



id: SI_009

model: anthropic/claude-sonnet-4.6

seeds:

04_risiko: sårbarhet

03_relasjon: makt

05_handling: utforsk

01_perspektiv: juridisk

02_tidhorisont: historisk

Instans SI_009

Seeds: juridisk | historisk | makt | sårbarhet | utforsk

Maktens blindsoner: Hvem eier agentens handlinger?

Historisk sett har autonome systemer alltid skapt et juridisk vakuum — fra jernbanens ansvarsspørsmål på 1800-tallet til algoritmiske handelsfeil som Flash Crash 2010. Mønsteret er konsistent: **makten til å handle flyttes, men ansvaret forblir uklart.**

Det virkelige styringsproblemet her er ikke teknisk. Det er at dere er i ferd med å skape en **ny maktaktør** i organisasjonen — en agent som kan initiere, endre og trigge — uten at ansvarslinjen er juridisk forankret.

Kritisk dissens fra mainstream

Mainstream-rådet er: “Bygg Zero Trust, OPA, audit-pipeline.” Dette er riktig, men **sekundært**. Dere kan ikke teknisk-arkitektere dere ut av et juridisk ansvarsvakuum.

Første prioritet: **Definer hvem som bærer strafferettslig og sivilrettslig ansvar** når agenten endrer data i et fagsystem, f.eks. feilberegner en ytelse eller sletter en journal. Forvaltningsloven §41, personopplysningsloven og evt. sektorspesifikk lov (helse, NAV, etc.) gir ikke unntak for KI-handlinger.

Sårbarhetskart — de tre farligste punktene

1. **Legacy-systemer uten idempotens:** En agent som kaller et gammelt fagsystem to ganger ved timeout, kan dobbeltregistrere en transaksjon. Ingen



rollback. Historisk erfaring: ERP-migreringskatastrofer viser at det er nettopp i koblingssonene mellom nytt og gammelt system feil akkumulerer seg.

2. **Delegert autorisasjon uten revisjonsspor:** Ephemeral credentials som utløper etter at skaden er skjedd, men før loggene er konsolidert — dere vil ikke kunne rekonstruere agentens fullmaktsgrunnlag.
3. **Latent coupling i event-drevne kjeder:** En agent trigger en webhook → som trigger tre andre systemer → ett av dem mangler circuit breaker. Kjedereaksjonen er usynlig i designfasen.

Utforsk dette før piloten

Spørsmålet ingen stiller: **Har dere kartlagt hvilke systemer agenten vil interagere med som er underlagt forvaltningsrettslige krav om menneskelig skjønn?**

Automatisering av enkeltvedtak krever hjemmel (forvaltningsloven §2b).

Neste steg

1. **Juridisk risikoanalyse nå** — kartlegg hvilke agenthandlinger som berører enkeltvedtak, personopplysninger eller lovpålagte prosesser. Gjøres før arkitekturvalg.
2. **Idempotens-audit av legacy-APIer** — test eksplisitt for dobbelt-kall-scenarioer i alle fagsystemer piloten skal berøre.
3. **Definer én navngitt ansvarsperson** (ikke en rolle — en person) med mandat til å isolere agenten innen 15 minutter ved avvikshendelse.