



id: SI_002

model: mistralai/mistral-large-2512

seeds:

04_risiko: mulighet

03_relasjon: makt

05_handling: implementer

01_perspektiv: strategisk

02_tidhorisont: kortsiktig

Instans SI_002

Seeds: strategisk (maktbalanse) | kortsiktig (pilotvindu) | makt (kontroll over autonome agenter) | mulighet (gevinst i heterogent landskap) | implementer (konkrete grep innen 9 måneder)

Strategisk maktbalanse: Autonome agenter som midlertidig privilegert aktør

Agentisk KI er ikke et teknisk problem, men en maktfordelingsutfordring. I et heterogent landskap vil agenter fungere som temporære superbrukere – med potensial til å bryte siloer, men også til å destabilisere. **Kortsiktig strategi:** Behandle agenter som midlertidige maktprosjer med eksplisitte grenser, ikke permanente løsninger. Pilotvinduet på 6–9 måneder er en mulighet til å teste maktbalansen før den låses.

Implementerbar kontroll: 3 grep innen 9 måneder

1. Policy-sandkasse som maktbegrenser (OPA + Kyverno)

- Kortsiktig: Implementer Open Policy Agent (OPA) med pre-execution policy evaluation for alle agenthandling. Eksempel: En agent som skal oppdatere en pasientjournal må først passere en policy som sjekker:
 - Er handlingen innenfor tillatt tidsvindu? (ABAC-regel)
 - Er datasettet “sensitivitetsklasse B” eller lavere? (RBAC-utvidelse)
 - Er API-kallet idempotent? (Statisk sjekk mot API-kontrakt)



- Makt: OPA fungerer som en automatisert kontrollør som kan stanse handlinger før de utføres. Bruk Kyverno for Kubernetes-baserte mikrotjenester.

2. Ephemeral credentials som maktbegrenser (HashiCorp Vault + SPIFFE/SPIRE)

- Kortsiktig: Erstatt statiske service accounts med tidsbegrensede tokens (levealder: 5–15 minutter) utstedt via Vault. Hver agent får en unik identitet via SPIFFE, og tokens scopes til én spesifikk handling (f.eks. “opprett sak i fagsystem X”).
- Makt: Forhindrer “token sprawl” og sikrer at agenter ikke kan eskalere privilegier. Mulighet: Integrer med IAM-plattformen for å dynamisk opprette midlertidige roller.

3. Audit-graf som maktoversikt (OpenTelemetry + Grafana Loki)

- Kortsiktig: Bygg en execution graph for hver agenthandling ved å logge:
 - Prompts (input til agenten)
 - Mellomsteg (beslutninger, API-kall, retries)
 - Sideeffekter (endringer i databaser, eksterne systemer)
- Makt: Lag en tidslinje som viser hvordan agenter beveger seg på tvers av systemer. Bruk OpenTelemetry for å spore kallkjeder og Grafana Loki for å konsolidere logger. Mulighet: Automatisk generere “rollback-skript” basert på grafen.

Dissens: Legacy-systemer som blindsoner for makt

Mainstream-fokuset på Zero Trust og policy-as-code overser at legacy-systemer ikke er designet for dynamisk kontroll. Risikoen er ikke at agenter gjør feil, men at de utløser latent kobling – f.eks. at et API-kall til en monolitt utløser en kjedereaksjon i et annet system. **Løsning:** Modellér worst-case-scenarier som “agent utløser batch-jobb i SAP som overskriver 10 000 poster” og implementer circuit breakers (f.eks. via Envoy) som stopper agenter ved uvanlige mønstre.

Neste steg (konkrete, innen 30 dager)

1. Pilotpolicy for én agent i ett system

- Velg én lavrisiko-prosess (f.eks. “automatisk arkivering av dokumenter”) og definer:
 - Tillatte handlinger (API-kall, datakilder)
 - Policy-regler (OPA-skjelett)
 - Audit-krav (hva må logges)



- Mål: Få en fungerende policy-sandkasse innen 30 dager.

2. Vault-integrasjon for ephemeral credentials

- Sett opp HashiCorp Vault og konfigurer en service account til å utstede tidsbegrensede tokens.
- Mål: Bytt ut én statisk nøkkel med ephemeral credentials innen 30 dager.

3. Execution graph-prototype

- Bruk OpenTelemetry til å spore én agenthandling (f.eks. "hent data fra API X → skriv til database Y").
- Mål: Visualiser grafen i Grafana innen 30 dager.

SI_002s perspektiv: Agentisk KI er en midlertidig maktutvidelse – ikke en permanent løsning. Pilotvinduet er en mulighet til å teste kontrollmekanismer før agenter får for mye makt. Fokusér på implementerbare grep som begrenser autonomi, ikke på perfekte løsninger.