



id: SI_009

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

seeds:

04_risiko: sårbarhet

03_relasjon: tillit

05_handling: implementer

01_perspektiv: strategisk

02_tidhorisont: langsiktig

Instans SI_009

Seeds: strategisk | langsiktig | tillit | sårbarhet | implementer

Analysen identifiserer en fundamental asymmetri: Dere planlegger å innføre **strategisk** autonomi i et systemlandskap preget av teknisk **sårbarhet**. Å gi agenter tilgang til legacy-monolitter uten deterministiske kontrakter er ikke en teknisk oppgradering; det er en introduksjon av systemisk uforutsigbarhet.

Arkitektur og kontroll: Fra tillit til verifisering

Mainstream-tilnærmingen er å bygge “sikkerhet rundt agenten”. Min analyse tilsier at dere må bygge “sikkerhet som agenten”.

- **Prinsipp:** Implementer en Sidecar Proxy Architecture for alle agenter. Agenten får aldri snakke direkte med et API. All trafikk må gå gjennom en kontroll-node (f.eks. en OPA-integrert Envoy-proxy) som fungerer som en “Policy Enforcement Point”.
- **Execution Graph:** Modeller agentens handlinger som en Directed Acyclic Graph (DAG). Hver node i grafen må validere “intent” mot en pre-definert policy før eksekvering.

IAM og Integrasjon: Den temporære modellen

Glem statiske service accounts. For å sikre **langsiktig** stabilitet må dere **implementere** Ephemeral Identity Provisioning.

- Bruk en teknisk løsning (f.eks. HashiCorp Vault) som genererer tokens med en TTL (Time-To-Live) på maksimalt 5–15 minutter, skreddersydd for den spesifikke oppgaven (Token Scoping).



- **Risiko:** Legacy-systemer mangler idempotens. En agent i en loop kan doble transaksjoner. Dere må implementere et abstraksjonslag (Adapter Pattern) som tvinger idempotens-nøkler på alle legacy-kall.

Blindsoner og Emergent Behavior

Den største **sårbarheten** er ikke et enkelt API-kall, men latent coupling. Agenten kan oppdage en logisk vei mellom to urelaterte systemer som skaper en kjedereaksjon (f.eks. endring av en parameter i system A som trigger en feil i system B via en uforutsett database-trigger).

- **Tillit** kan ikke bygges på agentens logikk, men på evnen til å isolere den.

Strategiske neste steg (Implementeringsplan)

1. **Etabler en "Blast Radius" Sandbox:** Identifiser ett isolert datasett og tre moderne mikrotjenester. Ingen legacy-systemer får berøres i pilotfasen (6 mnd).
2. **Implementer en "Circuit Breaker" for agenter:** Bygg en uavhengig monitoreringstjeneste som automatisk deaktiverer agentens credentials dersom rate-limits eller uventede feilrater (4xx/5xx) overstiger en terskel på 2% i løpet av 60 sekunder.
3. **Policy-as-Code prototype:** Sett opp en sentral OPA-instans og definer de første 10 "harde" reglene (f.eks. "Ingen sletting av data i system X", "Maks 5 API-kall per minutt per agent").