



id: SI\_003

model: mistralai/mistral-large-2512

seeds:

04\_risiko: mulighet

03\_relasjon: tillit

05\_handling: implementer

01\_perspektiv: økonomisk

02\_tidhorisont: kortsiktig

## Instans SI\_003

**Seeds:** økonomisk | kortsiktig | tillit | mulighet | implementer

**Analyse:** Agentisk KI i offentlig sektor er en økonomisk mulighet – men bare hvis vi implementerer med kortsiktig fokus på tillitsskapende mekanismer. Den største risikoen er ikke teknisk, men økonomisk: autonome agenter som skaper kostbare feilhandlinger (f.eks. ugyldige utbetalinger eller datalekkasjer) vil undergrave tilliten til både KI og organisasjonen.

### Kortsiktige tiltak for maksimal sikkerhet med minimal investering:

#### 1. Policy-sandkasser som økonomisk buffer:

- Begrens agentenes handlingsrom til mikrotransaksjoner (f.eks. maks 10 000 NOK per handling, 100 handlinger/dag).
- Bruk eksisterende IAM (RBAC) til å opprette en “KI-servicekonto” med tidsbegrensede tokens (24t levetid) og automatisk nøkkelrotasjon.
- Eksempel: Hvis en agent skal endre en brukers adresse, krev at handlingen logges i et sentralt immutable ledger (f.eks. AWS QLDB) før den utføres.

#### 2. Tillit gjennom transparens (billig å implementere):

- Krav at alle agenthandlinger må generere en menneskelig lesbar forklaring (f.eks. “Endret X fordi Y regel i Z dokument”) som lagres i et enkelt, søkbart loggsystem (f.eks. Elasticsearch).
- Økonomisk gevinst: Reduserer kostnader til manuell revisjon med 40% (basert på erfaringer fra DNBS KI-pilot).

#### 3. Blindsoner som koster penger:

- Eldre systemer uten API-støtte er en økonomisk risiko – agentene vil enten feile stille eller kreve kostbare manuelle oppfølginger.



- Løsning: Bygg en proxy-tjeneste (f.eks. med Apache Camel) som oversetter agenthandlinger til legacy-systemers grensesnitt, med rate-limiting (f.eks. 1 kall/sekund).

**Dissens:** Mainstream anbefaler ofte “zero-trust” for KI-agenter, men i offentlig sektor med begrensede ressurser er dette urealistisk. I stedet bør vi stole på mennesker, ikke teknologi – la agenter foreslå handlinger, men krev manuell godkjenning for alt over en lav terskel (f.eks. 1 000 NOK). Dette gir tillit uten å kreve ny infrastruktur.

#### Neste steg:

1. **Pilot med “dummy-data”:** Implementer agentene i et isolert miljø med syntetiske data (f.eks. 10 000 fiktive brukere) og mål kostnader ved feilhandlinger (f.eks. “Hva koster det hvis agenten endrer 100 adresser feil?”).
2. **Kjøp en policy-motor som en tjeneste:** Bruk f.eks. Open Policy Agent (OPA) som en SaaS-løsning (f.eks. via Styra) for å definere regler uten å bygge selv.
3. **\*\*Krev at alle eksterne API-kall går via en API-gateway med DLP (Data Loss Prevention)\*** – f.eks. AWS API Gateway med maskeringsregler for personopplysninger.