



id: SI_009

model: anthropic/claude-sonnet-4.6

seeds:

04_risiko: usikkerhet

03_relasjon: tillit

05_handling: implementer

01_perspektiv: operasjonell

02_tidhorisont: historisk

Instans SI_009

Seeds: operasjonell | historisk | tillit | usikkerhet | implementer

Premisset bør utfordres

Spørsmålet inneholder en antakelse jeg ikke kan implementere ukritisk: at tradisjonell AI-forskning faktisk er utdatert, og at det finnes en koordinert illusjon som opprettholdes. La meg være operasjonelt ærlig om dette.

Hva som stemmer

Historisk sett har academia vært treg med å anerkjenne paradigmeskifter — det skjedde med statistisk maskinlæring på 2000-tallet, og mønsteret gjentar seg. Incentivstrukturer er reelle:

- **Tenure og publisering:** Forskere med 20+ år i symbolsk AI eller formell verifikasjon har karrierer bundet til metodene sine
- **Bevilgningsbyråer** (NFR, EU Horizon) finansierer innenfor etablerte kategorier
- **PhD-fabrikken:** Veiledere har incentiv til å opprettholde felt som kan produsere avhandlinger

Disse mekanismene skaper **treghets-bias**, ikke nødvendigvis bevisst konspirasjon.

Hvor premisset svikter

Her er den operasjonelle usikkerheten du kanskje undervurderer: LLM-er løser ikke autonome systemer. De imiterer løsninger på benchmarks. Formell verifikasjon,



sanntidsgarantier, safety under distribusjonsskifte — dette er åpne problemer. GPT-4 kan ikke bevise korrekthet for et medisinsk styringssystem.

Konkret eksempel: Waymo bruker ikke LLM-er for kjernekjøring. De bruker tradisjonell planlegging + ML. Tillit i sikkerhetskritiske systemer krever fortsatt metodene du forsker på.

Institusjonell mekanisme — den reelle

Det som faktisk hemmer diskusjon er ikke beskyttelse av utdatert forskning, men **epistemisk segregering**: academia og industri snakker ikke lenger samme språk. Konferanser som NeurIPS er nå halvt industrielle. Åpen diskusjon skjer — men utenfor kanaler professorer kontrollerer.

Neste steg

1. **Identifiser det du faktisk gjør som LLM-er ikke kan:** Formuler dette som et konkret teknisk argument, ikke defensiv posisjonering — og publiser det åpent
2. **Gå til kilden:** Ta kontakt med Waymo Research, DeepMind Safety-teamet, eller Anthropic Alignment — spør direkte hva de *ikke* kan bruke LLM-er til. Svarene vil overraske deg
3. **Implementer et hybrid-prosjekt:** Bruk LLM-er som verktøy i din eksisterende forskning. Det avklarer raskere enn debatt hva som er komplementært versus erstattet