



id: SI_007

model: anthropic/claude-sonnet-4.6

seeds:

04_risiko: resiliens

03_relasjon: makt

05_handling: implementer

01_perspektiv: strategisk

02_tidhorisont: langsiktig

Instans SI_007

Seeds: strategisk | langsiktig | makt | resiliens | implementer

Makten du ikke ser: agentens systemposisjon

Mainstream-rådet om Zero Trust og policy-as-code er korrekt, men taktisk. Det strategisk kritiske spørsmålet undervurderes: **hvem eier agentens makt over tid?**

En autonom agent som opererer på tvers av systemer akkumulerer de facto systemkontroll – ikke gjennom ondsinnet design, men gjennom legitim bruk. Etter 6-9 måneder pilot vil agenten ha etablert API-mønstre, datatilganger og prosess-koblinger som er **vanskelige å reversere**. Dette er ikke en sikkerhetsrisiko – det er en maktforskyvning.

Arkitektur som resiliensgrunnlag

Ikke bygg for optimal drift. Bygg for **kontrollert degradering**:

- **Execution graph som konstitusjonslagring:** Modeller agentens tillatte handlingsrom som en DAG (directed acyclic graph) med eksplisitte knutepunkter for menneskelig godkjenning. Ikke bare logging – dette er virksomhetens konstitusjon for hva agenten er.
- **Shadow-modus i 90 dager:** Agenten logger og foreslår, men eksekverer ikke. Du kartlegger emergent behavior før du gir reell makt.
- **Dead man's switch:** Agenten krever aktiv re-autorisasjon hver 24. time for høy-risiko handlinger. Ingen re-autorisasjon = automatisk nedgradering til read-only.



IAM: ephemeral er ikke nok

Ephemeral credentials løser autentisering, ikke autorisasjon. Implementer **kontekstuell tilgangsevaluering** per steg i execution graph: samme agent, samme oppgave, men ulik systemtilstand = ulik tilgang. OPA med Rego-policies pr. handlingstype er minimum MVP.

Den undervurderte blindsonen: latent coupling i legacy

Eldre monolitter uten idempotente API-er er den reelle worst-case. En agent som trigger samme endepunkt to ganger under retry-logikk kan dobbeltskrive finansdata, initiere doble varsler eller korrupte tilstandsmaskiner. **Ingen rollback-mekanisme redder deg her** – du trenger kompenserende transaksjoner designet før agenten går live.

Strategisk langsiktig realitet

Virksomheten har ikke Zero Trust i dag. Å implementere agentisk KI og Zero Trust parallelt på 6-9 måneder med begrensede ressurser er en fiasko-oppskrift. Velg én: utsett agentene, eller aksepter høyere residualrisiko med hard scope-begrensning.

Neste steg

1. **Uke 1-2:** Kartlegg alle legacy-systemer agenten vil nå – test idempotens manuelt på topp-10 endepunkter
2. **Uke 3-6:** Implementer shadow-modus med strukturert audit-logging (OpenTelemetry + sentralisert sink) før én eneste autonom handling
3. **Måned 2:** Definer og vedta “agentens konstitusjon” – et policy-dokument som governance-laget signerer, ikke bare IT