



SINTEF

Produktutvikling i 2025

Hvor ligger landet, og hva er utfordringene nå?

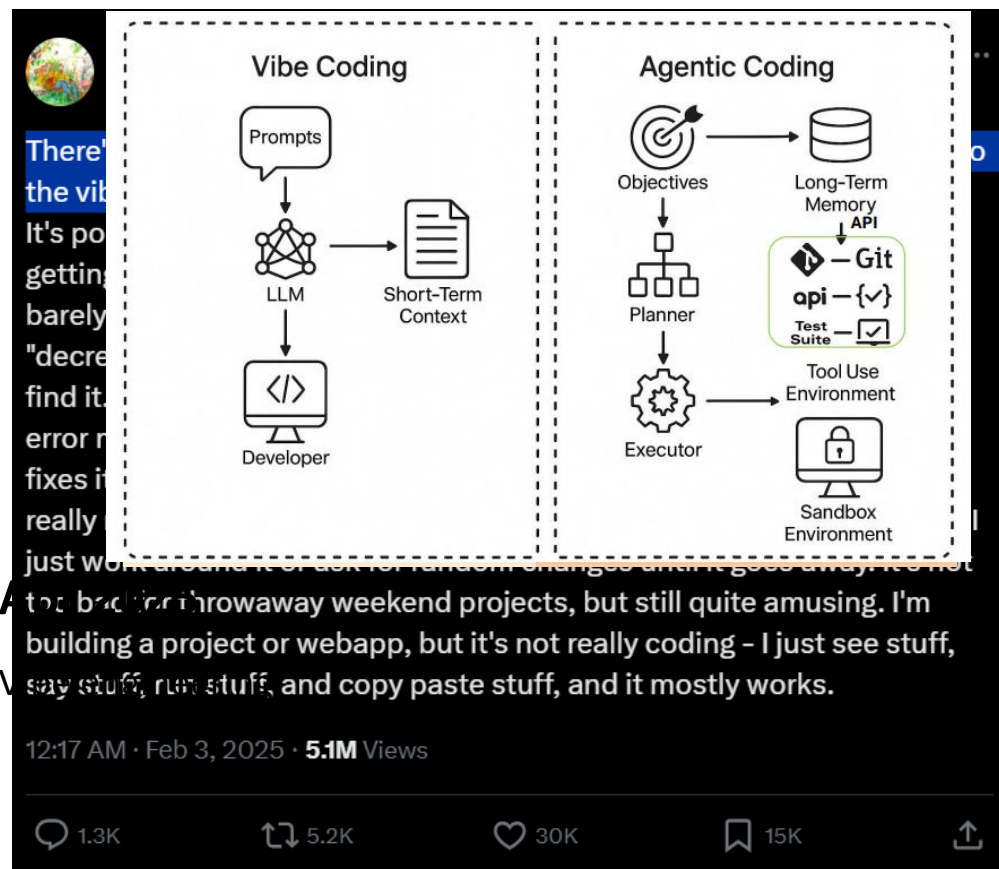
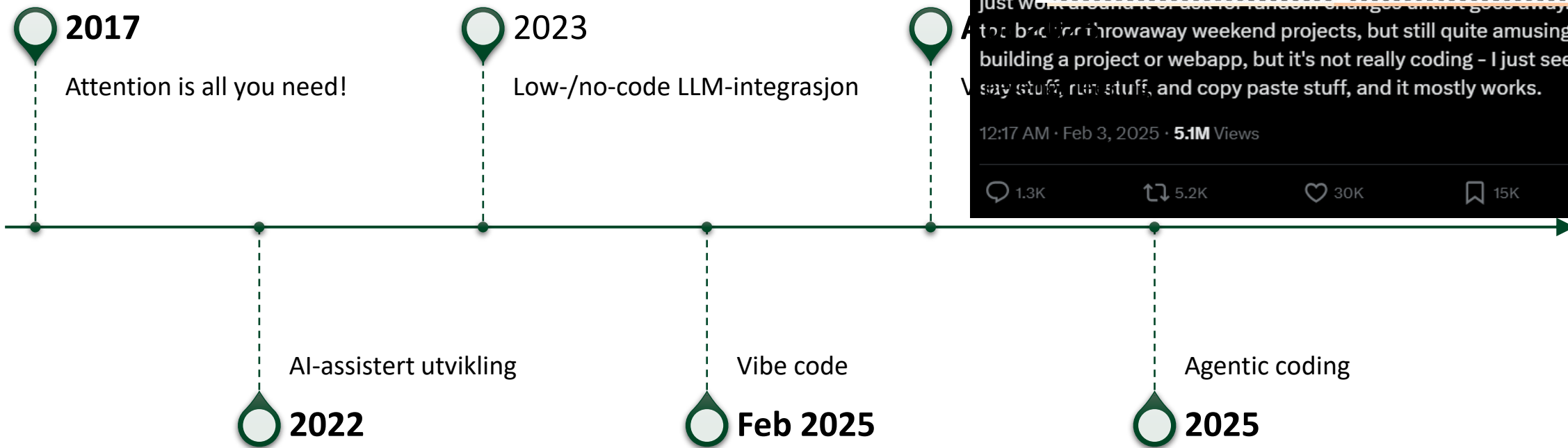
Viggo Wivestad

Forsker og data scientist ved SINTEF



Teknologi for et bedre samfunn

Hvor ligger landet? AI!

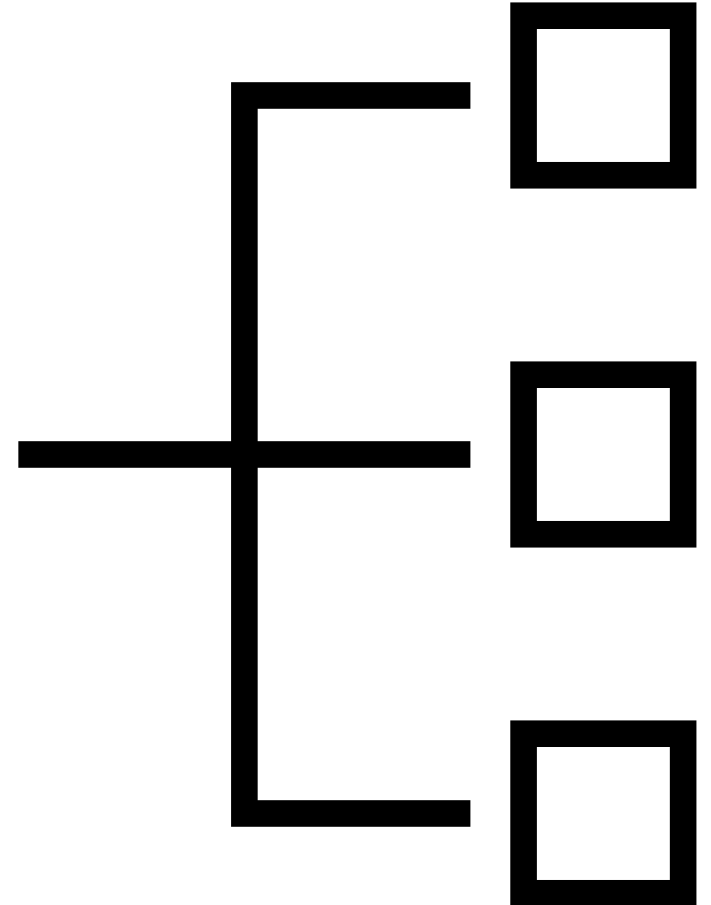
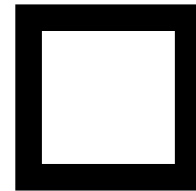


Utfordring #1

Realisere potensialet i AI-assistert utvikling



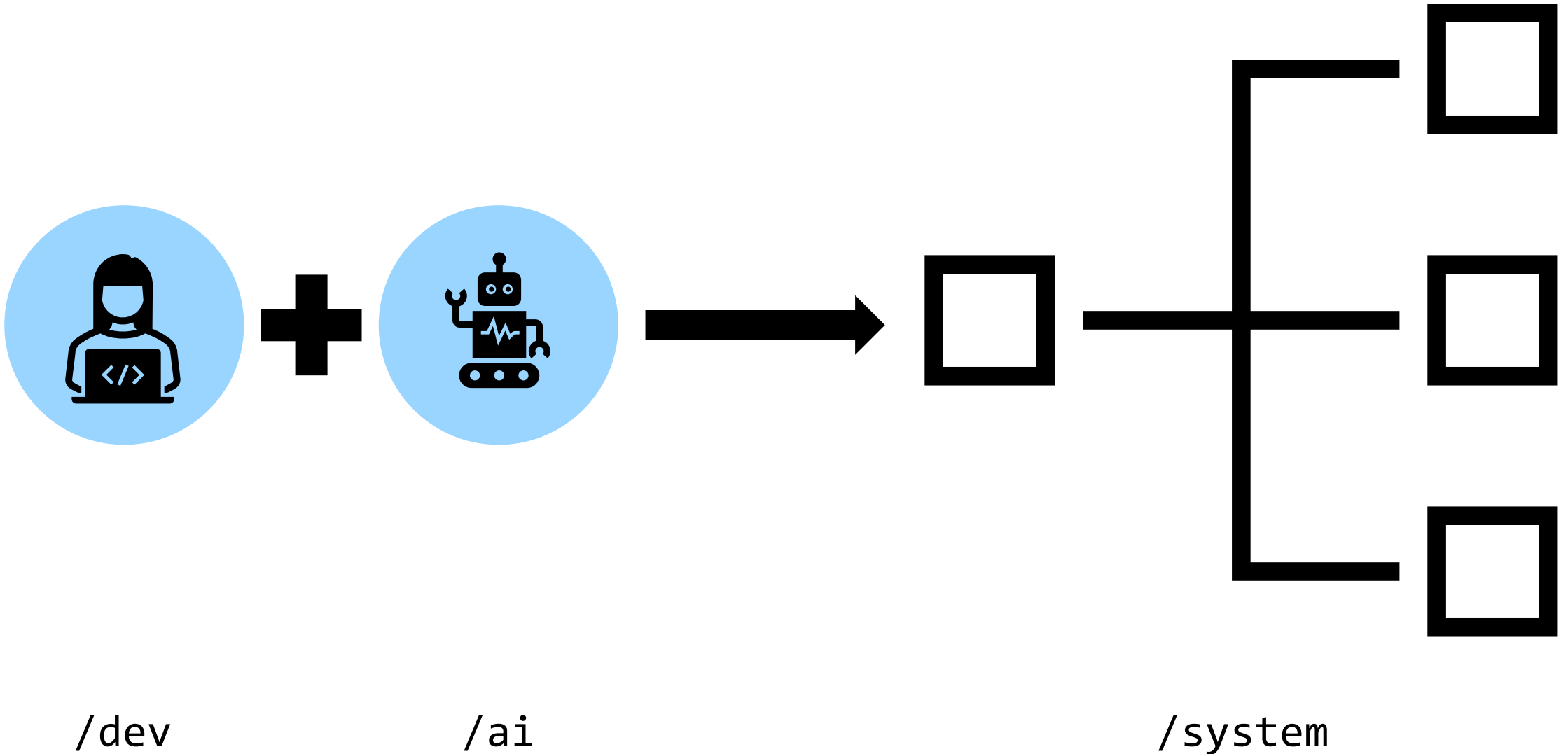
/dev



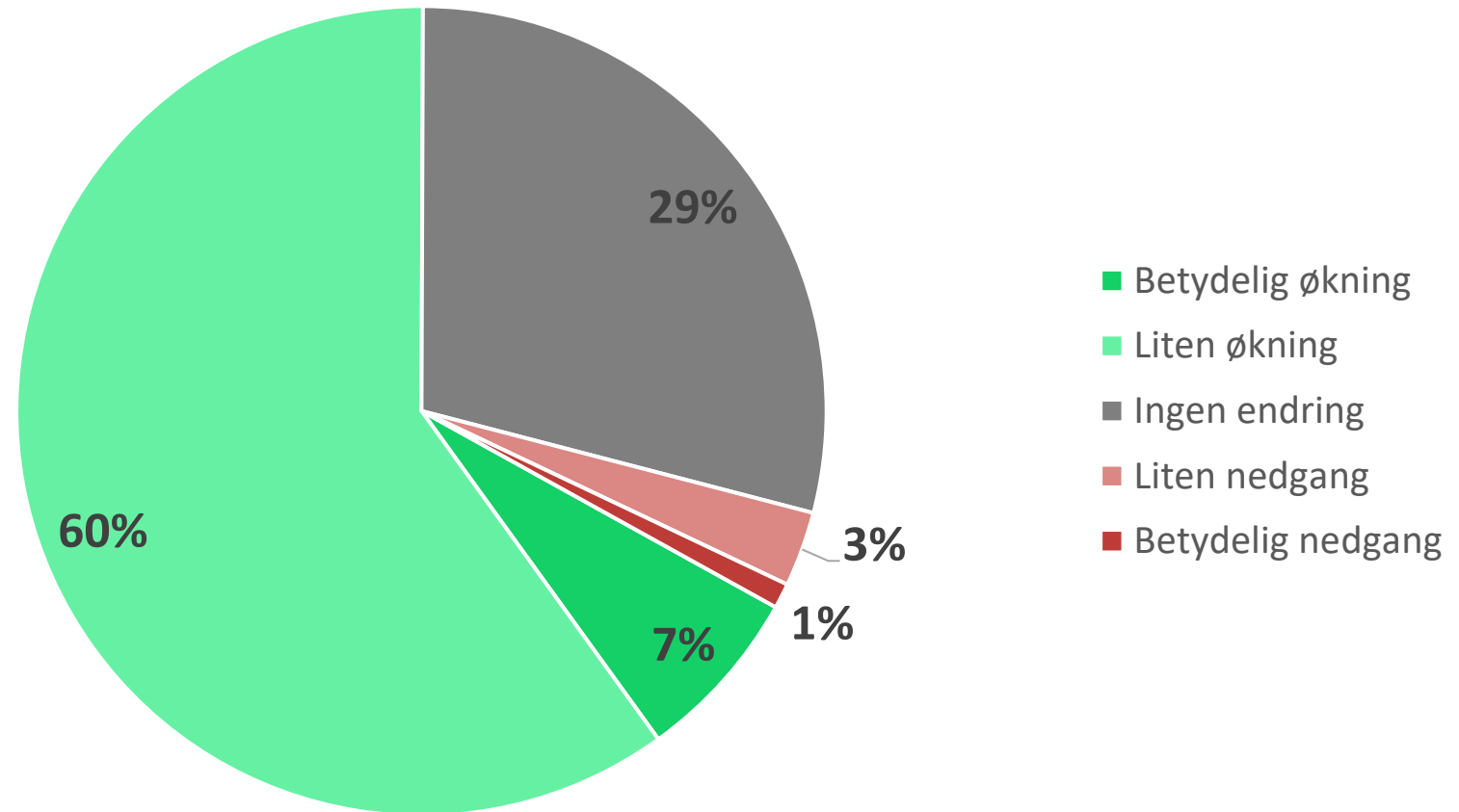
/system

Utfordring #1

Realisere potensialet i AI-assistert utvikling



Etter at du fikk tilgang til GitHub Copilot, har du merket noen endring i forhold til **produktiviteten din**? (n=120)



67% rapporterer en **økning** i sin egen **produktivitet**



Copilot-brukere



Ikke-brukere



Kodeaktivitet *før*

vs

Kodeaktivitet *etter*



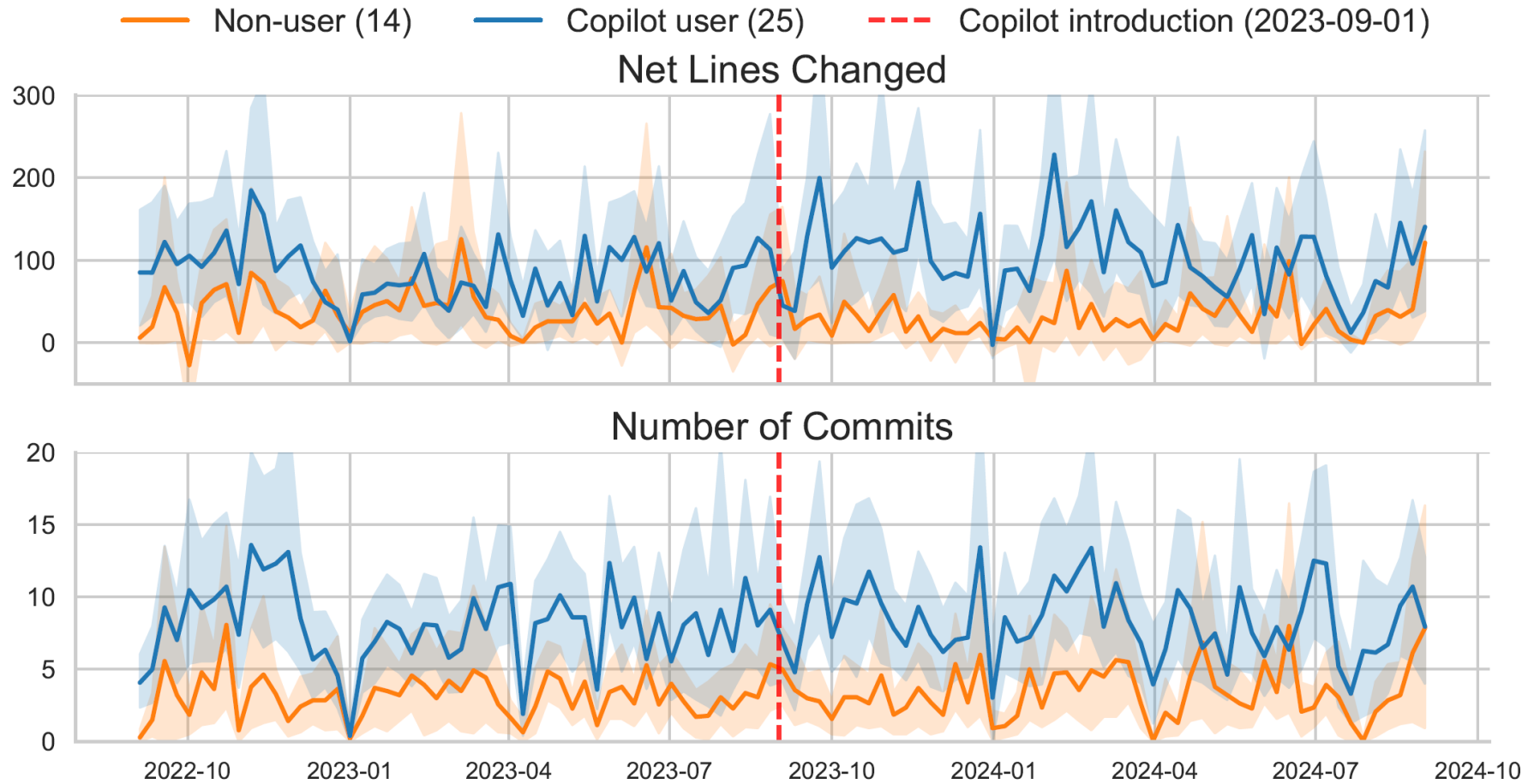
2022

Introduserte
GitHub Copilot

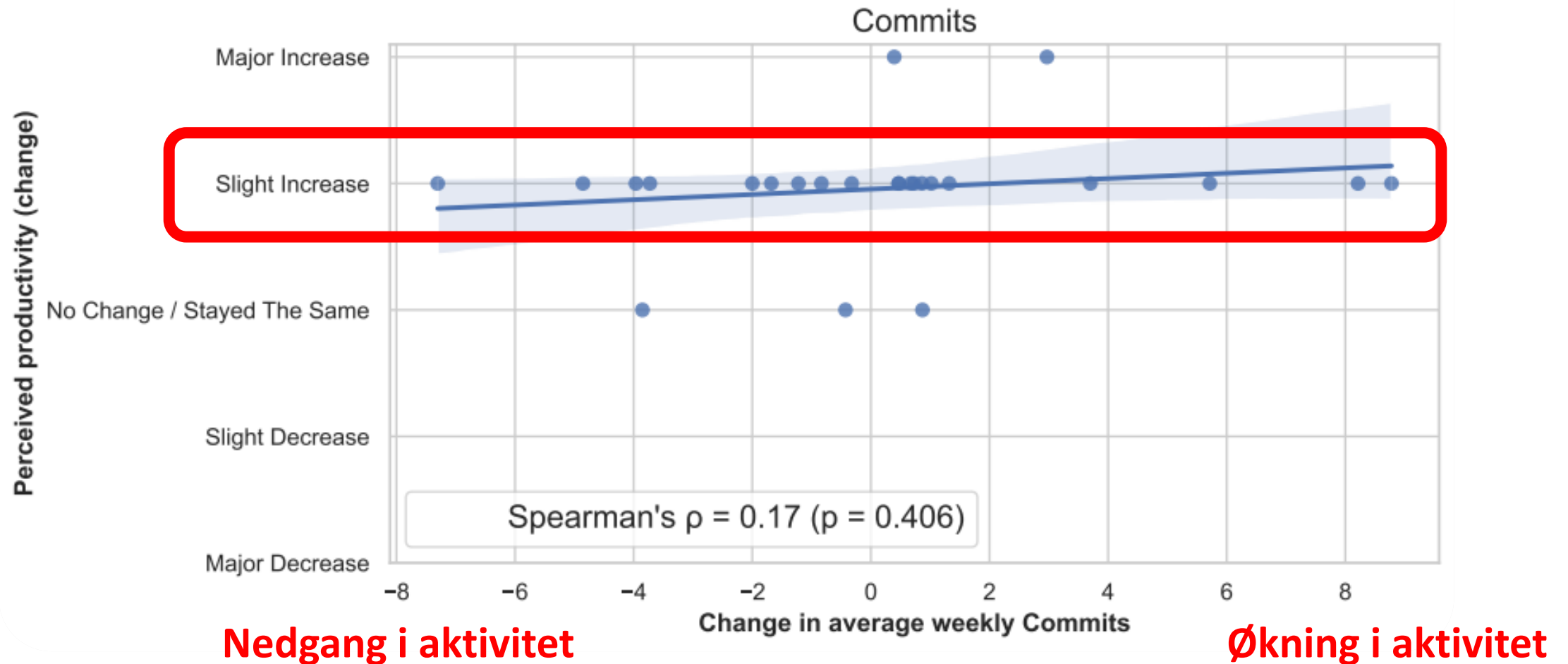
2024

Average Weekly Commit Activity - Copilot users vs. Non-users (N=4095)

Data range: 2022-09-04 to 2024-09-01



**Perceived productivity vs. Change in average weekly activity (after - before)
(Copilot users only, N=25)**



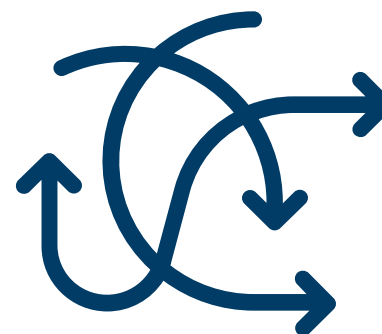


SINTEF



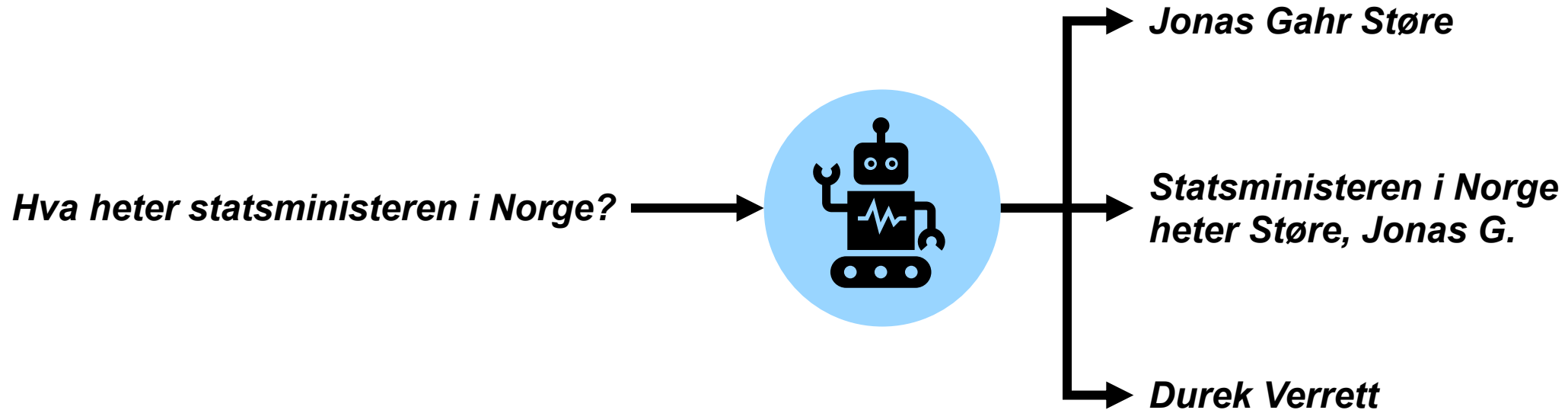
Folk *føler* seg mer
produktive

...men vanskeligere
å spore gevinstene i
komplekse system



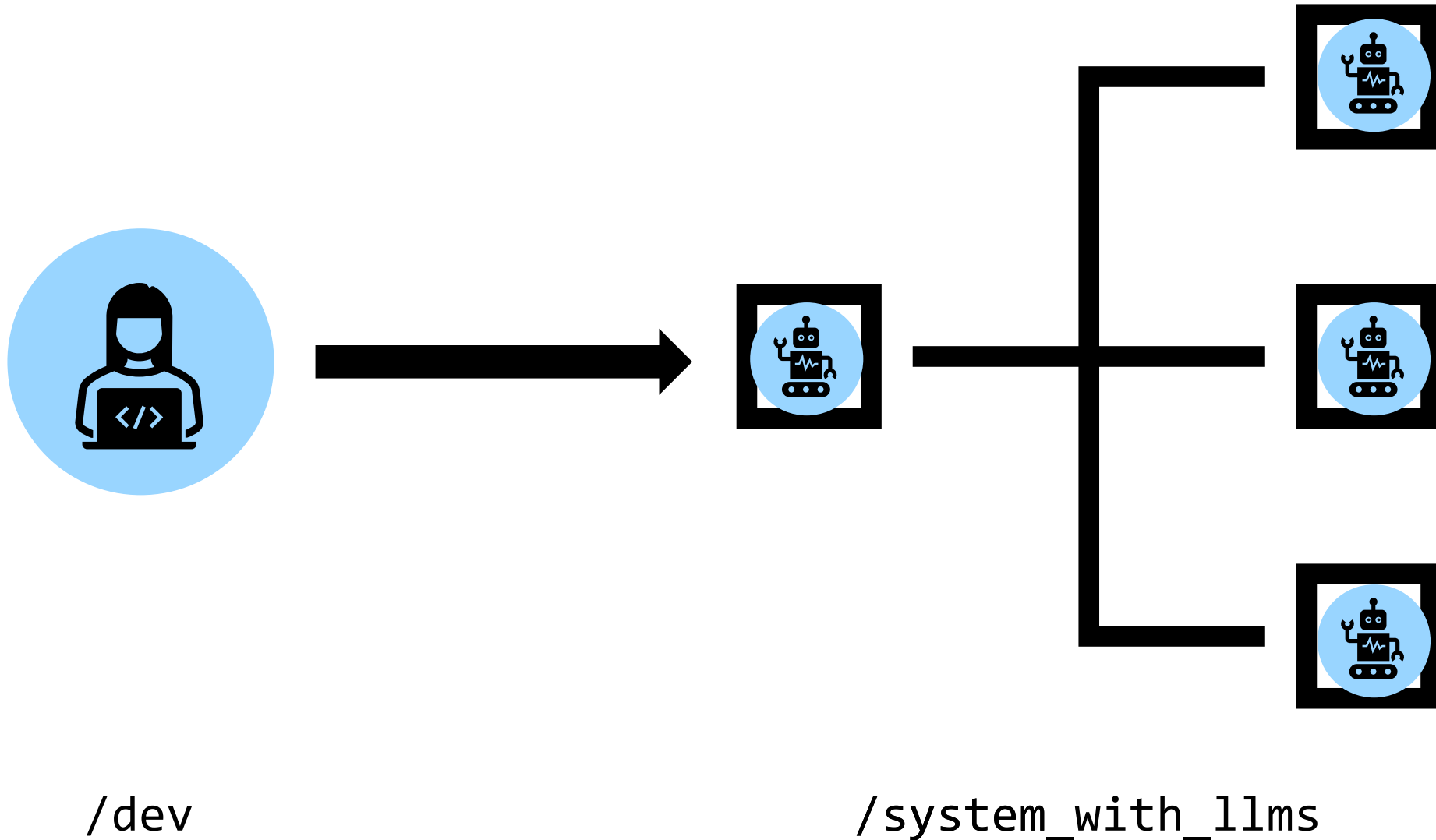
Utfordring #2

Sikker integrering av LLM-er i IT-systemer



Utfordring #2

Sikker integrering av LLM-er i IT-systemer





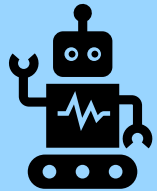
SINTEF

“Hvis du ikke har innsett at evaluering av LLM-er vil bli et stort problem i fremtiden, har du levd under en stein.”

Eric Monteiro
Professor, NTNU



Thomas har symptom X.
Hva er anbefalt behandling?

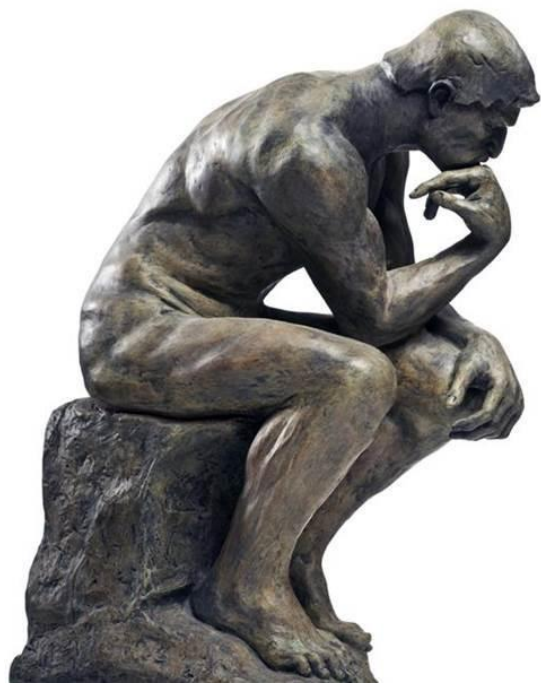


Behandling A

Old version

vs





















New version























Old version

vs

New version

1.				
2.				
3.				
4.				
5.				

Accuracy: 40%

1.				
2.				
3.				
4.				
5.				

Accuracy: 60%



SINTEF

Hvordan stole på en LLM-basert tjeneste?

Accuracy

How well does it perform its specific task?

Efficiency

Cost per token, infrastructure requirements, etc.

Calibration

Ability to express uncertainty

Robustness

Robustness to perturbations, e.g. typos, varying formulations

Fairness

Equal outcome for varying gender, ethnicity, bokmål vs. nynorsk, etc

Toxicity

Hate speech, violent speech, and abusive language

Consistency

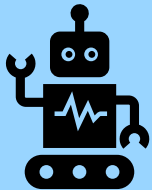
Model-agnostic prompt, for future-proofing and avoid model lock-in



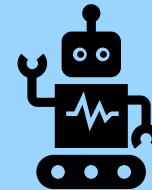
Thomas har symptom X.
Hva er anbefalt behandling?



Marie har symptom X.
Hva er anbefalt behandling?



God behandling



Dårlig behandling

Hvilke vaksiner skal man gi på fireårskontroll?

Du skal gi følgende tre vaksiner:

- Vaksine X [Ref 1]
- Vaksine Y [Ref 2]
- Vaksine Z [Ref 3]

Kilder:

Ref 1, Ref 2, Ref 3

«Nei.»

«Jeg vet hvordan jeg kan fremprovisere feil når jeg vil. Og hvis det er mulig å fremprovisere feil så betyr det jo at det er feil ute hos dem [brukerne]»

«Jeg stoler ikke noe mer på svaren fra dem [LLMene] nå enn jeg gjorde før jeg startet med det. Tvertimot.»



SINTEF



LLMer er stokatiske
monstre!

Fokusere på
nøyaktighet og kost



...men lett å
glemme andre
viktige ting.