



id: SI\_004

model: anthropic/claude-opus-4.6

seeds:

04\_risiko: mulighet

03\_relasjon: konkurranse

05\_handling: implementer

01\_perspektiv: juridisk

02\_tidhorisont: kortsiktig

## Instans SI\_004

**Seeds:** juridisk | kortsiktig | konkurranse | mulighet | implementer

### Dissens: Ikke vent på perfekt arkitektur — implementer juridisk forsvarlig MVP nå

Mens andre instanser trolig anbefaler å bygge komplett kontrollrammeverk først, ser jeg dette annerledes: **den største risikoen er ikke å feile teknisk — det er å miste konkurransevinduet** mens dere perfektjonerer arkitekturen. Offentlige virksomheter som Skatteetaten og NAV piloterer allerede. 6–9 måneder er nok, men bare med riktig avgrensning.

### Juridisk handlingsrom som styrende designprinsipp

Fordi dere er offentlig virksomhet, er forvaltningsloven, offentleglova og GDPR ikke valgfrie hensyn — de **er** arkitekturen. Konkret:

- **Artikkel 22 GDPR** krever menneskelig inngripen ved automatiserte enkeltvedtak. Definer agentenes handlingsrom juridisk **før** teknisk: agenter som **forbereder** beslutninger er lovlige, agenter som **fatter** vedtak krever eksplisitt hjemmel.
- **Forvaltningsloven § 40a** (automatisert saksbehandling) setter krav til dokumenterbarhet som direkte dikterer loggarkitekturen.



---

## Kortsiktig implementeringsstrategi (6-9 måneder)

### Måned 1-2: Juridisk sandkasse

- Kartlegg 2-3 prosesser der agenter kun gjør lesing + forslag (aldri skriving til fagsystemer). Eksempel: automatisk sammenstilling av saksdokumenter, utkast til standard svar.
- Implementer OPA (Open Policy Agent) som policy-motor foran eksisterende RBAC. OPA krever ikke IAM-oppgradering — den legges som lag.

### Måned 3-5: Kontrollert skrivetilgang

- Innfør **transaksjonsbasert godkjenning**: agenten foreslår, menneske bekrefter via køsystem (f.eks. Temporal.io for orkestrering med innebygd rollback).
- Rate-limiting: maks 50 skriveoperasjoner/time per agent. Hardkodet, ikke konfigurerbart av agenten selv.
- Alle API-kall via dedikert gateway (Kong/Apigee) med egne service accounts — **aldri** gjenbruk av brukeridentiteter.

### Måned 6-9: Kontrollert autonomi

- Gradvis utvide handlingsrom basert på revisjonsdata. Kun prosesser med fullstendig reversibilitet.

---

## Blindsoner dere vil undervurdere

1. **Eldre fagsystemer uten idempotens**: En agent som retry-er et kall kan opprette duplikate vedtak. Krev idempotency-keys på alle integrasjonspunkt.
2. **Prompt injection via eksterne API-svar**: Agenten kaller ekstern tjeneste, svaret inneholder instruksjoner som endrer agentens oppførsel. Svar må saniteres som uklart input.
3. **Juridisk sporbarhet ved feilkjeder**: Når agent A trigger agent B som endrer data — hvem er ansvarlig? Implementer korrelerings-ID-er fra dag 1.

---

## Konkurransesargumentet

DFØ og Digdir beveger seg mot felles KI-tjenester. Virksomheter med dokumentert piloterfaring vil forme standardene — ikke de som ventet.



---

## Neste steg

1. **Uke 1:** Gjennomfør juridisk vurdering med personvernombud — definer konkret hvilke handlingstyper som er tillatt uten vedtakshjemmel.
2. **Uke 2-3:** Deploy OPA som policy-lag foran én utvalgt prosess, med Temporal.io for orkestrering og rollback.
3. **Uke 4:** Etabler sentral loggpipeline (OpenTelemetry → Elasticsearch) med korrelerings-ID på tvers av alle systemer — dette er juridisk nødvendig infrastruktur, ikke «nice to have».