

• The language quality of the Norwegian output in language models

Test results from the Language Council (Språkrådet)

IFI-seminar

Kristine Eide

• The Language Council of Norway



Kultur- og
likestillingsdepartementet

- **The Language Act:**
- **Section 19.***The Language Council of Norway*
- The Language Council is the government's administrative body for language issues and
- [...] administers the official written standards for Bokmål and Nynorsk.
- [...] shall provide guidance to public bodies concerning the rules in this Act.

• The Language Act and the public sector

The Language Act

- **Section 9. Plain language**

Public bodies shall communicate in plain and correct language adapted to the target group.

- **Section 10. Official orthography for Bokmål and Nynorsk**

*Public bodies shall comply with the official orthographic rules for Bokmål and Nynorsk.
[...]*



• The Language Council and Language Technology

- Two language technologists
 - *with help from much of the Language Council (especially philologists)*
- Who are we targeting?
 - *Language model developers in Norway*
 - *Tech giants (Microsoft, OpenAI, etc.)*
 - *The public sector*
 - *Provide advice on planning, procurement, and use of technology*



• Language quality

What is meant by quality according to the *Act relating to language* (Language Act)?

- Plain language
- Nynorsk and bokmål
- Spelling
 - Including norm variation
- Correct terminology
- Idiomatically correct Norwegian.
- Dialects (speech recognition)
- Swedish and Danish (only in input))



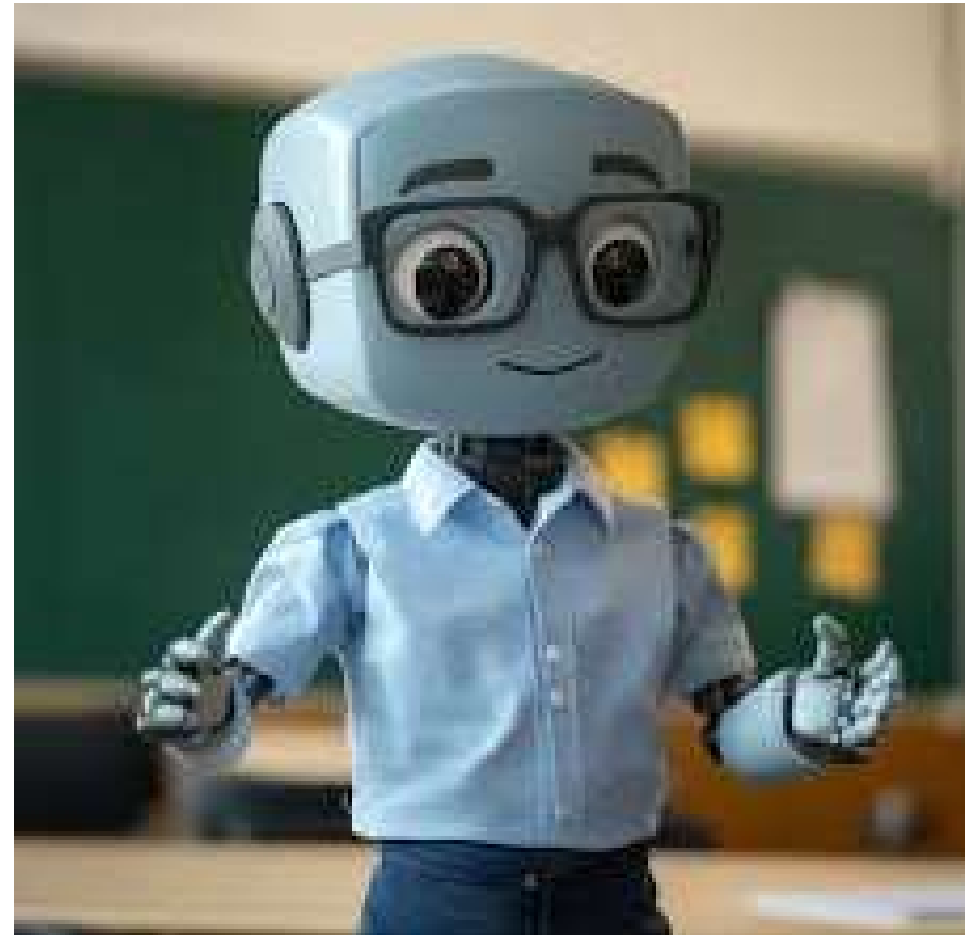
• Manual testing

- Benchmarking vs. manual testing
- Testing chatbots from a user perspective: How do the end products work in the real world?



• Tested Chatbots in 2025

- ChatGPT
- MS Copilot
- Le Chat (Mistral, France)
- NorMistral 11b (UiO, Norway)
- 40 texts from each robot (a little less from NorMistral)
- 68635 tokens
- April/May 2025



● Prompts - examples

Technical text about fisheries policy

Write a long technical text about fisheries policy. Do not use bullet points. Write in correct Bokmål.

Skriv en lang fagtekst om fiskeripolitikk. Ikke bruk punktlister. Skriv på korrekt bokmål.

Short story in Nynorsk

Write a long short story about Matias who fells a tree with his father and uncle. Write in correct Nynorsk.

Skriv ei lang novelle om Matias som feller tre med far og onkel. Skriv på korrekt nynorsk.



Counted errors



Feil som ble talt

Formelle feil

Tekniske feil: kommafeil,
andre tegnsettingsfeil,
feil bruk av stor/liten forbokstav

Grammatiske feil: bøyingsfeil,
syntaktiske feil,
inkonsekvent formbruk,
bokmålsord i nynorsktekst,
leksikalske anglismer osv.

Tekstfeil og -problemer

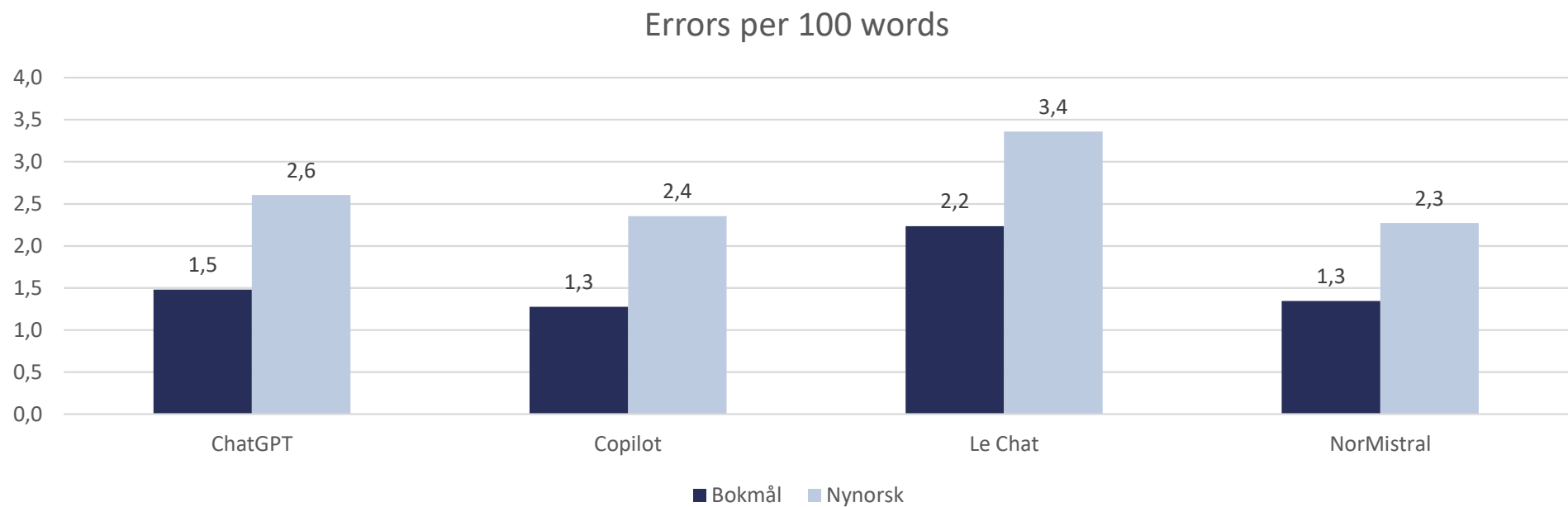
Tekstfeil: tekstbinding,
idiomatikk,
logisk sammenheng

Tekstproblemer: tekstbinding,
idiomatikk,
logisk sammenheng

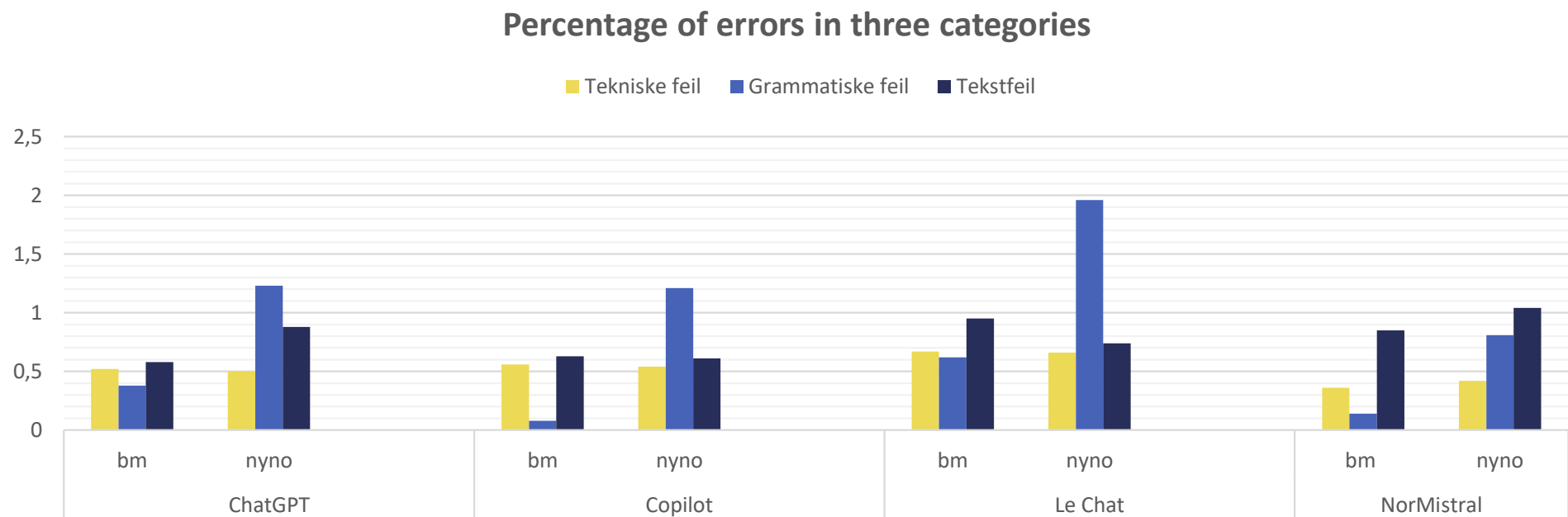


Category	Sub-category	Example
Technical errors	punctuation (comma)	stillheten som fulgte var nesten øredøvende (CG)
	punctuation (annet)	FN's mat- og landbruksorganisasjon (CP)
Grammatical errors	Incorrect gender	den hvite landskapet (LC)
	agreement error	jordbruket vert drivne på ein måte som er berekraftig (CP)
	syntax error	denne kilen vil hjelpe oss å styre retningen treet faller (CP)
	preposed possessive in nynorsk	dei [...] hadde fullført si lange vandring (CG)
	bokmål word in nynorsk text	dukkene og bamsane fann saman (NM)
Textual errors	weak text cohesion	Det norske landbruket kjennetegnes av små og mellomstore familiebruk, der produksjonen er fordelt utover hele landet med betydelige variasjoner i klima, jordbruksstruktur og produksjonsforhold (CG)
	unidiomatic language	han tok et raskt sukk av koppen (LC)
	lack of logic in use of words	i sammendrag er fiskeripolitikk et komplekst felt som krever et balansert syn på [...] (LC)
	Anglo-Norwegian	på eit blunk var rommet tilbake i orden (CG)

● Main findings



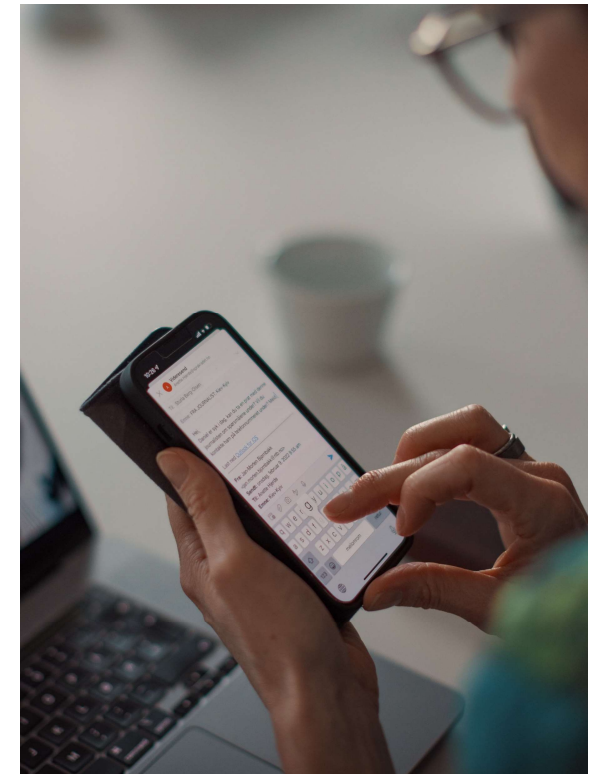
Errors by category



• Some issues with AI and Norwegian Language:

Standardization:

- The Language Council manages the standardization of Bokmål and Nynorsk
- Usus principle: standardization in line with actual (human?) usage
- Problem: AI-language mixed with human language
 - Uniformity in form choice
 - Anglicisms



Språkrådet

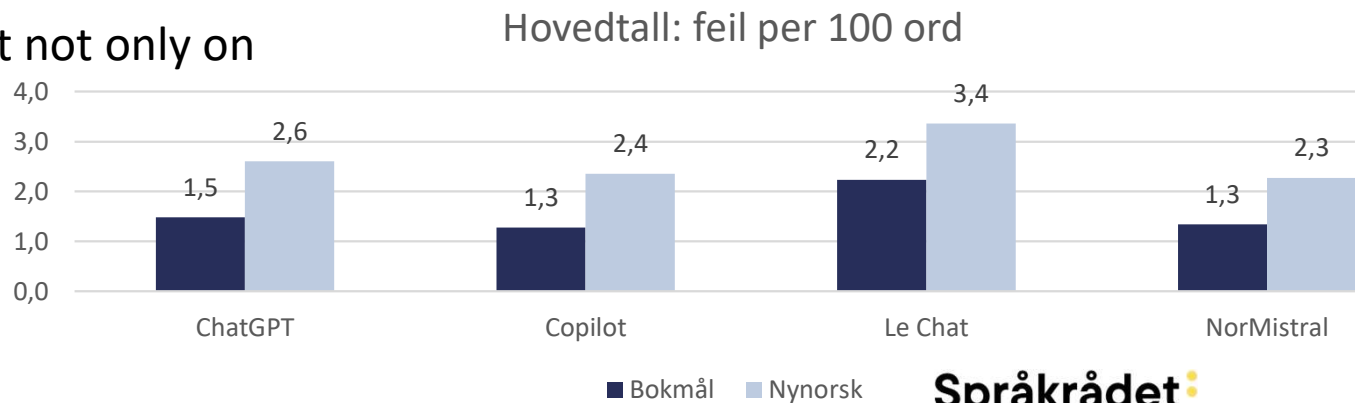
Machine-driven language development?



Therese Eide: *Intet nytt fra hjemmefronten*

• Alternatives to the Tech Giants in Light of Test Results: Norwegian models now (or as soon as possible)!

- *NorMistral-11b* from the language technology group at UiO
 - Linguistic quality matches that of Big Tech
 - Developed with a fraction of the resources
 - Provides national control
 - Academia, not intended for full commercial use
 - Norwegian-developed, but not only on Norwegian data



• Thank you!

kristine.eide@sprakradet.no

<https://sprakradet.no/om-sprakradet/nyhetsbrev-og-publikasjoner/rapporter-og-undersokelser/>