Standard for tillitvekkende kunstig intelligens

Hvordan håndtere KI-risiko?
En veiledning for norske virksomheter

Et samarbeid mellom EY & Langsikt

V.01 - Oppdatert per 11. juni 2024

EY
Building a better working world

Langsikt

Sammendrag

i. Kort om Standarden

Dette dokumentet («Standarden») gir offentlige og private virksomheter en oversikt over risikobildet i møte med kunstig intelligens («KI») og hva som kan gjøres for å håndtere relevante risikofaktorer.

Standarden er utarbeidet i fellesskap av tankesmien Langsikt og Ernst & Young Advokatfirma AS («EY»).[1] Langsikt er en politisk uavhengig tankesmie som kaster lys over de viktigste og mest neglisjerte problemene i vår tid. EY er et ledende advokatfirma innen bl.a. teknologi, og er en del av det globale EY-nettverket.

Standarden er tiltenkt som et verktøy for norske virksomheter, private og offentlige, for å hjelpe virksomhetene i å håndtere risiko forbundet med kunstig intelligens.[2] Standarden er ment å fungere som en praktisk manual for å navigere i kompleksiteten i teknologien, regelverk og etiske hensyn. Den er oppdatert i tråd med datoen på første siden og vil søkes holdt løpende oppdatert i tråd med den teknologiske, økonomiske og juridiske utviklingen.

ii. Hvordan bruke dette dokumentet

Standarden er strukturert etter de primære valgene en virksomhet står overfor i møte med KI og hvilke momenter virksomheten bør tenke på i den forbindelse. Den gir rettledning i møte med tre sentrale spørsmål: Hva er risikoene ved KI, hva er risikokildene og hva bør virksomheten gjøre for å håndtere risikoene?

Standarden beskriver hensyn man må ta i de ulike delene av livssyklusen til et KI-system.

1. Datainnsamling
2. Dataprosessering
3. Trening, fintrening og/eller brukstilpasning
4. Intern eller ekstern bruk
5. Vedlikehold og oppdatering av KI-modellen

Det vil være nyttig å lese dokumentet i sin helhet, men leseren kan også gå til delene som er mest relevante. Hvordan Standarden brukes vil avhenge av type virksomhet og rollen til leseren, se Vedlegg 1: Relevante brukergrupper– hvem retter Standarden seg mot?

Standarden inneholder strukturert metodikk, retningslinjer for risikostyring og anbefalinger for tillitvekkende bruk av KI. Den fremmer behovet for styringsstrukturer som kan takle uforutsette hendelser og opprettholde tillit til virksomheten. Se Vedlegg 2: Nøkkelbegreper og definisjoner som er relevante for forståelse av denne Standarden.

Standarden er ikke ment å være uttømmende og gir ikke råd for konkrete situasjoner, men veilederen er tenkt å gi virksomheter en veiledning i tenkning rundt hvordan risiko ved KI kan håndteres. Virksomheten må selv vurdere behovet for konkret juridisk rådgivning i situasjonen virksomheten står i, og EY og Langsikt tar ikke ansvar for hvordan Standarden benyttes og står ikke

---

[1] Fra Langsikt har Jacob Wulff Wold (rådgiver), Aksel Braanen Sterri (fagsjef) og Jakob Graabak (seniorrådgiver) bidratt til Standarden. Langsikt takker også for bidrag fra Julia Graham og Hanna Malm. Fra EY har Mads Ribe (assosiert partner / advokat og KI-leder EY Tax & Law), Andreas Bjørnebye (assosiert partner / advokat, EY Law) og Benjamin Green (advokat, EY Law) bidratt.

[2] Etisk og juridisk risiko, sammen med manglende kompetanse, nevnes i en undersøkelse, utført av EY, fra september 2023 som de mest vesentlige barrierene for bruk og skalering av generativ KI for norske virksomheter. Standarden legger derfor ekstra mye vekt på de hensynene.

til ansvar for eventuelle feil eller mangler enten i Standarden eller som følger av bruk av Standarden på konkrete situasjoner.

iii. Tillitvekkende KI er en konkurransefordel

Tillit er en bærebjelke i norsk samfunnsliv. Vårt siktemål er at Standarden skal hjelpe norske virksomheter til å opprettholde og forvalte deler av denne tilliten ved å identifisere, og sette i verk tiltak for å begrense risikoer ved utvikling og bruk av KI.

En systematisk tilnærming til KI og risiko vil kunne gi virksomheten en konkurransefordel. Det vil gjøre det lettere for virksomheten å møte regulatoriske krav, fremme ansvarlig innovasjon, og gjøre virksomheten i bedre stand til å håndtere risikoer ved KI i møte med fremtidige utfordringer, som i verste fall kan rokke ved virksomhetens eksistens.

iv. Fire prinsipper for tillitvekkende kunstig intelligens

Virksomheter bør vurdere følgende grunnleggende tilnærming ved utvikling og bruk av kunstig intelligens:

1. Formålstjenlig: Sørg for å bruke KI til å oppnå overordnede mål for virksomheten og at KI bare benyttes der fordelene ved bruken overstiger ulempene.

2. Risikostyring: Virksomheten bør identifisere risiko og ha en systematisk tilnærming til å redusere risiko i tråd med den forventede nytten.

3. Hensyn: Hensynta etiske og juridiske hensyn i virksomhetens utvikling og bruk av KI og sikre at bruken er i tråd med regelverk, bransjestandarder og rimelige forventninger fra brukere og samfunnet for øvrig.

4. Helhetlig: Sørg for at alle funksjoner i virksomheten og alle brukere får oppleve de positive effektene av KI.

INNHOLDSFORTEGNELSE

Vedlegg:

1 Bør virksomheten bruke kunstig intelligens?

1.1 Hva er kunstig intelligens?

KI-forordningen artikkel 3 definerer en KI-modell slik (vår oversettelse fra engelsk):

> Et "KI-system" er et maskinbasert system, som opererer med varierende grad av selvbestemmelse og kan tilpasses etter implementering. Et KI-system styrer etter mål og genererer resultater fra inputen det mottar (f.eks. et prompt), i form av prediksjoner, innhold, anbefalinger eller beslutninger som kan påvirke fysiske eller virtuelle miljøer.[3]

En skiller ofte mellom to typer KI:

1. Generativ KI er antakelig mest kjent gjennom ChatGPT, og har navnet sitt fra systemets evne til å generere ny tekst, nye bilder, lyd og videoer.[4] De mest generelle generative modellene utgjør i økende grad også en ny infrastruktur, kalt grunnmodeller. Grunnmodeller har gode generelle ferdigheter, som så kan spesialiseres etter behov.

2. Snever KI er tilpasset et snevert bruksområde, hvor den gjerne er svært god. Eksempelvis kan slike systemer brukes til bildediagnostikk, spille gitte spill eller regne ut 3D-formene til proteiner ut ifra en genetisk kode.

Begge typene KI lages stort sett ved maskinlæring. Før programmerte man smarte regler inn i KI-systemet. I dag lar vi modellen lage sine egne regler ved å belønne den for å komme med det resultatet utviklerne ønsker seg under treningen. For eksempel kan en modell lære å oppdage kreftsvulster ved å trene på bilder med og uten svulster, og gi belønning når modellen gjetter riktig. For å trene frem slike modeller trengs det data som modellen kan trene på og datakraft til å utføre regneoperasjonene til grunn for treningen.

1.2 Strategiske hensyn ved bruk av KI

Det overordnet viktigste virksomheten bør vurdere i møte med kunstig intelligens er:

1. Hvilke konkrete behov kan virksomheten tilfredsstille ved å utvikle eller ta i bruk KI? For eksempel automatisering av eksisterende arbeidsoppgaver eller bruke KI til å løse problemer en ikke har kunnet løse med eksisterende kompetanse.

2. Hvilket behov har virksomheten for å bygge opp kompetanse ved hjelp av KI, som på sikt vil hjelpe å tilfredsstille flere konkrete behov?

3. Hvilket behov har virksomheten for å tilpasse forretningsmodellen eller virksomhetsstrategien etter den teknologiske utviklingen?

Bruk av KI må tilpasses den enkelte virksomhet og oppgavene den utfører. Dette må vurderes løpende etter hvilke KI-løsninger som finnes og hvordan KI endrer markedet en opererer i. Siden teknologien og tilbudet av nye modeller endrer seg raskt, bør virksomheten ha oversikt over

---

[3] Originaltekst fra KI-forordningens artikkel 3: "'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."

[4] Utviklingen i bruk av KI, og især generativ kunstig intelligens, har skutt fart siden det amerikanske selskapet OpenAI lanserte ChatGPT november 2022. Siden har det kommet flere tilsvarende løsninger på markedet i form av Google Gemini, Anthropic Claud, Meta Llama m.fl.

nåsituasjonen og en plan for ulike scenarioer fremover. Se Vedlegg 3: Strategi-sjekkliste for strategiske valg.

2 Risikostyring

2.1 Fem typer virksomhetsrisiko

Risiko er produktet av skadepotensialet ved en negativ hendelse og sannsynligheten for at hendelsen inntreffer. Jo større sannsynlighet for at skaden inntreffer og jo mer alvorlig skaden er, jo større risiko.

Overordnet er det fem former for risiko virksomheten står overfor:

1. Juridisk risiko: risiko for brudd på rettslige forpliktelser, både nåværende og fremtidige, med sivilrettslige eller strafferettslige sanksjoner hvis virksomheten bryter forpliktelsene.

2. Etisk risiko: risiko for at en handler i strid med etiske normer.

3. Verdirisiko: risiko for at virksomheten ikke opptrer i tråd med virksomhetens verdier.

4. Oppgaverisiko: risiko for at virksomheten ikke løser virksomhetens oppgaver på en tilfredsstillende måte.

5. Disrupsjonsrisiko: risiko for at teknologiske og samfunnsmessige endringer undergraver virksomhetens eksistensgrunnlag.

Til sammen og hver for seg kan disse virksomhetsrisikoene true virksomhetens evne til å nå sine mål, eller utøve sine funksjoner.[5] Å utsette seg for de ulike typene risikoene kan også medføre skade på virksomhetens omdømme. Standarden vil ha mest å si om juridisk risiko, etisk risiko og oppgavespesifikk risiko.

Merk at de fem virksomhetsrisikoene er nøytralt formulert - de er ikke spesifikke for KI. Det reflekterer at KI både kan øke og redusere virksomhetsrisikoen.

KI kan brukes på måter som er i strid med lovverk og verdier og som svekker kvaliteten på produktet virksomheten leverer, men kan også brukes til å levere bedre produkter og bedre etterleve lover, regler og verdier.

KI kan gi nye muligheter som ikke tidligere var tilgjengelig for virksomheten, men også undergrave behovet for kjernevirksomheten. Dersom virksomheten ikke klarer å utnytte mulighetene i KI på en måte som gjør den relevant og effektiv, vil dette kunne innebære en betydelig konkurranseulempe.

2.2 Hvordan skal virksomheter hensynta risiko?

I møte med risiko er det sentralt at virksomheten:

1. Skaffer seg oversikt over og informasjon om risiko og kildene til risiko

2. Skaffer seg relevant kompetanse for å håndtere risikoen

3. Fordeler ansvar for å håndtere KI og tiltak

4. Setter i verk tiltak for å håndtere eller minimere risiko

[5] Alle fire kategorier utdypes lengre ned i dokumentet og kan splittes opp i underkategorier. En kan for eksempel snakke om modellrisiko og andre mer spesifikke risiko.

5. Identifiserer svakheter eller mangler i tiltakene, evaluerer og sikrer kontinuerlig forbedring i tråd med bruk og teknologisk utvikling

6. Får på plass gode styringssystemer for 1, 4 og 5 (governance) som reduserer risiko og skaper tillit hos eiere, brukere, ansatte og samfunn (noe som igjen reduserer risiko)

7. Kontinuerlig overvåker virksomhetens styringssystemer.[6]

2.3 Ta rimelige grep for å redusere risiko

Risikoer er generelt noe virksomheter lever med. Null risiko er ikke praktisk mulig, hverken i næringslivet, offentlig sektor eller i samfunnet for øvrig. Risiko knyttet til utvikling eller bruk av KI må derfor ikke elimineres, men håndteres forsvarlig.

Eiere, ansatte, kreditorer, kunder og resten av samfunnet vil forvente at virksomheten tar rimelige grep for å redusere risiko for at ulike hendelser inntreffer og for å håndtere situasjonen dersom risikoen inntreffer.

Hva som anses som rimelige tiltak vil i stor grad avhenge av en kost-nytte-analyse: Hvor mye vil et tiltak redusere risikoen, til hvilken kostnad for virksomheten og samfunnet?

Noen tiltak koster mer enn det smaker i forhold til reduksjon i risiko oppnådd. Slike disproporsjonale tiltak bør som regel ikke gjennomføres. Blant proporsjonale tiltak, bør de mest effektive tas i bruk.

2.4 Prinsipper for ansvarlig og etisk bruk av KI

For å forhindre etisk risiko, bør virksomheten handle i tråd med prinsipper for etisk og ansvarlig bruk av KI.

Det er til dels stort overlapp mellom det etiske og juridiske. Rettsreglene hviler gjerne på etiske hensyn, som rettferdighet, men overlappen er ikke perfekt: ikke alt som er umoralsk er ulovlig og ikke alt som er ulovlig er umoralsk. I listen over de mest relevante etiske hensynene vil det likevel være betydelig overlapp med rettsregler, som vil fordre juridiske, så vel som etiske, vurderinger.[7]

De mest relevante etiske hensynene er:

1	*Rettferdighet*	Alle har krav på like rettigheter, muligheter og rettferdig behandling. Virksomheten bør ikke forskjellsbehandle uten tilstrekkelig god grunn og bør sørge for at den ikke bruker KI på en måte som er urettferdig.
2	*Forhindre skade*	Enhver bør gjøre det som er rimelig for å forhindre skade på andre som følge av handlinger og valg en er ansvarlig for. Det gjelder også virksomheter, som bør forhindre at brukere eller tredjeparter skades av KI virksomheten tar i bruk.
3	*Selvbestemmelse*	Hvert myndig individ har selvbestemmelsesrett. De KI-systemer virksomheten tar i bruk bør styrke, ikke svekke, brukere og andre berørte parters evne til å ta gode valg. KI systemer bør for eksempel ikke utnytte personers sårbarheter.

[6] Se fullstendig sjekkliste i Vedlegg III
[7] Flere av punktene I listen overlapper med EUs Ethics by Design and Ethics of Use Approaches for Artificial Intelligence av 25.11.2021:.

4	*Privatlivets vern*	Alle har et privatliv som de skal få ha i fred. Der virksomheten tar i bruk KI bør det sikres at privatlivet til både brukere og tredjeparter respekteres.[8]
5	*Gode begrunnelser*	Avgjørelser som påvirker andre, må rettferdiggjøres. Virksomheten må derfor kunne begrunne og rettferdiggjøre bruk av KI-systemer overfor brukere, myndigheter og andre som påvirkes av systemet.
6	*Forståelse*	Gode begrunnelser forutsetter forståelse hos andre. KI-systemets virkemåte og beslutninger må kunne forstås både av brukere, myndigheter og andre som påvirkes av systemet.[9]
7	*Ansvar*	Alle – både fysiske og juridiske personer – har ansvar for egne valg og handlinger. Virksomheten må ta ansvar for beslutninger foretatt av KI-systemet og må derfor forstå, overvåke og kontrollere utformingen og driften av KI-baserte systemer.

2.5 Juridisk risiko: «Lovlig KI»

Ansvarlig KI innebærer et bevisst forhold til hvilke etiske hensyn som gjør seg gjeldende og i hvilken grad konteksten innebærer risiko for brudd på rettsregler. Som nevnt over er det til dels stort overlapp mellom det etiske og juridiske, hvor rettsreglene gjerne operasjonaliserer etiske hensyn, som innen personvern eller ansvar for skade. En forståelse for de grunnleggende prinsippene bak regelverk som er relevant for KI (selv de som ikke er implementert i Norge enda), kan bidra til å redusere juridisk risiko og styrke virksomhetens posisjon.[10]

Norge har, i stor grad, et teknologinøytralt lovverk: handlinger reguleres uavhengig av hvilken teknologi eller metode som benyttes.[11] Selv om KI ikke reguleres direkte i dag (juni 2024), vil det derfor likevel være lovgivning som treffer utvikling og bruk av KI.[12]

Utover dagens lovverk, er det EUs KI-forordning som er mest aktuelt å forholde seg til for norske virksomheter. KI-forordningen bryter med tradisjonen med et teknologinøytralt lovverk og regulerer direkte KI. Norske myndigheter har uttrykt at forordningen vil implementeres i norsk lov.

Foruten eksisterende eller fremtidig lovverk, er det nyttig for virksomheten å undersøke om det er relevante standarder som kan være nyttige for å enten sertifisere egen modell eller for å sjekke sertifisering av en utviklet modell virksomheten planlegger å bruke.[13] Virksomheten må selv vurdere hva som er aktuelt for sin utvikling eller bruk av KI, som hensyntatt bransje, geografisk lokasjon, risiko-profil og kunnskapsnivå.

---

[8] Se Datatilsynets oversikt over personvernprinsippene.
[9] Det er ulike måter å forklare et system på. At et system viser seg å fungere i mange sammenhenger kan i enkelte sammenhenger regnes som en adekvat forklaring.
[10] En forståelse for KI-forordningen vil øke bevisstheten rundt KI-risikoer, forberede virksomheten på overgangen til kravene som gjelder under forordningen, bygge tillit hos kunder rundt utvikling og bruk av KI, og økt kredibilitet overfor myndigheter. Gode forberedelser vil gjøre overgangen enklere og potensielt gi et konkurransefortrinn når forordningen blir implementert. Det kan også hjelpe virksomheten med å forutse styringsbehov og overholdelseskrav som kan bli aktuelle for virksomhetens spesifikke utvikling og bruk av KI.
[11] F.eks. forbud mot innbrudd, uavhengig av om innbruddet er digitalt eller fysisk og uavhengig av hvilken metode eller teknologi som benyttes.
[12] Dette inkluderer immaterielle rettigheter (som vern av åndsverk under åndsverkloven), personvernlovgivningen (f.eks. vern mot urettmessig bruk av persondata), IKT-forskriften for sikkerhet i IT-systemer, diskrimineringslovgivning (som vern mot diskriminerende adferd). Videre vil kontraktsforpliktelser også kunne påvirke bruk av teknologi (som utkontrakteringsavtaler for IT-tjenester).
[13] Se Standard Norges side om KI-standarder.

En virksomhet som opererer i flere rettslige områder må være seg bevisst at ulike jurisdiksjoner har ulike reguleringer. Vi har derfor inkludert en oversikt over regulatoriske trender internasjonalt i Vedlegg 4: Oversikt over reguleringer internasjonalt.

2.5.1 Klassifisering av KI-systemer under KI-forordningen

EUs KI-forordning regulerer KI etter risiko, roller og bruksområde. Virksomheten må ha et aktivt forhold til hvilken rolle den har i ulike situasjoner der KI er involvert, og samtidig kunne kategorisere sin bruk av KI opp mot de ulike risiko-kategoriene.[14]

Risiko-kategoriene er som følger:

Kategori	Regulering	Beskrivelse	Eksempler - KI-systemer som:
Uakseptabel risiko	Forbud	Systemer som utgjør uakseptable risikoer og som kan brukes til å undergrave en persons grunnleggende rettigheter, eller som kan utsette dem for fysisk eller psykisk skade.	• Utnytter personers sårbarheter (f.eks. barn, eldre, nedsatt funksjonsevne) og som omgår brukernes vilje • Poengsetting for sosialt ønskelig atferd • Tolkning av ansattes følelser på arbeidsplassen • Biometrisk kategorisering for å tolke sensitiv data • Ansiktsgjenkjenning som ikke er målrettet og som lagres i databaser • Prediktiv overvåkning av enkeltpersoner for å hindre lovbrudd som kanskje vil skje
Høyrisiko	Lovlig, men strenge vilkår	Systemer som utgjør betydelig risiko for skade på helse, sikkerhet eller grunnleggende rettigheter. De fleste forpliktelsene i KI-forordningen gjelder slike systemer, fordelt mellom leverandører og brukere.	• Brukes som sikkerhetskomponenter i andre produkter og systemer (f.eks. i medisinske enheter, biler, maskiner) • Utgjør betydelig risiko for skade på helse, sikkerhet eller grunnleggende rettigheter, inkl. biometrisk ID, styring og drift av kritisk infrastruktur, utdanning, ansettelse, tilgang til offentlige tjenester, visse former for myndighetsutøvelse (f.eks. politi, asyl), demokratiske prosesser (valgsystemer) m.m.
Begrenset risiko	Lovlig, men begrensede vilkår	Begrensede krav til åpenhet	• Interagerer med mennesker må tydelig informere om at et KI-system benyttes (f.eks. chatbots) eller at en ser på KI-generert materiale (f.eks. deepfakes). • Visse unntak foreligger for kunstneriske verk, eller innhold som er åpenbart KI-generert.

[14] Dette for å hensynta virksomhetens nåværende krav under eksisterende lovgivning, både generelt og sektor-spesifikke krav (f.eks. personvernlovgivning, åndsverkslovgivning, markedsføringslovgivning, datasikkerhet og IKT-forskriften, EUs maskindirektiv, osv.), virksomhetens kontraktsforpliktelser (i den grad bruk av KI påvirker virksomhetens kontrakts-rettigheter og forpliktelser, f.eks. knyttet til åpenhet om og bruk av resultatet/utfallet av KI-bruk i spesifikke situasjoner) eller fremtidige krav under KI-Forordningen og øvrig relevant lovgivning slik som Ansvarsdirektivet for KI-systemer og det oppdaterte Produktansvarsdirektivet (begge fra EU).

Kategori	Regulering	Beskrivelse	Eksempler - KI-systemer som:
Minimal risiko	Lovlig, men få eller ingen vilkår	Alle andre KI-systemer som ikke passer inn under de andre kategoriene	• Typisk øker effektiviteten i arbeidsprosesser uten at brukernes rettigheter påvirkes (f.eks. fotoredigering, produktanbefaling, spamfilter, oversettelsesverktøy m.m.)

Forpliktelser etter KI-forordningen avhenger videre av hvilken rolle virksomheten har i forhold til utvikling eller bruk av KI.

Rolle	Beskrivelse	Krav
Utvikler (Provider)	Fysisk eller juridisk person, offentlig myndighet, organisasjon eller annen enhet som har utviklet et KI-system som rulles ut på markedet, eller for å sette i drift under sitt eget navn eller varemerke, enten mot betaling eller gratis.	Krav til leverandører av høyrisiko KI-systemer inkluderer: • Etablering og vedlikehold av passende risiko- og kvalitetsstyringssystemer • Effektiv datastyring • Menneskelig tilsyn • Overholdelse av aktuelle standarder (f.eks. cyber-sikkerhet) • Registrering av systemet i EU-database, inkludert for kritisk infrastruktur i en særskilt ikke-offentlig del av databasen.
Bruker (Deployer)	En fysisk eller juridisk person, offentlig myndighet, organisasjon eller annen enhet som bruker et KI-system under sin myndighet.	Krav til brukere av høyrisiko KI-systemer inkluderer: • Vurdering av innvirkning på grunnleggende rettigheter (FRIA) før KI-systemet tas i bruk, hvis brukeren: – Er en offentlig instans eller privat enhet som tilbyr offentlige tjenester – Tilbyr essensielle private tjenester, inkl. vurdering av kredittverdighet, risikovurdering og prising i forhold til livs- og helseforsikring. • Menneskelig tilsyn av personer med passende opplæring og kompetanse • Sikring av at prompts / inngangsdata inn i systemet er relevant for systemets bruk • Stansing av bruk av systemet hvis det utgjør en risiko på nasjonalt nivå • Informering av KI-systemleverandøren om eventuelle alvorlige hendelser • Overholdelse av GDPR-forpliktelser for å utføre en vurdering av personvernkonsekvenser • Verifisering av at KI-systemet er i samsvar med KI-loven og at all relevant dokumentasjon er bevist • Informering av personer om at de kan være underlagt bruk av høyrisiko KI

Rolle	Beskrivelse	Krav
Distributør (Distributor)	Enhver fysisk eller juridisk person i leverandørkjeden, som ikke er leverandør eller importør, men som gjør et KI-system tilgjengelig på markedet i EU.	Krav til distributører ved distribusjon av høyrisiko KI-systemer inkluderer: • Kontrollere at systemet har nødvendig CE-merking, med kopi av EU-samsvarserklæring m.m. • Undersøke tilgjengelig informasjon og vurdere om KI-systemet er i tråd med KI-forordningen Om det ikke er i tråd med forordningen, skal korrigerende tiltak iverksettes eller gjennomføre tilbakekallelse. • Vurdere om KI-systemet utgjør en risiko basert på tilgjengelig informasjon, og samarbeide med relevante myndigheter for å redusere risikoen. • Gi all informasjon og dokumentasjon om systemet, ved begrunnet forespørsel fra relevante myndigheter.
Importør (Importer)	Enhver fysisk eller juridisk person som er lokalisert eller etablert i EU, og som plasserer et KI-system på markedet som bærer navnet eller varemerket til en fysisk eller juridisk person etablert utenfor EU.	Krav til importører ved distribusjon av høyrisiko KI-systemer inkluderer: • Sikre at leverandøren av KI-systemet har utført nødvendige prosedyrer og teknisk dokumentasjon som kreves. • Kontrollere at systemet har nødvendig CE-merking, med kopi av EU-samsvarserklæring m.m. • Ikke plassere systemet på markedet dersom det er grunn til å tro at systemet strider mot KI-forordningen, er forfalsket, osv. • Gi all informasjon og dokumentasjon om systemet, ved begrunnet forespørsel fra relevante myndigheter. • Samarbeide med relevante myndigheter for å redusere risikoer identifisert ved systemet.

Hvilke krav som stilles til virksomheten og bøtenivået ved brudd på KI-forordningen, vil avhenge av rolle og klassifisering av det aktuelle KI-systemet. For de fleste virksomheter er det “bruker” (deployer) som vil være mest aktuelt. Modellvalg og risiko-klassifisering av aktuell KI-løsning vil være vesentlig for virksomhetens risikoprofil og rammeverk for å håndtere slik risiko (governance).

2.6 Risiko ved modellvalg

Overordnet har en virksomhet som vil ta i bruk KI tre muligheter: 1) Utvikle en egen KI-modell, 2) anskaffe og bruke eksisterende modeller, 3) tilpasse eksisterende modeller. Vi vil gå nærmere inn på de tre mulighetene i resten av dokumentet.

Hvis virksomheten vil utvikle en egen KI-modell eller drive omfattende tilpasning av en eksisterende modell, må virksomheten ta stilling til risiko rundt tilgang til treningsdata, risiko ved trening av modellen, kostnader og risiko ved bruk av modellen.

Hvis virksomheten vil bruke en eksisterende modell, er det primært spørsmål rundt bruk som gjelder. I alle tilfeller vil både eksisterende regelverk og KI-forordningen måtte hensyntas.

Modell-valg sjekkliste:

1	Problemforståelse	Virksomheten må ha en klar problemforståelse og strategi for hva KI skal løse. Er modellen godt tilpasset problemet KI skal løse i virksomheten?
2	Ytelse versus forklarbarhet	Noen KI-modeller kan gi høy ytelse, men er ofte "black boxes" med lav forklarbarhet. I noen tilfeller kan det være

		nødvendig å ofre noe ytelse for å oppnå større forklarbarhet og tillit, spesielt i regulerte bransjer.
3	Ressurser og ekspertise	Har virksomheten de nødvendige ressursene og ekspertisen til å utvikle og vedlikeholde den valgte KI-modellen? Hvis ikke, kan det være lurt å velge hyllevare.
4	Tidsramme og kostnader	Implementering av teknologi tar tid og kan være ressurskrevende. Rask og kostnadseffektiv implementering av en KI-modell må veie opp mot betydningen KI vil kunne få for automatisering av oppgaver og strategisk betydning for virksomheten ved en mer helhetlig og skreddersydd tilnærming.
5	Skalerbarhet	Valg av KI-modell må la seg skalere på tvers av virksomhetens aktiviteter, både nå og i nær fremtid. Videre bør modellen være under konstant utvikling og forbedring.
6	Sikkerhet og personvern	Virksomheten må sørge for at KI-modellen overholder gjeldende sikkerhetsstandarder og personvernlovgivning, spesielt når det gjelder behandling av sensitive data.
7	Systemintegrasjon	KI-modellen må la seg integrere med virksomhetens eksisterende IT-systemer og prosesser for optimal bruk. Integrasjon vil avhenge av leverandørens standard integrasjoner og virksomhetens eksisterende IT infrastruktur.
8	Påvirkning og etikk	Virksomheten må vurdere den forventende påvirkningen av KI-modellen på kunder, ansatte og samfunnet.

Sjekklisten bør informere om man 1) utvikler egen KI-modell, 2) anskaffer og bruker eksisterende modeller, eller 3) tilpasser eksisterende modeller, og om man velger åpne eller lukkede modeller. Informasjon om tilpasning av eksisterende modeller og åpne og lukkede løsninger kan finnes i Tilpasning av grunnmodeller.

3 Treningsdata

Spørsmål knyttet til risiko ved treningsdata er først og fremst relevant ved to typer bruk:

1. Der virksomheten utvikler en ny KI-modell eller finjusterer eksisterende KI-modeller med konkret gjenkjennbare treningsdata for virksomhetens behov. Å utvikle eller finjustere en KI-modell krever at virksomheten anskaffer, håndterer og gjør tilgjengelig data for trening av modellen.

2. Der virksomheten går til anskaffelse av ekstern KI-modell som er trent på ukjente eller altomfattende data (typisk ved dataskraping fra internett). Å trene en modell på ukjente eller altomfattende data kan resultere i inngrep i andres (immaterielle) rettigheter.

Ved innsamling, håndtering og bruk av data må virksomheten ta tre typer hensyn:

i. Juridisk: Overholde gjeldende regelverk

ii. Etisk: Ta tilstrekkelige grep for å sikre at datainnsamlingen ivaretar rimelige hensyn

iii. Treffsikkerhet: Kvalitetssikre dataene, så modellen blir treffsikker nok til å oppnå formålet til virksomheten.

Som vi vil se kan de ulike hensynene kunne komme i konflikt med hverandre.

3.1 Juridisk risiko

Før virksomheten samler inn og/eller gjør data tilgjengelig for modellen i treningsøyemed bør virksomheten vurdere dataenes lovlige bruksområde. Dette kan følge av ulike rettslige grunnlag, typisk avtale med innehaver eller lovverket. Viktigheten av å vurdere lovlig bruk av dataene kan illustreres ved følgende: Dersom virksomheten ønsker å trene en modell som skal brukes i ett område og dataene kun lovlig kan brukes i et annet område, har virksomheten en rettslig utfordring ved bruk av dataene.

Virksomheten må sørge for en juridisk vurdering av dataene i det konkrete tilfellet. Det er imidlertid flere regelverk som regulerer ulike data.

3.1.1 Personopplysninger

Personopplysninger fra borgere i EU/EØS er regulert under General Data Protection Regulation (GDPR) og er gjennomført i norsk rett under personopplysningsloven. Personopplysninger er et vidt begrep og omfatter enhver opplysning om en identifisert eller identifiserbar fysisk person. Enhver bruk av personopplysninger må et behandlingsgrunnlag.

Enkelte data (sensitive personopplysninger, som opplysninger om en persons etniske opprinnelse, politiske ståsted, religion/filosofisk overbevisning, genetiske forhold, biometriske data, helse, seksuelle forhold og legning) er, som utgangspunkt, forbudt å bruke. Forordningen regulerer en rekke sider ved innsamling og bruk av personopplysninger og grundig forståelse av forordningen er viktig ved bruk av slike data.

3.1.2 Åpne-data direktivet

En del data er underlagt EUs åpne data-direktivet (Data Act). Det åpne data-direktivet skal legge til rette for tilgjengeliggjøring og deling og bruk av data mellom næringslivet og offentlig sektor. Dataforordningen regulerer hvem som kan bruke dataene som forordningen omfatter, eksempelvis data generert av «Internet of Things». Forordningen regulerer sider ved innsamling og bruk av data, slik som ulike kontraktsvilkår og datas interoperabilitet.

3.1.3 Immaterielle rettigheter

En del data kan være beskyttet av immaterielle rettigheter, slik som opphavsrett / databasevern og forretningshemmelighetsvernet under EUs bedriftshemmelighetsforordning, gjennomført i norsk rett ved forretningshemmelighetsloven. Som generell retningslinje kan virksomheten ta utgangspunkt i at data som ligger til grunn for en virksomhets forretningsmodell, eller som må antas å være sensitiv for en virksomhet som regel ikke kan antas å være fritt til å bruke uten nærmere avklaring. Ved såkalt «skraping» av data fra nett kan filen «robots.txt» gi verdifull informasjon om rettigheter til dataene, men dette innebærer ingen garanti for at dataene er til fri bruk.

3.1.4 Lisensvilkår

Selv om data kan være rettighetsbelagt kan de være tilgjengelige for bruk under ulike lisensvilkår, som enten følger generelle standarder eller som må fremforhandles konkret. Eksempler på generelle lisensvilkår er Creative Commons (CC), MIT og NLOD. Noen lisensvilkår tillater at dataene brukes til trening av en KI-modell, men dette må vurderes.

3.1.5 Diskriminering

Korrekt informasjon er ikke tilstrekkelig til å sikre at KI-modellen er i tråd med regelverket. Hvis data som dokumenterer diskriminering eller marginalisering av enkelte grupper eller individer brukes til trening av en KI-modell kan diskrimineringen forsterkes eller i det minste videreføres av modellen. Dette kan støte an mot lovverkets vern mot diskriminering.

3.2 Etisk risiko

Foruten de ulike juridiske hensyn som gjør seg gjeldende ved data til trening av en KI-modell, er det også ulike etiske hensyn som gjør seg gjeldende. Det kan være til dels overlapp mellom de juridiske og etiske hensynene. For eksempel er det viktig å sikre at den som personopplysningene gjelder har forstått hva dataene skal brukes til, slik at samtykket er informert og at dataene bukes i tråd med den registrertes rimelige forventninger.

Det er også problemer knyttet til skjevheter og svakheter i data, hvor data som dokumenterer diskriminering eller marginalisering kan videreføre eller forsterke slike trender i KI-modellen. Virksomheten har et ansvar for å motvirke dette.

3.3 Oppgaverisiko

3.3.1 Representativitet

Videre bør det vurderes om dataene som brukes er representative for det modellen skal brukes til. Dersom dataene som modellen trenes på er fra en annen populasjon enn det modellen skal anvendes på, risikerer virksomheten upresise eller feil resultat fra modellen. At dataene ikke er representative, betyr likevel ikke at de ikke bør brukes. Forventede forskjeller mellom trening og bruk burde kartlegges så dette kan hensyntas i trening, testing, bruk og kommunikasjon av modellen.

3.3.2 Dataforgiftning

Et annet hensyn som bør vurderes for en riktig anvendelse av dataene er hvorvidt treningsdataene kan være utsatt for såkalt «forgiftning» (training data poisoning). For eksempel kan sårbarheter i et spam-filter skapes for deretter å legge inn feilkategoriserte eksempler i treningsdatasettet. Slike svakheter kan utnyttes.

Det finnes flere teknikker for å introdusere feller og feil i datasett. Data som skrapes fra nett kan derfor ha redusert verdi. Dette bør alltid vurderes, og henger ofte sammen med spørsmålet om dataene er rettighetsbelagt (se over om juridiske hensyn).

4 Treningsmetode

Skal en KI-modell være nyttig for virksomheten og ikke ha uønskede og skadelige resultater, må den være treffsikker. Skal modellen bli treffsikker, må KI-modellens arkitektur, målfunksjon, og treningsdata, *til sammen*, være egnet. En må lage en teori om hva slags arkitektur og målfunksjon som best kan oppnå virksomhetens mål gitt treningsdataen man har.

4.1 KI-modellens arkitektur

Arkitekturen er designet til KI-modellen. Det beskriver egenskapene til en KI-modell som ikke endres under treningen. Valg av arkitektur bestemmer hvor god modellen kan bli, hvor mye regnekraft og data som kreves for å trene modellen og hva slags type feil modellen vil gjøre.

For eksempel har noen bildeanalysemodeller en innebygget antagelse av at piksler langt unna et område er irrelevante for å forstå hva som skjer i det området, og at de samme type relasjonene mellom nabopiksler gjelder i hele bildet. Andre bildemodeller ser alltid på hele bildet i sammenheng.

Enhver arkitektur har implisitte antagelser om formen på treningsdataen og problemet den skal løse.

Et godt arkitekturvalg er derfor tilpasset egenskaper ved dataen og målet med KI-modellen. KI-modellen kan da bli god på det som er viktig og dårlig på irrelevante ferdigheter ved å bruke treningsdataen, uten å kreve unødvendig mye regnekraft.

4.2 KI-modellens målfunksjon

En målfunksjon er et tall som definerer målet til en KI-modell. KI-modellen trenes til å enten maksimere eller minimere dette tallet. Store språkmodeller som ChatGPT trenes for eksempel til å minimere forskjellen på hva den tror er neste ord i en rekke tekster og hva som faktisk er neste ord. Det at en målfunksjon krever at man må redusere overordnede mål til ett enkelt tall gjør det vanskelig å sikre at målfunksjonen, og dermed KI-modellen, manifesterer virksomhetens ønsker.

Den valgte målfunksjonen kan føre til underprioritering av enkelte scenarioer. Det vil si at den fungerer bra i de fleste tilfeller, men feiler uproporsjonalt i enkelte situasjoner. Anvendt på personer, så kan det føre til indirekte diskriminering (disparate impact). Manglende presisjon vil også innebære en oppgaverisiko (f.eks. hvis modellen tar feil i avgjørelser som er spesielt viktige for virksomheten).

Foruten å sikre at modellen er treffsikker, må virksomheten også vurdere at all data og merkelapper (labels) som tilgjengeliggjøres for modellen er i tråd med gjeldende lovverk og etiske hensyn. Gitt formålet med modellen, hva er legitime beslutningsvariable? Er det for eksempel i tråd med diskrimineringslovgivning og etiske prinsipper å bruke data på rase til å trene modellen?

4.3 Behovet for testing

For å forbedre modellen og danne grunnlag for hensiktsmessig bruk, er god og løpende testing av modellen essensielt. Test teorien din for valg av arkitektur og målfunksjon, funker modellen slik du trodde den ville?

1. Kartlegg presisjon for ulik bruk: Sjekk om skjevhetene du kartla i datasettet har forplantet seg til modellen, og test for kjente feilmoduser for arkitekturen du valgte. Hvis en skal anvende algoritmen på persondata, må en teste for diskriminering basert på egenskaper som kjønn, etnisitet, nedsatt funksjonsevne, graviditet og permisjoner.[15]

2. Forklarbarhet: Prøv å forstå hvordan modellen opererer, for eksempel ved prinsipalkomponentanalyse. Er det mulig å forklare hvordan modellen kommer fram til resultatet? Hvis modellen brukes til å fatte avgjørelser som er relevante for personer, kan det

[15] For testing av diskriminering, se LDO.

være ulovlig, uetisk eller forretningsmessig risikabelt å ta avgjørelser basert på modellens resultat uten å forstå hvorfor eller hvordan modellen kom med sin vurdering. Bruk derfor trenings- og testfasen på å forstå modellen.

3. «Red teaming»: Prøv aktivt å bryte ned egen modell med målrettede angrep av mulige svakheter. Her burde det særlig fokuseres på å identifisere svakhetene i datagrunnlaget og kjente feilmoduser ved lignende modeller.

4. Standarder: Burde modellen sertifiseres etter en relevant standard? Standarder kan redusere behovet for selv å finne ut alt man burde teste for, gjøre det tryggere for kunder å bruke en KI-modell, samt bidra til mer gjennomsiktighet og bedre bransjepraksis.[16]

4.4 Tilpasning av grunnmodeller

En grunnmodell er en type KI-modeller som har gjennomgått omfattende generell trening, og derfor enkelt kan brukes eller tilpasses ulike formål. Mest kjent er store generative KI-modeller[17] som kan håndtere tekst, bilde, lyd og video. Det finnes også mer domenespesifikke grunnmodeller.

Hvis virksomheten tar utgangspunkt i en grunnmodell, burde en skaffe oversikt over hvordan modellen er trent. Modellkortet gir nyttig informasjon om modellen, treningsgrunnlaget, presisjon og anbefalt anvendelsesområde.[18] Tilpasning kan gjøres hovedsakelig på tre måter: fintrening, few-shot prompting og datatilgang.

4.4.1 Fintrening

Å fintrene en grunnmodell innebærer å trene modellen på mer data for å spesialisere modellen på akkurat det virksomheten ønsker å bruke den til. Siden grunnmodellen har trent så mye på forhånd, trenger den ofte bare litt ekstra trening for å spisses til virksomhetens formål. Fintrening kan derfor gi gode modeller for lite data og datakraft. Ofte låser en av store deler av arkitekturen i grunnmodellen før fintreningen, for å bevare kunnskapen modellen har fra før.

Om virksomheten fintrener modellen gjelder de samme hensynene som i seksjon 4.3, men sett i sammenheng med informasjonen om grunnmodellen. Målfunksjonen og treningsdataen for fintreningen og grunnmodellens arkitektur, presisjon og treningsgrunnlag må til sammen være egnet for å få en treffsikker modell.

4.4.2 Few-shot prompting

En enda enklere måte å tilpasse en grunnmodell på, er såkalt “few-shot prompting”. Før en brukerforespørsel kommer til grunnmodellen, limes det automatisk inn flere eksempler med forespørsler og gode svar. Eksemplene gjør at modellen kan forstå hva slags type svar den skal gi før den så forsøker å gi et slikt svar på forespørselen. Bare et par eksempler kan gjøre modellen mye bedre. Ingen trening kreves, kun et sett med eksempler på forespørsler og svar.

Few-shot prompting gir virksomheten økt kontroll over hva slags type svar modellen gir. Men man guider, ikke overstyrer, den underliggende modellen. Virksomheten bør derfor være bevisst på at den fortsatt vil oppføre seg som grunnmodellen er trent til. Eksemplene som brukes bør være gode, siden modellen vil benytte hvert eksempel for hver eneste forespørsel.

---

[16] Se Standard Norges side om KI-standarder.
[17] Se fotnote 4 for eksempler.
[18] Se oversikt fra Hugging Face.

4.4.3 Datatilgang

Å gi en grunnmodell datatilgang er nyttig for å inkorporere bedriftens egne data og kunnskap i en ellers generell modell. Dette kalles gjerne RAG ("Retrieval-Augmented Generation"). Det fungerer ved at en søkemotor slår opp i virksomhetens egen database ut fra brukerens prompt, og limer inn relevant informasjon sammen med promptet. Promptet som kommer til grunnmodellen inkluderer dermed både brukerforespørselen og kunnskapen som kreves for å svare på forespørselen. Modellen trenger bare å svare.

Med slik datatilgang blir det altså viktig å kontrollere at søkefunksjonen fungerer godt. Den må kun ha tilgang til data man ønsker å dele med brukeren, og burde ha en pålitelig oppslagsmetode. Det kan være nyttig for brukere å være klar over at det ikke er grunnmodellen selv som slår opp i databasen, men en separat søkefunksjon. Det hjelper altså ikke om grunnmodellen er smart om søkemotoren er dum. Modellen får bare tilgang til informasjonen søkemotoren limer inn.

4.4.4 Åpne eller lukkede modeller

Både fintrening, few-shot prompting og datatilgang kan anvendes på både åpne og lukkede modeller. Man bør imidlertid være svært nøye på hva man velger for å beskytte egne data, og bevare kontroll over egne systemer og kostnader.

Åpne modeller krever mer teknisk ekspertise og må driftes av virksomheten selv. Til gjengjeld har man kontroll over prosessen. Lukkede modeller kan trenes og driftes enkelt gjennom et API, men da har man ikke kontroll over databehandlingen eller tilgang til selve modellen. Med mindre det er eksplisitt utelukket av tilbyderen må man regne med at tjenester med lukkede modeller samler inn all dataen som går gjennom modellen. Lukkede modeller har gjerne en pay-per-use løsning, og blir raskt dyre ved mye bruk.

I KI-forordningen blir virksomheten juridisk regnet som tilbyder om virksomheten inkluderer modellen i et produkt under eget navn (se kapittel 1.5).

5 Bruk av KI

Før KI modellen tas i bruk, bør virksomheten ta stilling til mulighetene og risikoene knyttet til bruken. På det grunnlaget kan en vurdere hvordan det er hensiktsmessig å ta den i bruk og om den i det hele tatt bør brukes.[19]

For å sikre nyttig og ansvarlig bruk av KI er det viktig at:

1. Brukerne har god kunnskap om hva KI-modellene kan og ikke kan. Da er det mindre sjanse for at brukerne baserer seg for mye eller lite på modellen, eller bruker modellen på feil måte.

2. Måten KI brukes kan forsvares sikkerhetsmessig, etisk og juridisk. For eksempel bør virksomheten sette fokus på «etisk prompting», der brukere læres opp i hvordan spørsmål og kontekst fremstilles på en måte som reduserer risikoen for negativt utfall, som for eksempel diskriminerende resultat.

3. KI bare benyttes der fordelene ved anvendelse, som økt effektivitet, overstiger nedsidene, som kostnadene for mennesker, samfunn og miljø.

4. Virksomheten etablerer mekanismer for å kontrollere at KI brukes på måten de ønsker.

Ved ansvarlig bruk av KI kan samfunnets høye grad av tillit til næringsliv og offentlig sektor opprettholdes. Se for øvrig Vedlegg 5: Sjekkliste ved utvikling eller bruk av KI.

Vi kan skille mellom intern og ekstern bruk av KI-produkter. Intern bruk er for eksempel bruk i ansettelsesprosesser eller som analyseverktøy. Ekstern bruk er bruk i salg, lisensiering eller fronting av KI-produkter til selskaper og individer, inkludert brukergrensesnitt mot kunder som for eksempel når eksterne møter en chatbot.

Nedenfor følger liste over ulike risikokilder ved bruk av KI og hvilke tiltak virksomheten kan iverksette for å begrense effekten av den aktuelle risikokilden. En risikokilde kan medføre én eller flere av de fem typene risiko: juridisk risiko, etisk risiko, verdirisiko, oppgaverisiko eller disrupsjonsrisiko. Se for øvrig Vedlegg 6: Risiko longlist.

5.1 Risikokilder ved bruk av KI

Risikokilder	Beskrivelse	Tiltak
Ulovlig datahåndtering	Modellen samler inn data, prosesserer data eller genererer resultat ved bruk på en ulovlig måte eller som for øvrig påvirker virksomhetens juridiske risiko, f.eks. ved bruk av andres åndsverk, persondata, kontraktsforpliktelser m.m..	Sørg for rutiner for dataminimering, datatransparens, anonymisering, datasikkerhet, overholdelse av regelverk, opplæring av bruk, analyse av genererte resultat m.m. Vurder omdømmetap, erstatningsansvar, bøter og kontraktsrisiko ved bruk av slik KI-modell.
Uetisk datahåndtering	Modellen samler inn data, prosesserer data eller genererer resultat ved bruk på en måte som er lovlig ift. gjeldende regelverk og øvrig forpliktelser, men som likevel kan betegnes å være	Sørg for etisk og ansvarlig bruk av KI med tilhørende gode prosesser, definerte roller og oversikt, godt kunnskapsgrunnlag internt, god etterlevelse av KI-bruk opp mot regelverk, åpen KI-praksis overfor

[19] Se kapittel 1 for en oversikt over hvordan en bør foreta en slik risikovurdering

Risikokilder	Beskrivelse	Tiltak
	uetiske og i strid med virksomhetens eller samfunnets verdier.	kundene, beskyttelse av kundedata, leverandørkontroll, m.m.
Datalekkasje	Modellen fremviser uønsket datadeling, typisk som resultat av at modellen lekker treningsdata til bruker av modellen.	1. Kontrollere tilgang til modellen. 2. Kontrollere resultat fra modellen. 3. Se tilgjengeliggjøring av data under trening.
Ufrivillig datadeling	Virksomheten bruker en lukket modell hvor modellens tilbyder absorberer brukerens data ved bruk, som resulterer i at virksomheten deler data ved bruk som trener KI-modellen (eksempelvis åpen ChatGPT).	• Sjekk betingelsene for bruk. • Anskaff en versjon som tilbyr lukkede omgivelser for brukerens input/data (såkalt enterprise-versjon).
Forklaringssvikt	Manglende forståelse av modellen fører til at bruker ikke kan produsere juridisk eller etisk påkrevde forklaringer av beslutninger tatt av eller ved hjelp av modellen.	Gi brukere (og potensielle brukere) av modellen bedre forståelse for modellens funksjoner og resultat, hvordan disse bør tolkes, settes i sammenheng og ettergås av mennesker, og kommuniser viktigheten av å kunne forklare beslutningene som tas. Vurder alternativ KI-modell. Sjekklister for begrunnelser kan vurderes.
Svekket konkurranseposisjon	Potensielle negative konsekvenser som en virksomhet kan møte hvis konkurrentene utnytter KI mer effektivt, (f.eks. tap av markedsandel, redusert inntjening, osv.)	Vurder hvordan KI påvirker selskapets konkurransesituasjon og etabler en strategi for å møte konkurransen (f.eks. ved å øke kompetanse og forbedre produkter/tjenester for å sikre egen konkurranseevne).
Sårbarheter knyttet teknologien	Omfatter potensielle utfordringer ved implementering og drift av KI-systemer, inkludert feil eller mangler i systemene, sikkerhetssårbarheter og teknologiavhengighet.	Vurder risikoene nøye og etabler robuste risikostyringsstrategier for å sikre pålitelig bruk av KI i samsvar med gjeldende lover og etiske normer.
Tredjepartsrisiko	Virksomheten får dårlig omdømme, risikerer økonomisk tap eller svekket kredittverdighet ved å bruke KI (selv ved ansvarlig bruk), enten på grunn av feil, ulykker eller uansvarlig bruk av KI hos andre (f.eks. hos KI-leverandør).	Vær åpen og tydelig på hvordan virksomheten bruker KI, og hvilke retningslinjer og sikkerhets-mekanismer virksomheten har på plass for å sikre ansvarlig bruk. Sørg for grundige risikovurderinger, sikre at KI-system som brukes er robust og godt testet, øk kompetansenivået, følge nøye med på markedsforhold og stresstest modellen jevnlig.
Leverandør	Leverandøren av KI-løsningen vil kunne skape driftsforstyrrelser i virksomhetens operasjon og dets IT-system. Valg av KI-leverandør kan også gi en *lock-in* virkning, som binder virksomheten til valgt leverandør. Det kan påvirke selskapets kritiske	Vurder nøye hvilke(n) leverandør som velges, åpne eller lukkede modeller, hvordan de(n) passer inn med virksomhetens leverandører / systemer ellers og sårbarhet. Hensynta leverandørvilkårene og åpenhetslovens krav i vurderingen.

Risikokilder	Beskrivelse	Tiltak
	funksjoner og være vanskelig å endre ved en hendelse i fremtiden.	Se Tilpasning av grunnmodeller og Modellvalg.
Utilsiktet KI-oppførsel eller KI-bruk	Implementering av KI-løsninger fører til diskriminering, krenkelse av immaterielle rettigheter, uønskede kontraktsforpliktelser m.m.	Etabler klare retningslinjer og prosesser for etisk KI-bruk, sørg for oversikt over og god dokumenterbar etterlevelse av relevant lovgivning, hold oversikt og etabler ansvaret for KI-bruk, dokumenter rutiner og etterlevelse, og planlegg for implementering av KI-forordningen.
Bærekraft	Potensielle negative virkninger KI kan ha på miljømessige, sosiale og styringsmessige (ESG) aspekter av bærekraft, inkludert kraft- og ressursbruk, sosial påvirkning og styringsutfordringer.	Inkluder bærekraftsrisiko i KI-strategier og inkluder eksternaliteter i regnskap og rapportering, der bl.a. klimapåvirkning og ressursbruk (f.eks. bruk av datakraft), sosial påvirkning (f.eks. diskriminering og jobbtap), og påvirkning på beslutningsprosesser må inkluderes i ESG-rapportering.
Radikal disrupsjon	Virksomhetens eksistensgrunnlag kan endres eller i verste fall forsvinne: 1. fordi etterspørselen etter virksomhetens produkter eller tjenester svikter gjennom radikal disrupsjon, eller 2. transformative endringer i det marked og den kontekst virksomheten opererer i vil gjøre det umulig å fortsette som før.	Inkluder KI inn i strategien til virksomheten og vurder både kortsiktige og langsiktige implikasjoner av KI. Ta KI inn på agendaen til styret og vurder tiltak for å begrense disrupsjonsrisiko, slik som radikalt skifte av virksomhetens karakter, endring av tilbud av produkter og tjenester, systemiske endringer i infrastruktur, fokus på F&U, konsolidering og partnerskap, m.m.

5.2 Intern bruk

Ved intern bruk av en KI-modell bør det vurderes hvorvidt modellen vil bidra til å forbedre virksomhetens prosesser, og hvilke negative virkninger bruk av modellen kan føre med seg.

Risikokilder	Beskrivelse	Tiltak
Overforbruk og avhengighet *("Over reliance")*	For høy tillit til KI-modellen på ulike områder fører til systematisk dårligere beslutninger. Overforbruk av modellen fører til en avhengighet med anvendelser som KI-modellen er mindre egnet til og sårbarhet ved systemfeil. KI-modellen brukes til å fatte avgjørelser uten at bruker har tilstrekkelig forståelse for hvorfor modellen tar en gitt vurdering.	Sørg for kontinuerlig opplæring for bruk av KI og gi brukere (og potensielle brukere) av modellen og bedre forståelse for modellens svakheter og begrensninger. Hensynta viktigheten av menneskelig oppsyn og etterprøving av KI-genererte resultater. Diversifiser leverandørkjeden med alternative systemer og prosesser. Foreta kontinuerlige evalueringer og risikovurderinger, overvåk teknologien, sikre god datastyring, og etisk og juridisk etterlevelse.

Risikokilder	Beskrivelse	Tiltak
Underforbruk og undervurdering ("Under-reliance")	For lav tillit til KI-modellen fører til lite bruk eller dårligere bruk. Underforbruk kan skyldes overdreven eller ugrunnet skepsis, manglende forståelse, eller manglende trening eller erfaring.	Gi brukere (og potensielle brukere) av modellen (i) bedre forståelse for modellens funksjoner, kapasiteter og hvordan modellen kan brukes; (ii) innføring, trening og praksis i bruk av modellen, (iii) grunnlag for å vurdere misoppfatninger og begrensninger knyttet til KI.
Feiltolkning	Bruker kan misforstå hvorfor modellen tar de avgjørelsene modellen tar og hvordan modellens resultat skal tolkes. Bruker kan også oppfatte at KI-modellen tar avgjørelser modellen egentlig ikke tar og dermed reagere negativt på sviktende eller feilaktig grunnlag.	Gi brukere (og potensielle brukere) av modellen bedre forståelse for modellens funksjoner og leveranser, hvordan disse bør tolkes og settes i riktig sammenheng, og ettergås av mennesker.
Negativ spillover-virkning	Selv om modellen brukes riktig kan bruken innebære utilsiktet negativ systematisk påvirkning på virksomheten eller på samfunnet (f.eks. jobbtap, uro, komprimerende bruk av persondata, forsterkninger av diskriminerende praksis og miljøpåvirkning)	Løpende vurder de bredere konsekvensene av KI-strategien og bruk av KI-modeller, øk kompetansen blant ansatte, utvikle retningslinjer, kontroller for potensielle negative effekter og iverksett passende tiltak.
KI-avhengighet	Virksomheten blir avhengig av KI-modellen for å ta gode beslutninger. Det kan resultere i sårbarheter i form av at virksomheten varer eller tjenester eller interne prosesser kan forstyrres dersom KI-modellen skulle bli utilgjengelig.	Sørg for en balansert bruk av KI, vurder alternative KI-modeller og backup-løsninger, opprettholde menneskelig kunnskap, etabler robust kontroll- og overvåkningssystemer, m.m.
Operasjonelle problemer	Potensialet for tap som følge av feil, svikt eller mangler i interne prosesser, mennesker eller systemer, der KI-systemer er involvert (f.eks. tekniske feil, driftsavbrudd, sviktende datakvalitet og håndtering, sikkerhetsbrudd, manglende kunnskap og dårlig endringsledelse)	Sørg for grundig testing og validering av KI-systemer, kontinuerlig overvåkning og vedlikehold, effektiv datastyring, robuste sikkerhetstiltak, og god opplæring og støtte til ansatte. Sørg for et kontrollerbart KI-domene for virksomheten (f.eks. i egen beskyttet skyløsning) og ha en intern gruppe («*task force*») som raskt kan forstå oppståtte problemer og iverksette tiltak.
Systemsvikt	KI kan påvirke virksomhetens systemer negativt eller brukes til å avdekke svakheter som kan lede til datainnbrudd, nettverkseffekter med skadepotensiale på tvers av funksjoner/infrastruktur, m.m.	Virksomheter og myndigheter må samarbeide for å utvikle standarder og protokoller for KI-sikkerhet, overvåkning og kontroll av KI-systemer, samt beredskapsplaner for håndtering av mulige systemiske feil.

5.3 Ekstern bruk

Når en KI-modell tilbys som del av en tjeneste til tredjeparter (marked eller til brukere eller borgere i samfunnet for øvrig (f.eks. offentlige tjenester), kan det oppstå særlige utfordringer.

Risikokilder	Beskrivelse	Tiltak
Redusert kvalitet på vare / tjeneste	KI vil kunne redusere kvaliteten på en vare eller tjeneste ved at den blir mer generisk eller får lavere kvalitet med tilhørende produktansvar og svekket markedsposisjon.	Vurder virkning ved bruk av KI og hvordan dette påvirker varen eller tjenesten. Utarbeid oversikt over virksomhetens viktige særegenheter og risiko for forflatning i disse. Vurder også hvilket ansvar virksomheten har etter forbrukerlovgivningen og EUs produktansvarsdirektiv m.fl.[20]
Feil-informasjon	Virksomheten har et ansvar for at korrekt informasjon om produktet eller tjenesten gis, som forsterkes med KI-forordningen.	Vurder hvordan produktet eller tjenesten omtales utad / markedsføres og hvilket ansvar virksomheten har etter bl.a. markedsføringsloven. Merk KI-genererte resultat (*watermark*).
Negative samfunns-virkninger	Det kan skyldes at den misbrukes, eller at bruken har utilsiktede konsekvenser, eller at den inngår i en bruk som gir systemiske problemer.	Få på plass et system for å overvåke om systemet fyller funksjonen som ment og hvordan risikoen for misbruk kan reduseres.
Negativ påvirkning på personer	Produktet eller tjenesten vil ved bruk av KI kunne påvirke personer negativt, f.eks. gjennom falske bilder eller nyheter (deepfakes)	Få oversikt hvordan produktet eller tjenesten vil kunne (mis)brukes og hvilke risiko-reduserende tiltak som kan iverksettes for å hindre slik bruk.
Redusert cyber-sikkerhet	Bruk av KI-modeller kan gi opphav til nye sikkerhetssårbarheter. Noen eksempler på dette kan være: • "Prompt injections": Tredjepartsangrep ved hjelp av spesialdesignede prompts (enten direkte eller indirekte via en nettside/fil systemet leser), som resulterer i uønskede konsekvenser. • "Model denial of service": Tredjepart sender forespørsler til KI-modellen som tar så mye ressurser at det hindrer andre i å bruke tjenesten. • "Supply chain vulnerabilities": KI-modeller (særlig språkmodeller) medfører nye avhengigheter som er vanskelig å oppdage. Dette kan gjelde alt fra treningsdataen, til grunnmodellen eller plugins. • For mer omfattende oversikter over slike former for risiko, se OWASPs topp 10	Utviklere og tilbydere bør danne seg oversikt over nye sårbarheter og hvordan de kan motvirkes. National Cyber Security Center (NCSC) har utformet retningslinjer for sikker bruk av KI i samarbeid med Norge og 17 andre land.[22]

[20] Forslaget er til behandling i EU

[22] se Guidelines for secure AI system. Se også OWASPs AI Security and Privacy Guide development.

Risikokilder	Beskrivelse	Tiltak
	trusler for LLMs og for informasjonssikkerhet generelt.[21]	
Modelltyveri	Tyveri, eller kopiering av virksomhetens KI-modell.	Sørg for å vurdere behovet for utvikling av robuste sikkerhetsrutiner rundt tilgang til modellen, oversikt over svakheter som kan resultere i modelltyveri, samt tiltak dersom dette skulle inntreffe. Vurder sammenligning og eventuelt samkjøring med virksomhetens rutiner for immaterielle rettigheter og forretningshemmeligheter
Datatyveri	Tyveri av virksomhetens data, eller data i eller til grunn for modellen.	Vurdere behovet for utvikling av robuste sikkerhetsrutiner rundt tilgang til data/forretningshemmeligheter og oversikt over svakheter som kan resultere i datatyveri, samt tiltak dersom dette skulle inntreffe. Vurder sammenligning og eventuelt samkjøring med virksomhetens rutiner for personopplysninger.
Nedetid	En tjeneste kan grunnet integrasjon med en KI-modell bli mindre robust, og dette kan resultere i økt nedetid, eller dårligere kvalitet	Vurder kartlegging av risiko for svakheter som følge av integrasjoner, oversikt over tiltak som kan forebygge eller reparere slike svakheter dersom de skulle oppstå

De ulike risikokildene og avbøtende tiltak må vurderes i sammenheng med den faktiske utvikling eller bruk av KI. Dagens fokus på store språkmodellene og generativ KI, gir et behov for å identifisere spesifikke kilder til risiko for denne teknologien. Dette kan være:

i) Misbruk: uetisk eller ulovlig utnyttelse av generativ KI (f.eks. svindel eller bruk av deepfakes for spredning av desinformasjon)

ii) Feilbruk: ufullstendige eller feil resultat (f.eks. der modellen hallusinerer)

iii) Uriktig fremstilling: resultatet fra generativ KI brukes og spres til tross for usikkerhet eller manglende troverdighet (f.eks. videoer med usikkerhet knyttet til kilde og bruk av KI),

iv) Uhell: resultatet fra generativ KI spres uten at brukeren er klar over feil eller manglende troverdighet (f.eks. spredning av video som tilsynelatende er ekte, men som viser seg i ettertid å være laget ved bruk av KI).[23]

Virksomheten må i lys av den raske teknologiske utviklingen sørge for kontinuerlig oversikt over hvordan KI påvirker selskapet og i hvilken grad de fem ulike risikoene materialiserer seg. En slik fremgangsmåte vil kunne gi en ansvarlig bruk av KI.

[21] Top 10 trusler for LLMs og Top 10 informasjonssikkerhet.
[23] Oversikt fra Öykü Isik, Amit Joshi, and Lazaros Goutas: "*4 Types of Gen AI Risk and How to Mitigate Them*" Harvard Business Review (31.5.2024), se henvisning i ressurslisen vedlegg 7.

Vedlegg 1: Relevante brukergrupper– hvem retter Standarden seg mot?

Vi mener Standarden vil være relevant for et bredt spekter av norske virksomheter, både private og offentlige aktører, i deres eget arbeid med å utvikle og bruke KI ansvarlig. Vi har identifisert følgende funksjoner som kan ha god nytte av Standarden:[24]

(i) *Bedriftsledere, styremedlemmer og andre beslutningstakere*: Denne gruppen trenger å forstå de bredere implikasjonene ved å implementere KI i sine operasjoner, inkludert etiske overveielser, virksomhetsmessige risikoer og samsvar med regelverket.

(ii) *Compliance- og juridiske team*: Disse funksjonene må sørge for at KI-applikasjoner overholder relevante lover, standarder og etiske normer.

(iii) *Etikk- og KI-styringsteam*: Organer innenfor organisasjoner dedikert til å håndheve etisk bruk av KI.

(iv) *IT-sjefer / Produktsjefer*: De har oversikt over utviklingen av IT-relaterte systemer / produkter, inkludert KI-produkter/tjenester, fra konsept til lansering. De trenger å forstå hvordan KI-funksjoner og -risikoer påvirker produktstrategi og brukeropplevelse, og hvordan KI-systemer påvirker IT-arkitekturen for øvrig og systemrisiko. Dette inkluderer cyber-sikkerhet og reguleringer som EUs Digital Operational Resilience Act (DORA) og EUs NIS 2 Direktiv.

(v) *KI-utviklere og -ingeniører*: Dette er de tekniske teamene som designer, bygger og vedlikeholder AI-systemer. De trenger retningslinjer for hvordan de kan implementere etiske prinsipper i koden sin og forstå risikoene forbundet med KI-systemene de utvikler.

(vi) *Spesialister innen risikostyring*: De har ansvaret for å identifisere, vurdere og redusere risikoer forbundet med KI-systemer. De søker etter retningslinjer om potensielle KI-risikoer og hvordan å integrere risikovurderinger i forretningsprosesser.

(vii) *Dataforskere*: Personer involvert i dataanalyse og bygging av maskinlæringsmodeller. De krever en forståelse av implikasjonene av dataene de bruker og hvordan man kan designe modeller som er rettferdige og åpne.

(viii) *Forretningsfunksjoner*: De kan bruke KI i sine funksjoner i virksomheten (f.eks. finansfunksjonen, HR-funksjonen, salgs- og markedsføringsteam, m.m.), og må være oppmerksomme på hvordan teknologien brukes, potensielle hallusinasjoner og bias'er (dvs. fordommer og diskriminerende praksis) og i hvilken grad persondata og / eller personers rettigheter blir påvirket (som krever vurderinger etter GDPR og KI-forordningen).

(ix) *Sluttbrukere og kunder*: De endelige mottakerne av KI-drevne produkter og tjenester. De drar nytte av å forstå hvordan KI fungerer, åpenhet om at KI brukes i tjenesten de benytter (f.eks. ved bruk av kundesupport), KIs begrensninger og potensielle risikoer (f.eks. knyttet til persondata eller andre rettigheter) slik at de kan bruke disse produktene ansvarlig.

(x) *Frivillige organisasjoner:* Organisasjoner som fokuserer på forbrukerrettigheter og samfunnsmessige effekter av teknologi, som søker å forstå og påvirke utviklingen av etiske AI-retningslinjer, samt bruk av teknologi for å nå sitt formål mer effektivt (f.eks. effektiv fordeling av midler til organisasjonens målgruppe).

[24] Listen er ikke ment å være uttømmende

Vedlegg 2: Nøkkelbegreper og definisjoner

Nedenfor er en liste med nøkkelbegreper og definisjoner som er relevante for forståelse av denne Standarden:[25]

Begrep	Definisjon
Kunstig intelligens (KI) & KI-drevne systemer: *(«AI systems»)*	Kunstig intelligens, eller et KI-system, er et maskinbasert system, som opererer med varierende grad av selvbestemmelse og kan tilpasses etter implementering. Et KI-system styrer etter mål og genererer resultater fra inputen det mottar (f.eks. et prompt), i form av prediksjoner, innhold, anbefalinger eller beslutninger som kan påvirke fysiske eller virtuelle miljøer.[26]
Generelle KI-modeller: *(«General Purpose AI»)*	KI-modell som viser betydelig generell evne og er i stand til kompetent å utføre et bredt spekter av forskjellige oppgaver.[27]
Stor språkmodell: *(«Large Language Model – LLM»)*	En generativ KI-modell som er i stand til å forstå og generere naturlig språk og andre typer innhold for å utføre et bredt spekter av oppgaver. LLM er trent på enorme mengder data, noe som gjør dem i stand til å gjenkjenne komplekse mønstre i eksisterende innhold og generere nytt innhold. En LLM vil i noen tilfeller kunne klassifiseres som en generell KI-modell, men ikke nødvendigvis.
Generativ KI	En type KI som basert på store mengder treningsdata kan generere nytt innhold, som tekst, bilder, videoer, kode og annen type data, ved basert på forespørsler fra brukere.[28]
KI-resultat *(«output»)*	Resultat eller «output» som genereres av en KI-modell basert på forespørsel fra brukeren (såkalte «prompts»).[29]
«Prompts» / «input»	Forespørsel fra brukeren («input») til en KI-modell om å generere et resultat («output») basert på KI-modellens treningsdata.
Hallusinasjoner	Når en stor språkmodell (LLM) skaper KI-resultat som er unøyaktige eller feil.[30]
Skjevhet *(«bias»)*	Mangelfulle eller feil/skjeve resultater som stammer fra de opprinnelige treningsdataene eller KI-algoritmene.[31] Skjevet kan føre til brudd på forbudet mot likhets- og diskrimineringsprinsippet forankret i Grunnloven § 98 og i likestillings- og diskrimineringsloven.[32]
«Deep Fake»	KI-generert eller manipulert bilde, lyd eller videoinnhold som ligner på eksisterende personer, objekter, steder, enheter eller hendelser og som vill fremstå for en person som autentisk eller sannferdig.[33]
Ansvarlig KI	Sett med prinsipper som gir veiledning om utvikling, implementering og bruk av KI for å redusere risiko, som i EYs rammeverk er

[25] Ikke en uttømmende liste over relevante begreper for KI-systemer og ikke alfabetisk rekkefølge.
[26] EUs AI Act artikkel 3 (1)
[27] EUs AI Act artikkel 3 (63)
[28] Se f.eks.: Generative Artificial Intelligence (AI) | Harvard University Information Technology
[29] Basert på OECD: Updates to the OECD's definition of an AI system explained - OECD.AI
[30] Basert på IBMs artikkel: What Are AI Hallucinations? | IBM
[31] Basert på IBMs artikkel: What Is AI Bias? | IBM
[32] Se også Likestillings- og diskrimineringsombudets KI veileder
[33] EUs AI Act artikkel 3 (60)

Begrep	Definisjon
	ansvarlighet, rettferdighet, tillit, åpenhet, forklarbarhet, datasikkerhet, juridisk etterlevbarhet og bærekraft.[34]
Etisk KI	For eksempel: (i) Respekt for menneskets selvbestemmelse, ii) hindre skade, iii) rettferdighet og iv) forklarbarhet.[35]
Risiko	Kombinasjonen av sannsynligheten for forekomst av skade og alvorlighetsgraden av skaden.[36]
KI-forordningen	EUs AI Act som oppnådde politisk enighet i desember 2023 og som endelig vedtatt i EUs parlament mars 2024 og av EU Kommisjonen 21. mai 2024, med siste språklige versjon av 19. april 2024.
KI-kyndighet *(«AI literacy»)*	Ferdigheter, kunnskap og forståelse som lar leverandører, brukere og berørte personer, som hensyntar deres respektive rettigheter og forpliktelser etter KI-forordningen, gjøre en informert implementering av KI-systemer, så vel som å være bevisst mulighetene og risikoene ved KI og mulig skade som det kan forårsake.[37]
Immaterielle rettigheter *(«IPR»)*	Immaterielle rettigheter kan forstås som *kreativt tankearbeid*, og retten til å bestemme over og ta eierskap til verdien i arbeidet.[38] Immaterielle rettigheter inkluderer åndsverk[39] med tilhørende opphavsrett, patenter, varemerker, forretningshemmeligheter, m.m.
Virksomhet	Juridisk person, som produserer eller tilbyr varer eller tjenester. En skiller mellom private og offentlige virksomheter, kommersielle, ikke-kommersielle og ideelle.[40]

[34] Basert på EYs *'Responsible AI Framework'*: EY's commitment to ethical and responsible AI principles | EY - Global
[35] Basert på EUs Ethics Guidelines for Trustworthy AI fra 2019: Ethics guidelines for trustworthy AI - Publications Office of the EU (europa.eu)
[36] Se EUs AI Act artikkel 3 (2)
[37] EUs AI Act artikkel 3 (56)
[38] Se Patentstyret med henvisninger: Hva er immaterielle verdier og rettigheter? - Patentstyret
[39] Se Åndsverksloven
[40] Som definert i Store Norske Leksikon: virksomhet – Store norske leksikon (snl.no)

Vedlegg 3: Strategi-sjekkliste

1	Formålet	Virksomheten må klart definere hva den ønsker å oppnå med KI, og hvordan utvikling eller bruk av KI henger sammen med virksomhetens strategi. Dette kan inkludere effektivisering eller automatisering av oppgaver, forbedring av kundeservice, innovasjon av produkter eller tjenester, eller å skape nye forretningsmuligheter.
2	Utvikle, kjøpe eller vente	Virksomheten må bestemme om den skal bygge KI-kompetanse internt, kjøpe eksterne tjenester fra KI-leverandører eller vente til KI i enda større grad blir hyllevare. Dette valget vil avhenge av virksomhetens eksisterende kompetanse, ressurser og langsiktige mål.
3	Datastrategi	KI er avhengig av data for å lære og fungere effektivt. Virksomheten må utvikle en datastrategi som sikrer tilgang til høykvalitetsdata, samtidig som den overholder personvernlovgivning og etiske retningslinjer.
4	Modellvalg	Dersom virksomheten ønsker å bruke en modell som er utviklet av andre, finnes det ulike alternativer: en kan lisensiere en lukket (closed) KI-modell, laste ned en åpen KI-modell (open source), integrere (embedded) KI direkte i annen maskinvare eller programvare, eller abonnere (SaaS) på KI-løsninger via en nettside eller applikasjon uten å integrere løsningen i selskapets skytjeneste. Se kapittel X for mer om avveiningene i denne forbindelse.
5	Etikk og ansvarlighet	Det er viktig å vurdere de generelle etiske implikasjonene av KI-bruk, inkludert diskriminering og personvern. Virksomheten bør etablere retningslinjer for ansvarlig bruk av KI og mekanismer for åpenhet og ansvarlighet.
6	Regulatorisk overholdelse	Virksomheten må vurdere den regulatoriske konteksten av utvikling eller bruk av KI, både i forhold til eksisterende teknologi-nøytral lovgivning og fremtidig spesifikk lovgivning knyttet til kunstig intelligens.[41] For offentlige virksomheter vil dette punktet også innebære spørsmålet om de har nødvendig kompetanse og ressurser til å utforme regler, overvåke og håndheve reguleringene.
7	Kompetanse og opplæring	Virksomheten må investere i opplæring og utvikling av ansatte for å bygge nødvendig kompetanse for å utvikle, administrere og vedlikeholde KI-systemer.
8	Integrasjon med eksisterende systemer	KI kan og muligens bør integreres sømløst med virksomhetens eksisterende tekniske infrastruktur. Dette krever strategisk planlegging for å sikre kompatibilitet og ivaretakelse av IT-sikkerheten.
9	Skalerbarhet	Virksomheten må vurdere hvordan KI-løsninger kan skaleres for å møte fremtidig vekst og endrede behov.

[41] Som eksempel vil bruk av KI for å prise forsikringsprodukter kunne være i strid med forsikringsavtaleloven fordi forsikringsselskaper skal sikre "korrekt pris" fremfor høyest mulig pris.

10	Måling av suksess	Effekten av KI-initiativer bør måles og strategien justeres basert på resultatene.
11	Langsiktig visjon	KI-teknologi utvikler seg raskt, og virksomheten bør ha en langsiktig visjon for hvordan den kan tilpasse seg nye muligheter og risikoer som oppstår.

Vedlegg 4: Oversikt over reguleringer internasjonalt

EU har vedtatt den mest omfattende KI-reguleringen. Sammenholdt med Personvernforordningen (GDPR) søker EU å beskytte borgernes fundamentale rettigheter og personvernet. Se mer om EUs KI-Forordning over under punkt

USA har på sin side så langt vedtatt et «Blueprint for AI Bill of Rights» fra oktober 2022. Dette danner grunnlag for et fremtidig lovforslag for regulering av KI i USA. Det er basert på et sett med retningslinjer for ansvarlig KI, basert på følgende fem prinsipper: i) Trygge og effektive systemer, ii) beskyttelse mot algoritme-diskriminering, iii) personvern, iv) informasjon og forklarbarhet, v) menneskelige alternativer, vurderinger og sikkerhetsnett.

Videre fremmet syv ledende KI-selskaper i USA i juli 2023 en frivillig forpliktelse til amerikanske myndigheter om å fokusere på trygg, sikker og åpen utvikling av KI-teknologi.[42] Amerikanske myndigheter har i etterkant (oktober 2023) utstedt en presidentordre[43] som reflekterer forpliktelsen til selskapene om trygg, sikker og tillitfull utvikling og bruk av KI, som ikke innebærer noen juridiske krav overfor selskaper men danner grunnlag for utvikling av retningslinjer og standarder relatert til KI. Det er likevel ingen forventning om vedtakelse av omfattende KI-reguleringer i USA med det første, som gjør at enkelte delstater har begynt å forberede egne KI-reguleringer på delstatsnivå, slik som Virginia.[44]

Andre land som Storbritannia, Canada, Kina, Japan, Korea, Singapore, m.fl. har vedtatt ulike former for KI-reguleringer. Basert på reguleringene kan vi kan identifisere noen tydelige regulatoriske trender:

- Reguleringene og retningslinjene fra myndigheter er generelt i tråd med prinsipper for ansvarlig KI utviklet av OECD og bekreftet av G20-landene.
- De ulike jurisdiksjonene har generelt tatt en risiko-basert tilnærming til KI-reguleringer. Dette betyr at de skreddersyr reguleringene mot de forventede KI-risikoene, slik som personvern, ikke-diskriminering, åpenhet og sikkerhet.
- På grunn av de ulike bruksområdene til KI, fokuserer noen jurisdiksjoner på sektor-spesifikke regler i tillegg til sektor-agnostisk fremgangsmåte.
- KI-regler utformes i kontekst av andre digitale prioriteringer, slik som cyber-sikkerhet, personvern, immaterielle rettigheter – med EU som ledende aktør.
- Flere jurisdiksjoner tar i bruk regulatoriske sandkasser som et verktøy for privat sektor til å samarbeide med politiske beslutningstakere. Dette for å utvikle regler som både møter målet om ansvarlig KI og samtidig hensyntar implikasjonene ved innovasjoner med høy risiko forbundet med KI.

I tillegg bør regulerende myndigheter, så langt som mulig, engasjere seg i multilaterale prosesser for å gjøre KI-regler mellom jurisdiksjoner harmoniserte og sammenlignbare for å minimere risikoen forbundet med regulatorisk arbitrasje. Dette er spesielt viktig når man vurderer regler som styrer bruken av en global teknologi som KI. Som eksempel ser vi betydelige forskjeller mellom EU, USA og Kina hva gjelder bruken av persondata ved utvikling og bruk av KI-modeller.

[42] Dette er Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI, se mer her.
[43] Presidentordren kan leses her.
[44] Som fokuserer på bruk av KI hos offentlige myndigheter, les mer her.

Vedlegg 5: Sjekkliste ved utvikling eller bruk av KI

Relevante spørsmål virksomheter bør stille seg ved utvikling eller bruk av KI:[45]

1. Har virksomheten oversikt over og informasjon om teknologien og dets risiko-kilder?
2. Har virksomheten rett kompetanse til å utvikle eller ta i bruk teknologien?
3. Har virksomheten organisert seg hensiktsmessig for hensynta utvikling, bruk og påvirkning som treffer virksomheten som følge av KI?
4. Hvilken strategi har virksomheten for utnyttelse av KI, og hvordan henger denne sammen med strategien for virksomheten som helhet?
5. Er ansvar/roller for å sikre oversikt over utvikling og bruk av KI fordelt i virksomheten?
6. Påvirker bruk av KI virksomhetens markedsposisjon?
7. Hva er formålet ved bruk av KI; effektivisering internt eller øke salg mot kunder?
8. Veier de antatte fordelene ved bruk / utvikling av KI tyngre enn de potensielle risikoene og ev. risikoreduserende tiltak som må / bør iverksettes?
9. Hvilken KI-modell skal benyttes som basis for virksomhetens bruk eller videreutvikling?
10. Skal open eller closed source modell benyttes, og hvilke forhold må tas i betraktning i lys av dette valget?
11. Hvordan bør virksomheten skille på interne- og tredjepartsmodeller for å kontrollere usikkerheten?
12. Hvilke reguleringer / krav gjelder for virksomhetens bruk eller utvikling av KI, hensyntatt bransje, type virksomhet, produkter / tjenester og bruk av data?
13. Hvordan klassifiseres virksomhetens bruk / utvikling av KI under KI-forordningen, og hvilke krav gjelder for tjenestene som brukes / utvikles?
14. Har virksomheten reflektert over de sikkerhetsmessige implikasjonene bruk av KI har?
15. Har virksomheten vurdert hvor data lagres og hentes fra?
16. Hvilke muligheter får en trusselaktør med tilgang til deres KI-verktøy?
17. Hvem har, og kan få, tilgang til informasjonen KI-modellene prosesserer?
18. Hvilke kontrollmekanismer ligger på plass for å avdekke feil / misbruk / mangelfull / ulovlig / diskriminerende bruk av KI?

[45] Ikke en uttømmende liste

19. Hvilke prosesser er på plass for å hensynta løpende risikoer ved utvikling eller bruk av KI (governance) og er denne tilpasset faktisk bruk / utvikling?

    (i) Hvilke tiltak er satt i verk for å begrense risiko? Hvordan sikre kontinuerlig forbedring i tråd med bruk og teknologisk utvikling for å avdekke mangler eller svakheter ved tiltakene?

    (ii) Vil KI kunne skape et avhengighetsforhold til en enkelt KI-leverandør som påvirker virksomheten og dets langsiktige leverandørrisiko? Hva skjer ved eventuell nedetid på KI-løsningen / modellen?

20. Hvem sikrer passende opplæring og forståelse hos brukerne?

21. Kan vi skape en kultur for etisk prompting / bruk av KI for å begrense potensielle risiko ved diskriminering?

22. Har vi hensyntatt brukere i virksomheten som har nedsatt funksjonsevne og hvordan denne gruppen kan få ta del i gevinstene ved bruk av KI, og eventuelt om tilpasninger er nødvendig?

23. Vil bruk av KI ha en negativ påvirkning på miljøet / klimamål virksomheten har satt?

24. Er det landspesifikke / kulturspesifikke forhold som bør hensyntas ved bruk av KI?

Vedlegg 6: Risiko longlist

Følgende liste over risikoer kan identifiseres ved utvikling eller bruk av KI:[46]

1. Manglende eller feil kunnskap om bruk eller effekten av KI
2. Manglende eller feil kommunikasjon om bruk eller effekten av KI
3. Manglende eller feil datagrunnlag (KI-systemet kan være trent på manglende eller feil / ikke-representative data)
4. Manglende forståelse av KI-systemets resultat
5. Datalekkasje (via hack eller at data på andre måter kommer på avveie)
6. Tillitbrudd (bruk av KI utover det systemet kan svare for og / eller sine funksjoner/styrker)
7. Feiltolkning (tolkning av KI-systemets genererte resultat på en gal måte)
8. Systemfeil og Hallusinasjoner (KI-systemet kan ha systemfeil, som bl.a. kan lede til hallusinasjoner og andre feil-genererte resultat)
9. Tap av arbeidsfunksjoner (effektivisering kan lede til tap av arbeidsfunksjoner og skape uro blant enkelte arbeidsfunksjoner / yrker og fagforeninger)
10. Overvåkning (AI blir et overvåkningssystem, som kan være vanskelig for både samfunnet og den enkelte å kontrollere)
11. Maktkonsentrasjon (for mye makt konsentreres hos enkelte selskaper / personer)
12. Plagiat (AI-systemer kan lede til brudd på opphavsrett og andre immaterielle rettigheter)
13. Diskriminerende resultat / bias (KI-resultater kan forsterke historiske data / biaser i generering av nye resultat)
14. Diskriminerende bruk (bruk av KI tilpasses ikke personer med nedsatt funksjonsevne som potensielt mister effektivitetsgevinster som tilgjengeliggjøres for resten av virksomheten / samfunnet)
15. Objektivfunksjon som ikke overlapper med bedriftens mål
16. Omdømmerisiko (utvikling eller bruk av KI kan gi betydelig omdømmetap, ved f.eks. feil bruk av KI-resultat, brudd på opphavsrett, bruk av personers/kunders data, m.m.)
17. Etisk risiko (f.eks. bruk av KI som bryter samfunnets eller selskapets verdier og / eller normer, uten at dette nødvendigvis bryter gjeldende rettsregler)
18. Juridisk risiko (ved brudd på relevant lovgivning risikeres både administrative reaksjoner og straff, samt sivilt erstatningsansvar både for virksomheten og for sentrale personer slik som daglig leder og styremedlemmer). Dette vil kunne innebære brudd på:
    - Personvernlovgivningen

[46] Ikke uttømmende og i tilfeldig rekkefølge

- Eiendomsrett til data / databasevernet
- Vern om konfidensiell informasjon / forretningshemmeligheter
- Immaterielle rettigheter og åndsverk
- Diskrimineringslovgivningen
- Forbrukerlovgivning
- Markedsføringslovgivning/informasjonsansvar og produktansvarslovgivning
- Arbeidsrettslovgivning (f.eks. hvordan KI brukes, opplæring av de ansatte, diskriminering i ansettelsesprosessen m.m.)
- Konkurranselovgivning (f.eks. prissettingsalgoritmer, markedsdominans, misbruk av datasett m.m.)
- Erstatningsrett
- Finanslovgivning (f.eks. ved av eksterne KI-løsninger / modeller i kritiske funksjoner uten notifikasjon til Finanstilsynet og nødvendige kontrollrutiner)
- Sikkerhetslovgivning (f.eks. cybersikkerhet m.m.)
- Straffelovgivningen
- Fremtidig lovgivning som EUs KI-forordning som forventes implementert i norsk rett, og EUs foreslåtte Ansvarsdirektivet for KI-systemer som muligens vil vedtas og implementeres i fremtiden
- Kontraktsvilkår

Vedlegg 7: Ressurser

Vi har brukt følgende ressurser for å utvikle denne standarden:

Norske kilder:

- Norges posisjonsnotat til EU Kommisjonen vedrørende KI-forordningen
- Digitaliseringsdirektoratets veileder
- Likestillings- og Diskrimineringsombudets veileder

EU-kilder:

- Artificial intelligence act | Think Tank | European Parliament (europa.eu)
  - Se også: The AI Act Explorer | EU Artificial Intelligence Act
- Liability Rules for Artificial Intelligence - European Commission (europa.eu)
- ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf (europa.eu)
- Policy and investment recommendations for trustworthy AI - Publications Office of the EU (europa.eu)
- Digital Operational Resilience Act (DORA) - European Union (europa.eu)
- The NIS2 Directive: A high common level of cybersecurity in the EU | Think Tank | European Parliament (europa.eu)
- GDPR og Personopplysningsloven
- Data Act

Andre internasjonale kilder:

- UNESCOs anbefalinger for etisk kunstig intelligens
- OECD:
  - Prinsipper for kunstig intelligens
  - Rammeverk for klassifisering av AI-systemer
- NIST AI Risk Management Framework (AI RMF)
- DNVs anbefalinger for KI-systemer (DNV-RP-0671)
- Task Force on responsible AI | MiT, Stanford & EY
- OWASP (Sikkerhet):
  - Top 10 informasjonssikkerhet
  - Top 10 for LLMs
  - AI Security and Privacy Guide
- National cyber security center (NCSC) + 18 land: Guidelines for secure AI system development
- Öykü Isik, Amit Joshi og Lazaros Goutas: 4 Types of Gen AI Risk and How to Mitigate Them (31.5.2024, Harvard Business Review)

Annen bakgrunnslitteratur:

- Inga Strümke: Maskiner som tenker - algoritmenes hemmeligheter og veien til kunstig intelligens (2023, Kagge Forlag)