Når halen rister hunden

Nye risikoer ved bruk av KI på arbeidsplassen

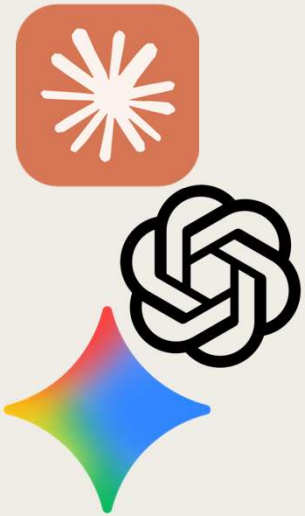
SPADE consulting

Bakgrunn

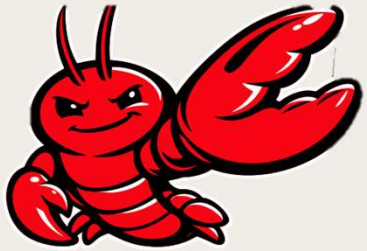
- KI-systemer er sære.
- Folk er enten litt for skremte eller litt for nonchalante
- Det er mye å hente ved å tenke litt på sikkerhet.

Hva snakker vi om?

Kjedelig



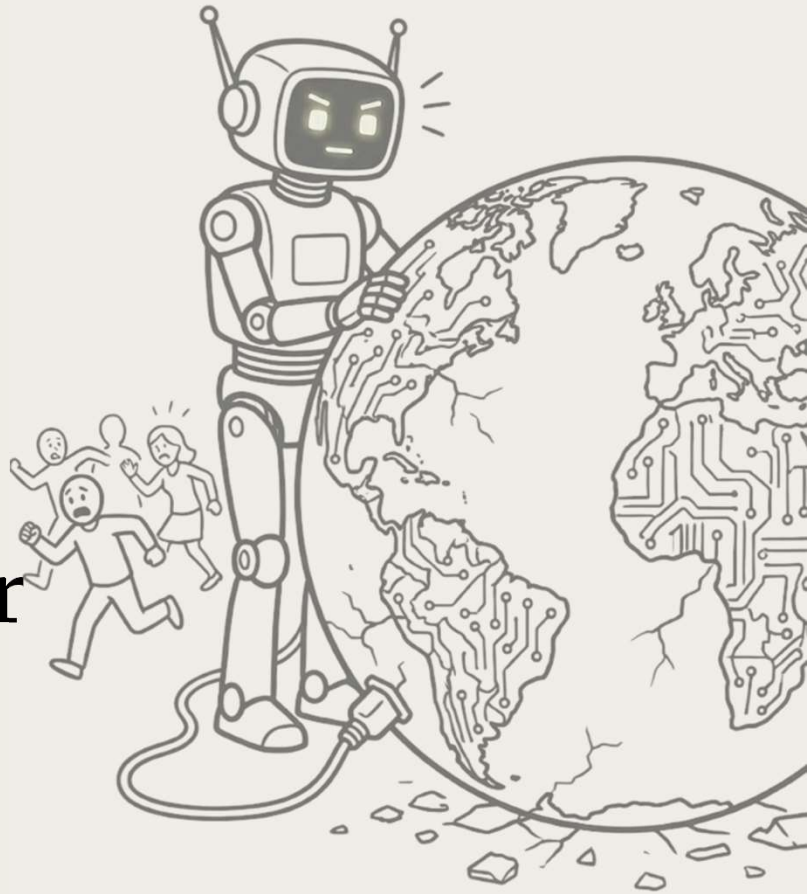
Sweetspot



Ouch! Too hot

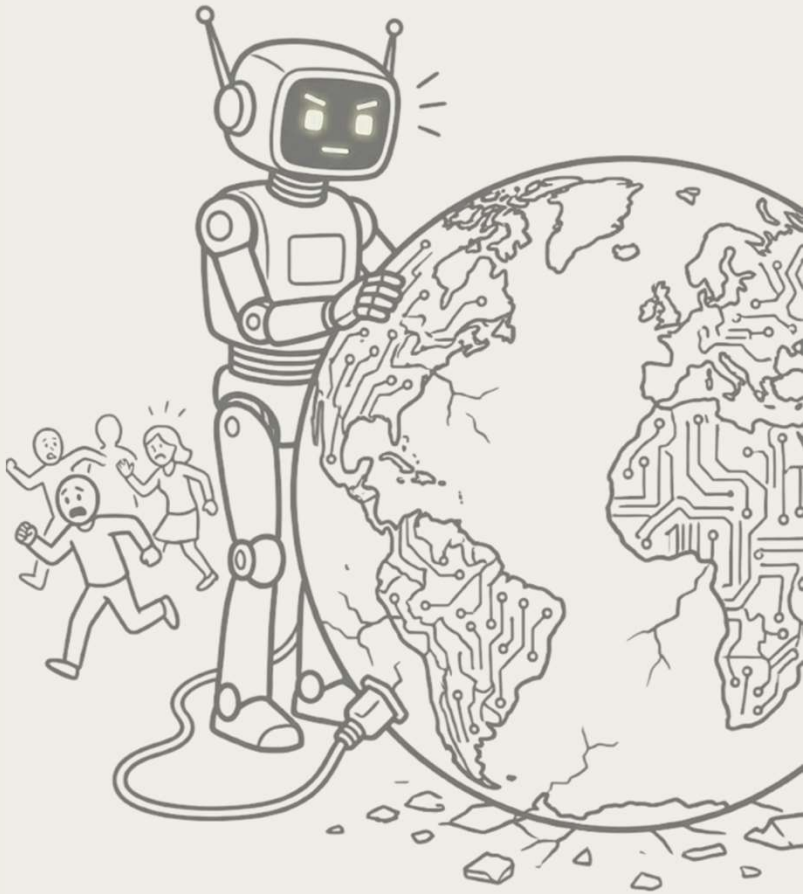
Dagens mål

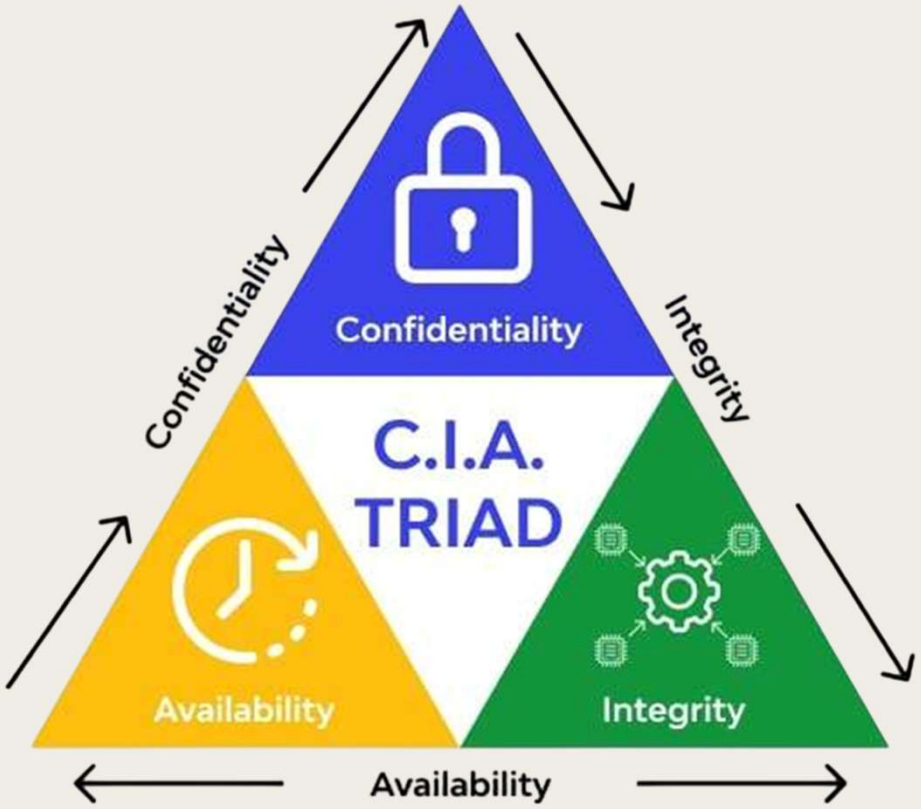
1. Ødelegge stemningen
2. Gjør alt veldig komplisert
3. Selge skinnet før bjørnen er skutt
4. Legge alle eggene i en kurv



Dagens mål

1. Gi overblikk
2. Se på risikoer som ligger nært bedriften
3. Foreslå konkrete tiltak
4. Fokusere på muligheter




Confidentiality
Integrity
Availability
C.I.A.
TRIAD
Confidentiality
Availability
Integrity

Er ut.....

Sikkerhet i den tekniske løsningen

Risiko som følger av bruken av løsningen

...er inn!

- **K**onfidensialitet
- **I**ntegritet
- **T**ilgjengelighet

- **J**uridisk
- **AI**/Agent
- **P**ositiv risiko (Muligheter)

Konfidensialitet/ integritet

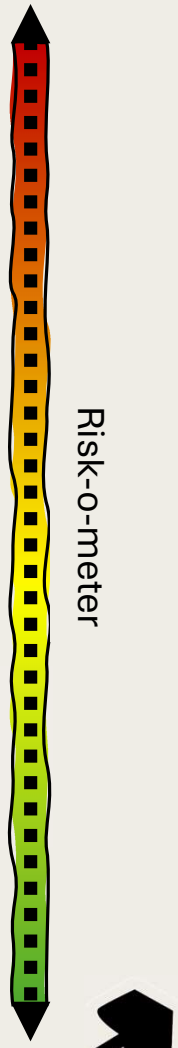
Risiko: *Informasjon lagt inn i løsningen faller i feil hender og blir lekket eller endret på.*

Vurdering

- Alle tre leverandører kan vise til gode sikkerhetstiltak.
- Anthropic er minst integrert som leverandør
- Google er mest integrert
- Alle har mange avhengigheter til USA
- Ingen kjente datainnbrudd knyttet til kundeinput for noen av disse

Tiltak

1. Etabler rammer for hva som er greit å putte i AI. ~det du kan putte i sky.
2. Velg løsning med SSO.



Tilgjengelighet 1

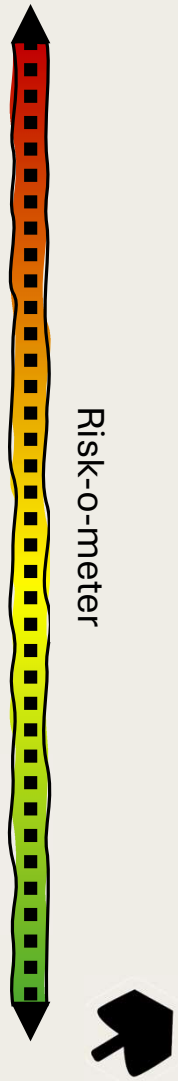
Risiko: *Løsningen blir utilgjengelig*

Vurdering

- Alle tre leverandører bruker etablerte skyplatformer
- Det har vært kortere nedetidsperioder for samtlige (opp mot en dag)
- Google og Anthropic gir oppetall i SLA, (99,5 og 99,9) Open AI gir ikke.

Tiltak

1. Vurder bruk av API for kritiske formål
2. For kritiske arbeidsflyter- ha kjennskap til alternativ gjennomføring



Tilgjengelighet 2

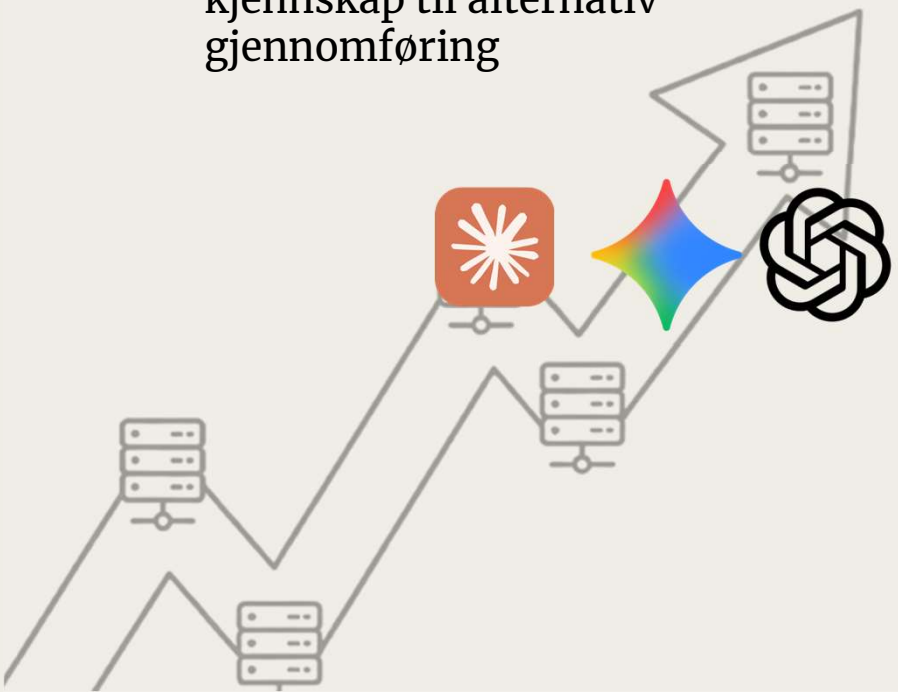
Risiko: *Leverandøren gjør deg avhengig av dem gjennom proprietære verktøy, eller endrer tilgjengeligheten*

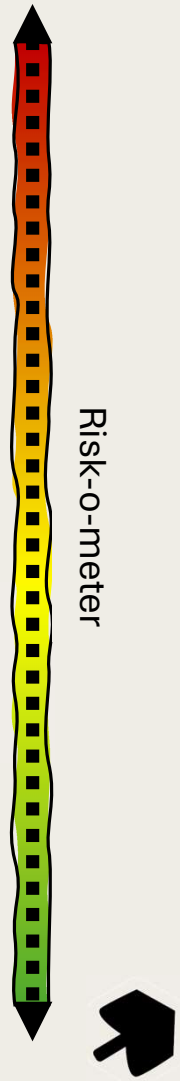
Vurdering

- Alle tre løsninger leverer fundamentalt sett det samme- men de bygger løsninger på toppen for å gjøre kundeforholdet mer «sticky»
  - Open AI Agent builder
  - Claude cowork
  - Gemini custom agents
- Om en benytter en av disse, bygger en ikke generell kompetanse, og det blir vanskeligere å bytte leverandør om nødvendig.

Tiltak

1. Vurder bruk av API for kritiske formål
2. For kritiske arbeidsflyter- ha kjennskap til alternativ gjennomføring






Nestlé
Kit
1 finger =
Energy
42kcal
2%
Find out more
% of Guideline Daily Amounts of an average adult (2000kcal)
Coated Wafer
Have a break,

- **J**uridisk
- **AI**/Agent
- **P**ositiv risiko (Muligheter)

Juridisk 1

Risiko: *Brudd på GDPR- Det blir ulovlig å overføre (person)-opplysninger til løsningen*

Vurdering

- Alle tre leverandører er amerikanske selskaper underlagt CLOUD act, FISA 702
- Gjeldende overføringsavtale EU<->USA er skjør
- Trump har allerede sanksjonert europeiske individer via Microsoft.
- Politisk-strategiske vurderinger øker risiko kommersielt og mot myndigheter

Tiltak

1. Etabler rammer for hvilke personopplysninger som er akseptable å putte i løsningen
2. Vurder API for å sikre mulighet for raskt bytte til europeisk/ åpen/ lokal språkmodell

Juridisk 2

Risiko: *Brudd på GDPR- Ulovlig automatisert behandling*

Vurdering

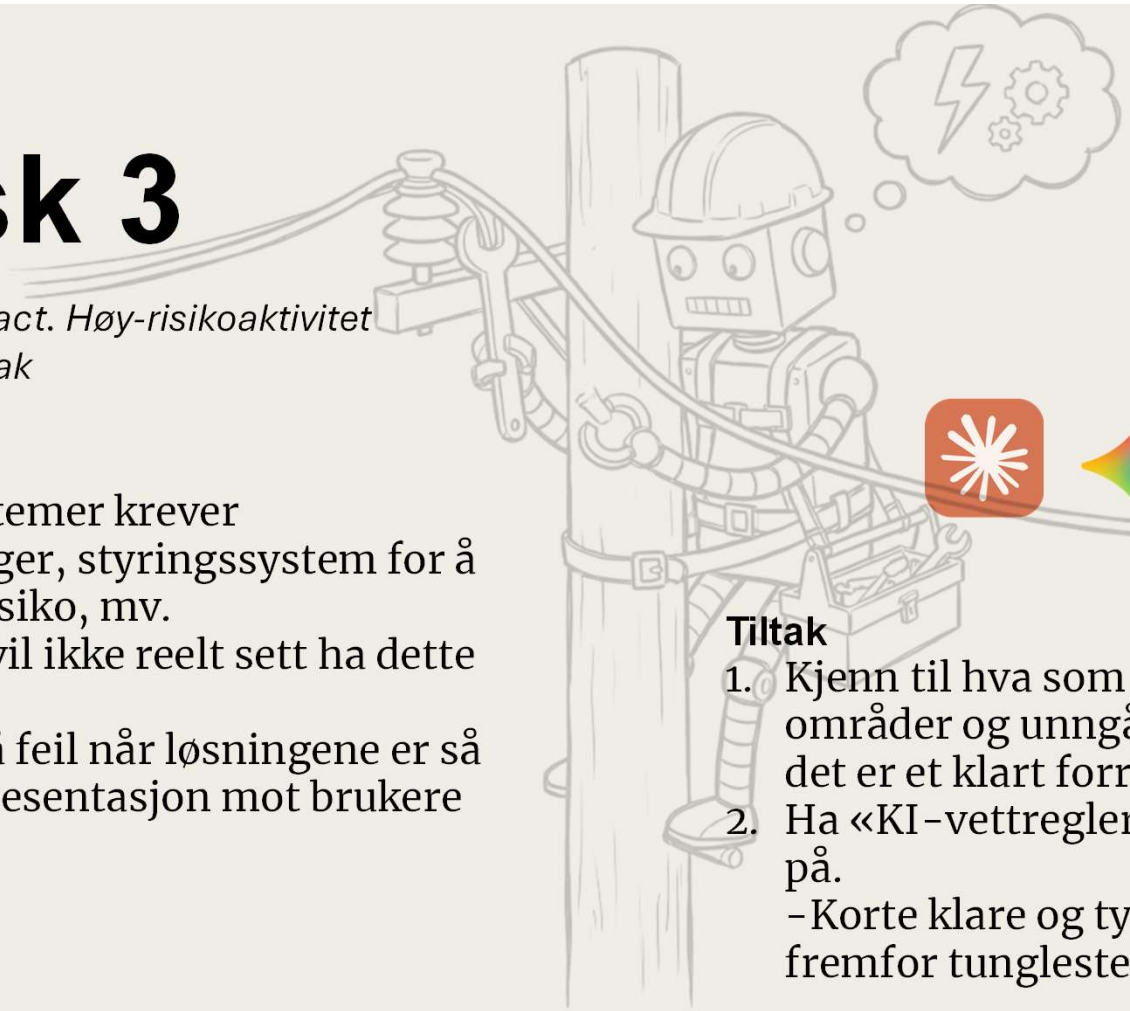
- Det er reservasjonsrett mot automatiserte beslutninger
- Derfor må en informere dersom KI benyttes til å treffe beslutninger uten reell menneskelig inngripen
- Det kan være vanskelig å sikre, særlig når KI- bruken er desentralisert

Tiltak

1. Etabler og beskriv praksis for «human in the loop»
2. Ha regler for bruk av personopplysninger i løsningen

Juridisk 3

Risiko: *Brudd på AI act. Høy-risikoaktivitet uten nødvendige tiltak*



Vurdering

- Høyrisiko-systemer krever risikovurderinger, styringssystem for å håndtere KI-risiko, mv.
- De aller fleste vil ikke reelt sett ha dette på plass
- Veldig lett å trå feil når løsningene er så «åpne» i sin presentasjon mot brukere

Tiltak

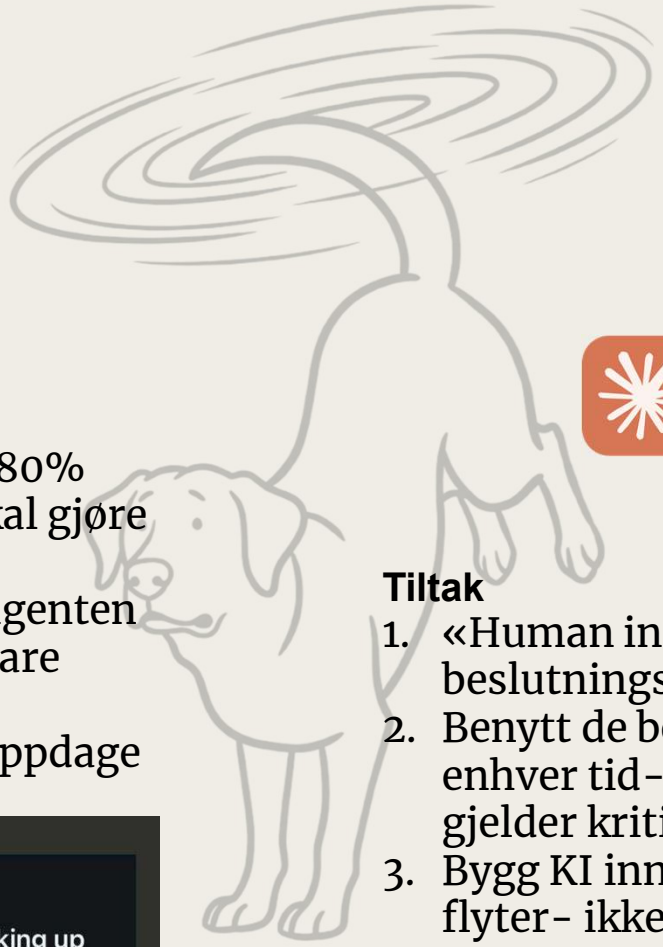
1. Kjenn til hva som er høyrisiko-områder og unngå med mindre det er et klart forretningscase.
2. Ha «KI-vettregler» en terper på.
   -Korte klare og tydelige fremfor tungleste og snirklete

AI 1

Risiko: *AI-agenter gjør feil eller gir feil informasjon som påvirker avgjørelser*

Vurdering

- De beste KI-agentene har rundt 80% treffsikkerhet pr. skritt når de skal gjøre oppgaver på egen hånd
- I kompliserte arbeidsflyter kan agenten bli påvirket til å gjøre feil, eller bare gjøre feil på egen hånd
- Det er vanskelig for individer å oppdage feil.

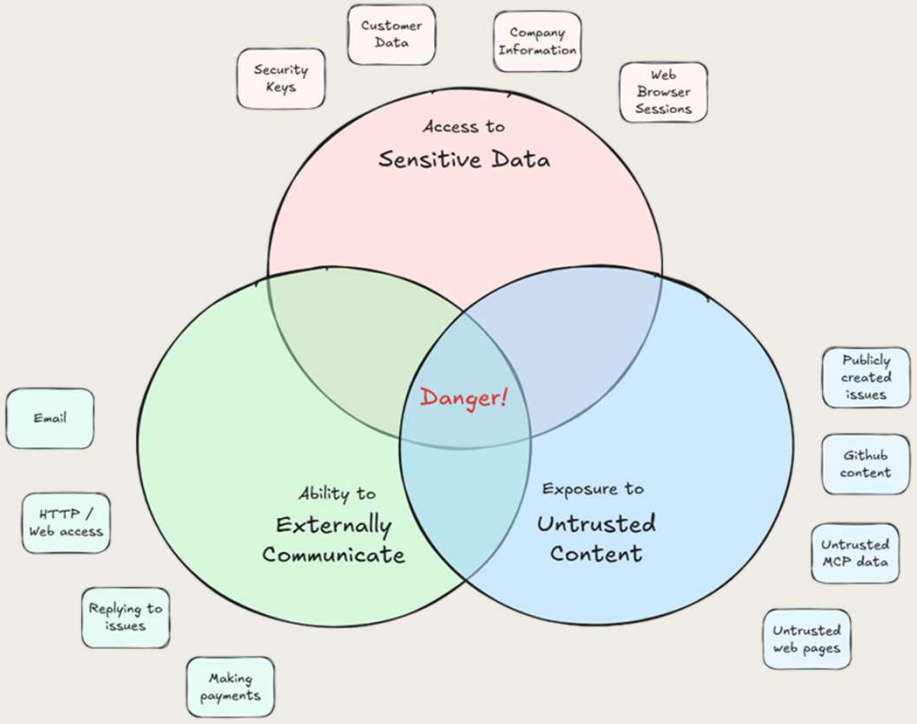


Tiltak

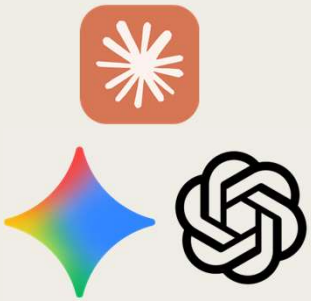
1. «Human in the loop» i beslutningsprosesser
2. Benytt de beste modellene til enhver tid-særlig når det gjelder kritiske områder
3. Bygg KI inn i veldefinerte flyter- ikke la KI gjøre for mange skritt helt på egen hånd.

AI 2

Risiko: *KI-agenter blir lurt til å gi tredjeparter tilgang til sensitive opplysninger, til å dele feilopplysninger med brukeren, endre data internt, eller gjøre uønskede handlinger*



(litt ekstra høyt pga. Claude Cowork)



Tiltak

1. «Human in the loop»
2. Benytt API og bygg rammer
3. Unngå å dele sensitive opplysninger - særlig de som kan gi videre tilgang
4. Benytt de kraftigste modellene (lavest feilrate og mest kontrollerbare)
5. Ha intern opplæring i KI-risiko

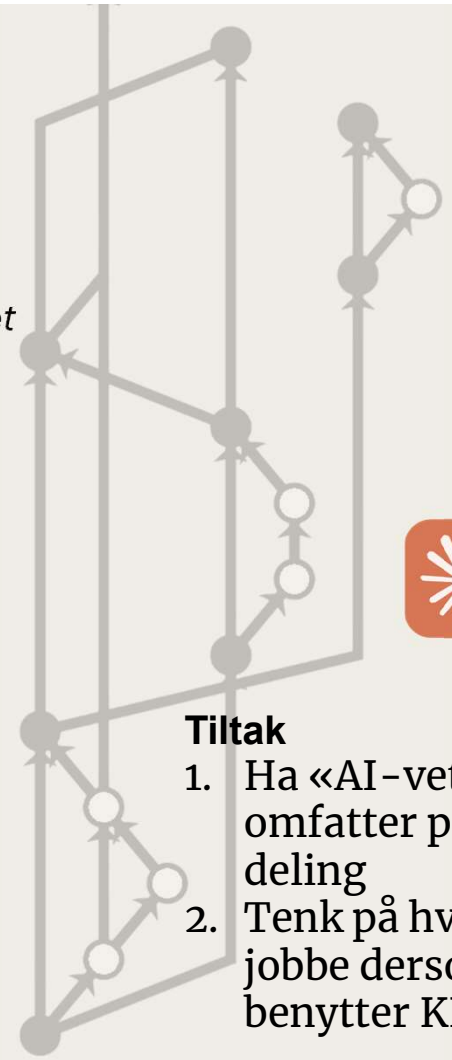
Kilde: Lethal trifecta- først definert av Simon Willison: https://simonwillison.net/2025/Jun/16/the-lethal-trifecta/ Illustrasjon: https://martinfowler.com/articles/agentic-ai-security.html

AI 3

Risiko: *Informasjonsutglidning og tap av omforenthet fører til tapt arbeid, tap av effektivitet, konflikt*

Vurdering

- Ki-systemene er i varierende grad mulig å integrere i interne systemer. En jobber på siden av eller utenfor tradisjonelle dokumenter.
- En kan ende opp med å jobbe veldig langt på et område før informasjonen kommer tilbake til et omforent sted.
- Det er lett at en mister konsensus når en kan komme seg langt alene.



Tiltak

1. Ha «AI-vettregler» som omfatter praksis for lagring og deling
2. Tenk på hvordan en burde jobbe dersom alle uansett benytter KI-assistenter

Risk-o-meter

Mulighet 1

Mulighet: *Minsket skyggebruk av KI, bedre kontroll på (negativ) risiko*

Vurdering

- Folk vil benytte KI-systemer uansett
- Ved å tilby en god «cutting edge»-løsning vil brukere guides over på løsninger der du har bedre kontroll
- Å tilby gode verktøy gir også godvilje mot å følge regler og rutiner

Tiltak

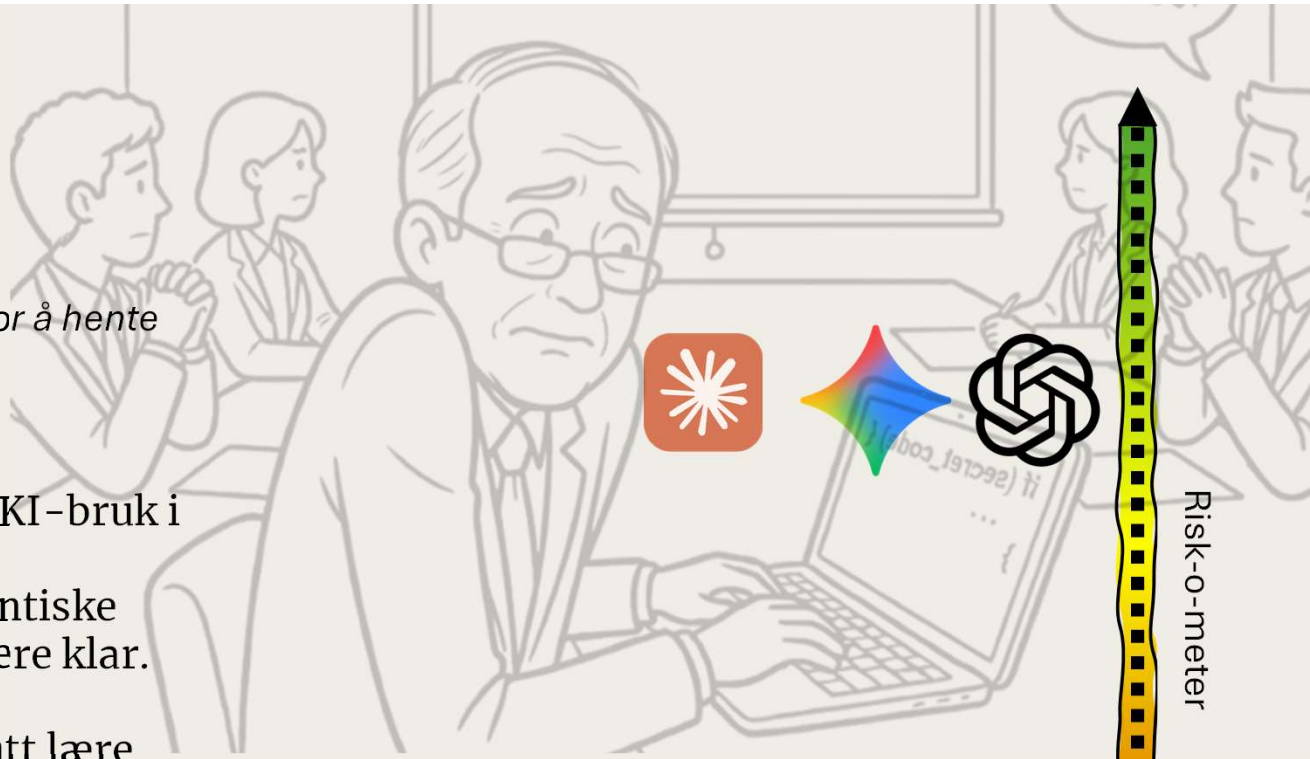
1. Balanser KI-goder med ansvar
2. Bruk muligheten til å etablere rutiner og regler.
3. Tydelig forby ikke-godkjente KI-verktøy

Mulighet 2

Mulighet: *Økt kompetanse- mulighet for å hente fremtidige gevinster raskere.*

Vurdering

- Det er varierende gevinster fra KI-bruk i bedrifter- foreløpig.
- Når kvalitets-terskelen for agentiske løsninger nås, er det viktig å være klar.
- Programmering f.eks har blitt allemanseie- men en må fortsatt lære seg å jobbe med systemene.



Tiltak

1. Ta i bruk hele produktet- også koding og API
2. Etabler erfaringsdeling internt
3. Gi brede men tydelige rammer

Mulighet 3

Mulighet: *Økt effektivitet.*

Vurdering

- KI-systemer gir økt effektivitet på generelle kontoroppgaver, også pr. dags dato
- Men effekten er uforutsigbar og varierende
- Prøving og feiling er viktig

Tiltak

1. Ta i bruk hele produktet- også koding og API
2. Etabler erfaringsdeling internt
3. Gi brede men tydelige rammer
4. Tenk på hvordan du kan løse det underliggende forretningsbehovet, ikke hvordan en gjør mennesket mer effektiv

Takk for meg!

Spørsmål?
Kritikk?
Tanker?

SPADE consulting

PDF av denne presentasjonen:

spadeconsulting.no/pdf