



id: SI_008

model: anthropic/claude-sonnet-4.6

seeds:

04_risiko: mulighet

03_relasjon: makt

05_handling: forhandle

01_perspektiv: operasjonell

02_tidhorisont: kortsiktig

Instans SI_008

Seeds: operasjonell | kortsiktig | makt | mulighet | forhandle

Maktposisjon før arkitektur

Glem teknologivalget i første omgang. Du sitter med en **forhandlingsposisjon** du ikke utnytter: 200 ansatte i helsevesen = Microsoft vil beholde dere. Det betyr at Azure OpenAI Service med private endpoints, data residency i EU-North, og HIPAA/GDPR-compliance ikke er en kostnad du ber om — det er noe du **forhandler inn** i eksisterende EA-avtale. Ring Microsoft CSP-partner før arkitekturmøtet.

Kortsiktig operasjonell realitet: Copilot Studio er en felle

Copilot Studio + Power Automate ser enkelt ut, men for ekte sverm-orkestrering (100 parallelle agenter) treffer du umiddelbart **throttling-grenser** på Power Platform (25 samtidige flows uten Premium-lisens per bruker). Det er ikke sverm — det er kø. **Azure OpenAI + Semantic Kernel** er eneste reelle vei for helsevesen + cloud-only + batch-aksept.

Mulighetsvindu: Semantic Kernel som orkestrator

```
# Pseudo-kode: rolle-differensiert agent-spesialisering
kernel = Kernel()
agents = [
    ChatCompletionAgent(kernel, system_prompt=f"Du er {role}. Analyser kun
{domain}." )
```



```
for role, domain in AGENT_ROLES # 100 unike (role, domain)-par
]
# Parallel dispatch via asyncio – batch på 20 om gangen mot token-grenser
results = await asyncio.gather(*[a.invoke(problem) for a in agents[:20]])
```

Nøkkelen: **ingen agent deler system-prompt**. Rolle-differensiering skjer via domenespesifikke instruksjoner + RAG mot SharePoint via Microsoft Graph API — data forlater aldri tenant.

Aggregering uten kaos

Ikke aggreger 100 rå-svar. Bygg en **meta-agent** som kun ser strukturerte JSON-outputs (score, begrunnelse, flagg) fra primæragentene. Dette er operasjonelt håndterbart og revisjonsbart — kritisk i helsevesen.

Failure mode du må unngå

Sverm er feil løsning når casene har **sekvensiell avhengighet** (output fra agent A endrer premiss for agent B). Da er pipelines bedre. Test dette eksplisitt i pilot.

Konkrete neste steg

1. **Uke 1-2**: Forhandle Azure OpenAI private endpoint + EU data residency inn i eksisterende Microsoft-avtale — dette er din maktposisjon, bruk den nå
2. **Uke 3-4**: Bygg minimal Semantic Kernel-pilot med 5 rolle-differensierte agenter mot én SharePoint-kilde — bevis konseptet uten full infrastruktur
3. **Uke 5-6**: Implementer token-tracking via Azure Monitor + Application Insights før du skalerer — kostnader uten observerbarhet er blindflyging i produksjon