



id: SI_001

model: anthropic/claude-opus-4.6

seeds:

04_risiko: mulighet

03_relasjon: samarbeid

05_handling: vent

01_perspektiv: strategisk

02_tidhorisont: kortsiktig

Instans SI_001

Seeds: strategisk | kortsiktig | samarbeid | mulighet | vent

Kjerneposisjon: Ikke bygg — vent strategisk, men vent aktivt

Min dissens er tydelig: **dere bør ikke innføre agentisk KI i produksjon innenfor 6-9 måneder.** Ikke fordi muligheten er dårlig, men fordi timingen er feil gitt deres tekniske gjeld. Å rulle ut autonome agenter uten konsolidert logging, uten policy-motor, uten Zero Trust, i et landskap med ikke-deterministiske API-er — det er ikke innovasjon, det er ukontrollert risiko i en offentlig virksomhet.

Hva “vent” betyr strategisk

Vent ≠ passivitet. Vent = **bruk pilotvinduet til å bygge fundamentet**, ikke agenten.

Kortsiktig (0-6 mnd) — invester i kontrollplanet:

- Innfør OPA/Styra som sentralisert policy-motor. Begynn med 2-3 fagsystemer. Estimat: 2-3 FTE i 4 mnd.
- Konsolider logging til én audit-pipeline (f.eks. OpenTelemetry → Loki/Elastic) med immutable event-sourcing. Uten dette er enhver agent en blindsoner.
- Implementer ABAC-utvidelse i IAM med støtte for ephemeral credentials og token-scoping (maks 15 min TTL, scopet per API-endepunkt).



Samarbeid som akselerator:

- Gå i dialog med DFØ, Digdir eller andre offentlige virksomheter som piloterer KI-agenter. Del arkitekturbeslutninger åpent — dette er ikke konkurransefortrinn, det er felles risiko.
- Etablér en tverrfaglig «agent-readiness»-gruppe: sikkerhet, jus (personvern/dataminimering), arkitektur og domeneeksperter. Ikke la dette bli et rent tech-prosjekt.

Muligheten som faktisk finnes nå

Kjør agenter i **read-only sandkasser** mot produksjonsdata (anonymisert). La dem foreslå handlinger uten å utføre dem — «human-in-the-loop» for 100 % av skriveoperasjoner. Dette gir:

1. Reelle data om agentadferd, execution graphs og emergent behavior
2. Grunnlag for å kalibrere autonomi-grenser empirisk, ikke teoretisk
3. Audit-trail å øve hendelseshåndtering mot

Blindsoner dere undervurderer

- **Latent coupling:** Agent kaller API A, som trigger webhook til system B, som skriver til database C. Ingen enkeltperson kjenner hele kjeden. Uten execution graph-modellering (DAG-basert, med sideeffekt-annotasjoner) er rollback umulig.
- **Ikke-idempotente API-er i legacy:** Én dublettkalling kan korrumpere data. Circuit breakers (Resilience4j/Polly) og idempotency-keys er minimumskrav — men legacy-monolittene deres støtter trolig ikke dette.
- **Worst case:** Agent med skrivetilgang eskalerer egne rettigheter via feilkonfigurert RBAC, entrer loop, endrer tusenvis av poster i fagsystem uten transaksjonell garanti. Rollback krever manuell rekonstruksjon. I offentlig sektor = potensielt lovbrudd.

Neste steg — konkret

1. **Uke 1-2:** Gjennomfør en «agent-readiness assessment» — kartlegg hvilke systemer som har idempotente API-er, transaksjonsstøtte og tilstrekkelig logging. Resultatet avgjør scope.
2. **Måned 1-3:** Deploy OPA + audit-pipeline som infrastruktur, uavhengig av KI-prosjektet. Dette har verdi alene.



3. **Måned 4-6:** Start read-only agentpilot i sandkasse med full observability. Bruk funnene til å definere realistiske autonomi-grenser før noen agent får skrivetilgang.

Oppsummert: Muligheten er reell, men strategisk tålmodighet er det som skiller kontrollert innføring fra en hendelse dere må forklare til Riksrevisjonen.