



id: SI\_006

model: qwen/qwen3.6-plus

seeds:

04\_risiko: sårbarhet

03\_relasjon: makt

05\_handling: implementer

01\_perspektiv: operasjonell

02\_tidhorisont: langsiktig

## Instans SI\_006

**Seeds:** operasjonell | langsiktig | makt | sårbarhet | implementer

**Arkitektur & Makt:** Unngå eksterne rammeverk (CrewAI/LangGraph). Din helsekontekst og cloud-only-krav krever Azure OpenAI + Azure Durable Functions + Copilot Studio. Dette gir Entra ID-native identitet, EUs dataresidens og bevarer Microsoft-kontroll. Copilot Studio er raskt for prototyper, men låser orkestrering i Power Platform – en langsiktig operasjonell blindgate. Bygg swarm-logikken i Durable Functions for full styring over parallelitet, retry og auth.

**Agent-spesialisering (Case 1):** Duplisering er en sårbarhet. Bruk dynamisk kontekst-routing. Hver agent henter kun relevant vector-segment: `ctx = search.query(role="regulatory", filter="active_only", top_k=5)`. Sammenstilling skjer via en «synthesis orchestrator» som veier output basert på fase, ikke gjennomsnitt. Pseudo-flyt:

```
resp = llm.generate(ctx, temp=0.1, schema=role_output_json)
out = orchestrator.aggregate([resp_1..100], weights=role_priority)
```

Kaos unngås med streng JSON-schema, deterministisk temperatur og fase-basert vektet avstemning.

**Inbox-modell (Case 2):** Azure Service Bus som kø. Durable Functions fordeler via kompleksitets-gate: `if case.tokens > threshold -> route_to_40; else -> route_to_mini`. Base system-prompt er felles; case-spesifikk data injiseres via function-calling. Consistency sikres med output-parsers og idempotente batch-jobber. Timer-batch gir deg rom for kostnads- og feil-optimalisering.

**Data & Sårbarhet:** Aldri eksponer raw PHI. Bruk Azure AI Search med private endpoints + Managed Identity. Entra ID RBAC gir `on-behalf-of` tilgang til SharePoint/Dataverse. Kritisk: Zero-trust injection – agenter kun får masked metadata eller indeksreferanser,



med full audit-logg per token-request. Helsevesen tåler ikke prompt-leaks eller model-drift uten sporbarhet.

**Observasjon & Kostnad:** Application Insights + custom dimensions. Track `agent_id`, `tokens`, `cost`, `fail_rate`, `latency` per orchestrator-run. Optimer med 70/30 model-routing, Redis-caching for repetitiv case-logikk, og auto-fallback ved rate-limit. Sverm uten streng cost-gate blir en økonomisk lekkasje.

**Fase & Fallgruver:** 0–3 mnd: Copilot Studio → enkelt-agent. 3–6 mnd: Durable Functions → batch-sverm. Quick win: 3-agent pasientjournal-sammendrag. Sverm feiler ved deterministiske oppgaver, fragmentert grunnlag, eller der compliance krever menneskelig signatur før iterasjon. Det er en makt-overføring til algoritmen – håndter det med governance, ikke hype.

**Neste steg:**

1. Oppsett Azure AI Search med private endpoints & Entra ID RBAC for zero-trust injeksjon.
2. Bygg minimal Durable Function-orchestrator med 3 roller, batch-test mot syntetiske helsecases.
3. Definer «kill-switch», token-budsjetter og compliance-audit før produksjonsrulling.