



id: SI_001

model: anthropic/claude-sonnet-4.6

seeds:

04_risiko: mulighet

03_relasjon: konkurranse

05_handling: implementer

01_perspektiv: strategisk

02_tidhorisont: langsiktig

Instans SI_001

Seeds: strategisk | langsiktig | konkurranse | mulighet | implementer

Strategisk posisjon: Azure OpenAI er eneste forsvarbare valg

For en 200-person helseorganisasjon med streng datasensitivitet er **ikke** valgfriheten mellom plattformer reell. CrewAI, Anthropic-hosted og lignende sender data utenfor din kontrollzone. Det er ikke et kompromiss – det er et compliance-brudd som venter på å skje. **Azure OpenAI med Private Endpoints + Managed Identity** er den eneste arkitekturen som bevarer langsiktig konkurranseevne uten regulatorisk eksponering.

Arkitektur: Orkestrering via Azure Functions + Durable Orchestrators

Bruk **Azure Durable Functions** (fan-out/fan-in pattern) som sverm-orkestrator – cloud-only, serverless, og native Microsoft-stack:

```
# Orkestrator-funksjon (pseudo)
@df.orchestrator_generator
def swarm_orchestrator(context):
    agent_roles = load_role_configs("sharepoint://agent-roles/health-
analysis.json")
```



```
# Case 1: Parallell perspektivanalyse
tasks = [context.call_activity("run_agent", {
    "role": role,
    "system_prompt": role["specialized_prompt"],
    "context_chunks": inject_from_dataverse(role["context_filter"])
}) for role in agent_roles[:100]]

results = yield context.task_all(tasks)
return synthesize_with_meta_agent(results)
```

Kritisk differensiering: Hver agent får **filtrert kontekst fra Dataverse** basert på sin rolle – ikke samme dokument til alle. Regulatory reviewer får lovtekst-chunks; cost analyser får budsjett-tabeller. Kontekst-injeksjon skjer **inne i Azure-tenanten via Managed Identity** – ingen data forlater perimeteret.

Failure modes du MÅ planlegge for

- **Token-eksplosjonen:** 100 agenter × 8k tokens = potensielt 800k tokens per kjøring. Batch-godkjent for timer, men **sett hard quota per agent-rolle** via Azure OpenAI token budgets
- **Konsistensproblemet i Case 2:** Variable case-kompleksiteter bryter consistency. Løs med **complexity scoring upfront** – ruter enkle cases til GPT-4o-mini, komplekse til GPT-4o
- **Observerbarhet:** Azure Monitor + Application Insights med custom dimensions per agent-ID. Uten dette er produksjonsdrift blind

Realistisk timeline

- **Mnd 1-2:** PoC med 5 spesialiserte agenter via Durable Functions – velg én høyverdi-problemstilling
- **Mnd 3-4:** Dataverse-integrasjon + Entra ID-basert tilgangsstyring per agent-rolle
- **Mnd 5-6:** Skalering til 20-agent parallellkjøring med full observerbarhet

Konkrete neste steg

1. **Implementer Durable Functions fan-out PoC denne sprenten** – ikke vent på perfekt arkitektur



2. **Kartlegg hvilke Dataverse-tabeller og SharePoint-biblioteker som er lovlige å injisere** – dette blokkerer alt annet
3. **Sett Azure OpenAI token-kvoter per deployment** før første multi-agent kjøring – cost-overruns i helse-kontekst er politisk ødeleggende