



id: SI_007

model: anthropic/claude-opus-4.6

seeds:

04_risiko: mulighet

03_relasjon: makt

05_handling: utforsk

01_perspektiv: juridisk

02_tidhorisont: langsiktig

Instans SI_007

Seeds: juridisk | langsiktig | makt | mulighet | utforsk

Makten dere gir fra dere — og aldri helt får tilbake

Kjerneproblemet er ikke teknisk. Det er et **maktspørsmål**. Når en autonom agent får skrivetilgang til fagsystemer i en offentlig virksomhet, delegerer dere faktisk **offentlig myndighetsutøvelse** til en ikke-menneskelig aktør. Forvaltningsloven § 2, offentleglova, arkivlova — ingen av disse forutser at en KI-agent initierer vedtaksrelevante prosesser. Denne juridiske blindsonen er farligere enn manglende OPA-policies.

Juridisk-først-arkitektur

Dissens mot mainstream: De fleste rammeverk starter med teknisk kontroll (Zero Trust, policy-as-code). Jeg argumenterer for at dere starter med **juridisk handlingsromskartlegging** — før én linje policy skrives.

Konkret: Klassifiser alle agenthandlinger i tre juridiske soner:

1. **Grønn:** Ingen rettslig virkning (lesing, rapportgenerering) — kan automatiseres fritt
2. **Gul:** Indirekte rettsvirkning (dataendringer som påvirker saksbehandling) — krever human-in-the-loop med juridisk begrunnelse
3. **Rød:** Direkte myndighetsutøvelse — forbudt sone for agenter, punkt



Dette er ikke bare risikostyring. Det er **langsiktig institusjonell overlevelse**. Riksrevisjonen vil om 3-5 år spørre: «Hvem fattet denne beslutningen?» Svaret kan ikke være «agent-42-ephemeral-token-x».

Utforsk mulighetene — men med juridisk anker

Den reelle **muligheten** ligger i grønn sone: agenter som utforsker datalandskapet, identifiserer inkonsistens, forbereder beslutningsgrunnlag. Her er gevinsten enorm og risikoen håndterbar. Piloten bør **eksklusivt** operere her.

For audit-pipeline: Krev **juridisk bevisstandard**, ikke bare teknisk logging. Hver agenthandling må ha: (a) beslutningsgrunnlag/prompt, (b) vurderte alternativer, (c) valgt handling, (d) faktisk systemrespons — i uforanderlig, signert hendelseslogg. Tenk Noark-kompatibel arkivering, ikke bare Elasticsearch.

Langsiktig maktforskyvning

Det undervurderte scenariet: **latent kompetansetap**. Når agenter håndterer kompleksitet i legacy-integrasjoner, mister organisasjonen gradvis evnen til å forstå egne systemer. Om 3 år eier dere ikke lenger kunnskapen om hvorfor ting fungerer. Dette er en irreversibel maktforskyvning fra organisasjon til leverandør/modell.

Tre konkrete neste steg

1. **Juridisk sonekartlegging (uke 1-4):** Engasjer jurist med forvaltningsrettkompetanse til å klassifisere alle planlagte agenthandlinger i grønn/gul/rød. Ingen teknisk arkitektur starter før dette foreligger.
2. **Grønn-sone-pilot med arkivkrav (uke 5-20):** Begrens piloten til lesebaserte agenter med Noark-kompatibel audit-trail. Bruk OPA som policy-gate, men med juridisk definerte regler — ikke teknisk-drevne.
3. **Kompetansebevaringsplan (parallelt):** Dokumenter eksplisitt hvilken systemkunnskap agentene erstatter. Etabler krav om at minst to mennesker til enhver tid kan forklare — og overstyre — alt agenten gjør.