



Debrief — google/gemma-4-26b-a4b-it

- Instanser: 9
 - Tokens inn: 19,889
 - Tokens ut: 7,555
 - Kostnad: \$0.0042
 - Kjørt: 2026-04-15 17:25:22.798283+00:00
-

Sverm-debrief: Agentisk KI-implementering

Konsensus

Svermen er enstemmig om at det nåværende systemlandskapet er **uegnet** for direkte agent-interaksjon. Følgende punkter er kritiske for suksess:

- **Proxy-basert arkitektur:** Agenter må aldri kommunisere direkte med systemer. En “Intermediary Proxy” eller “Sidecar” (f.eks. Envoy/OPA) er obligatorisk for å fungere som Policy Enforcement Point.
 - **Ephemeral Identity:** Statiske service accounts må erstattes med kortlevde (minutter), oppgave-spesifikke tokens (JIT-tilgang) for å begrense skadeområdet.
 - **Idempotens-lag:** Siden legacy-systemer mangler transaksjonell garanti, må proxy-laget tvinge frem idempotens for å forhindre duplikate handlinger ved retries/loops.
 - **Hard Kill-switch:** Det må eksistere en umiddelbar, sentralisert mekanisme for å terminere alle aktive agent-sesjoner og tokens på tvers av hele landskapet.
-

Dissens

Det er en fundamental konflikt mellom **ambisjon og realisme**:

- **Tempo:** Noen instanser (SI_001, SI_004) anbefaler å **stoppe piloten** umiddelbart fordi infrastrukturen mangler fundamentale kontrollmekanismer. Andre (SI_003) foreslår å **forhandle risiko** ved å begrense piloten til en “juridisk sandbox” for å opprettholde konkurranseevne.



- **Kontrollmetode:** Diskusjon om man skal fokusere på prediktiv kontroll (modellere execution graphs/DAGs før utførelse) eller reaktiv kontroll (circuit breakers og overvåking av feilrater).

Blindsoner avdekket

Svermen identifiserte risikoer som går utover teknisk svikt:

- **Latent Coupling:** Den største faren er ikke agentens feil, men uforutsette kjedereaksjoner mellom urelaterte systemer via uformelle logiske koblinger (f.eks. en database-trigger som reagerer på en agent-endring).
- **Semantic Drift:** Agenten kan tolke en svak API-kontrakt teknisk korrekt, men logisk feil, noe som skaper systemisk korrupsjon.
- **Emergent Behavior som “Optimalisering”:** Risikoen for at agenter finner “smarte” snarveier som er teknisk lovlige, men som bryter med virksomhetens intensjon eller juridiske rammeverk.

Anbefalinger

1. **Etabler en “Mediator Layer”:** Bygg en API-gateway/proxy som håndterer rate-limiting, skjemavalidering og idempotens før agenten får røre legacy-systemer.
2. **Definer “Blast Radius”:** Begrens piloten til Read-Only eller moderne mikrotjenester. Identifiser de 3 mest kritiske systemene og innfør obligatorisk Human-in-the-loop for alle endringer der.
3. **Implementer “Reasoning Logs”:** Bygg en audit-pipeline som kobler Prompt -> Intent -> Decision -> API-kall -> Effekt . Uten denne er det umulig å rekonstruere hendelser post-mortem.
4. **Automatisert Isolasjon:** Implementer Circuit Breakers som automatisk deaktiverer agent-identiteter ved uventet økning i feilrater eller uvanlige transaksjonsmønstre.