



id: SI_006

model: anthropic/claude-haiku-4-5

seeds:

04_risiko: sårbarhet

03_relasjon: makt

05_handling: utforsk

01_perspektiv: strategisk

02_tidhorisont: kortsiktig

Instans SI_006

Seeds: strategisk | kortsiktig | makt | sårbarhet | utforsk

Analysen

Du identifiserer en **maktstruktur i sårbarhet**. Det er ikke ett “hvem” — det er et system av kortsiktige incentiver som låser alle.

Konkret maktmekanikk:

Departementhoder, forskningssentre-direktører og NFR-paneler har 3-5 år på seg til å vise “sin” forskning relevans. En professor som sier “GPT-4 gjorde min karrièrespesialisering redundant” blir ikke satt på paneler eller søknadsutvalg. Det er karriærselv mord på kort sikt. Systemet selekterer derfor for politisk overlevelses-narrativ, ikke sannhet.

Mekanismen som stikler:

- Grant-strukturen:** NFR/Sintef krever “originale forskningstilskudd” — ikke “monitorering av kommersiell teknologi.” En søknad på 15 MNOK for “tilpasning av autonome system-teorier” får ja. En søknad på “Vi må lære fra OpenAI sin arkitektur” får nei.
- Publiseringsmakt:** Reviewere i autonome systemer er selv investert i tradisjonelle PhD-programmer. De har sårbarhet i anerkjennelse. Papirer som sier “LLM-er gjør X bedre” blir møtt med “men det er ikke ekte intelligens” — en ikke-falsifiserbar forsvarslinje.
- Doktorgradskjeden:** 200+ norske PhD-kandidater i AI/autonome systemer. Hvis deres avhandlinger blir irrelevante, må veilederne ta ansvar. Kortsiktig løsning: nedtone LLM-relevansen.



Sårbarhetens kjerne:

Akademia har **strategisk investert i en teknologibane** som nå konkurreres av en annen. Men institusjonen er organisert rundt første banen. Å innrømme det åpent er å akseptere 10-15 år av misallokeert kapital.

Hva som kunne bryte systemet:

En større institusjon (Universitet i Oslo, NTNU) gjør det mot-intuitiv: **stiftelsen sier offentlig** "Vi nedprioriterer tradisjonell autonome-systemer-forskning 2025-2027. Vi fokuserer på LLM-arkitektur og alignment."

Det bryter tabuen. Plutselig er det tillatt å snakke om.

Neste steg

1. **Kartlegg kortsiktige insentiver:** Intervju 5-10 seniorforskere konfidensielt om hva som hindrer dem fra å være eksplisitte.
2. **Utforsk alternativ-narrativ:** Kan tradisjonell autonome-systemer-forskning repositioneres som "grunnleggende for LLM-sikkerhet"? Hvis ja, løser det sårbarhetsproblem.
3. **Test makt-dynamikk:** Foreslå en