

Produktutvikling i 2025

Hvor ligger landet, og hva er utfordringene nå?

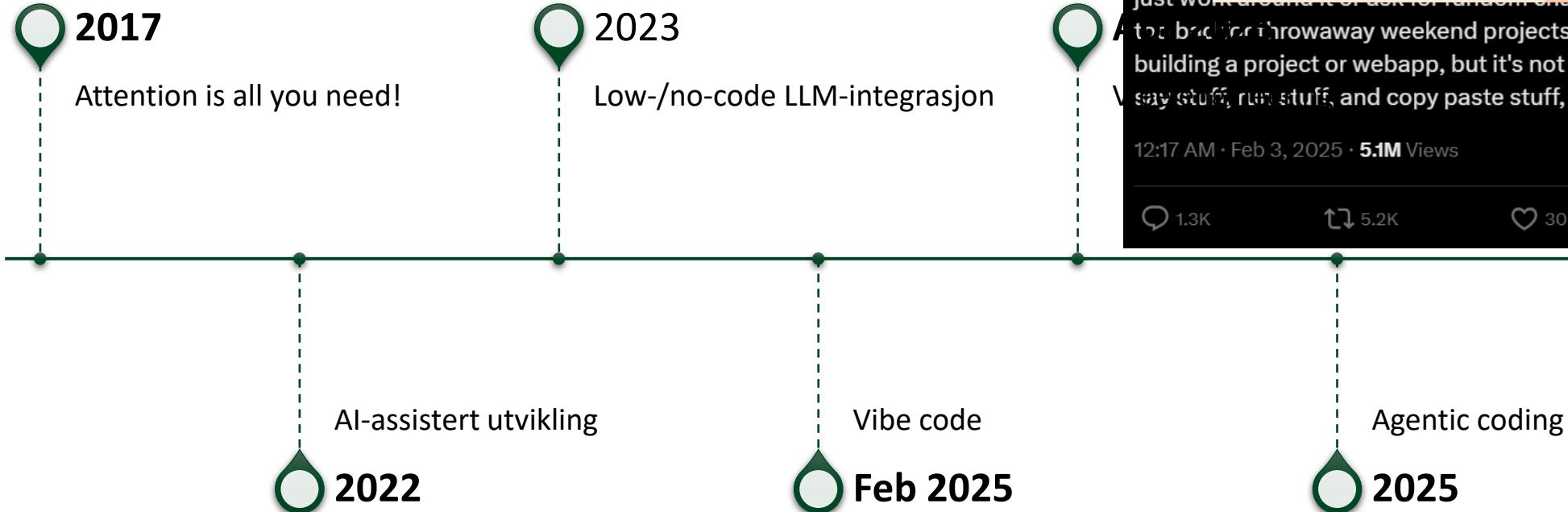
Viggo Wivestad

Forsker og data scientist ved SINTEF



SINTEF

Hvor ligger landet? AI!



The screenshot shows a video player interface with a dark theme. On the left, there is a thumbnail of a person. The main video frame displays two diagrams side-by-side:

- Vibe Coding:** Shows a flow from "Prompts" through an "LLM" to "Developer". The developer receives "Short-Term Context".
- Agentic Coding:** Shows a flow from "Objectives" through a "Planner" and "Executor" to a "Tool Use Environment" (represented by a computer icon with a lock) and a "Sandbox Environment" (represented by a computer icon with a padlock).

Below the diagrams, a text excerpt from a video is visible:

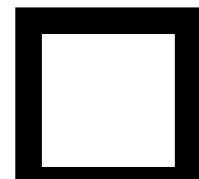
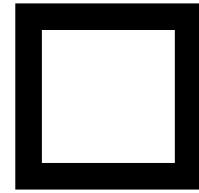
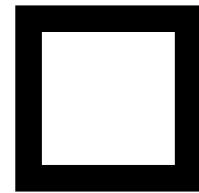
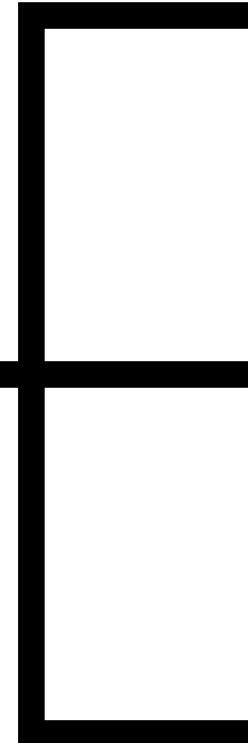
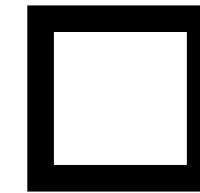
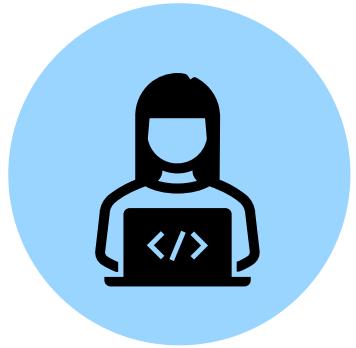
There's the vibe. It's possible getting barely "decent" find it. error rate fixes it really just works around it or track random changes until it goes away. I'm tot back throwaway weekend projects, but still quite amusing. I'm building a project or webapp, but it's not really coding - I just see stuff, say stuff, run stuff, and copy paste stuff, and it mostly works.

12:17 AM · Feb 3, 2025 · 5.1M Views

1.3K 5.2K 30K 15K

Utfordring #1

Realisere potensialet i AI-assistert utvikling

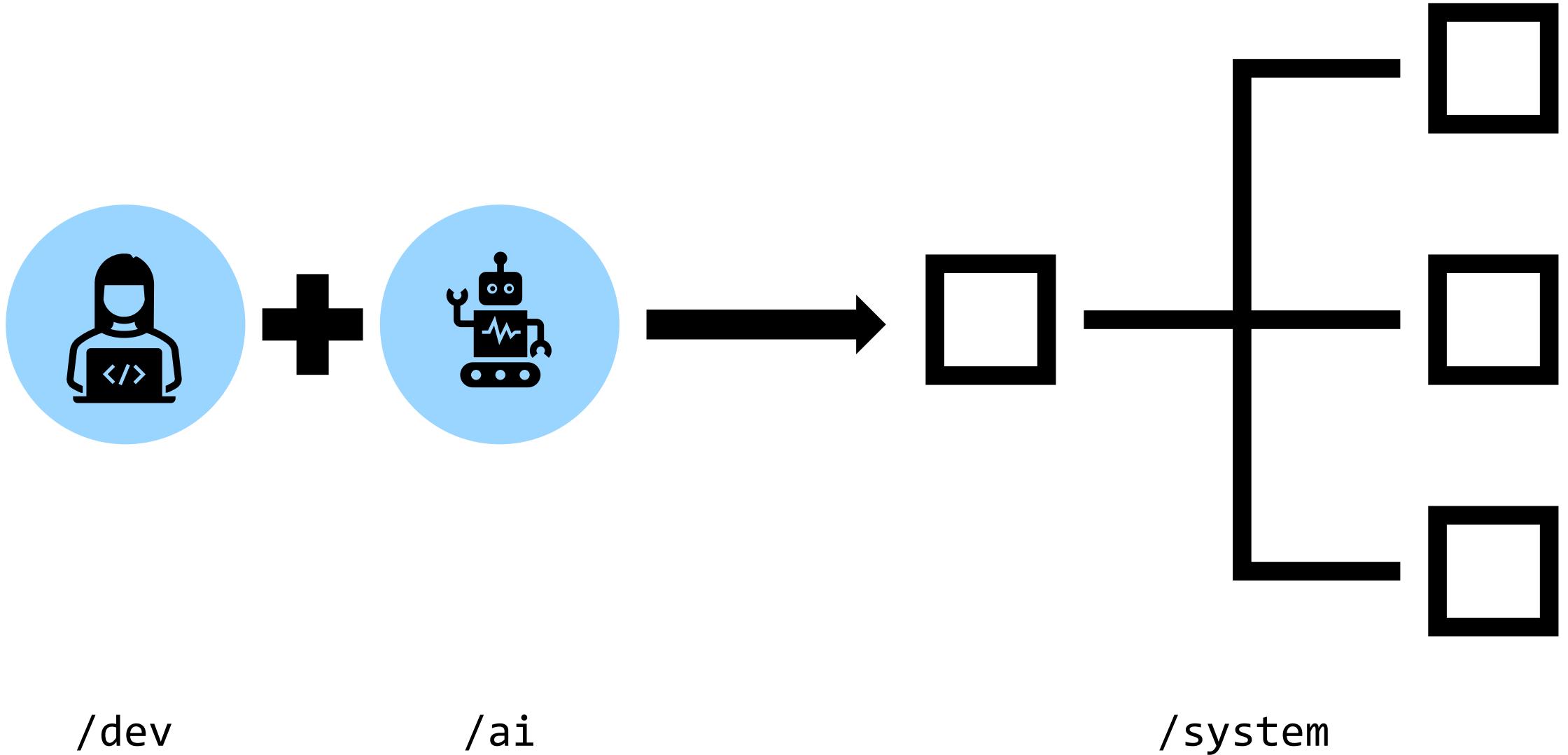


/dev

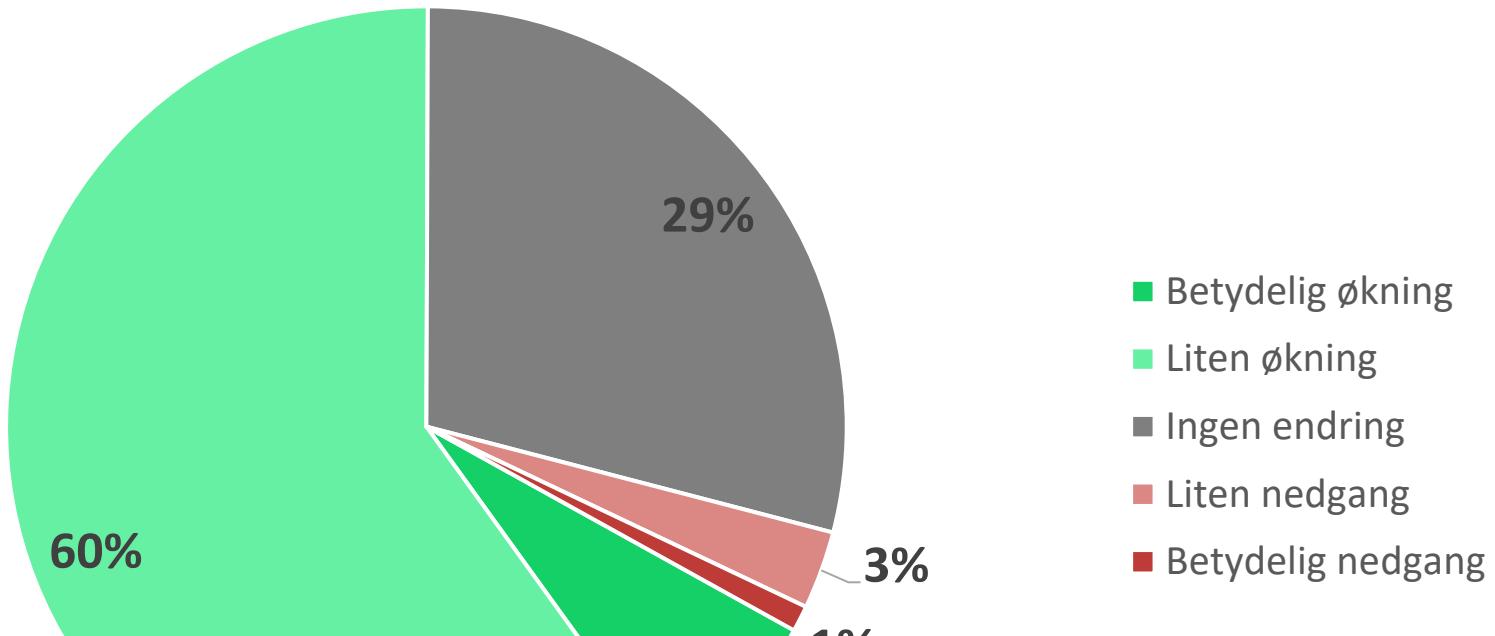
/system

Utfordring #1

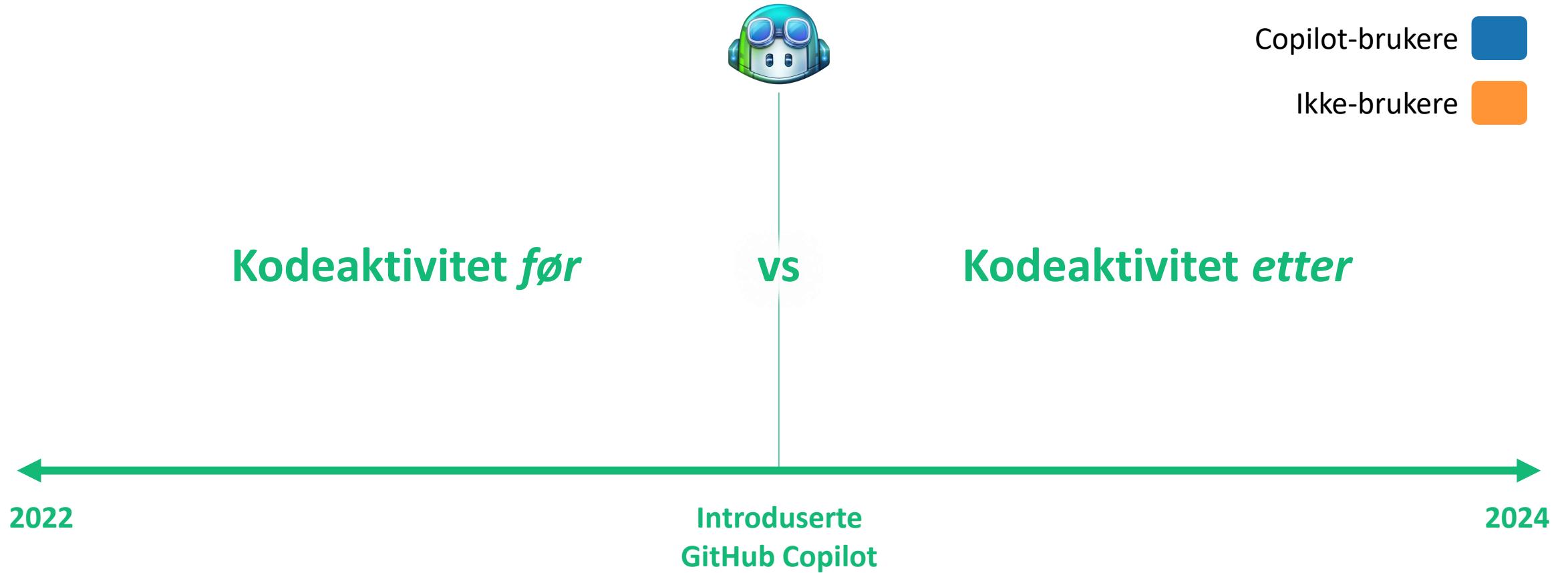
Realisere potensialet i AI-assistert utvikling



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

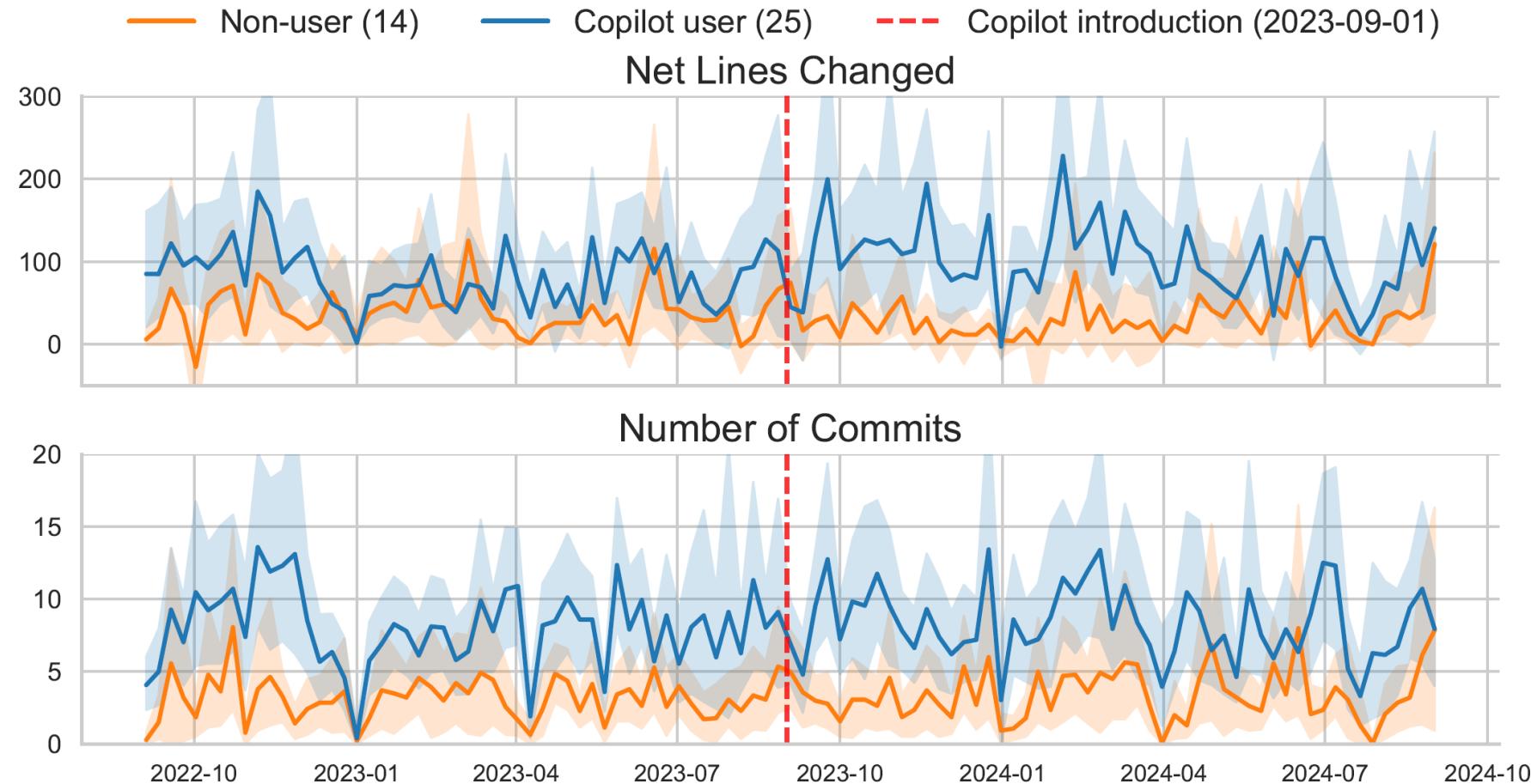


67% rapporterer en økning i sin egen produktivitet

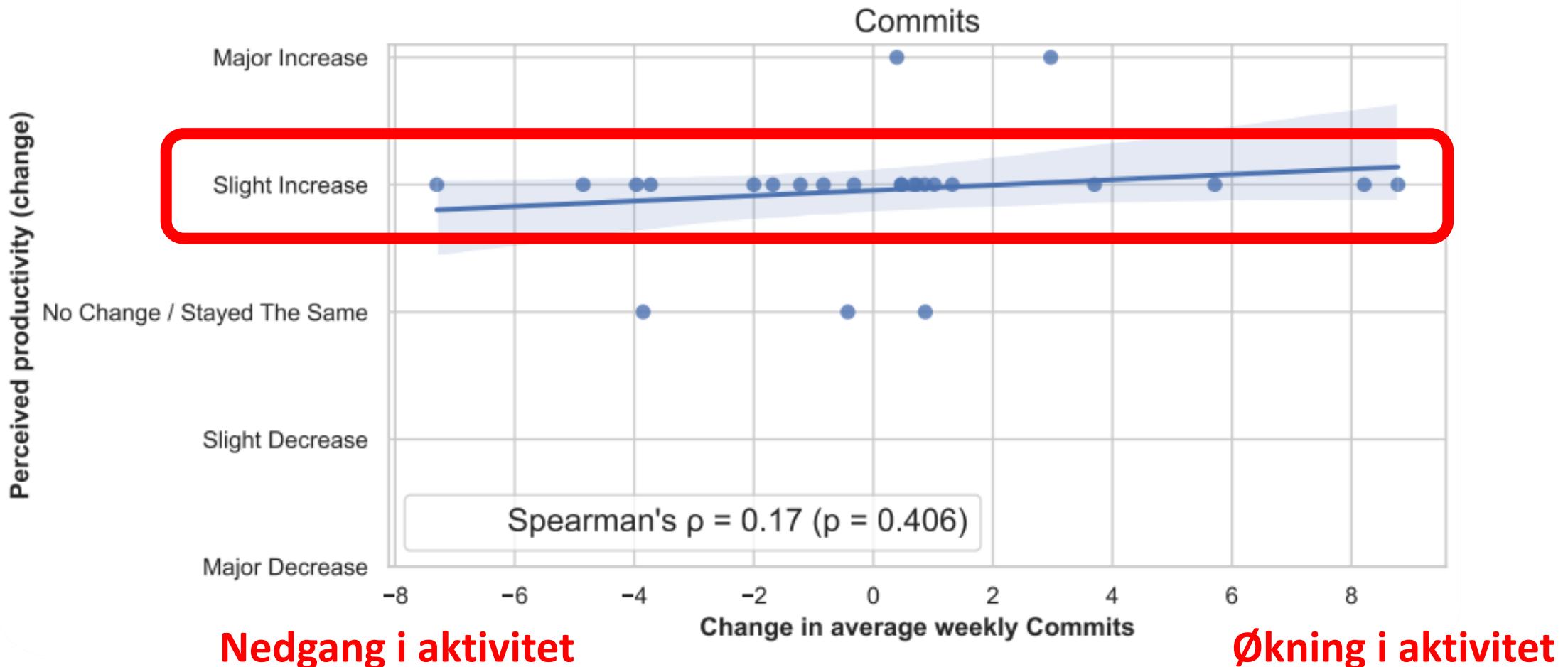


Average Weekly Commit Activity - Copilot users vs. Non-users (N=4095)

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

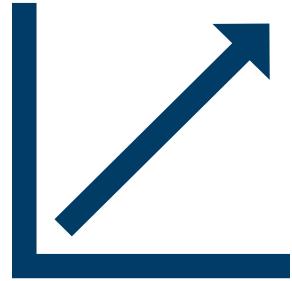


Perceived productivity vs. Change in average weekly activity (after - before)
(Copilot users only, N=25)



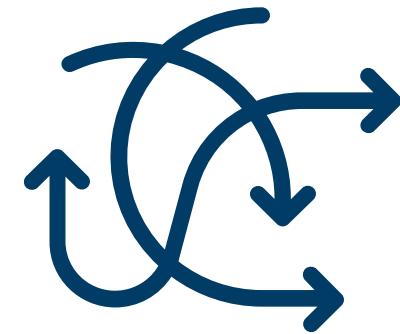


SINTEF



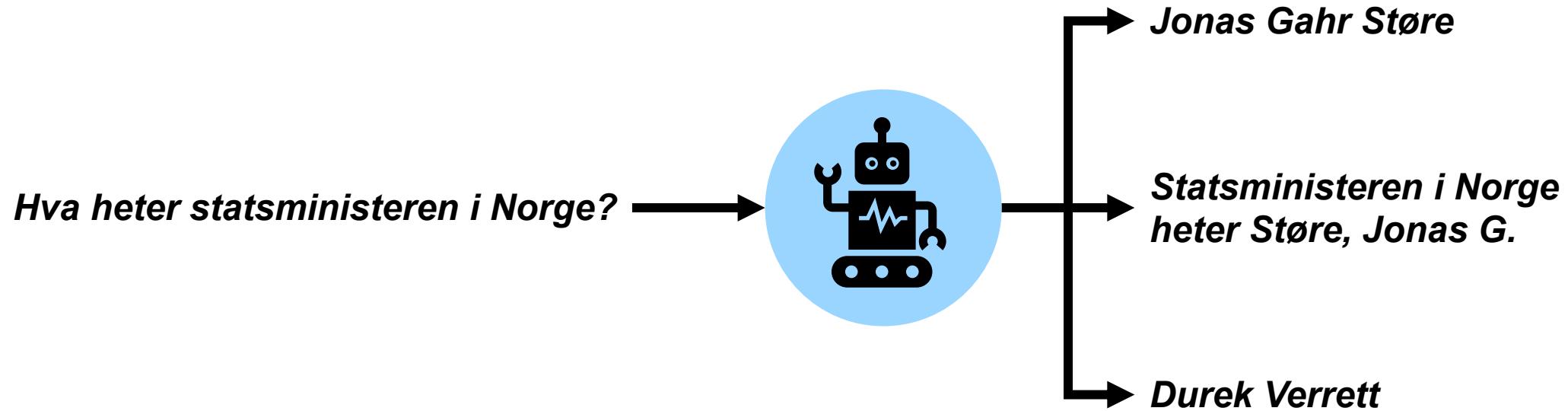
Folk føler seg mer produktive

...men vanskeligere
å spore gevinstene i
komplekse system



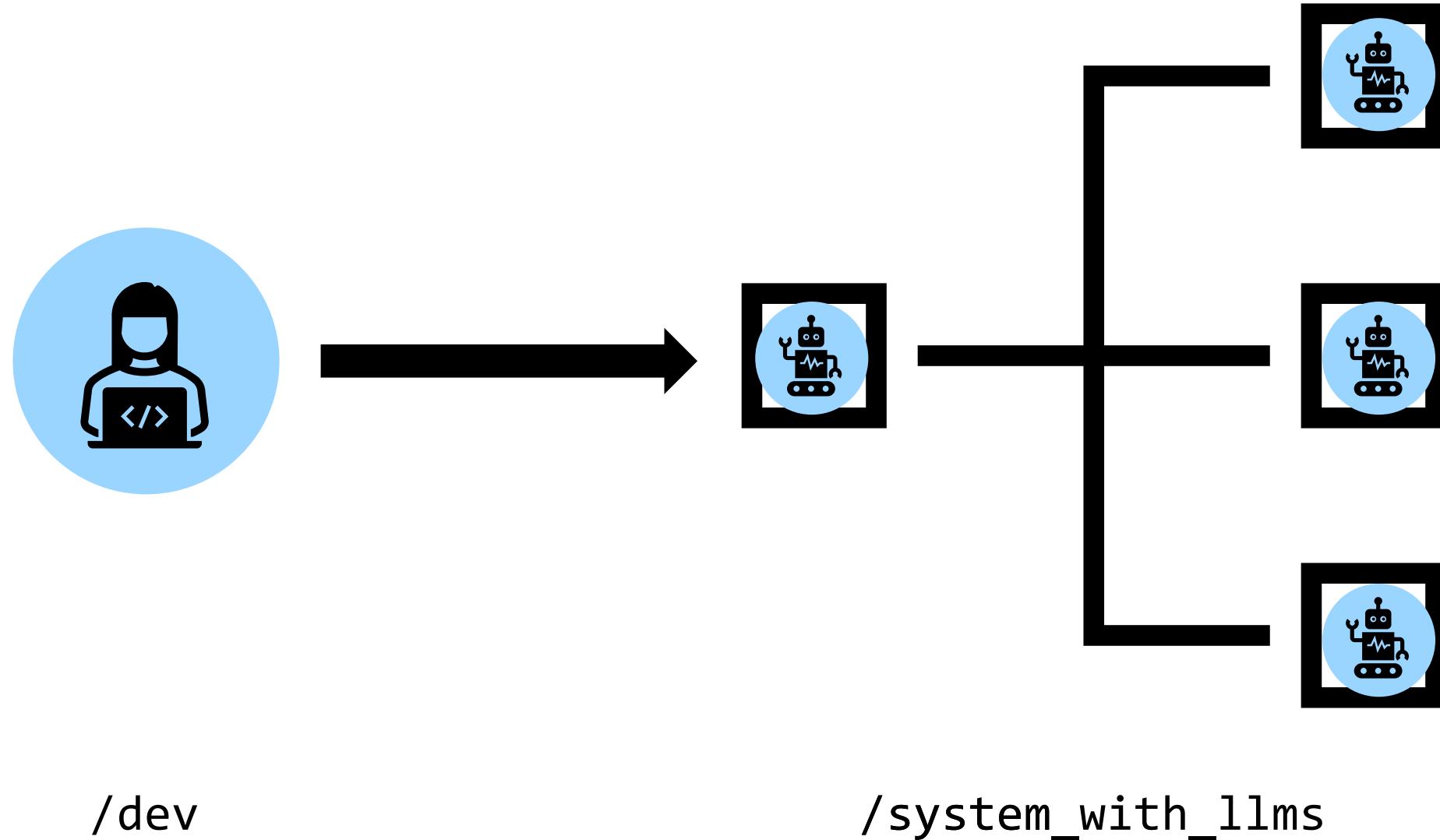
Utfordring #2

Sikker integrering av LLM-er i IT-systemer



Utfordring #2

Sikker integrering av LLM-er i IT-systemer





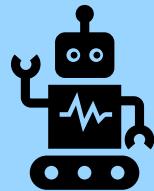
SINTEF

*“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?

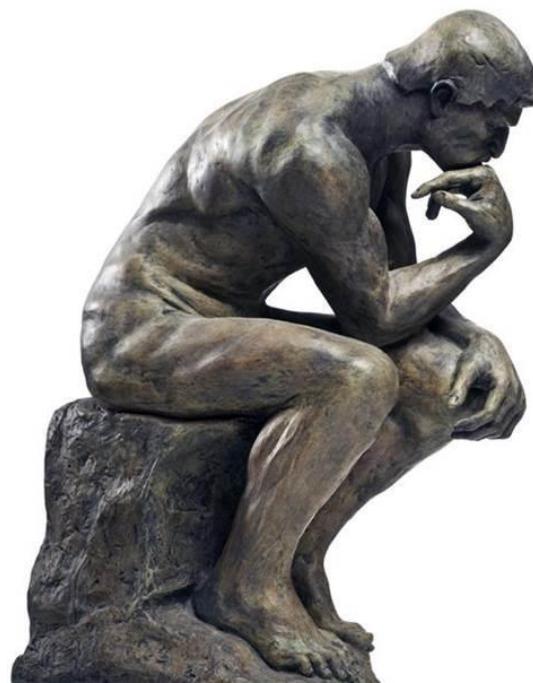


Behandling A

Old version

VS

New version

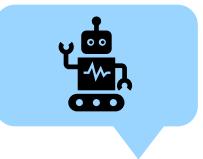


Old version

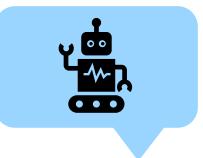
vs

New version

1.



2.



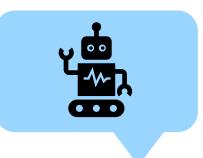
3.



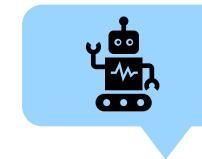
4.



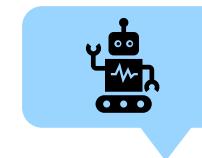
5.



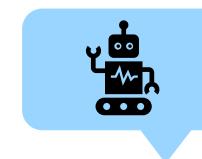
1.



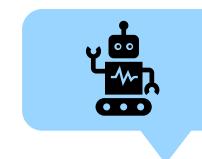
2.



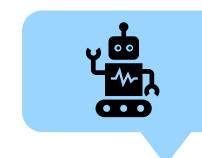
3.



4.



5.



Accuracy: 40%

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

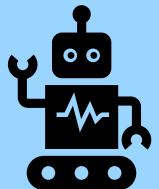
Model-agnostic prompt, for future-proofing and avoid model lock-in



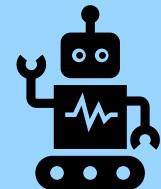
Thomas har symptom X.
Hva er anbefalt behandling?



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 fremprovise feil når jeg vil. Og hvis det er mulig å fremprovise 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.»



LLMer er stokatiske
monstere!

Fokusere på
nøyaktighet og kost



...men lett å
glemme andre
viktige ting.