvikles få mer oppmerksomhet.

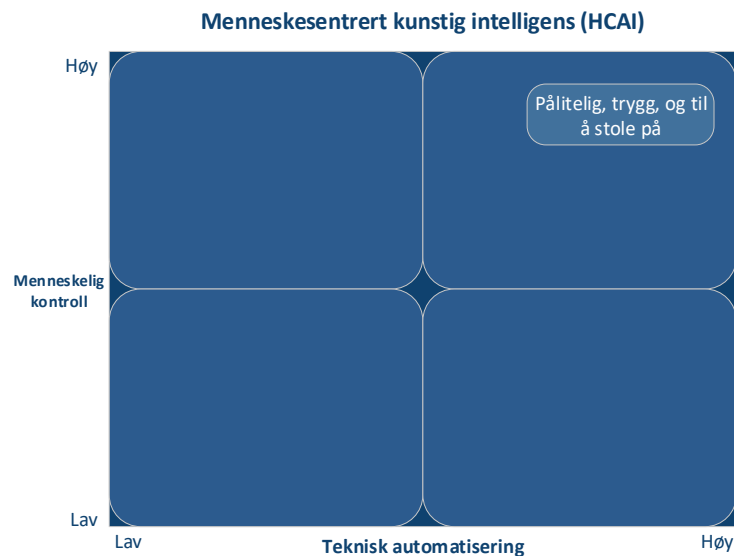
Et forslag fra forskningslitteraturen som kan bidra til begrepsavklaring er å trekke ut «cyber»-delen for seg selv slik at software- og datamodellering får dedikert oppmerksomhet, og følgelig kalle rammeverket for «cyber-sosioteknisk system» /19/.

#### **Kognitiv teknologi i et HCAI rammeverk for systemdesign**

Menneskesentrert kunstig intelligens («Human-centered artificial intelligence», HCAI) er et rammeverk som kan brukes til systemdesign. Sluttbrukers meninger om hvilken datamodell som kan brukes i operasjonen bør komme frem i design, noe som er viktig for at det kognitive systemet skal være pålitelig, trygt og til å stole på («Reliably, Safe and Trustworthy», RST) /20/. Sluttbrukers krav er mindre håndfaste enn mange tekniske krav, og prioriteres derfor ofte i andre rekke. Likevel har sluttbrukers krav direkte implikasjoner på hvor effektiv den kognitive teknologien (hvor datamodellen er en sentral del) er i praksis.

Shneiderman anbefaler HCAI-rammeverket til bruk også i sikkerhetskritiske operasjoner /21/. Dette rammeverket er anbefalt for å utvikle RST-systemer, og utfordrer det tradisjonelle synet på at økt automatisering bør føre til mindre menneskelig kontroll. Det trekkes et eksempel til den profilerte ulykken der to Boeing 737 MAX fly havarerte, hvor automatisering var et sentralt tema. Et sentralt poeng i HCAI er å sikre både høy grad av menneskelig kontroll samtidig med en høy grad av automatisering og at utviklere ikke må velge mellom å gi kontroll til tekniske- eller til menneskelige aktører. Viktige temaer for å fremme RST-systemer er gjennomsiktighet, tillit og integrering med sluttbruker (som diskuteres ytterligere i kapittel 4). Et eksempel på RST-systemer er hvitevarer som vaskemaskiner og oppvaskmaskiner. Slike maskiner tillater at sluttbrukeren velger programmer som spesifiserer hva sluttbruker ønsker. Når sluttbruker har valgt ønsket program gir sluttbruker deretter kontrollen til sensorer og tidtakere som styrer prosessen. Dersom vaskemaskinen er designet på en

god måte har den et forutsigbart og forståelig brukergrensesnitt. Dette kan for eksempel være at maskinen kan stoppes av menneskelig kontroll mens den er i gang.



Figur 3-1 HCAI som viser hvordan systemdesign bør fokusere både på høy grad av menneskelig kontroll (via f.eks. gjennomslittige modeller, sluttbrukerintegrering) og teknisk automatisering. Lav-lav referer til systemer med lav grad av automatisering og menneskelig kontroll, som f.eks. landminer eller spilledåser. Lav automatisering og høy menneskelig kontroll er f.eks. sykling eller pianospilling. Høy automatisering og lav menneskelig kontroll er f.eks. pacemaker eller airbag i bil. Høy automatisering, høy grad av kontroll er f.eks. heiser eller kameraer, ref./21/.

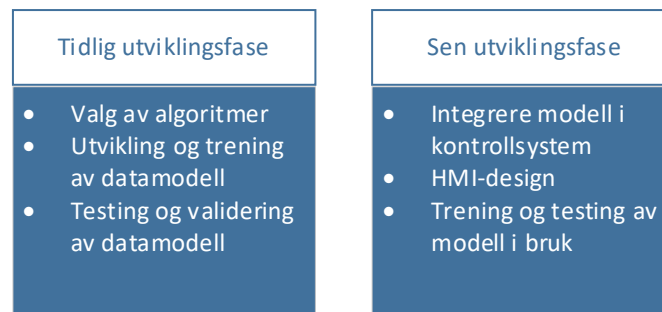
Rammeverket setter mennesket i fokus og fremmer kriteriene som sluttbrukere har til valg, utvikling og trening av datamodell, i tillegg til kriterier for det tekniske systemet. I kvadrantene i figuren over sammenlignes datamodellens kriterier og menneskets kriterier for å optimalisere systemdesign for effektive og trygge operasjoner (se Figur 3-1) /21/. I den nederste kvadranten til høyre kreves det rask handling at teknisk automatisering. Dette kan for eksempel være utløsning av kollisjonsputer i biler, pacemakere eller defensive våpensystemer. Den øverste kvadranten til venstre krever høy menneskelig kontroll, men lav teknisk automatisering. Dette er for eksempel sykling, pianospilling eller baking. Eksempler på enheter som tilhører i den nederste kvadranten er klokker, musefeller og dødelige enheter som landminer. I slike systemer er det lav menneskelig kontroll og lav teknisk automatisering.

*Funn 4: En tilnærming basert på et sosioteknisk rammeverk bør ytterligere nyanseres slik at cyberdelen i systemet, som inkluderer utvikling av datamodell og software, får økt fokus i arbeid med kognitive teknologier. HCAI-rammeverket kan legges til grunn for å fremme brukersentrert utvikling av kognitive teknologier. Særlig i systemer der det er både høy grad av automatisering og menneskelig kontroll.*

### 3.2 Involvering av sluttbruker

Når kognitiv teknologi skal brukes i sikkerhetskritiske operasjoner er det viktig at sluttbruker er integrert i hele utviklingsfasen. For kognitive teknologier innebærer dette at sluttbruker bidrar både i valg og trening av datamodell (her kalt tidlig utviklingsfase) og implementering av datamodell inn i kontrollsystem/HMI (her kalt sen utviklingsfase). Vi har forenklet utviklingsfasen i tidlig og sen fase for å strukturere og problematisere involveringen av sluttbruker (se Figur 3-2).





Figur 3-2 Et forenklet skille mellom arbeid i tidlig og sen utviklingsfase slik det er benyttet i denne rapporten.

### Involvering av sluttbruker i tidlig utviklingsfase

Tidlig utviklingsfase innebærer utvikling av datamodell. Arbeidsoppgaver i tidlig utviklingsfase inkluderer rydding og kategorisering av tilgjengelig data, valg og oppsett av ML-algoritmer, samt trening og utvikling av datamodellen. I denne fasen inkluderes eventuell «feature engineering», som innebærer ytterligere kategorisering av trekk som man finner i datasettet basert på kunnskap om operasjonen.

I tidlig utviklingsfase kom det frem i intervjustudien at bruk av HF-metoder for brukersentrert design i stor grad ikke benyttet. Det var eksempler på at sluttbruker involveres i deler av den tidlige utviklingsfasen, som f.eks. å gi navn på datamodellens output (labeling), men at denne involveringen er usystematisk. Brukerinvolveringen ble ikke knyttet til noe konkret metodikk, standarder, guidelines eller regelverk. Det ble fortalt at krav til involvering av sluttbruker av og til settes av bestiller, men at dette i praksis ikke ble fulgt opp. Fra bestiller sitt perspektiv ble dette forklart med at sluttbrukerkompetanse er en knapp ressurs som gjerne blir prioritert først i senere designfaser, som f.eks. implementering av etablerte datamodeller inn i kontrollsystemet/HMI. Det ble også problematisert i intervjuene hvorvidt det eksisterte gode HF-metoder til bruk i tidlig utviklingsfase.

En konkret utfordring fra intervjuene som ble diskutert går på å la sluttbruker definere eksemplene som skal brukes til trening av datamodellen. Med andre ord, hvordan sikre at sluttbruker vet hvordan han eller hun kan fortelle datamaskinen hva som er- og ikke er «en katt». Andre viktige områder der sluttbruker kan og bør bidra som ble nevnt, er definering av testprosedyrer og oppsett av treningsdatasettet. I tillegg bør sluttbruker si noe om kontekst til der datamodellen skal brukes, f.eks. områdets layout.

De fleste selskapene rapporterer å bruke workshops eller tilsvarende møter om sluttbrukere er med i utviklingsprosessen. Workshopene varierer i omfang, gjerne fra et par timer til noen dager. I tillegg varierer hyppigheten fra at det skjer kun én gang til at det er gjentakende møter, f.eks. hver tredje måned. Dersom workshop med sluttbrukere er en del av prosjektet, så avhenger hyppighet og workshopens lengde av kompleksiteten til datamodellen som utvikles, tilgang på sluttbrukere, budsjett, og størrelse på prosjektet. Det ble generelt forklart at bestiller av kognitiv teknologi hovedsakelig stoler på at leverandør gjør de testene som er tilstrekkelig for at teknologien er i henhold til krav.

I intervjuene ble det diskutert om brukermedvirkning i tidlig utviklingsfase. Flere av informantene var usikre på hvordan sluttbrukere skulle være med i denne fasen, og om de i det hele tatt skulle være med. Fra intervjuene er det identifisert fem argumenter til hvorfor sluttbruker ikke bør være med i tidlig utviklingsfase:

- 1) Dette er en fase hvor utvikling av algoritmer er av en såpass høy vanskelighetsgrad at sluttbruker ikke har kompetansen som kreves.
- 2) Sluttbruker kan øke faren for feilprogrammering og menneskelige feil, ved å feilinformere de som utvikler matematiske formler på grunn av manglende forståelse for hva oppgaven egentlig går ut på.
- 3) Det er ikke utviklet human factors-metoder som er egnet til bruk for samhandling mellom utvikler og sluttbruker i denne fasen.
- 4) Det er knapphet på sluttbrukere hvilket gjør at underleverandører ikke har tilgang på sluttbrukere til sin del av utviklingen. Deltakelse av sluttbrukere prioriteres ikke i denne fasen, men heller i senere faser i utviklingen.
- 5) I et fremtidsscenario ble det fremhevet at sluttbrukere for de teknologiene som utvikles ikke finnes ennå.

Vi ble fortalt at tidlig utvikling av datamodell gjerne blir outsourcet til underleverandører med lite eller ingen erfaring og kjennskap til offshore operasjoner. Flere bruker arbeidskraft fra utlandet for å få ned kostnader. I tillegg ble vi gjort kjent med at flere informanter mente det ikke var krav til å ta med sluttbruker i denne fasen. Det ble påpekt at de som bestiller teknologi ikke setter krav til underleverandører/leverandører om involvering av sluttbruker. Fravær av krav, standarder og spesifikasjoner om sluttbrukerinvolvering fra kunde, kan bidra i forklaring om hvorfor utviklingsteam i denne fasen ofte står uten deltakelse av sluttbrukere.

Forskningslitteraturens perspektiv på involvering av sluttbruker presiserer at tilnærminger til brukerinvolvering må tilpasses sluttbrukers kompetanse i datavitenskap og dataingeniørens tverrfaglige kompetanse. Denne vurderingen er viktig for å sikre at tilnærmingen til brukerinvolvering tilpasses hvor mye sluttbruker forstår av datamodellering (f.eks. styrker og svakheter med ulike ML-algoritmer) og hvor mye dataingeniøren forstår av fagfeltet og operasjonen som skal modelleres. Riktig tilnærming til brukerinvolvering er viktig for at alle i utviklingsteamet skal bidra med sine synspunkter.

I lys av dette kan det være behov for å oppdatere verktøykassen av metoder for brukerinvolvering til å reflektere utfordringene knyttet til å jobbe med tverrfaglig spisskompetanse i tidlig utviklingsfase av ML-baserte datamodeller. To dominerende tilnærminger i forskningslitteraturen med fokus på dette er AutoML (når sluttbruker har kompetanse i datavitenskap) og interaktiv maskinlæring (når sluttbruker ikke har kompetanse i datavitenskap), der sistnevnte ikke er funnet benyttet og validert for høyrisikoindustri og førstnevnte krever at sluttbruker har datavitenskapelig kompetanse.

### **AutoML og interaktiv maskinlæring (IML)**

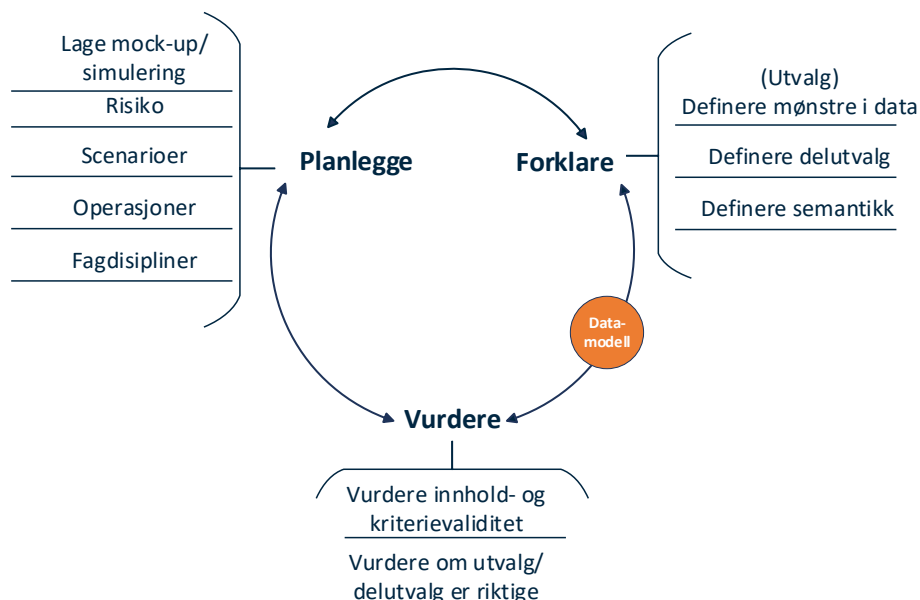
Flere forsøk er gjort på å gjøre bruken av ML-algoritmer tilgjengelig for de som ikke er dataingeniører, og AutoML er en slik plattform. AutoML en samlebetegnelse for ulike plattformer som automatiserer flere deler av definering og design av ML-baserte modeller. Fordelen med AutoML er at sluttbruker kun må fokusere på å definere outputs (labels) til modellen, og derfor ikke må tenke på avanserte problemstillinger som valg av algoritme, parametre/hyperparametre og features (inputs) til modellens beregninger. En ulempe med denne tilnærmingen er at datamodellene blir lite gjennomsiktlige for brukere og andre utviklere av systemet, inkludert sluttbrukere.

En alternativ tilnærming til design av modeller er «interaktiv maskinlæring (IML)». Det er en menneskesentrert tilnærming som har til hensikt å fremme menneskelig kognisjon og persepsjon i designprosessen fra start. Denne tilnærmingen forsøker å presisere ansvarsforholdet fra «machine learning» til «machine *teaching*», slik at sluttbruker (eller fagekspert) blir en aktiv part i å utvikle datamodellen. Hensikten med IML er å forankre sluttbruker til datamodellen på en annen måte enn

om den serveres til sluttbruker gjennom treningsprogram etter at den er utviklet. Gutzwiller og Reeder /22/ fant at IML som designmetode gir mer tillit til ML-modeller og at ML-modeller kjennes igjen.

IML er en iterativ prosesslignende til design og består av 1) planlegging, 2) forklaring, og 3) vurdere, slik beskrevet i Ramos, et al. /23/:

1. Planlegging innebærer at en fagekspert identifiserer og vurderer hva som behøves for å gjennomføre treningen av datamodellen. Dersom modellen skal brukes til å klassifisere status i et farlig område, kan fageksperten komme opp med nødvendige beskrivelser av miljøet som modellen skal fungere i. Dette kan være faktorer som layout og lys, eller mer komplekse trekk som interaksjon med andre kognitive teknologier eller mennesker. Her må det defineres en bred og gyldig liste med eksempler som er relevant til bruksområdet.
2. Forklaring innebærer at fageksperten gir sin eksplisitte fagkompetanse til modellen. Denne oppgaven består av tre deler. Den første er å beskrive for ML-algoritmen hva som må gjøres i spesifikke scenarier – f.eks. klassifisere en ventil som åpen eller igjen. Denne prosessen kalles for «å merke ulike mønstre i datasettet (labeling)». Den andre er å gi mening til mønstrene i datasettet ved å bruke sin kjennskap om sammenhenger og avhengigheter i systemet ML-algoritmen skal fungere i. Den tredje delen er å etablere et felles språk for læring som er til for å gjøre kommunikasjonen presis og effektiv, som også bør inkludere mentalisering, altså en forståelse for at andre personer har sine egne versjoner av sannheten som kan avvike fra ens egen. En viktig del av å definere dette språket og konsepter er å sikre at egenskaper og merkelapper ikke er tvetydige eller overlapper hverandre, og istedenfor bruke en «usikker»-lapp for tilfellene der det er ikke-distinkte mønstre i datasettet.
3. Vurdere algoritmens prestasjon. Dette steget innebærer å vurdere om prediksjonene er riktige, kvaliteten av prediksjonene, identifisere problematiske eksempler, samt identifisere feil eller ufullstendige merkelapper som er tidligere gitt til ML-algoritmen. For at fagekspert skal kunne gjøre sin vurdering er det viktig med et felles språk og begrepsapparat. Vurdering av datamodellen er også sentralt for at sluttbrukeren skal ha tillit til at systemet fungerer.



Etablere et felles språk for å utvikle og implementere datamodell

Figur 3-3 Visualisering av IML-prosessen, adaptert fra Ramos et al. /23/.

Innenfor IML er det tydelige avgrensninger mellom hva en fagekspert skal bidra med og hva dataingeniøren skal bidra med i design av modellen. Denne avgrensningen innebærer blant annet at fageksperten ikke trenger å tenke på å manipulere parameterne eller annen optimalisering av algoritmen og arkitekturen. Det gir mulighet til å fokusere på en brukersentrert designprosess, der fagekspertene kan rette oppmerksomheten på sitt eget fag. Det er gjort mye forskning og utvikling på hvordan sluttbruker uten datakompetanse kan medvirke i bruken av ML-algoritmer.

For eksempel, at sluttbruker medvirker i utviklingsprosessen gir fordeler for det endelige resultatet. Dietvorst et al. /24/ fant i sin studie at enkle feil som datamodellen gjør i trening og operasjon fort bryter tilliten mellom bruker og kognitiv teknologi. På den andre siden, dersom samme feil som er gjort av den samme modellen under designutvikling, ser ikke dette ut til å ha den samme negative effekten på tillit. Ved å benytte utviklingsprosesser som IML, hvor sluttbruker er med som designer, er det rimelig å anta at det vil gi økt tillit til modellen for sluttbruker.

Lignende effekter er reproduisert i senere eksperimenter. Gutzwiller og Reeder /25/, fant i sitt eksperiment at å utvikle datamodeller med bruk av IML-prosessen fører til mer identifiserbare, troverdige modeller og foretrukket atferd grunnet erfaring og kjennskap til modellen. Resonnementet er at analyser og beslutninger iverksatt eller foreslått av modellen vil være gjenkjennbar for sluttbruker dersom det har vært interaktiv i utviklingen av denne, slik IML legger til rette for. Det ble imidlertid ikke diskutert hvorvidt denne gjenkjennbarheten også gjelder for sluttbrukere som ikke var involvert i utviklingen.

*Funn 5: Det er behov for metoder og tilnærminger for involvering av sluttbruker i tidlig utviklingsfase av datamodeller til bruk i sikkerhetskritiske operasjoner. AutoML og interaktiv maskinlæring (IML) er to tilnærminger til å involvere sluttbruker og som kan fungere som et utgangspunkt til videre arbeid.*

### **Involvering av sluttbrukere i sen utviklingsfase**

I denne fasen skal datamodellen testes og implementeres i det øvrige kontrollsystemet. Dette arbeidet inkluderer integrering med HMI, symbolbruk, og generelt arbeid med datamodellens interaksjon med sluttbruker.

Behov for sluttbrukers kompetanse i datavitenskap er et moment som er forskjellig i tidlig- og sen utviklingsfase. Der den sistnevnte fasen ikke stiller de samme kravene til datavitenskapelig kompetanse hos sluttbruker ettersom datamodellen da allerede er utviklet. Det antas å gi en merverdi i denne fasen dersom sluttbruker har vært involvert også i tidlig utviklingsfase.

Prinsippet om brukermedvirkning står allerede sentralt for sen utviklingsfase i design av menneske-maskingrensesnitt (HMI) i petroleumsindustrien, både i forskning og praksis. Dette prinsippet trekkes tydelig frem i standarder for ergonomisk design av kontrollsentre, ISO 11064 /25/ og NORSOK S-002. Det er flere ulike HF-metoder som benyttes for brukersentrert design som er verifisert, validert og godt kjente innenfor petroleumsindustrien. Eksempler på slike metoder er: funksjons- og oppgaveanalyse, arbeidsbelastningsanalyser, STEP-analyser, og kvantitative eller kvalitative menneskelig pålitelighetsstudier (HRA). Det samme bekjentskapet til HF-metoder for tidlig utviklingsfase ble ikke funnet blant informantene.

I intervjuene ble det fortalt at brukerinvolvering øker jo lengre ut i utviklings- og implementeringsprosessen man kommer. Det ble fortalt at føringer for brukerinvolvering er hovedsakelig regelverk, standarder, samt krav satt av bestiller. For sistnevnte erfarte informantene at det er variasjoner mellom ulike bestillere, og at disse er avhengig av hvert selskaps egne prosesser for utvikling og implementering av ny teknologi. Hos noen selskaper ble det fortalt at det er satt strenge

krav til å sikre gode og robuste prosesser for utvikling av kognitiv teknologi, men at fokuset på krav om brukerinvolvering oftest blir ikke fremmes før sen utviklingsfase. Generelt sett erfarer næringen at det er flere og mer konkrete krav til brukerinvolvering i sen utviklingsfase, noe som påvirker prioritering av sluttbruker i sen fase.

Valget om å først ta med sluttbruker i sen utviklingsfase ble forklart med at tilgangen til sluttbrukere er begrenset og at de større selskapene som har sluttbrukere in-house har lettere tilgang på sluttbrukerkompetanse. Mindre underleverandører har utfordringer med å få tilgang på den nødvendige kompetansen de trenger for å trene og utvikle sine datamodeller, spesielt når de må konkurrere med større selskaper om denne.

*Funn 6: Involvering av sluttbruker fungerer i sen utviklingsfase. Det ble understreket at sluttbrukere som en knapp ressurs prioriteres i sen utviklingsfase av selskapene. Blant annet grunnet mer konkrete krav satt av regelverk og standarder og usikkerheter knyttet til metoder for brukerinvolvering i tidlig fase.*

### **Trening og bruk etter implementering**

Sluttbruker får teste teknologien som er utviklet i simulatorer og komme med tilbakemeldinger til både form og funksjon. Tilbakemeldingene sluttbruker kan komme med begrenses av prosjektets fase og med hensyn til tid og kost. Tilbakemeldinger handler som regel om tilpasninger i brukergrensesnitt. I intervjuene ble følgende elementene trukket frem som sentrale i sen utviklingsfase:

- Sen utviklingsfase er ofte første gang sluttbruker får medvirke i utviklingen av datamodellen.
- Sluttbruker har begrenset begrepsapparat til å prate om og gi tilbakemeldinger om teknologien og brukergrensesnittet som modellen integreres i.
- Sluttbruker har ikke en godt utviklet mental modell av hvordan datamodellen fungerer, dette påvirker sluttbrukers forståelse av hva han eller hun kan gi tilbakemeldinger om i denne fasen.
- Involvering av sluttbruker i sen utviklingsfase gjøres i henhold til hvordan standarder som NORSOK S-002 og ISO 11064 blir praktisert.

Etter at teknologien er installert offshore kommer sluttbrukere med tilbakemeldinger om bruk og forbedring av teknologien. De fleste informantene påpekte at dette anses å være en kontinuerlig forbedringsfase hvor sluttbruker hele tiden har mulighet til å komme med innspill i hvordan teknologien kan forbedres og at dette er noe det tas hensyn til. Produkter vedlikeholdes og utvikles i tråd med tilbakemeldingene som kommer fra driftsorganisasjonen. Dette inngår som regel i en serviceavtale som man betaler ekstra for. Større selskaper opplever denne fasen som viktig, og at tilbakemeldingene fra sluttbruker er sentrale i videre produktutvikling.

*Funn 7: Tilbakemeldinger fra sluttbrukere i sen utviklingsfase er viktig for videre oppdatering og utvikling av modellen. Tilbakemeldingene baseres på trening med kognitiv teknologi i simulatorer og bruk i drift etter at teknologien er implementert offshore.*

## **4 SAMHANDLING MENNESKE OG KOGNITIV TEKNOLOGI**

I dette kapittelet diskuteres viktige faktorer for sikker og effektiv samhandling mellom menneske og kognitiv teknologi. Faktorene som vi trekker frem er mentale modeller, gjennomsliktige datamodeller, tillit til teknologi, og funksjonsallokering. Selv om faktorene diskuteres hver for seg, så henger de sammen i et kognitivt system bestående av både menneskelige- og tekniske kognitive aktører. Faktorene må derfor vurderes både individuelt og sammen for sikre effektiv samhandling.

### **4.1 Mentale modeller viktig for samhandling**

Det er flere viktige komponenter som er viktig for et godt samarbeid. I litteraturen beskrives «mentale modeller» som særlig relevant, også for samhandling mellom menneskelige og tekniske aktører. En mental modell er en person eller systems forståelse av hvordan et objekt, maskin, teknologi eller fenomen fungerer. I samarbeid er det viktig å ha en tilstrekkelig felles forståelse av systemet, altså en delt mental modell.

Forståelsen av hvordan et objekt, maskin, teknologi eller et fenomenet fungerer inkluderer hvordan hver aktør i samarbeidet fungerer. Det innebærer å kjenne til hverandres styrker og svakheter. Sluttbruker bør kjenne til grensene til kognitiv teknologi. Kognitive teknologi bør kjenne til og tilpasse seg menneskets grenser, slik at mennesket og datamodellen gjensidig tilpasser og styrker hverandre i operasjonen.

Mennesker er (fortsett) den mest tilpasningsdyktige aktøren i dette samspillet og er derfor ofte den som må tilpasse seg. Samtidig er det et økt oppmerksomhet på at teknologien skal tilpasse seg mennesket. Et enkelt eksempel på en slik tilpasning er når HMI automatisk endrer farge og lysstyrke etter tiden på døgnet. En mer avansert variant er at teknologien overtar mer styring når arbeidspresset øker. Utfordringene i dag er at det motsatt gjerne skjer: når arbeidspresset øker – eller er uvanlig – så er systemets grenser nådd og mennesket må ta over. Det er komplisert å utvikle et teknisk system som er allsidig og kan overta når operasjonen er unormal, blant annet grunnet fravær av statistikk på unormale situasjoner som kan brukes til å trene datamodellen.

I dag fungerer teknologi best som lagspiller når mennesket kan overta ansvaret for operasjonen i god tid og med god situasjonsforståelse. Denne overtakelsen av ansvar krever at mennesket har en god mental modell av hvordan teknologien fungerer. Dette er enklere å oppnå når datamodellene som benyttes er enkle å tolke (det er en tydelig sammenheng mellom input og output). Med introduksjonen ML-baserte datamodeller er det vanskeligere for sluttbruker å få en god mental modell. Dette gjør at det må settes nye krav til datamodellens gjennomsliktighet.

*Funn 8: Sluttbrukers forståelse av hvordan kognitiv teknologi fungerer er viktig for samhandling. Viktige faktorer for å danne seg en god mental modell er trening og bruk, i tillegg til at datamodellen er gjennomsliktig.*

### **4.2 Gjennomsliktige datamodeller**

Det er flere dimensjoner av gjennomsliktighet. Viktige dimensjoner er forklarbarhet, tolkbarhet, reproducerbarhet, sporbarhet samt generelle drøftelser knyttet til etikk, tillit og rettferdighet. Hvilke av disse dimensjonene som er mest relevant for en gitt datamodell må vurderes mot tiltenkt bruk. Det er kommet forslag fra EU til juridiske krav som bør settes til datamodeller, og det er mye forskning på hvilke standarder og veiledninger som kan og bør brukes. Se Tabell 4-1 under for en liste og beskrivelse av de ulike dimensjonene.

Tabell 4-1 liste over dimensjoner som må vurderes for gjennomsiktighet og beskrivelse.

Dimensjon	Beskrivelse
Forklarbarhet	<p>Dersom datamodellen består av funksjoner som er for kompliserte for mennesket å forstå trenes verktøy, metoder, teknikker for å forstå datamodellen (<i>a posteriori</i> forklaring) /26/. Slike modeller kalles Black Box-modeller /27/.</p> <p>Det er et omfattende og raskt voksende fagfelt å utvikle slike verktøy. Eksempler er SHAP (Shapley additive explanations), LIME (Local Interpretable Model-agnostic Explanations), og det er flere andre.</p>
Tolkbarhet	<p>Tolkbarhet er en passiv karakteristikk ved datamodellen (den forklares <i>a priori</i>). For at en modell skal være tolkbar må det være mulig å forstå sammenhengen mellom inputvariablene og outputvariablene som datamodellen foreslår. Tolkbare modeller har tydelige årsakssammenhenger mellom input og output, noe som er viktig for at sluttbruker skal stole på modellen /26/.</p>
Reproduserbarhet	<p>Datamodellen gjengir (omtrent) de samme resultatene når den brukes på samme datasett.</p>
Sporbarhet	<p>Det bør være mulig å få informasjon om datamodellens historie, inkludert når den ble utviklet og av hvem. Det bør også være mulig å si hvordan den er trent, hvilken kontekst som er benyttet, samt hvilke grenser og erfaringer som ble gjort med hensyn til feil, akseptkrav og modellens generelle ytelse.</p>
Etikk, tillit og rettferdighet	<p>Generelle drøftelser knyttet til etikk, tillit og rettferdighet kan være relevante. Stereotypier og bias kan representeres i datamodellen.</p>

I tillegg til teoretisk gjennomsiktighet av datamodell må utviklere og sluttbrukere forstå hvordan modellen fungerer og konteksten som den opererer i. Det kan være mange forskjellige typer brukere, ofte på forskjellige tidspunkter i utviklingen av systemet og når systemet skal brukes /28/. For eksempel kan brukere av samme system være ulike kombinasjoner av flere roller med ulike mandat og bruksformål, som analytikere, myndigheter og eller operatører. Det kan også være andre brukere som har behov for å forstå systemet, som utviklere eller testoperatører som trenger å forstå hvor i modellen eller systemet det er mulig eller nødvendig med forbedring.. Hver brukergruppe kan ha egne preferanser for forklaringstype for hvordan informasjon kommuniseres på mest effektiv måte for deres formål og bruk.

En effektiv tilnærming til å forstå datamodellen er å vurdere den med et perspektiv fra sluttbrukere. Dette inkluderer en vurdering av sluttbrukers bakgrunnskunnskap, behovet for forklarbarhet og hvordan modellen kan og bør forklares /29/.

I intervjuene ble det presisert at sluttbrukere i sikkerhetskritiske operasjoner må forstå hvorfor datamodellen kommer med et gitt forslag, noe som krever gjennomsiktighet. Sluttbruker må forstå hvorfor, ettersom en datamodell kan komme med en korrekt prediksjon, men er utydelige på hvilken årsak som bidrar til prediksjonen. Når modellen er utydelig på sammenhengen mellom inputvariablene og output blir det vanskelig å diagnostisere feilen. Tre momenter ble diskutert:

- Det ble diskutert at ikke alle situasjoner krever fullstendig forklarbarhet, og at man må gå i dybden av dette i hvert enkelt tilfelle. Begrepet «forklarbarhet» ble forstått ulikt blant informantene. Forskjellene var ikke knyttet til selskap, bransje eller fag.

- I intervjuene ble det gjort et skille mellom operasjonell og teoretisk forklarbarhet. Operasjonell forklarbarhet kan forstås som at en datamodell kan være vanskelig å forklare i operasjonen, spesielt med lite tid, selv om det er mulig å diagnostisere hvorfor datamodellen kom frem til sin prediksjon med bedre tid til analyser. Det ble nevnt at selv komplekse deterministiske fysikkbaserte datamodeller kan oppleves å ha lav grad av forklarbarhet i operasjonen, selv om disse er mulig å diagnostisere og forklare i etterkant.
- I samtalene med næringen kom det frem at forklarbarhet ikke har vært et problem for ML-modeller, slik de brukes i dag, ettersom input til og vektingen av datamodellens parametere er lett tilgjengelig (ML-modellene er tolkbare). Det er et generelt prinsipp om at slike modeller er å foretrekke der hvor det er mulig.
- Om modeller som ikke er forklarbare må brukes, så ble det diskutert at bestiller av datamodellen bør presisere (og legge inn i budsjett) å få gjennomført «feature engineering» sammen med sluttbruker for å sikre at inputverdiene blir meningsfulle. Dersom datamodellen bruker flere lag med analyser (f.eks. nevrale nettverk), bør det benyttes verktøy for å øke forklarbarheten.
- Ikke alle datamodeller som er mulige å forklare i dag. Forskning på forklarbarhet er i rask utvikling, og nye verktøy for å forklare black box-modeller lages. Etter hvert som nye verktøy kommer til, så kan det bli mulig å bruke mer komplekse datamodeller også i sikkerhetskritiske operasjoner.

Forklarbarhet og tolkbarhet er de vanligste temaene knyttet til en modells gjennomskiktighet og er det som ble diskutert mest i intervjuene. I tillegg fremmer forskningslitteraturen at vurderinger knyttet til reproduserbarhet, sporbarhet og etiske vurderinger også er viktige, slik beskrevet i 4.2. Alle de ulike dimensjonene av gjennomskiktighet bidrar til at sluttbruker har muligheten til å utvikle gode og robuste mentale modeller av teknologien, noe som er viktig for samhandlingen.

*Funn 9: Datamodeller som er tolkbare for sluttbruker er å foretrekke. Dersom black box-modeller må brukes bør det benyttes verktøy for å øke modellens tolkbarhet.*

### 4.3 Tillit til kognitiv teknologi

Tillit kan defineres ut fra tre hoveddimensjoner; velvilje, integritet og evne /30/. I sammenheng med teknologi baseres tillit på brukerens egne erfaringer med systemet eller fra andre kilder som sier noe om omdømmet til teknologien /30/, blant annet fordi tekniske systemer ikke har en egen indre intensjon. Tillit, litt forenklet, handler om troen på at «den delen av samhandling som ikke er gjennomskiktig», har en positiv velvilje, integritet og evne. Det vil si den delen som sluttbruker ikke har oversikt og kontroll på, må han eller hun ha tillit til at fungerer slik den skal.

Tillit er også en dynamisk prosess og endres over tid /31/. Det påpekes at det er operatørens oppfatning av systemets prestasjon over tid, og gapet mellom operatørens oppfatning og den faktiske prestasjonen til systemet, som er avgjørende for tillit. For kognitiv teknologi er det faktorer som datamodellen i seg selv, brukerens kompetanse og kognitive bias, og det tekniske systemets ansvarsområder som virker inn på sluttbrukers tillit til den kognitive teknologien.

Sluttbruker bør ha et riktig nivå av tillit til teknologien for å bruke den effektivt /30/; 32/, og at sluttbruker ikke stoler blindt på teknologien /10/. Et kritisk blikk på kognitiv teknologi gjør at sluttbruker kan tidlig oppdage utilsiktede konsekvenser. Likevel er det behov for å sikre at sluttbruker adopterer teknologien på en effektiv måte, det bør tilstrebes en balanse mellom tillit og det å ha et kritisk blikk på teknologien.



## Riktig tillit er viktig i sikkerhetskritiske operasjoner

Tillit som konsept er omdiskutert i næringen. I intervjuene var det tydelig konsensus i at tillit er viktig for samhandling mellom mennesker og teknologi. Derimot var det varierte meninger om hvordan man oppnår denne tilliten. To perspektiver fra intervjuene på hvordan man kan oppnå riktig tillit var:

1. Tillit gjennom gjentakende vellykket bruk. I noen av intervjuene var det fokus på at tillit oppnås gjennom at sluttbrukere får erfare at datamodellen fungerer over tid.
2. Tillit gjennom å øke forståelse. Andre pratet om at tillit økes ved å forbedre forståelse for hvordan datamodellen fungerer. En viktig del av at datamodellen skal fungere, er at sluttbrukere har trening og forståelse for hvordan den skal brukes, samt hva den ikke skal brukes til. Noen fremhevet derfor at trening og opplæring i bruk av datamodell som viktig for å skape tillit. Det innebærer at sluttbruker har oversikt over hvilke inputs datamodellen bruker, samt hvordan disse prosesseres.

Dette ble problematisert videre med at ML-baserte modeller ikke nødvendigvis gir en tydelig oversikt over sammenhenger mellom inputvariablene og output. I noen intervjuer ble det diskutert at å vite hvorfor en datamodell predikerer slik den gjør er et absolutt krav for å kunne bruke teknologien i sikkerhetskritiske operasjoner. Dersom det stilles et absolutt krav om å kunne forstå datamodellene, setter det strenge krav til datamodellens gjennomsiktighet i sikkerhetskritiske operasjoner.

Et krav om at sluttbruker skal forstå datamodellen kan også handle om at en menneskelig aktør skal ha kontroll i operasjonen, samtidig med at det tekniske systemet får en høyere grad av automatisering. Den menneskelige kontrollen i automatiserte systemer ble også fremmet som viktig i HCAI-rammeverket som er diskutert i Kapittel 3.1. For sikkerhetskritiske operasjoner, som i økende grad baseres på kognitiv teknologi, vil det være viktig å ta stilling til hvor mye av operasjonen som skal baseres på tillit og hvor mye som skal overvåkes og kontrolleres av mennesket.

*Funn 10: Å balansere riktig grad av tillit i sikkerhetskritiske operasjoner er viktig og krevende. En del av denne balansen handler om at den menneskelige aktøren må forstå sammenhengen mellom inputvariablene og output, som igjen setter føringer for datamodellens gjennomsiktighet.*

## 4.4 Funksjonsallokering

Når deler av systemet automatiseres må det gjøres en vurdering om hvilke oppgaver som skal allokeres til det tekniske systemet og når i operasjonen dette skal gjøres. For å gjøre denne vurderingen reises tre spørsmål knyttet til funksjonsallokering /33/: 1) Metode, hvordan skal funksjoner mellom mennesker og maskiner allokeres, 2) Ansvar, hvem skal gjøre allokeringen? og 3) Autoritet, hvem kan autorisere en allokering?

I litteraturen skilles det mellom to ulike kategorier av funksjonsallokering, statisk og adaptiv allokering. Statisk allokering er at mennesket og kognitiv teknologi gjør separate oppgaver. En slik allokering kan f.eks. enten være å la beste aktør i systemet ha ansvar for funksjonen eller la kognitiv teknologi ha ansvar for alt som kan automatiseres og mennesket har ansvar for resten. Statisk allokering er basert på et teknologisentrert design og tar ikke hensyn til at endringer som forekommer i løpet av operasjoner. En slik allokering kan være farlig fordi uønskede hendelser kan oppstå uten at mennesket er forberedt på å overta ansvaret for en funksjon som i utgangspunktet var allokert til teknologien.

Dette er bakgrunnen for utviklingen av adaptive funksjonsallokeringsmetoder. Disse metodene tar utgangspunkt i et menneskesentrert design. Et eksempel på en slik metode er «automation allocation logic (AAL)» /33/. I denne metoden fremmes tre ulike strategier for å styre funksjonsallokeringen:

- Kritisk hendelseslogikk: automatisering overfører en spesifikk funksjon til mennesket når det er en kritisk hendelse.
- Målingsbasert logikk: gjensidig overføring av funksjoner mellom maskin og kognitive teknologier basert på kontinuerlig måling av ulike faktorer, som for eksempel arbeidsbelastning.
- Modelleringsbasert logikk: predikere fremtidig arbeidsbelastning for mennesket som underlag for funksjonsallokering.

AAL-strategiene er tilpasningsdyktig og tar hensyn til at ansvaret mellom mennesker og kognitive teknologier endres i operasjonsforløpet. Selv om adaptiv funksjonsallokering tar hensyn til at ansvaret for funksjonen kan endres underveis, så er det likevel mangler ved en slik tilnærming. Den kan distrahere mennesker, forstyrre situasjonsbevisstheten og følgelig påvirke sikkerhet og hvordan oppgaven utføres. Feilmålinger kan også forstyrre oppgaven som mennesket er involvert i og kan føre til allokeringseil.

CoCyS (cognitive-cyber-symbiosis) er en videreutviklet teori av AAL som prøver å håndtere begrensninger gjennom bruk av funksjonsallokeringsagenter /33/. En funksjonsallokeringsagent omfordeler funksjoner mellom mennesker og kognitiv teknologi, og gir deretter en årsaksforklaring til hvorfor maskinen ønsker å gjøre en handling. CoCyS teorien er basert på tillit og gir en allokeringssagent handlingsrom til å omfordele funksjoner mellom mennesker og kognitive teknologier basert på tillitsindikatorer. Indikatorene bidrar til at funksjonsallokeringsagentene kan forvalte forholdet mellom mennesker og kognitive teknologier.

Tillitsindikatorer som allokeringssagentene analyserer er f.eks. hjerneaktivitet, muskelbevegelser, hjertefrekvens, hudtemperatur, øyesporing og stemmeleie, målt fra mennesket under observasjon. Allokeringssagenten lærer deretter av denne informasjonen, og bruker den for å identifisere når og hvordan funksjonsallokeringsstrategien bør tilpasses, f.eks. hvordan mennesket endrer seg som en funksjon av kontekst. Spørsmålet om hvilken aktør som skal ha myndighet til å ta beslutningen om å utføre allokeringssagentenes anbefalte fremgangsmåte gjenstår likevel. Hovedsakelig kan sluttbruker gjøre dette ved bruk av forhåndsdefinerte regler og veiledninger /34/. Sluttbruker kan dermed også vurdere en situasjon og velge en strategi for når de ønsker å skyve en funksjon til kognitiv teknologi eller motsatt.

Funksjonsallokering er et viktig tema, spesielt dersom kognitiv teknologi skal overta avanserte funksjoner som tradisjonelt er utført av menneskelige aktører. Det er gjort mye forskning på ulike måter å allokere funksjoner mellom menneske og kognitive teknologier.

### **Funksjonsallokering i praksis**

I intervjuene ble det diskutert at allokeringer må vurderes i sammenheng med hvilken funksjon som plasseres på mennesket. Typer funksjoner kan være a) manuell utførelse av en gitt funksjon, b) fjernstyring av en gitt funksjon, c) monitorering av automatiserte funksjoner og oppgaver eller d) beslutningstaking, der hvert ledd flytter den menneskelige aktøren lengre bort i fra den direkte utførelsen av oppgaven. Innenfor sikkerhetskritiske operasjoner, har tekniske systemer innenfor petroleumsindustrien allerede overtatt mange typer manuelt utførte oppgaver, i tillegg til en del fjernstyrings- og overvåkingsoppgaver.

I intervjuene kom det frem to viktige prosjekter som knyttes til hvordan kognitive teknologier flytter mennesket enda et ledd bort i fra manuell utførelse av oppgaven. I det ene prosjektet er kognitiv teknologi brukt til automatisk retningsboring. I dette tilfellet flyttes mennesket fra å fjernstyre retningsboringen fra drillers kontrollkabin, til kun å overvåke (og beslutte ved behov) hvordan

teknologien utfører retningsboringen, litt som en autopilot. I det andre prosjektet, som ikke kan detaljeres, tar kognitiv teknologi også over overvåkingsansvaret slik at den menneskelige aktøren kun tar beslutninger for teknologien ved behov. Det ble ikke identifisert noen prosjekter eller oppgaver knyttet til at kognitiv teknologi også overtar beslutningsansvaret, hverken i bruk eller i konseptfase.

I det første prosjektet vil retningsborer få mindre eksponering til å utføre boring direkte. Det vil bli nødvendig å sikre at kompetansen med å styre boret ivaretas over tid ettersom den naturlige trening og vedlikehold av ferdighet gjennom arbeidet ikke lenger er til stede. I det andre prosjektet vil det bli nødvendig å sikre at den menneskelige kognitive aktøren er integrert i operasjonen, slik at han eller hun kan ta gode beslutninger dersom det blir behov for det.

Begge prosjektene er eksempler der kognitiv teknologi tar over funksjoner som tidligere var utført av menneskelige aktører. Over tid kan det tenkes at tilgangen til sluttbrukere med ferdigheter og kompetanse knyttet til å utføre disse funksjonene manuelt reduseres, ettersom det blir færre situasjoner der disse ferdighetene benyttes i utføring av arbeid. Det er viktig å vurdere hvilke implikasjoner dette kan ha for sikkerhet og for design- og utvikling av slike systemer hvor brukermedvirkning er et viktig verktøy for effektivt og sikker design.

Det kom ikke frem noen konkrete funksjonsallokeringsmetoder brukt i disse prosjektene i intervjuene. Det var en generell usikkerhet knyttet til hvilke metoder og tilnærminger som skal og bør brukes for å allokere funksjoner sikkert og effektivt. Dette kan ha en sammenheng med at kognitive teknologier som benyttes i dag kun baseres på enkle tolkbare modeller slik at avanserte metoder for dynamisk funksjonsallokering ikke er relevante. Tre viktige spørsmål ble reist i intervjuene angående funksjonsallokering og fremtidig kognitiv teknologi med høyere automatisering i dag:

1. Hvordan sikre at sluttbruker (her: borer) ivaretar kompetansen over tid, ettersom han/hun ikke lengre holder spakene hele dagen, men istedenfor følger med på automatisert utstyr?
2. Hvordan sikre at mennesket beholder en reell beslutningskompetanse i situasjonen?
3. Hvilke implikasjoner vil funksjonsallokering ha over en lengre tid, med tanke på at sluttbrukerkompetansen som er viktig for utviklingen av kognitiv teknologi, kan viskes ut?

*Funn 11: Metoder og tilnærminger for funksjonsallokering må passe til den faktiske bruken av kognitiv teknologi. Det finnes mange metoder og tilnærminger til temaet som antar mer avansert teknologi enn det som brukes i praksis.*

## 5 PREMISSER FOR ARBEID MED KOGNITIV TEKNOLOGI

Et viktig spørsmål for Ptil er hvilke premisser som er viktige for arbeid med kognitiv teknologi. Blant annet ved å vurdere hva som kreves for å håndtere både mulighetene og utfordringene som kognitive teknologier byr på. Premissene er diskutert både med petroleumsnæringen og andre myndigheter som møter på lignende problemstillinger innenfor sitt område.

### 5.1 Usikkerhet knyttet til sluttbrukers fremtidige kompetansekrav

Et tema som ble omfattende diskutert i intervjuene er hvilke kompetansekrav som bør settes til sluttbruker. Generelt fant vi at det er stor usikkerhet i næringen knyttet til hvilke kompetansekrav som skal og bør stilles til sluttbruker når det kommer til å utnytte kognitive teknologier. En inndeling som ble brukt var å skille mellom krav satt til operasjonell kompetanse og krav satt til kompetanse innenfor dataanalyse.

For operasjonell kompetanse kom det frem at tiltak må gjøres for å sikre at sluttbrukers operasjonelle kompetanse ivaretas dersom tekniske systemer blir mer automatiserte. Det ble argumentert med at intervensjon fra menneskelige aktører kan bli nødvendig dersom en komplisert og vanskelig situasjon oppstår og det tekniske systemet ikke lengre strekker til. En utfordring er da at mennesket ikke nødvendigvis har den mengdetreningen som kan være nødvendig for å løse en komplisert og vanskelig situasjon. Denne mengdetreningen får mennesket i dag gjennom naturlig utførelse av operasjonen. Mer konkret ble det foreslått i intervjuene at:

- Mer trening vil være nødvendig for å sikre kompetanse i de tilfellene der teknologien ikke strekker til.
- Utvikle motiverende og engasjerende treningsprogram, ettersom treningen også inkluderer repeterende og «kjedelige» arbeidsoppgaver, ikke kun sikkerhetskritiske oppgaver eller nødsituasjoner.
- Tilgang til simulatortrening vil være viktig for å ivareta kompetanse for sluttbruker.
- Et operasjonelt design som krever at mennesket periodisk skal utføre oppgaver selv om kognitiv teknologi *kan* gjøre det, kan bidra til å forfriske og vedlikeholde operasjonell kompetanse hos menneskelig sluttbruker.

Når det gjaldt datateknisk kompetanse for sluttbruker ble det diskutert at det er viktig å forstå hvordan datamodeller utvikles og fungerer. Det ble argumentert med at dette er viktig både for å kunne bidra effektivt inn i utviklingsprosesser, slik diskutert i kapittel 3.2, og for å kunne diagnostisere og forstå hvorfor datamodellen kommer med en gitt output, slik diskutert i kapittel 4.2. Mer konkret ble det diskutert at:

- Det er behov for at operatører forstår hvordan datamodeller fungerer for å kunne tolke output fra modellen på en god måte.
- Det er behov for å ha et begrepsapparat for kognitiv teknologi som kan brukes på tvers av sluttbrukere, dataingeniører og andre fagdisipliner.
- Det vil ikke være nødvendig med spesialistkompetanse i datavitenskap ettersom valg og trening av ML-algoritmer kan gjøres i dialog med dataingeniører, men sluttbruker må likevel ha nok kompetanse til å kunne være med i denne dialogen.
- Det er vanskelig å si konkret hva som er viktig å kompetanse for en sluttbruker angående datavitenskap.
- Sluttbrukers mentale modell av hvordan teknologi fungerer blir enda viktigere. Det blir f.eks. viktig å forstå grensene for hva teknologien kan gjøre og ikke gjøre.

- Økt kompetanse i dataanalyse kan gjøre det enklere å utvikle tillit for datamodellen.

Det er altså viktig med både operasjonell- og datateknisk kompetanse for å sikre effektiv bruk av fremtidig kognitiv teknologi. Samtidig er det vanskelig å presisere og konkretisere i hvilken grad denne kompetanse vil være nødvendig å ha for sluttbruker. Det ble diskutert i intervjuet at det kan være vanskelig å konkretisere kompetansekrav fordi kognitiv teknologi er bredt definert, ikke i utstrakt bruk i praksis, og at grensesnittet med sluttbruker er forskjellig fra operasjon til operasjon (f.eks. boring kontra prosessovervåking).

*Funn 12: Det er usikkerhet i næringen knyttet til sluttbrukers fremtidige kompetansebehov dersom kognitiv teknologi blir mer avansert eller utbredt.*

## 5.2 Behov for å avklare datadelingsfilosofi

I intervjuene diskuterte vi begrensninger på datatilgang for utvikling av kognitiv teknologi. Data som er nødvendig for å trene ML-modeller deles ofte mellom kunden og leverandør, og noen ganger på tvers av aktører i næringen. Uten data er det ikke mulig å lage datadrevne modeller. Tilgjengeligheten av en tilstrekkelig mengde relevante data av tilstrekkelig kvalitet kan være avgjørende for at teknologien som utvikles vil fungere effektivt og sikkert i de ulike situasjonene den står overfor.

For leverandører av datamodeller kan en stor del av første fase innebære kategorisering og rydding i store datasett før selve treningen av modellen kan begynne. Det kan også være at mengden data i en gitt situasjon ikke er tilstrekkelig for å kunne trene modellen slik man egentlig ønsker. Tilgang til en database hvor lignende statistikk kan brukes til trening av modellen kan være et suksesskriterie for å trene datamodellen.

Åpen deling av data mellom selskapene kan være et viktig premiss for utvikling av sikker og effektiv kognitiv teknologi. Det vil kreve en rekke avklaringer omkring hvordan en slik database skal se ut, hvem som har rettigheter og tilgang til denne og hvem som skal forvalte eller eie databasen.

I intervjuene med flere av selskapene ble behovet for en datadelingsfilosofi diskutert. Dette er en filosofi som næringen kan bruke til å formulere metoder for datadeling, lagring av data, og i hvilket format data deles i. Filosofien bør også si noe om forvaltning av databasen. Et annet spørsmål som bør avklares er hvem som kan og bør forvalte en slik database, om det er myndigheter eller om det er noe som næringslivet bør organisere og håndtere selv, eller om eierskapet av databasen bør desentraliseres.

*Funn 13: Det er behov for å lage en datadelingsfilosofi for å sikre/få til en åpen og rettferdig deling av data innen næringen. Hensikten med datadelingsfilosofien vil være å utvikle sikre og effektive datamodeller.*

## 5.3 Maskinlæring som begrep er utsatt for effektiv markedsføring

Analyse av resultatene fra intervjuene indikerer at bruk av kognitiv teknologi er betydelig mindre utbredt enn det kan fremstå fra markedsføring. I diskusjonene med dataanalytikere fra næringen ble det poengtert at det sjeldent er tilfellet at den faktiske teknologien som utvikles er slik som beskrevet i ulike markedsføringskanaler. Det kan henge sammen med at selskapene ønsker å fremstille en fremtidsrettet profil, blant annet for å være attraktive som investering og som arbeidsgiver.

*Funn 14: Maskinlæring er et begrep som benyttes vidt i markedsføringsøyemed. Det kan være vanskelig å få et riktig bilde av den kognitive teknologien ulike selskaper faktisk anvender, basert bare på hva som beskrives i markedsføring.*

## 5.4 Innovasjon påvirkes av samfunnets interesser i O&G

De fleste informantene vi snakket med, var enige om at kognitiv teknologi ga tilsynsmyndighetene nye utfordringer, selv om det ikke var enkelt å beskrive utfordringene, eller hvordan de skulle håndteres. Det ble videre diskutert hvorvidt disse utfordringene faktisk vil bli gjeldende, ettersom investeringsviljen i ny teknologi kan endres i takt med samfunnets interesser i olje og gass. I flere intervjuer ble det understreket en usikkerhet knyttet til intensiteten i videre utvikling av kognitiv teknologi.

*Funn 15: Innovasjonsgraden til kognitive teknologi kan påvirkes av samfunnets interesser i O&G. Det er en usikkerhet knyttet til om utfordringene med fremtidens kognitive teknologi blir gjeldende.*

## 5.5 Koordinering med andre myndigheter

Vi diskuterte flere av prosjektets problemstillinger i intervjuene med andre myndigheter. Hensikten var å få perspektiv på de samme utfordringene fra andre domener. Vi pratet med organisasjoner knyttet til både maritim industri og bilindustri.

I intervjuer med maritim industri kom det frem at både eksisterende lovverk i Norge og internasjonale føringer satt av Internasjonal Maritim Organisasjon (IMO) må vurderes for ivaretagelse av menneskelige faktorer ved innføring av kognitiv teknologi. Det ble fortalt at også i denne industrien er det hyppig utvikling av teknologi, som f.eks. autonome fartøy, og at det er stor oppmerksomhet på at regelverk og standarder må være treffende for den teknologiske utviklingen.

I diskusjon med bilindustrien ble det fortalt at internasjonale lovverk ivaretar utfordringene knyttet til kognitive teknologier. Ifølge EU-regelverket, for eksempel, står det at automatisert teknologi må sørge for at den menneskelige bilføreren er inkludert i det automatiserte systemet. Et tiltak for å sikre at bilføreren er inkludert kan være å måle dreiemoment på rattet. Videre ble det fortalt at verdens internasjonale bilforum har kommet med et forslag (ECE/TRANS/WP.29/2020/81) om å stille krav til at HMI sikrer at bilfører ikke misforstår eller kan misbruke teknologi for automatisert kjøring. Lovverket vil ha direkte implikasjoner for utviklingen av kognitiv teknologi. Det ble også fortalt at det finnes flere standarder som gir retningslinjer for hvordan utviklere av automatiserte støttesystemer kan ivareta bilføreren.

Selv om maritim-, bil-, og petroleumsindustrien har hver sine utfordringer knyttet til introduksjonen av kognitiv teknologi, så deles også flere av utfordringene på tvers av industriene når det kommer til fremtidig bruk av kognitiv teknologi. Et eksempel på en delt utfordring kan være hvordan ansvar bør plasseres når kognitiv teknologi har beslutningsmyndighet i sikkerhetskritiske operasjoner og ingen menneskelig aktør er en del av det kognitive systemet.

I intervjuene med organisasjonene fra andre myndigheter ble det fortalt at en omforent tilnærming til å både identifisere og håndtere delte utfordringer kan være hensiktsmessig for både de respektive industriene og for samfunnet ellers.

*Funn 16: Det er hensiktsmessig med samarbeid for å løse felles problemer knyttet til bruk av kognitiv teknologi på tvers av industrier og myndigheter.*

## 6 KONKLUSJON

Studien har sett på hva forskningen sier om samhandlingen mellom mennesker og kognitiv teknologi og hvordan mennesket bør integreres med denne. Det er videre samlet inn empiriske data for å få innsyn i hvordan næringen stiller seg til utvikling, implementering og bruk av kognitiv teknologi i dag. Informasjon fra litteratur- og intervjustudien er diskutert i lys av problemstillingene i dette prosjektet:

- Hva er kognitiv teknologi i petroleumsnæringen i Norge i dag?
- Hvordan samhandler mennesket og kognitiv teknologi?
- Hva er sluttbrukers rolle i utviklingen av kognitiv teknologi?
- Hvilke premisser gjelder for arbeid med kognitiv teknologi i dag?

Basert på diskusjon av problemstillingene har Safetec konkludert med følgende 16 funn listet i Tabell 6-1:

Tabell 6-1 Liste over temaer og funn fra studien

Tema	Funn	Nr
Kognitiv teknologi i petroleumsindustrien i dag.	Kognitiv teknologi benytter i dag enkle tolkbare modeller. Tilnærminger og metoder for å jobbe med samhandlingsproblematikk, som f.eks. funksjonsallokering, må gjøres i sammenheng med den faktiske bruken av kognitiv teknologi.	1
	Kognitiv teknologi bør utvikles for å løse definerte problem, ikke utvikles som løsninger som kan være nyttige for å løse problemer man ikke har definert. Behovet for teknologien må være drevet av problemløsning for å sikre god forankring i hele utviklingsprosessen.	2
	Sluttbrukers kriterier inkluderes for lite ved valg av datamodell. Praksis og erfaring viser at det er vanskelig å inkludere sluttbrukers kriterier i valg av datamodell på grunn av usikkerhet knyttet til hvordan operasjoner vil bli utført etter introduksjonen av kognitiv teknologi.	3
Utvikling og implementering av kognitiv teknologi.	En tilnærming basert på et sosioteknisk rammeverk bør ytterligere nyanseres slik at cyber-delen i systemet, som inkluderer utvikling av datamodell og software, får økt fokus i arbeid med kognitive teknologier. HCAI (Human Centered Artificial Intelligence)-rammeverket kan legges til grunn for å fremme brukersentrert utvikling av kognitive teknologier. Særlig i systemer der det er både høy grad av automatisering og menneskelig kontroll.	4
	Det er behov for metoder og tilnærminger for involvering av sluttbruker i tidlig utviklingsfase av datamodeller til bruk i sikkerhetskritiske operasjoner. AutoML og interaktiv maskinlæring (IML) er to tilnærminger som involverer sluttbruker og som kan fungere som et utgangspunkt til videre arbeid.	5
	Involvering av sluttbruker fungerer i sen utviklingsfase. Det ble understreket at sluttbrukere, som en knapp ressurs, prioriteres i sen utviklingsfase av selskapene. Blant annet grunnet mer konkrete krav satt av regelverk og standarder og usikkerheter knyttet til metoder for brukerinvolvering i tidlig fase.	6

	Tilbakemeldinger fra sluttbrukere i sen utviklingsfase er viktig for videre oppdatering og utvikling av modellen. Tilbakemeldingene baseres på trening med kognitiv teknologi i simulatorer og bruk i drift etter at teknologien er implementert offshore.	7
	Sluttbrukers forståelse av hvordan kognitiv teknologi fungerer er viktig for samhandling. Viktige faktorer for å danne seg en god mental modell er trening og bruk, i tillegg til at datamodellen er gjennomskiktig.	8
Samhandling med kognitiv teknologi.	Datamodeller som er tolkbare for sluttbruker er å foretrekke. Dersom black box-modeller må brukes bør det benyttes verktøy for å øke modellens tolkbarhet.	9
	Å balansere riktig grad av tillit i sikkerhetskritiske operasjoner er viktig og krevende. En del av denne balansen handler om at den menneskelige aktøren må forstå sammenhengen mellom inputvariablene og output, som igjen setter føringer for datamodellens gjennomskiktighet.	10
	Metoder og tilnærminger for funksjonsallokering må passe til den faktiske bruken av kognitiv teknologi. Det finnes mange metoder og tilnærminger til temaet som antar mer avansert teknologi enn det som brukes i praksis.	11
Premisser for arbeid med kognitiv teknologi.	Det er usikkerhet i næringen knyttet til sluttbrukers fremtidige kompetansebehov dersom kognitiv teknologi blir mer avansert eller utbredt.	12
	Det er behov for å avklare en datadelingsfilosofi for å sikre/få til en åpen og rettferdig deling av data innen næringen. Hensikten med datadelingsfilosofien vil være å utvikle sikre og effektive datamodeller.	13
	Maskinlæring er et begrep som benyttes vidt i markedsføringsøyemed. Det kan være vanskelig å få et riktig bilde av den kognitive teknologien ulike selskaper faktisk anvender, basert bare på hva som beskrives i markedsføring.	14
	Innovasjonsgraden til kognitive teknologier kan påvirkes av samfunnets interesser i O&G. Det er derfor en usikkerhet knyttet til om utfordringene med fremtidens kognitive teknologi blir gjeldende.	15
	Det er hensiktsmessig med samarbeid for å løse felles problemer knyttet til bruk av kognitiv teknologi på tvers av industrier og myndigheter.	16



## 7 MULIG VIDERE ARBEID MED KOGNITIV TEKNOLOGI

Basert på arbeidet i dette prosjektet foreslår vi å rette oppmerksomheten til oppdatering av standarder og metoder for integrering av sluttbruker i den tidlige utviklingen av kognitiv teknologi, slik definert i kapittel 3.2. I dette prosjektet ble det kartlagt at bruk av kognitiv teknologi i dag fortsatt krever at menneskelige kognitive aktører er en del av det kognitive systemet. Vi foreslår også en strategi for å være forberedt på introduksjonen av reelle autonome systemer.

### 7.1 Standarder og metoder for brukersentrert utvikling av kognitiv teknologi

Tradisjonelle prinsipper beskrevet og lagt til grunn for brukersentrert utvikling, som i f.eks. ISO 11064, er fortsatt gjeldende for samhandling mellom mennesker og teknologi slik teknologien fungerer i dag. Det er samtidig identifisert et behov for å utvikle og validere metodene som kan brukes for å integrere sluttbruker enda tidligere enn det som er vanlig i dagens utviklingsprosesser av kognitiv teknologi.

Et mantra i de fleste designstandarder er at sluttbruker må tidlig inn designprosessen. I dette prosjektet er vi blitt oppmerksomme på at definisjonen av «tidlig» trenger tydeliggjøring. «Tidlig» bør referere til den fasen hvor datamodellen etableres, trenes og testes. Dette er kalt tidlig utviklingsfase i denne rapporten. Altså er det behov for å tydeliggjøre at tidlig involvering referer til når datamodellen etableres, ikke når datamodellen integreres med kontrollsystemet. Fasen hvor datamodellen integreres med kontrollsystemet er referert til som sen utviklingsfase i denne rapporten.

Oppdateringen av begrepet «tidlig» medfører et behov for å vurdere gyldigheten til etablerte HF-designmetoder som i dag brukes hovedsakelig til å sikre brukersentrert design i sen utviklingsfase. I kapittel 3.2 foreslår vi to tilnærminger til dette basert på forskningslitteraturen, spesifikt AutoML og interaktiv maskinlæring. Disse tilnærmingene kan fungere som et utgangspunkt for videre arbeid, i tillegg til å vurdere gyldigheten til de allerede etablerte HF-metodene som er i bruk i dag.

### 7.2 Strategi for autonome systemer

Denne rapporten handler om samhandlingen mellom kognitiv teknologi og mennesket der kognitive menneskelige aktører er en del av det kognitive systemet, i tillegg til de kognitive tekniske aktørene. I vårt arbeid er det ikke funnet empiri eller eksempler på kognitiv teknologi som har evnen til å utføre funksjoner i sikkerhetskritiske operasjoner uten at menneskelige aktører er en del av det kognitive systemet. Først når et kognitivt system kan operere uten at menneskelige aktører er en del av systemet vil problemstillingene knyttet til utvikling og styring av autonome systemer bli fullt ut relevante. Det er usikkerhet knyttet til når, eller om, denne teknologien vil introduseres i sikkerhetskritiske operasjoner i petroleumsindustrien, men basert på intervjuer med næringen i dag, er det ikke planlagt prosjekter for autonome sikkerhetskritiske operasjoner.

Det er likevel viktig å være forberedt på hva som er nødvendig å gjøre i det autonome systemer introduseres. En strategi for håndtering av autonome systemer kan være et nyttig arbeid. En slik strategi kan for eksempel eies av aktørenes interesseorganisasjon, både nasjonalt og internasjonalt. Strategien vil kunne kontinuerlig oppdateres etter hvert som autonom teknologi introduseres i sikkerhetskritiske operasjoner. Der kan det også tas stilling til roller, aktører, begrepsbruk, definisjoner og rammeverk som kan bidra til at næringen får en omforent og felles tilnærming til de utfordringene som vil fortløpende bli presentert mens teknologien utvikles.

## 8 REFERANSER

- /1/ Bainbridge, L., 1982. Ironies of Automation. IFAC Proceedings Volumes 15, 129–135. [https://doi.org/10.1016/S1474-6670\(17\)62897-0](https://doi.org/10.1016/S1474-6670(17)62897-0)
- /2/ Santoni de Sio, F., van den Hoven, J., 2018. Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI* 5, 15. <https://doi.org/10.3389/frobt.2018.00015>
- /3/ Woods, D.D., Hollnagel, E., 2006. Joint cognitive systems: Patterns in cognitive systems engineering. CRC Press, Boca Raton FL.
- /4/ Nemseth, C.P., 2004. Human factors methods for design - making systems human-centered. CRC Press Taylor & Francis Group, Boca Raton FL.
- /5/ Carayon, P., 2006. Human factors of complex sociotechnical systems. *Applied Ergonomics* 37, 525–535.
- /6/ Gualtieri, J.W., Szymczak, S., Elm, W.C., 2005. Cognitive System Engineering - Based Design: Alchemy or Engineering. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 49, 254–258. <https://doi.org/10.1177/154193120504900309>
- /7/ Bennett, K.B., Flach, J., 2019. Ecological Interface Design: Thirty-Plus Years of Refinement, Progress, and Potential. *Human Factors* 61, 513–525. <https://doi.org/10.1177/0018720819835990>
- /8/ Challenger, R., Clegg, C.W., Shepherd, C., 2013. Function allocation in complex systems: reframing an old problem. *Ergonomics* 56, 1051–1069. <https://doi.org/10.1080/00140139.2013.790482>
- /9/ Woods, D.D., Dekker, S., Cook, R.I., Johannesen, L., Sarter, N.B., 2016. Behind human error. Second Edition. Taylor & Francis, New York.
- /10/ Sætren, G.B., Hogenboom, S., Laumann, K., 2016. A study of a technological development process: Human factors—the forgotten factors? *Cognition, Technology & Work* 18, 595–611. <https://doi.org/10.1007/s10111-016-0379-x>
- /11/ Hoffman, R.R., Klein, G., Mueller, S.T., 2018. Explaining Explanation For “Explainable Ai.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62, 197–201. <https://doi.org/10.1177/1541931218621047>
- /12/ Dybå, T., Kitchenham, B.A., Jørgensen, M. (2005). Evidence-based software engineering for practitioners. *IEEE software*, 2005-01, Vol.22 (1), p.58-65
- /13/ B.A. Kitchenham, T. Dybå, M. Jørgenson (2004) Evidence-Based Software Engineering *Proceedings of ICSE 2004*, IEEE Computer Society Press, pp. 273-281
- /14/ Tranfield, D., Denyer, D. & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14, 207-222
- /15/ Brereton, P., Kitchenham, B.A., Budgen, D. Turner, M. & Khalil, M. (2007). Lessons from applying the systematic literatur review process within the software engineering domain. *The Journal of Systems and Software*, 80, 571-583.
- /16/ Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- /17/ Hotvedt, M., Grimstad, B., & Imsland, L. (2020). Developing a hybrid data-driven, mechanistic virtual flow meter-a case study. *IFAC-PapersOnLine*, 53(2), 11692-11697.
- /18/ Miletić, T. Human-Artificial Intelligence Symbiosis: the Possibility of Moral Augmentation.
- /19/ Patriarca, R., Falegnami, A., Costantino, F., Di Gravio, G., De Nicola, A., & Villani, M. L. (2021). WAX: An integrated conceptual framework for the analysis of cyber-socio-technical systems. *Safety science*, 136, 105142.
- /20/ Modarres, M., Kaminskiy, M. P., & Krivtsov, V. (2016). Reliability engineering and risk analysis: a practical guide. CRC press.

- 
- /21/ Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495-504.
  - /22/ Gutzwiller, R. S., & Reeder, J. (2020). Dancing with algorithms: Interaction creates greater preference and trust in machine-learned behavior. *Human factors*, 0018720820903893.
  - /23/ Ramos, G., Meek, C., Simard, P., Suh, J., & Ghorashi, S. (2020). Interactive machine teaching: a human-centered approach to building machine-learned models. *Human-Computer Interaction*, 35(5-6), 413-451.
  - /24/ Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114.
  - /25/ Standard Norge (2000). NS-ISO 11064:2000, Ergonomisk utforming av kontrollsentre - Del 1-7.
  - /26/ Rudin (2019).
  - /27/ Zednik, C., & Boelsen, H. (2021). Preface: Overcoming Opacity in Machine Learning. *Overcoming Opacity in Machine Learning*, 1.
  - /28/ Kulesza, M., Burnett, W., Wong, S., Stumpf, Principles of explanatory debugging to personalize interactive machine learning, in *Proceedings of the 20th International Conference on Intelligent User Interfaces (ACM, 2015)*, pp. 126–137.).
  - /29/ Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, 4(37), 7120.
  - /30/ Asan, O., Bayrak, A. E., & Choudhury, A. (2020). Artificial intelligence and human trust in healthcare: focus on clinicians. *Journal of medical Internet research*, 22(6), e15154.
  - /31/ Atoyan, H., Duquet, J. R., & Robert, J. M. (2006). Trust in new decision aid systems. In *Proceedings of the 18th Conference on l'Interaction Homme-Machine* (pp. 115-122).
  - /32/ Van Den Broek, J., Van Diggelen, J., Van Der Kleij, R., Hueting, T. F., Van Der Waa, J., Van Schendel, J. A., & Langefeld, J. J. (2018). *Adaptive Maritime Automation*. Soesterberg: TNO.
  - /33/ Abbass, H. A. (2019). Social integration of artificial intelligence: functions, automation allocation logic and human-autonomy trust. *Cognitive Computation*, 11(2), 159-171.
  - /34/ Scholtz, J. (2003, January). Theory and evaluation of human robot interactions. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the* (pp. 10-pp). IEEE.