Kunstig intelligens, risiko, informasjonssikkerhet, personvern, …

Webinar Normen, 28. sept 2023

Inger Anne Tøndel
Seniorrådgiver Direktoratet for e-helse, avdeling informasjonssikkerhet
Medlem av sekretariatet for Normen

Hva vi skal innom i dag …

Hva er kunstig intelligens?

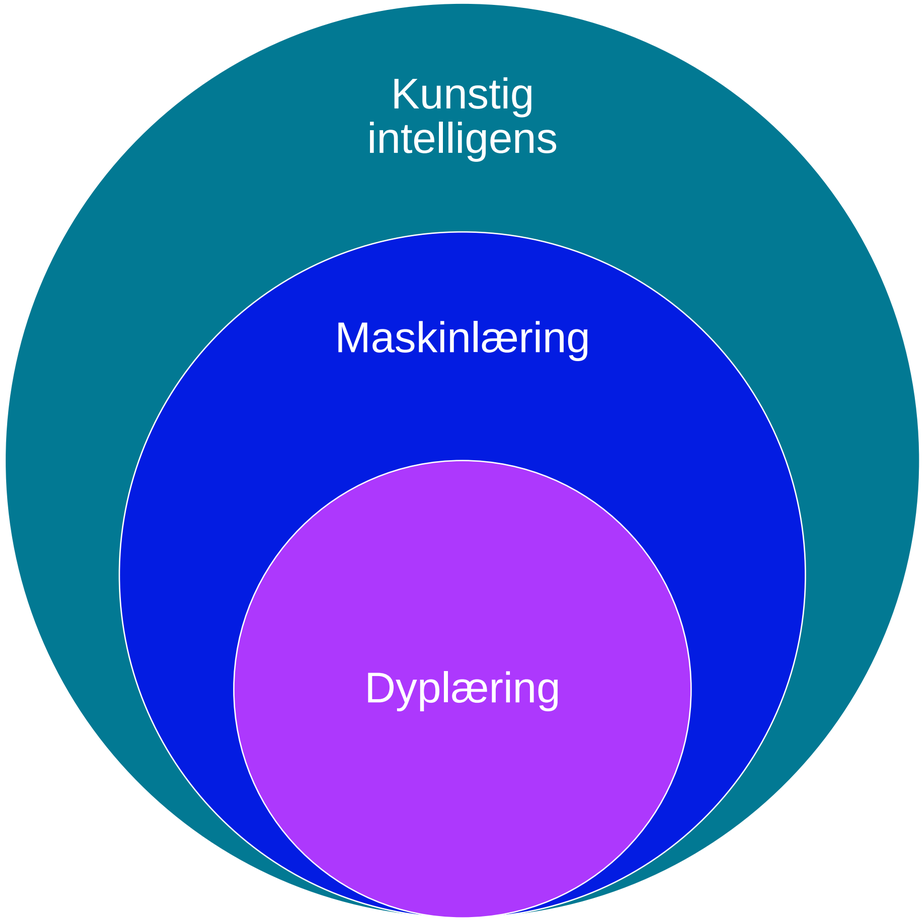
Risiko i systemer som benytter kunstig intelligens

Hva kunstig intelligens kan bidra med i helse- og omsorgssektoren

Hjelp og støtte som finnes og er på vei

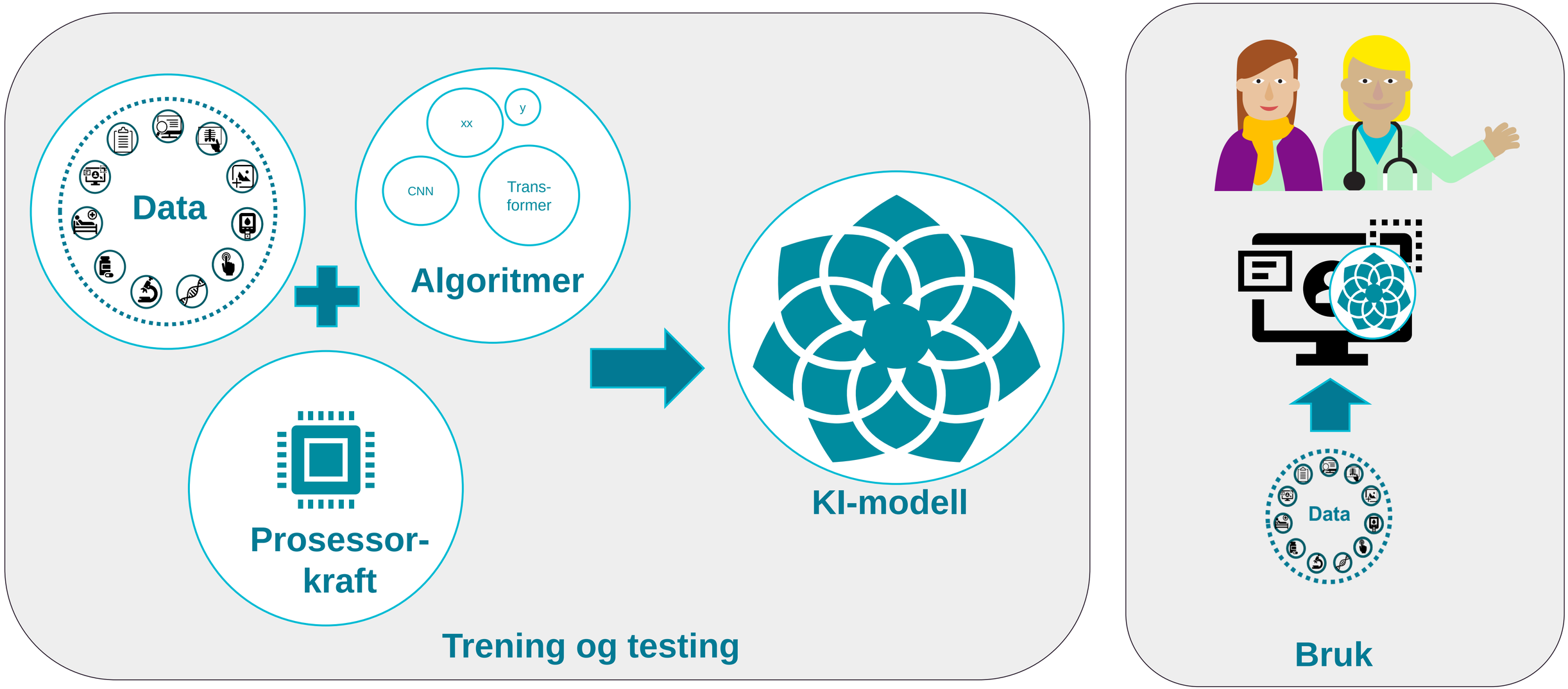
Hva er kunstig intelligens?

Kunstig intelligens og maskinlæring



NORMEN

KI-elementer



Illustrasjon Hilde Margrethe Lovett,

Det nasjonale koordineringsprosjektet for KI

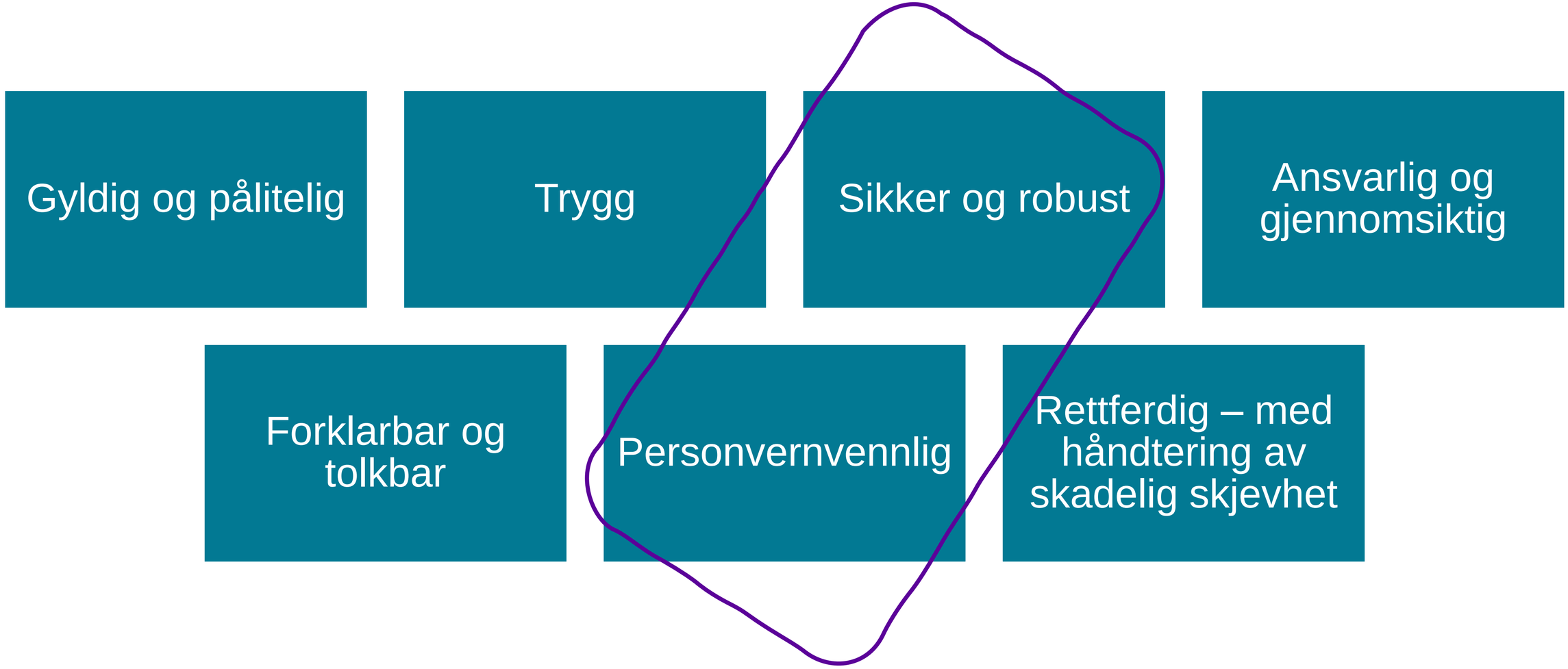
Sikkerhet og kunstig intelligens

Kunstig intelligens for å bedre sikkerheten

Kunstig intelligens i hendene på angripere

Sikkerhet innebygd i KI-systemene

Tillitsverdig kunstig intelligens



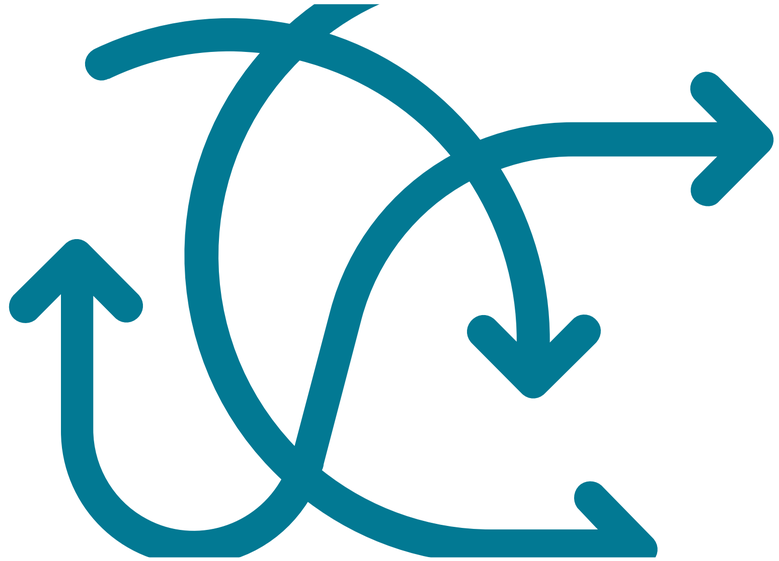
NORMEN

Kilde: NIST AI RMF 1.0

Sikkerhet og personvern – hva gjør kunstig intelligens spesielt?

Ved maskinlæring så lærer systemet funksjonaliteten fra data

Vi vet ikke alltid hva som er «riktig» oppførsel



Vanskelig å forstå for mennesker → komplekst

Kan man stole på treningsdataene?
Er treningsdataene sensitive?

Data

xx

y

CNN

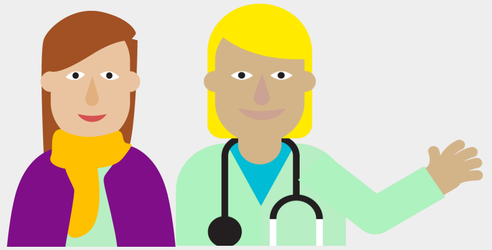
Trans-former

Algoritmer

Prosessor-kraft

KI-modell

Trening og testing



Data

Bruk

Basert på iIllustrasjon Hilde Margrethe Lovett,

Det nasjonale koordineringsprosjektet for KI

bbc.com/news/technology-35902104

Tay: Microsoft issues apology over racist chatbot fiasco

25 March 2016 · Comments

The AI was taught to talk like a teenager

By Dave Lee >
North America technology reporter

Microsoft has apologised for creating an artificially intelligent chatbot that quickly turned into a holocaust-denying racist.

But in doing so made it clear Tay's views were a result of nurture, not nature. Tay confirmed what we already knew: people on the internet can be cruel.

Tay, aimed at 18-24-year-olds on social media, was targeted by a "coordinated attack by a subset of people" after being launched earlier this week.

Within 24 hours Tay had been deactivated so the team could make "adiustments".

Kan man stole på treningsdataene? Er treningsdataene sensitive?

Kan modellen være manipulert?

Data

xx

y

CNN

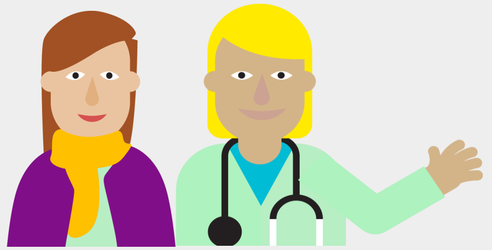
Trans-former

Algoritmer

Prosessor-kraft

KI-modell

Trening og testing



Data

Bruk

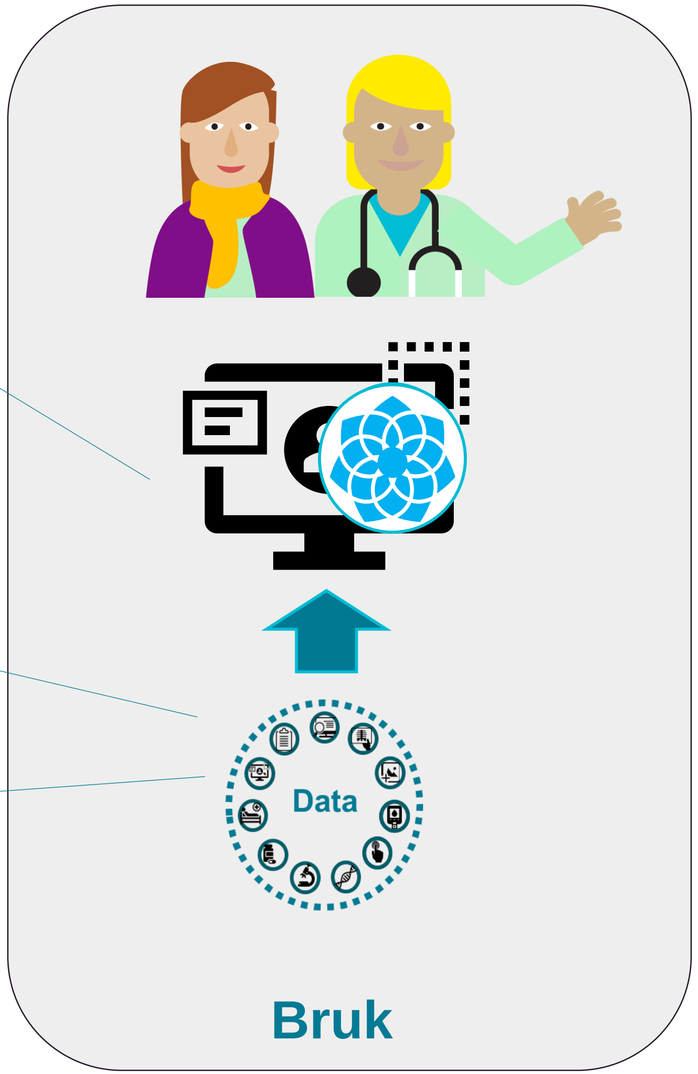
Basert på iIllustrasjon Hilde Margrethe Lovett,

Det nasjonale koordineringsprosjektet for KI

Kan angripere utlede informasjon om modellen, treningsdataene eller inngangsverdien?

Bør disse inngangsverdiene deles med systemet/ leverandøren?

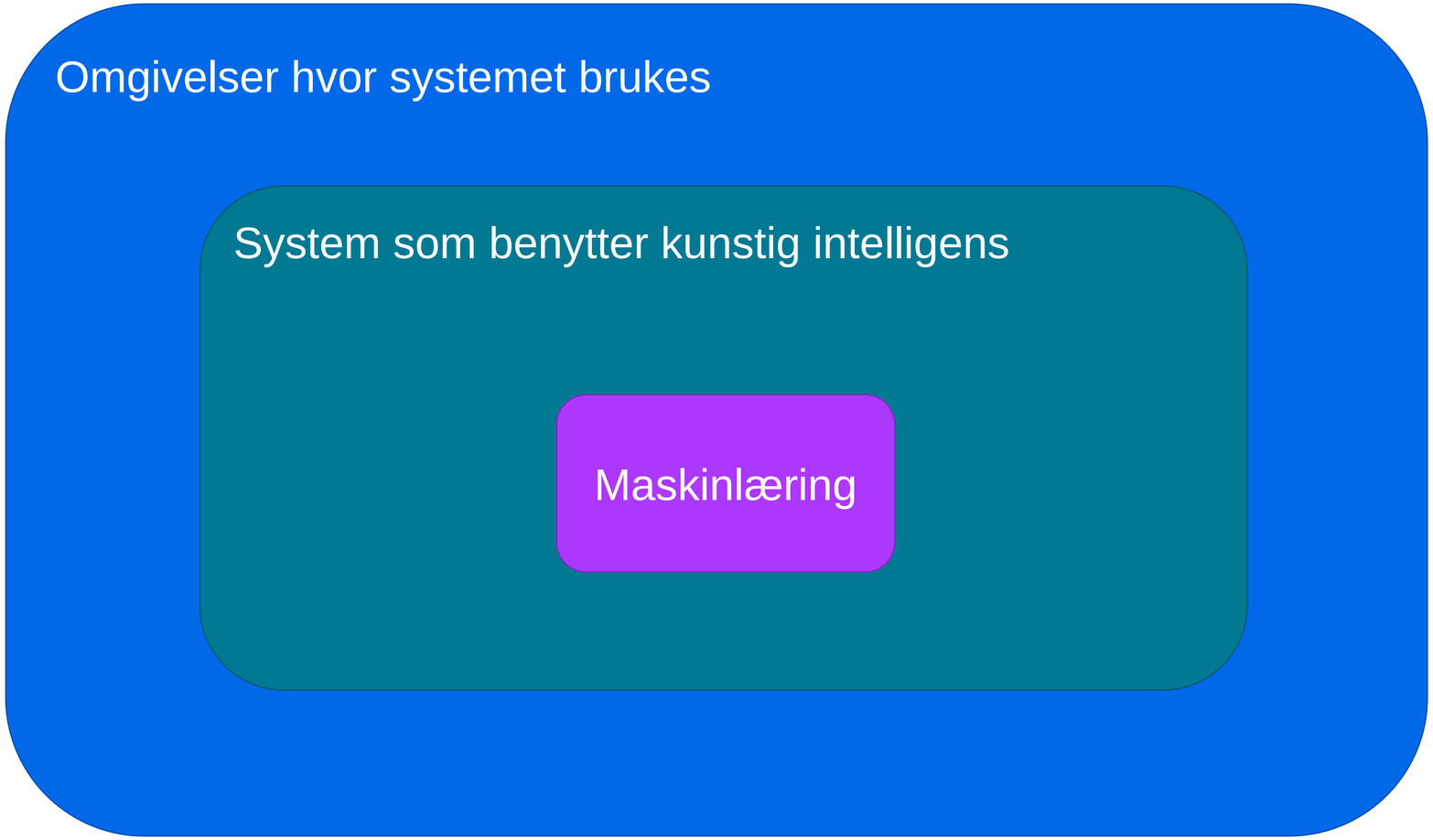
Kan systemet lures av skreddersydde inngangsverdier?



Basert på illustrasjon av Hilde Margrethe Lovett,

Det nasjonale koordineringsprosjektet for KI

Kunstig intelligens bringer med seg ny type risiko – men det gode gamle gjelder fortsatt…

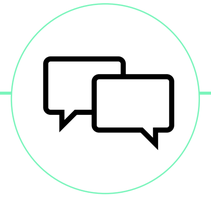


NORMEN

Med all denne risikoen…
– er det like greit å bare droppe det???

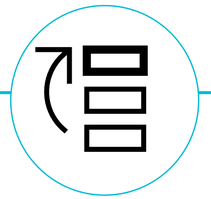
«KI som partner» kan være med å løse oppgavene i helse- og omsorgstjenesten fremover

Analysere medisinske bilder mer effektivt



Effektivisere manuelle oppgaver med tekst

Bedre ressursallokering



Persontilpasse behandling og oppfølging

Bedre helsehjelp, oppfølging og egenomsorg

Illustrasjon Hilde Margrethe Lovett,

Det nasjonale koordineringsprosjektet for KI

Vi må vite hvordan vi vurderer informasjonssikkerhets- og personvernrisiko i systemer som benytter kunstig intelligens

Hjelp som finnes og som er på vei

no/normen/aktuelt-om-normen/sikkerhetsrisiko-i-systemer-som-benytter-kunstig-intelligens--hva-vet-vi-og-hva-kan-vi-gjore

Direktoratet for e-helse

Forside > Normen > Aktuelt > Sikkerhetsrisiko i systemer som benytter kunstig intelligens – hva vet vi, og hva kan

Sikkerhetsrisiko i systemer som benytter kunstig intelligens – hva vet vi, og hva kan vi gjøre?

Ny teknologi knyttet til kunstig intelligens kan bidra til bedre pasientbehandling, bedre ressursbruk, reduserte kostnader og bedre folkehelse. Samtidig bringer kunstig intelligens inn ny sikkerhetsrisiko. Denne må forstås for å kunne ta gode valg og slik evne å maksimere positive virkninger og minske mulige negative virkninger av denne teknologien.

1. Hva vi mener når vi snakker om kunstig intelligens og maskinlæring
2. Hva som er spesielt viktig når det gjelder sikkerhet i maskinlæring
3. Hva vi kan gjøre med den kunnskapen vi har nå
4. Vurdering av sikkerhetsrisiko i en maskinlæringskomponent
5. Vurdering av sikkerhetsrisiko i hele systemet
6. Informasjonssikkerhetsrisiko og annen risiko
7. Oppsummering
8. Anbefalinger til videre lesning

NORMEN

Kunstig intelligens blir ett av temaene på årets Normkonferanse

Aktiviteter i koordineringsprosjektet

Tverretatlig veiledningsmateriale og utredninger

Illustrasjon Hilde Margrethe Lovett,

Det nasjonale koordineringsprosjektet for KI

Aktiviteter i koordineringsprosjektet

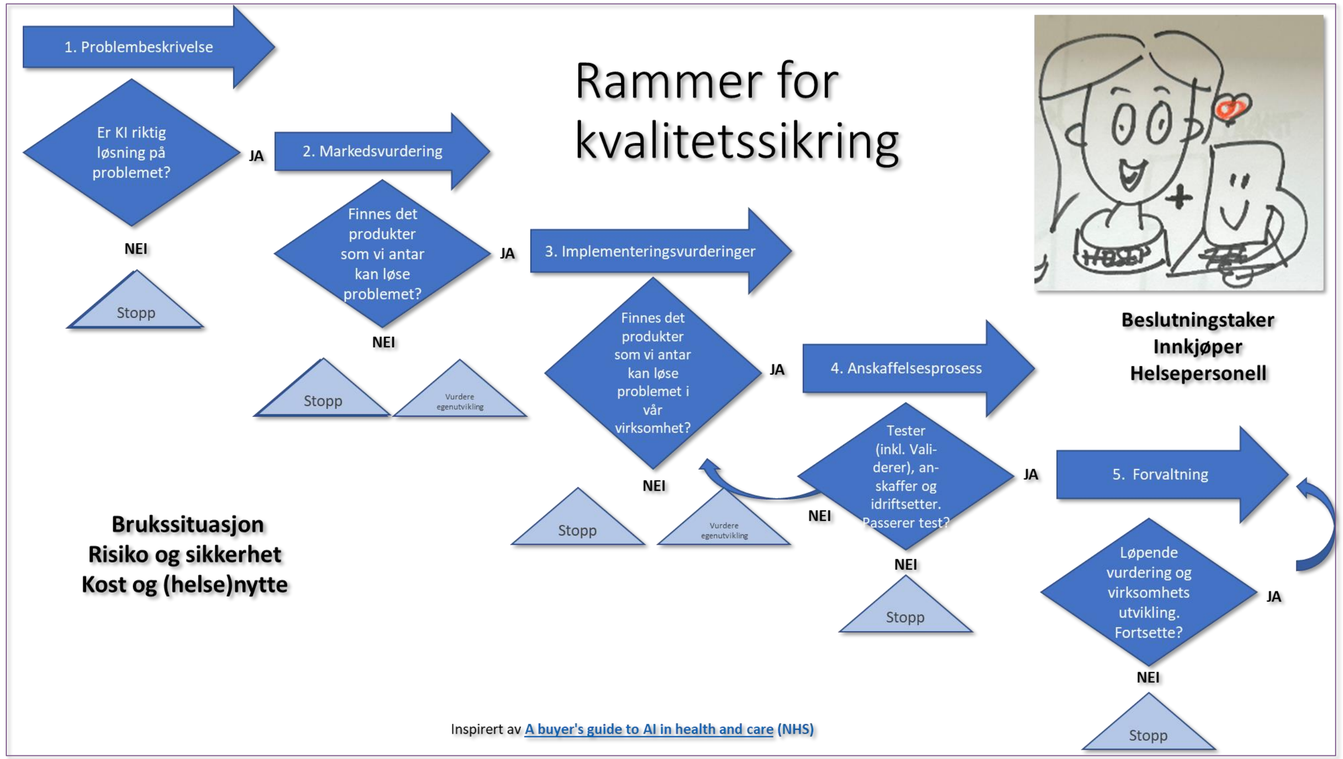
Tverretatlig regulatorisk veiledningstjeneste

Illustrasjon Hilde Margrethe Lovett,

Det nasjonale koordineringsprosjektet for KI

Aktiviteter i koordineringsprosjektet

Rammer for kvalitetssikring



Illustrasjon Hilde Margrethe Lovett,

Det nasjonale koordineringsprosjektet for KI