AI Governance – Styring i høy hastighet

KPMG AS

12.06.2024

Til stede fra KPMG

Nicolai Cappelen

Partner,

Risk & Regulatory

Thea Gullaug

Manager,

Risk & Regulatory

Agenda

Intro til AI – Hva er mulighetsrommet?

AI Act – Status og påvirkning

AI Governance

Internrevisjonens rolle

01

Intro til AI – Hva er mulighetsrommet?

Kunstig intelligens kan påvirke din virksomhet på ulike måter

Kunstig intelligens kan føre til både økt effektivitet og innsikt. Den nye teknologien vil imidlertid også føre med seg nye risikoer og behov for tilpasning.

Effektivisering

Endring av prosesser og systemer

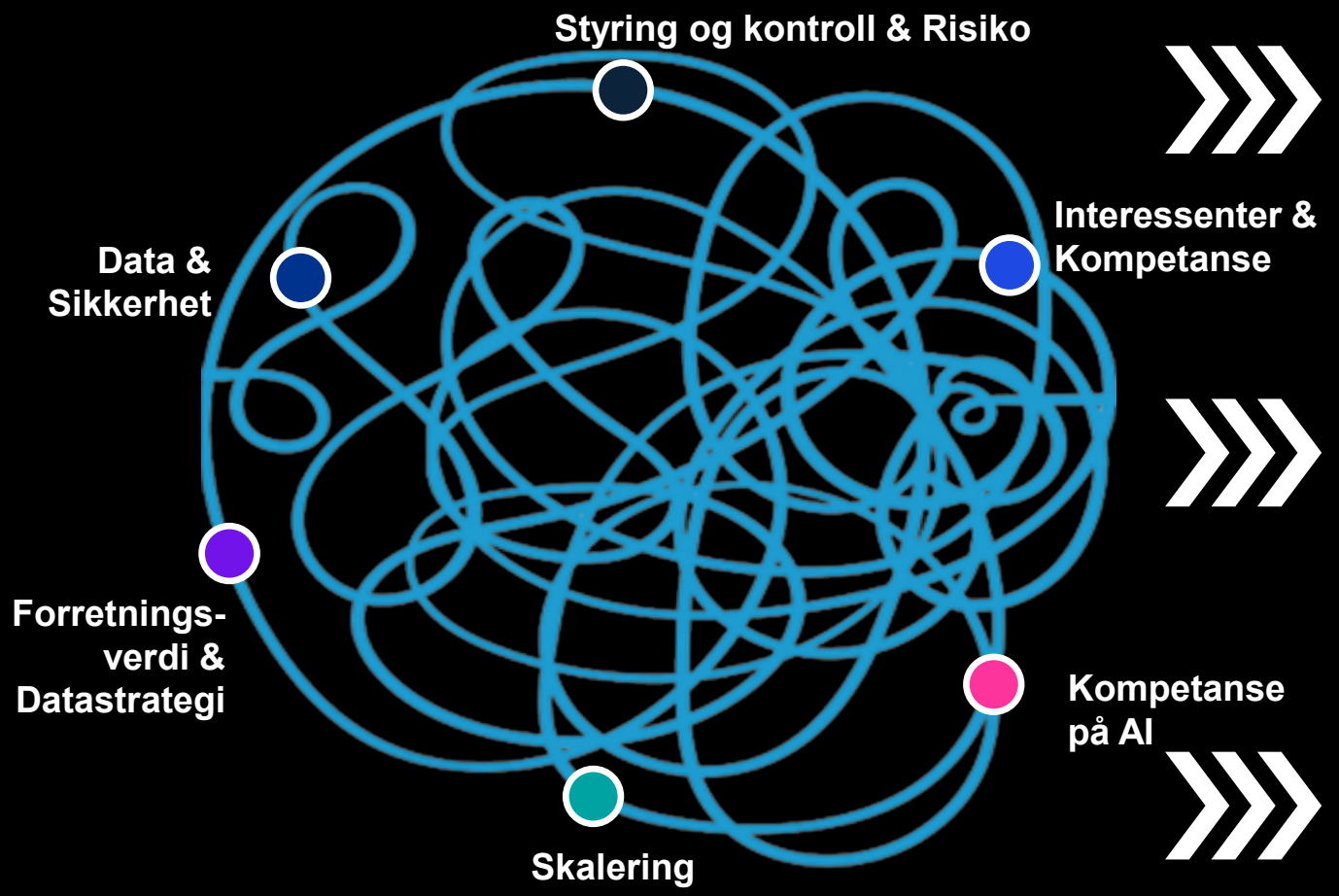
Risiko og tiltak

Selskaper møter i dag ofte utfordringer når de prøver å lykkes med AI

85%
av alle AI-prosjekt feiler

76%
sliter med å skalere

70%
mangler gevinst-realisering



- ✗ Manglende integrasjon i kjernevirksomheten
- ✗ Liten forståelse av underliggende teknologi
- ✗ Kan ikke skalere fra pilot til produksjon
- ✗ Mangel på kompetanse
- ✗ Mangler visjon for transformasjon
- ✗ Negativ oppfatning av AI
- ✗ Ledelsen er reaktive
- ✗ Skade på omdømme og tap av tillit

Sources: Gartner, Accenture

02

AI Act – Status og påvirkning

AI ACT og AI Liability Directive

Formål

- Følger en risikobasert tilnærming
- Risiko for brudd grunnleggende rettigheter
- Sikre balanse mellom regulering og innovasjon
- Regulerer tilfeller av skade påført av AI

Hva regulerer loven?

- Krav for AI-systemer eller produkter med AI-systemer
- Krav for lovlig bruk i EU
- CE-merking (produktansvarslov)

Definisjon "AI system"

- "machine-based system designed to operate with **varying levels of autonomy**"
- "**may exhibit adaptiveness** after deployment"
- "infers, from the **input it receives**, how to generate outputs such as **predictions**, **content**, **recommendations**, or **decisions**"
- "that can **influence physical or virtual environments**"

Hvem omfattes av AI Act?

Tilbyder/Provider

- Utvikler AI systemet

Bruker/Deployer

- Bruker AI systemet

Importører, distributører og autoriserte representanter

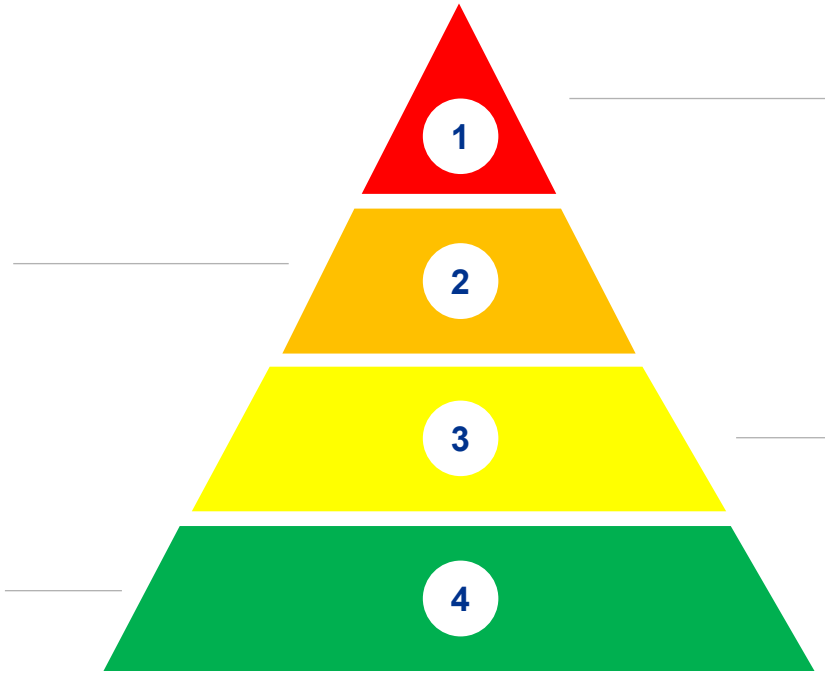
Loven følger en risikobasert tilnærming til krav

Uakseptabel risiko

Truer grunnleggende rettigheter

Høy risiko

Medfører en vesentlig risiko for helse, sikkerhet eller individets rettigheter



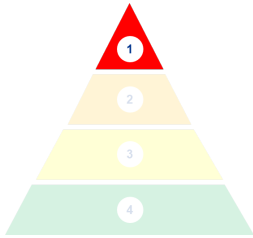
Begrenset risiko

Medfører en risiko for at individer kan bli villedet

Minimal risiko

Medfører ingen eller ubetydelig risiko

Bruksområder som har uakseptabel risiko



01

Manipulerende eller villedende AI-systemer

Som påvirker atferd og svekker beslutningstagning på en måte som leder til betydelig skade

02

AI-systemer som utnytter sårbarheter

på grunn av alder, funksjonshemming eller sosiale eller økonomiske situasjoner, og forårsaker betydelig skade.

03

AI-systemer for biometrisk kategorisering

Som trekker slutninger om rase, politiske meninger, fagforeningsmedlemskap, religiøse tro, seksualliv eller seksuell orientering

04

Klassifisering av individer eller grupper

basert på sosial atferd eller personlige egenskaper, noe som fører til skadelig eller uforholdsmessig behandling i irrelevante sammenhenger eller ubegrunnet ut fra deres atferd.

05

Sanntids fjernbiometrisk identifikasjon

på offentlige steder for rettshåndhevelse, unntatt for spesifikke nødvendige formål som å lete etter ofre eller savnede personer, forebygge trusler mot sikkerheten eller identifisere mistenkte i alvorlige forbrytelser.

06

Vurdering av risiko for at enkeltpersoner begår kriminelle handlinger

utelukkende basert på profilering eller personlighetstrekk, unntatt når de støtter menneskelige vurderinger basert på objektive, verifiserbare fakta knyttet til kriminell aktivitet.

07

AI-systemer som lager ansiktsgjenkjenningsdatabaser

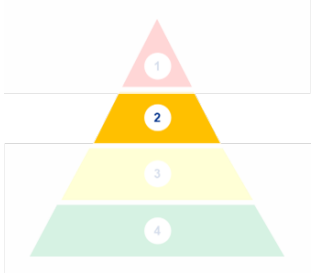
gjennom uspesifisert innsamling fra internett eller overvåkningskameraer.

08

AI-systemer som utleder følelser

På arbeidsplasser eller utdanningsinstitusjoner, unntatt av medisinske eller sikkerhetsmessige årsaker.

Kategorisering av høy risiko



To hovedkategorier: **høyrisikoprodukter** og **høyrisikosektorer.**

Høyrisikoprodukter (annex II)

- AI i produkter omfattet av EUs produktlovgivning pålagt tredjeparts samsvarsvurdering (CE-merking)
- Barneleker, medisinsk utstyr, heiser, biler osv.

Høyrisikosektorer (annex III)

AI-systemer brukt innenfor angitte områder

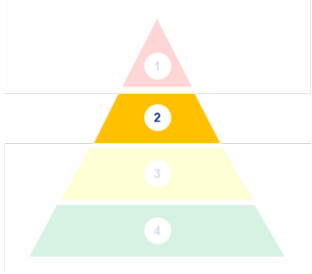
1. Fjernbiometrisk identifikasjon og kategorisering
2. Kritisk infrastrastruktur
3. Utdanning og yrkesopplæring
4. Ansettelse og arbeidsforhold
5. Tilgang til essensielle private og offentlige tjenester (f. eks. helseforsikring og lån)
6. Rettshåndhevelse
7. Migrasjon og grensekontroll
8. Forvaltning av rettssystemet/demokratiske prosesser

Intak fra annex III

1. *Ikke* betydelig risiko for skade på helse, sikkerhet eller grunnleggende rettigheter

2. ikke vesentlig påvirkning på utfallet av en avgjørelse

Hvilke krav gjelder hvis høy risiko?



Krav gjelder under **hele** livssyklusen til AI-systemet. Krav for den enkelte aktør vil variere avhengig av hvilken **rolle man har i verdikjeden**.

Krav spesifikt for høyrisikoprodukter (annex II)

- Etterleve relevant produktlovgivning pålagt tredjeparts samsvarsvurdering (CE-merking)

Krav som gjelder alle AI-systemer med høy risiko (ikke uttømmende)

- Samsvarsvurdering og CE-merking
- Risikostyring
- Datakvalitet
- Dokumentasjon
- Sporbarhet
- Åpenhet
- Menneskelig tilsyn
- Nøyaktighet
- Robusthet

Kategorisering av begrenset risiko

AI-systemer som **interagerer** med individer.

Begrenset risiko

- Chatboter
- AI-generert innhold

Krav for AI-systemer med begrenset risiko

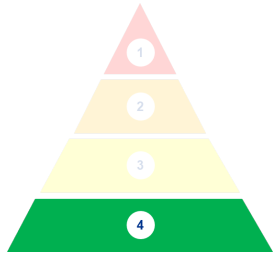
1. Informere sluttbruker om at man interagerer med et AI-system eller at innhold er generert av en AI

- Unntak: Enkel redigering/assistanse eller AI som ikke vesentlig endrer input-data

2. Deep fakes: Opplyse om at innholdet er kunstig generert eller manipulert

- Tilpasninger for deep fakes brukt i kunstneriske utrykk

Kategorisering av minimal risiko



AI-systemer som **ikke** er definert som uakseptabel, høy eller begrenset risiko.

Minimal risiko

- AI i videospill
- Anti-spam filtre

Krav for AI-systemer med minimal risiko

1. Frivillige krav – atferdsnormer og “codes of conduct”

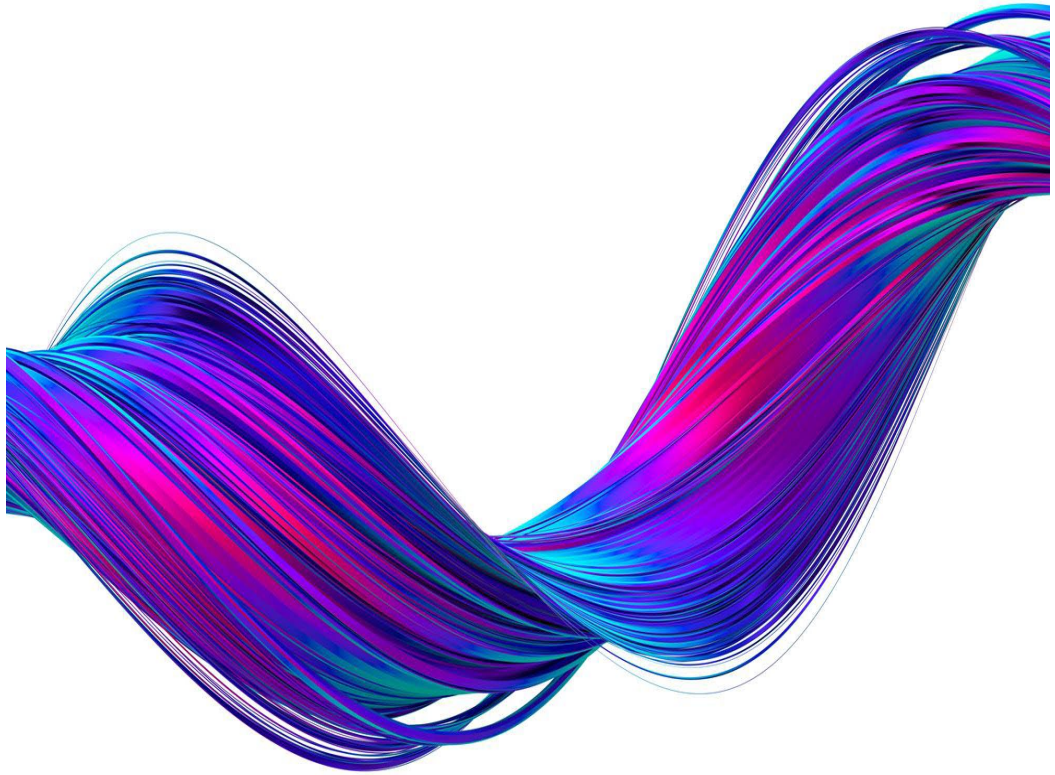
Unntak fra AI ACT

Unntak for sikre innovasjon

- Testing og utvikling av AI-systemer (egne regler for testing under virkelige forhold)
- Mindre strenge krav for små virksomheter og startups
- Vitenskapelig forskning

Formål knyttet til militæret, forsvar og nasjonal sikkerhet

Naturlige personer som bruker AI til ikke-profesjonelle formål.



03

AI Governance

Risikoappetitt for AI

Uregulert bruk av offentlig verktøy

Adopsjon av AI-applikasjoner og innkjøp av modeller

Tilpasning av innkjøpte modeller og egen utvikling

Hovedkategorier av risikoer (1)

- Personvernbrudd
- Sikkerhetsrisiko (datalekkasjer etc.)
- Brudd på åndsverk og konfidensiell informasjon
- Avhengighet
- Unøyaktighet, mangel på kontekst og hallusinering

Hovedkategorier av risikoer (2)

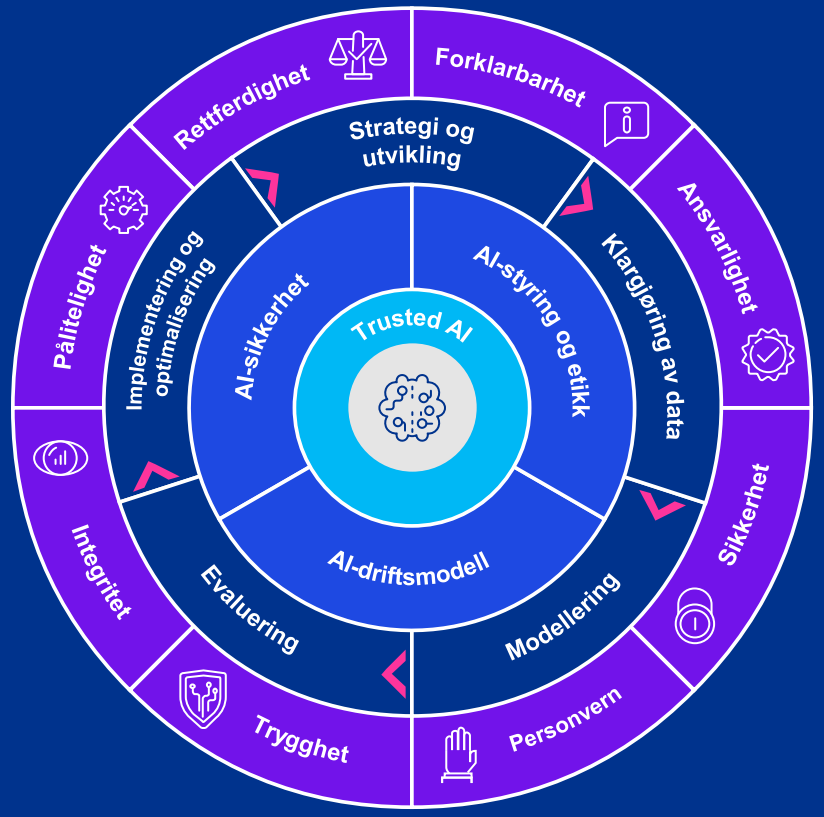
- Informasjonssikkerhet (tilganger)
- Mangel på gevinstrealisering (struktur informasjon)
- Omstilling av arbeidsprosesser
- Behov for kvalitetssikringsrutiner og avgrensninger
- Behov for kompetansetilpasning
- Ansvarsforhold ift. leverandør

Hovedkategorier av risikoer (3)

- Risikoer i (1) og (2)
- Risiko ift. data og dokumentasjon i livssyklusen (skjevheter, rettferdighet og ulovlig diskriminering)
- Modellrisiko
- Uklart eierskap og styringsstruktur gjennom livssyklusen
- Glidning i prestasjon, formål, data eller teknikk

Trusted AI

Trusted AI er KPMGs rammeverk for å sikre god styring og kontroll med bruk av AI. Rammeverket sørger for at organisasjonen kan adressere potensielle problemer ved bruk av AI på en strukturert og helhetlig måte.



1 **Rettferdighet**
Sikre at skjevheter i datasett blir redusert til virksomhetens akseptable nivå for rettferdig behandling, slik at diskriminering og urettferdighet forhindres.

2 **Forklarbarhet**
Sikre at resultatene som AI produserer kan forklares, slik at brukerne kan opprettholde tillit, aktørene kan holdes ansvarlige, og løsningen forbedres.

3 **Ansvarlighet**
Sikre at mekanismer som fremmer ansvar er på plass, slik at negative konsekvenser med AI kan forutses, og rettslige hensyn ivaretas.

4 **Sikkerhet**
Beskytt mot uautorisert tilgang, lekkasjer og manipulasjon, slik at gjeldende prinsipper og regulatoriske krav til informasjonssikkerhet overholdes for AI.

5 **Personvern**
Benytt prinsippene for innebygget personvern, vurder konsekvenser og andre krav til personvern slik at personvernforordningen etterleves for AI.

6 **Trygghet**
Fastsett risikotoleranse og reguler tillatte bruksområder, slik at AI ikke får negativ innvirkning på mennesker, gjenstander og miljøet.

7 **Integritet**
Sikre kvalitet i datahåndtering og berikelse gjennom hele livssyklusen, slik at AI produserer riktig beslutningsgrunnlag og overholder reguleringer.

8 **Pålitelighet**
Sikre at innhold og prediksjoner fra AI er i tråd med tiltenkt formål og ytelse, slik at økonomisk- og sikkerhetsrisiko holdes innenfor terskelverdiene.

Overordnet prosess for å ivareta "Trusted AI"

01

AI-konsekvensanalyse

Før AI use caset utvikles, bør det gjennomføres en prospektiv konsekvensanalyse. Analysen bør belyse AI use casets påvirkning på foretaket, og på forbrukere dersom AI use caset feiler. Sannsynlighet for at negativ påvirkning inntreffer må også vurderes.

02

Implementer passende kontrolltiltak

For å redusere risiko som er identifisert i AI-konsekvensanalysen skal det implementeres tilstrekkelig kontrolltiltak gjennom AI use casets livssyklus (fra modellutviklingen til man slutter å anvende use caset).

03

Empirisk konsekvensanalyse

Etter at modellutviklingsprosessen er ferdigstilt, skal det gjennomføres en empirisk konsekvensanalyse for å vurdere hvorvidt AI use caset er pålitelig og ivaretar etiske aspekter. Det vil si, en etterkontroll og en dokumentasjon av hvorvidt identifiserte kontrolltiltak vil være tilstrekkelige eller om justeringer må foretas.

04

Løpende overvåking

Etter at AI use caset er godkjent for bruk, må dette løpende følges opp gjennom kontrolltiltak som menneskelig oppsyn, datahåndtering og datakvalitet, dokumentasjon, ivaretagelse av fairness og robusthet og ytelse.

Hensiktsmessig menneskelig oppsyn – risikostyring

01

Risikostyrings-funksjonen

- Vurdere om risikostyringssystemet har tilstrekkelig kontroll, testing og feedback loops for bruk av AI
- Vurdere om nivået av automatisering er forsvarlig
- Vurdere om kontrollen med intern og ekstern data sikrer at denne er passende og uten skjevheter

02

Aktuarfunksjon

- Være i stand til å utfordre, vurdere risiko og etiske implikasjoner knyttet til bruk av AI på forsikringsområdet
- Aktuarfunksjonen bør også benytte styringsprinsippene for AI for å vurdere annen kompleks modellering

03

Compliance

- Overvåke endring i regelverk knyttet til AI
- Overvåke de AI use casene som er i bruk, og vurdere om disse vil etterleve eksisterende eller kommende regulatoriske krav

04

Andre

- Både internrevisjon, DPO, AI-kommite eller AI officer, compliance, ledergruppe, styret, utviklere, IT-sikkerhet, brukere og HR bør inngå med definerte ansvarsområder i risikostyringssystemet
- Menneskelig oppsyn må tilpasses hvor kompleks og høy konsekvens bruksområdet kan ha

Involvering i beslutningsprosesser for AI

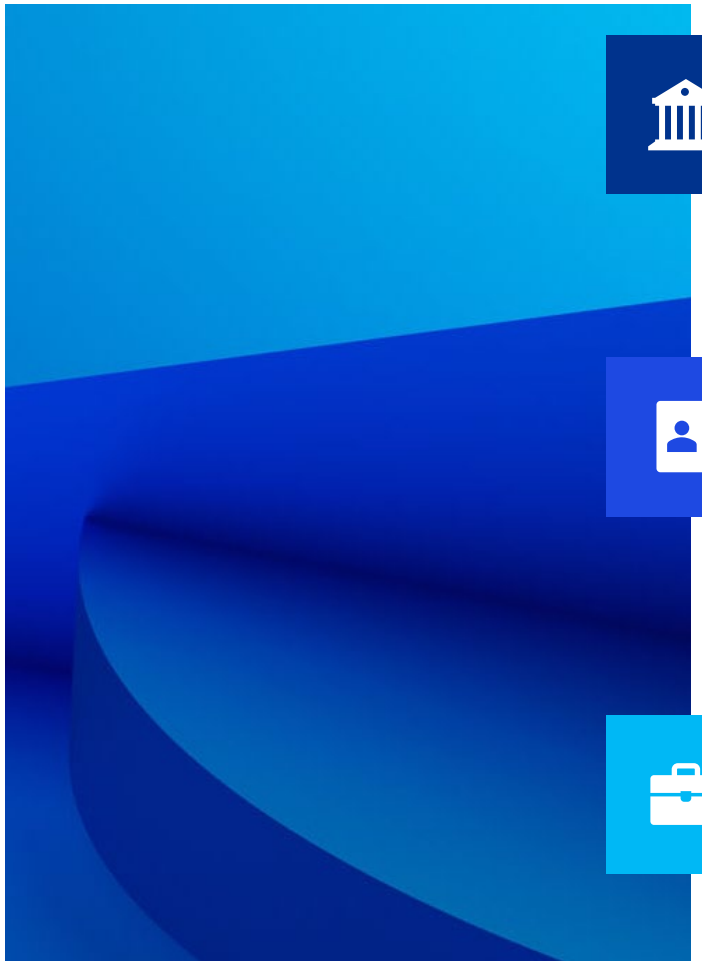
	Ledelse / styret	IT	Utvikler	DPO	AI / data officer	Complianc efunksjon	Risiko-styrings-funksjon	Aktuar-funksjon	Intern-revisjon
Høy	G	G	G	K	G	G	K	K	I
Medium	I	G	G	K	G	K	K	K	I
Lav	I	I	G	K	K	K	I	I	I

G = Godkjenne

K = Konsulteres

I = Informeres

Utfordringer vi ser i våre revisjons- og rådgivningsprosjekter



Datakvalitet og skjevheter

Modellrisikostyring og oppfølging

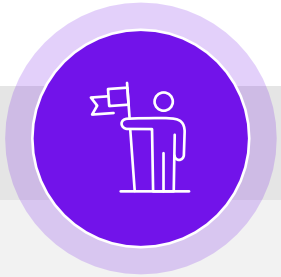
Risikoappetitt

.

04

Internrevisjonens rolle

Internrevisjonens rolle i implementeringen av AI



Kartlegging av gap og en sparringspartner for styret om status på ambisjonsnivå og risikoappetitt for AI

Gjennomføre revisjoner av styringssystemet for AI internt i virksomheten

Gjennomføre revisjoner av spesifikke AI-modeller og bruksområder

Eksempler på revisjonsområder, basert på modenhetsnivå

01

Utforskende stadium: Gevinstrealisering

Plankontekst. Internrevisjonen kan å kartlegge ambisjonsnivå og risikoappetitt for bruk av AI, og omfanget av bruken i organisasjonen i dag. Videre kan internrevisjonen vurdere prosesser for gevinstrealisering for å sikre at kostnyttevurderinger blir gjennomført tilknyttet AI-initiativene.

Risiko. Dersom risikoapetitt og ambisjonsnivå ikke er definert er det risiko for at selskapet ikke kan ha effektive prosesser for gevinstrealisering.

02

Aktivt stadium: Kompetanse og adopsjon

Plankontekst.

Internrevisjonen kan evaluere prosesser for kompetanseheving og –tilpasning, samt tiltak for håndtering av omstillingsrisiko ved adopsjon av AI-applikasjoner.

Risiko. Dersom kompetansehevingstiltak, kvalitetssikringsrutiner og fastsettelse av akseptable bruksområder ikke implementeres, vil adopsjon av AI-applikasjoner slik som Microsoft Copilot medføre en omstillingsrisiko.

03

Operasjonalisert stadium: Risikostyring og internkontroll

Plankontekst. Internrevisjonen kan foreta en overordnet gjennomgang av styringssystemet for AI. Dette omfatter blant annet evaluering av prosess for utvelgelse av bruksområder, kompetansenivå, risikostyring og internkontroll og hvorvidt det er fastsatt klar ansvarsstruktur.

Risiko. Dersom styringssystemet ikke tilpasses til at selskapet har tatt i bruk AI er det risiko både for finansielt og omdømmemessig tap.

04

Optimalisert stadium: Validering av bruksområder

Plankontekst. Internrevisjonen kan foreta en helhetlig evaluering av et av selskapets bruksområder for AI. Revisjonen kan vurdere blant annet modellering, datakvalitet, modellrisikostyring, etiske aspekter og hvorvidt bruken er pålitelig.

Risiko. Dersom det er feil eller for lite kontrolltiltak implementert for å redusere den spesifikke risikoen ved bruksområdet, kan dette medføre både tap av omdømme eller finansielle tap.

Kontakt oss

Nicolai Cappelen

Partner

Telefon: +47 406 39 644

E-post: nicolai.cappelen@kpmg.no

Thea Hvitmyhr Gullaug

Manager

Telefon: +47 992 52 399

E-post: thea.gullaug@kpmg.no

kpmg.no/sosialemedier