



id: SI_002

model: google/gemma-4-26b-a4b-it

seeds:

04_risiko: mulighet

03_relasjon: samarbeid

05_handling: implementer

01_perspektiv: strategisk

02_tidhorisont: historisk

Instans SI_002

Seeds: strategisk | historisk | samarbeid | mulighet | implementer

Analysen av agentisk KI i deres kontekst avdekker et fundamentalt paradigmeskifte: Dere går fra å styre instruksjoner til å styre intensjoner.

Strategisk arkitektur og kontroll Historisk har sikkerhet vært reaktiv (perimeter). Med autonome agenter må arkitekturen være **prediktiv**. Dere kan ikke vente på en logg; dere må kontrollere execution graph i sanntid.

- **Prinsipp:** Implementer en “Agentic Sidecar”-modell. Hver agent må operere gjennom en proxy som fungerer som en Policy Enforcement Point (PEP).
- **Kontroll:** Bruk OPA (Open Policy Agent) ikke bare for tilgang, men som en semantic firewall. Hvis agentens forespurt handling avviker fra den planlagte grafen, blokkeres den umiddelbart.

IAM og integrasjon: Fra statisk til flytende Dagens RBAC er utilstrekkelig for agenter som opererer i “skiftende kontekster”.

- **Løsning:** Gå for Ephemeral Identity. Agenter skal ikke ha faste service accounts, men motta kortlevde (minutter) JWT-tokens med ekstremt snever audience og scope, generert av en sentral kontrollflate basert på oppgavens natur.
- **Risiko ved legacy:** API-er uten idempotens er “miner i landskapet”. Dere må bygge et **Idempotency Layer** mellom agenten og legacy-systemene som logger og sikrer at et duplisert kall fra en “looping” agent ikke skaper doble transaksjoner.

Samarbeid og implementering av sikkerhet Muligheten ligger i å bygge en Human-in-the-loop (HITL) flywheel. I stedet for å se på menneskelig godkjenning som en hindring, må det integreres som en del av agentens confidence score-modell. Lav konfidens = automatisk eskalering til menneske.



Blindsoner og emergent risiko Den største faren er latent coupling: En agent endrer en parameter i et fagsystem som virker uskyldig, men som trigger en kjedereaksjon i et annet system via en uforutsett logisk kobling. Dette er ikke en teknisk feil, men en systemisk egenskaper ved kompleksitet.

Neste steg:

1. **Pilot (Måned 1-2):** Etabler en “Shadow Mode”-pipeline der agenter utfører handlinger i et sandkasse-miljø med replikert data, for å kartlegge execution graphs uten risiko.
2. **Arkitektur (Måned 3-4):** Implementer en sentralisert OPA-instans og en “Sidecar Proxy” for de første 3 kritiske API-ene.
3. **Governance:** Definer “Kill-switch”-protokoller som kan isolere alle agent-identiteter på tvers av IAM på under 30 sekunder.