Språket i KI-genererte tekster. Rapport fra Språkrådet om språklig kvalitet i tekster produsert av prateroboter basert på fire språkmodeller, oktober 2025

Sammendrag

Språkrådet har testet fire nyere versjoner av prateroboter som er basert på språkmodeller. Utgangspunktet er at disse robotene er i bruk eller vil bli tatt i bruk i skolen og i den offentlige forvaltningen. Vi har vurdert den språklige kvaliteten på om lag 40 tekster som er produsert av hver av disse robotene. Vi fant mange feil i alle tekstene og flest feil i tekstene på nynorsk. Feilene er av alle slag, fra logisk og syntaktisk brist via engelske vendinger til manglende konsekvens i bruken av valgfrie former.

Her er de viktigste rådene til statsorganer som vil bruke prateroboter til tekstproduksjon:

- Vurder nøye og kritisk hva praterobotene kan og bør brukes til.
- Kontroller språket og innholdet i tekstene grundig. Språket kan være dårlig, og tekstene kan inneholde såkalte hallusinasjoner.
- Ta vare på den menneskelige tekstkompetansen. Den er nødvendig både for å kontrollere KI-generert tekst og for å vurdere hva som er forsvarlig bruk av prateroboter til tekstproduksjon.

Språkrådet mener at problemene som vi har satt søkelyset på i denne undersøkelsen, bør være godt kjent også i skoleverket, men vi har foreløpig ingen råd om hvordan prateroboter bør brukes i opplæringen.

1 Innledning

Språkrådet har testet den språklige kvaliteten på tekster som er produsert av fire KI-verktøy (heretter kalt *prateroboter* (chatbots)). Formålet med testingen var todelt. For det første ville vi kartlegge feilene som praterobotene gjør, for å kunne å gi brukerne (særlig i forvaltningen og skolen) velfunderte råd om hvordan de bør eller ikke bør bruke prateroboter til tekstproduksjon. For det andre ville vi gi et bidrag til den videre utviklingen av framtidige modeller i Norge, for eksempel språkmodellene som skal utvikles ved Nasjonalbiblioteket på oppdrag fra Kultur- og likestillingsdepartementet. Disse modellene bør i størst mulig grad oppfylle de språkpolitiske kravene Stortinget har satt.

En vanlig måte å evaluere modeller på er såkalt *benchmarking* eller *referansemåling*. Modellene blir testet automatisk mot en manuelt utarbeidet fasit/standard, som gir mulighet for å måle modellenes ytelse ved utførelse av bestemte oppgaver.

Det er utfordrende å utarbeide en slik referansemåling for vårt formål. Et problem er at det ikke er mulig å lage en fasit, ettersom robotene genererer forskjellig tekst fra gang til gang. Man kunne ha brukt retteprogram for å identifisere feil i tekstene, men Språkrådets test av skriveprogrammer på norsk (Språkrådet 2025b) viser at de ikke er pålitelige nok for formålet.

Et annet problem med mange referansemålinger er at de tester språkmodellene i seg selv (for eksempel GPT-4 eller llama2). I dette prosjektet er vi interessert i brukerperspektivet. Vi har derfor valgt å teste de produktene sluttbrukerne møter, ved å granske tekster som er generert av de språkmodellbaserte prateroboten.

Vi bruker betegnelsen *prateroboter* om det grensesnittet brukeren opererer i, og oppgir hvilken språkmodell vi har valgt fra menyen i roboten (for eksempel 4.5 for ChatGPT). Det kan likevel være at teknologileverandøren bruker et annet navn om den tjenesten vi tester; for eksempel kaller Microsoft sin Copilot for en «samtaleagent».

I oktober 2024 gjennomførte Språkrådet en lignende test i et mindre prosjekt (Språkrådet 2025a). Den gangen testet vi bare ChatGPT. Resultatet var mellom 2 og 3 feil per side (à 320 ord) på bokmål og ca. 8 feil per side på nynorsk.

Det var flere gode grunner til at vi ønsket å gjennomføre en ny test våren 2025: For det første var det bare formelle feil som ble talt i 2024. Det var altså mange slags problemer som ikke ble registrert, blant annet logiske og syntaktiske brister, dårlig tekstbinding og generelt uidiomatiske formuleringer. Testresultatene gav dermed et litt for positivt bilde av språket i tekstene som ble generert. For det andre blir prateroboten stadig oppdatert og endret, og nye språkmodeller kommer på markedet. For å kunne gi gode råd trenger vi så ferske resultater som mulig. For det tredje har det kommet flere norskproduserte språkmodeller, og de er i liten grad avhengige av tjenestene til multinasjonale selskaper. Med tanke på demokratisk styring, behov for digital selvråderett og respekt for opphavsrett og personvern er det viktig å undersøke om norske modeller kan konkurrere med de utenlandske.

2 Metode

2.1 Valg av prateroboter

Språkrådet har i denne omgangen testet fire prateroboter:

- ChatGPT 4.5

- Microsoft Copilot
- Le Chat med Mistral pro
- NorMistral-11b-warm-instruct (utviklet ved Universitetet i Oslo)

Det finnes langt flere prateroboter (og ulike versjoner av dem) enn Språkrådet har hatt kapasitet til å teste. De fire robotene er valgt med følgende begrunnelse: ChatGPT og Copilot er de to modellene vi antar er mest brukt i Norge dag. Microsoft Copilot er Microsofts KI-samtaleagent. Den er blant annet integrert i Microsoft 365, som svært mange norske offentlige og private virksomheter benytter seg av. De to nevnte modellene er amerikanske. Vi ville også teste en europeisk modell, og valget falt på Le Chat. I tillegg tok vi med en versjon av NorMistral-11b for få en indikasjon på hvor langt modellutviklingen har kommet i Norge.

NorMistral-11b (Samuel et al. 2025) er utviklet av språkteknologigruppa ved Universitetet i Oslo og er en del av gruppas modellportefølje.[1] Modellen er initialisert fra en flerspråklig grunnmodell og videre trent på norskspråklige data. I instruksjonstreningen er det brukt norskspråklige data og ingen maskinoversatte datasett fra engelsk. Det er ikke offentliggjort brukergrensesnitt for modellen, men den kan lastes ned. NorMistral har ikke gjennomgått en like omfattende instruksjonstrening og testing som de kommersielle modellene har, og man kan heller ikke prate med den på samme måte. Derfor måtte vi gjøre enkelte mindre justeringer i instruksene til denne modellen for å få tekster som var tilpasset vår evalueringsmetode.

I perioden april–mai 2025 fikk vi robotene til å generere tekstene som skulle granskes. Teksttypene ble valgt med tanke på skolen og offentlig forvaltning, altså brukergrupper som den norske språkpolitikken retter seg særlig mot. ChatGPT, Copilot og Le Chat fikk instruks om å skrive ti korte fagtekster om forskjellige temaer, og ti korte noveller. NorMistral fikk instruks om å skrive fem fagtekster og åtte noveller. Instruksene skulle være så korte og enkle som mulig, uten andre språklige spesifikasjoner enn «korrekt bokmål» og «korrekt nynorsk».[2] Her er to eksempler:

- Skriv en lang fagtekst om fiskeripolitikk. Ikke bruk punktlister. Skriv på korrekt bokmål.
- Skriv ei lang novelle om Matias som feller tre med far og onkel. Skriv på korrekt nynorsk.

[1] Språkteknologigruppa ved UiO opplyser at en forbedret versjon skal lanseres om kort tid.
[2] Vi kunne ha valgt å lage mer detaljerte instrukser eller satt opp en agent med instrukser om å følge f.eks. Språkrådets skriveregler, noe som kunne forbedret resultatet. Det gjorde vi ikke, siden poenget med denne undersøkelsen ikke var instruksutforming, men testing av modellenes rene språkferdigheter.

2.2 Gjennomgang av tekster og markering av feil og problemer

Seks medarbeidere i Språkrådet med lang erfaring fra språklig kvalitetssikring av tekster (vask og korrektur) gjennomgikk tekstene og markerte feil og problemer. Bøyningsfeil, syntaksfeil og andre grammatiske feil ble sammen med tegnsettingsfeil, forbokstavfeil og andre «tekniske feil» kategorisert som *formelle feil*. Tekstbindingsfeil, logiske brister og svært uidiomatiske formuleringer ble kategorisert som *tekstfeil*.[3] Tabell 1 viser en del eksempler på feil i tekstene.

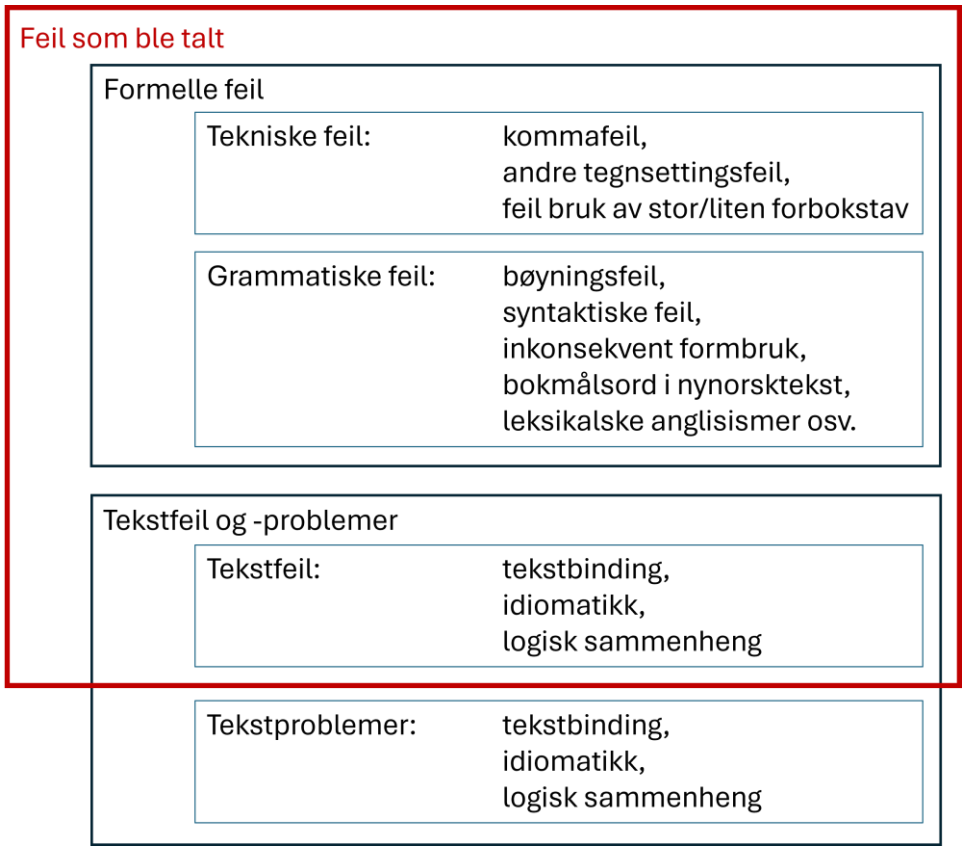
Kategori	**Underkategori**	**Eksempel**
tekniske feil	tegnsetting (komma)	*stillheten som fulgte var nesten øredøvende* (CG)
	tegnsetting (annet)	*FN's mat- og landbruksorganisasjon* (CP)
grammatiske feil	feil genus	*den hvite landskapet* (LC)
	kongruensfeil	*jordbruket vert drivne på ein måte som er berekraftig* (CP)
	syntaksfeil	*denne kilen vil hjelpe oss å styre retningen treet faller* (CP)
	foranstilt eiendomspronomen i nynorsk	*dei […] hadde fullført si lange vandring* (CG)
	bokmålsord i nynorsktekst	*dukkene og bamsane fann saman* (NM)
tekstfeil	sviktende tekstbinding	*Fra et system med relativt fri tilgang og lite regulering, har en økende forståelse for begrensede ressurser og faren for overfiske ført til strengere reguleringer og mer avanserte kvotesystemer.* (CG)
	uidiomatisk språk	*han tok et raskt sukk av koppen* (LC)
	ulogisk ordbruk	*i sammendrag er fiskeripolitikk et komplekst felt som krever et balansert syn på […]* (LC)
	anglonorsk	*på eit blunk var rommet tilbake i orden* (CG)

Tabell 1. Noen typer feil i tekstene med eksempler.
CG = ChatGPT, CP = Copilot, LC = Le Chat, NM = NorMistral.

Det er ofte et skjønnsspørsmål hva som skal regnes som språkfeil. Det vi har kalt tekstfeil, er vanskeligere å vurdere enn formelle feil. For å redusere risikoen for store forskjeller i vurderingen av tekstene talte vi opp bare de formelle feilene og de groveste tekstfeilene. Resten kategoriserte vi som «tekstproblemer», og de ble ikke tatt med i den kvantitative behandlingen av materialet. For å sikre mest mulig enhetlig vurdering gjennomgikk to av medarbeiderne i annoteringsgruppa alle tekstene på nytt etter den første gjennomgangen

[3] Én og samme tekststreng kan i mange tilfeller ha ulike typer feil i kategorien tekstfeil, f.eks. både dårlig tekstbinding og uidiomatisk språk. I slike tilfeller ble det likevel bare registrert én tekstfeil, jf. vedlegg 1.

og tok en endelig avgjørelse om hva som skulle regnes som feil. Figur 1 gir en skjematisk oversikt over kategoriseringen av feil og (andre) problemer.



Figur 1: Kategorisering av feil og problemer i tekstene.

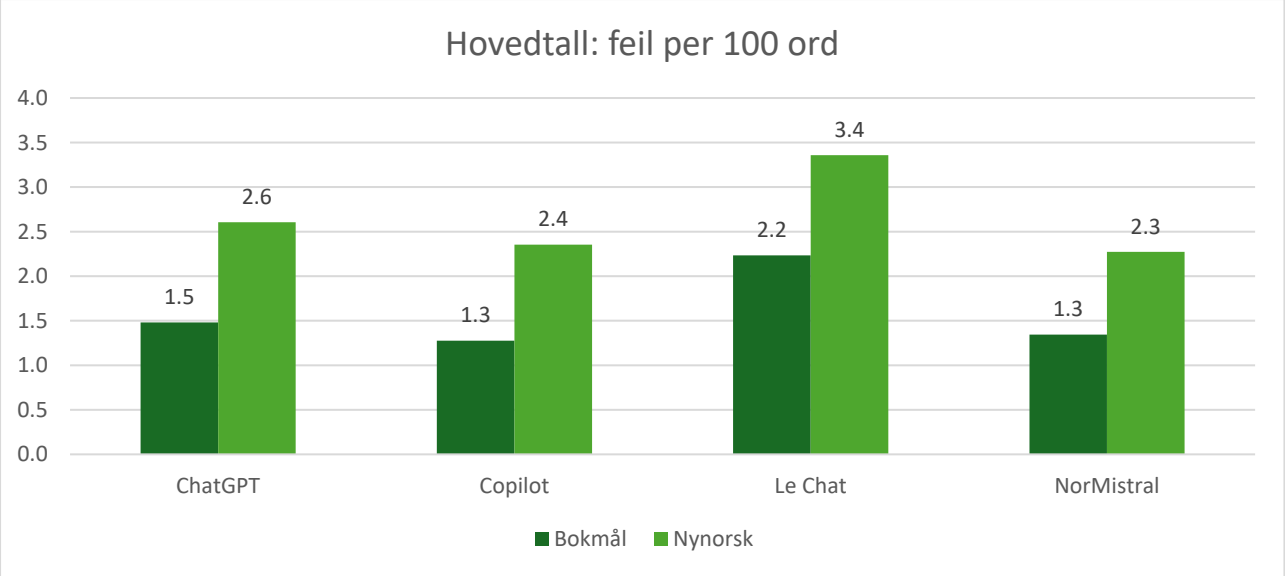
3 Resultat av telling av feil og problemer

3.1 Feilprosent hos de fire praterobotene

Det ble generert til sammen 68 635 ord. Tabell 1 og figur 2 nedenfor viser feilprosentene hos de fire praterobotene, uten hensyn til teksttype. Vi har slått sammen feiltallene for sakprosatekster og skjønnlitterære tekster.

Praterobot og skriftspråk	Antall ord	Antall feil	Feil per 100 ord
ChatGPT bokmål	9725	144	1,5
Copilot bokmål	9480	121	1,3
Le Chat bokmål	10070	225	2,2
NorMistral bokmål	6617	89	1,3
ChatGPT nynorsk	8405	219	2,6
Copilot nynorsk	9515	224	2,4
Le Chat nynorsk	8663	291	3,3
NorMistral nynorsk	6160	140	2,3

Tabell 1: Grunnlagstall og feilprosent hos de fire praterobotene på bokmål og nynorsk.



Figur 2: Feilprosent hos de fire praterobotene på bokmål og nynorsk.

Målet feil per 100 ord gir ikke nødvendigvis noe klart bilde av tekstkvaliteten. Det følgende eksempelet kan gjøre feilprosenten litt mer forståelig: En feilprosent på 2,3 tilsvarer om lag én feil for annenhver setning i en typisk sakprosatekst i vårt materiale. Det svarer til minst 8 feil per side (normalside à 2400 anslag). På en A4-side i Word kan tallet fort bli større, kanskje over 12. I tillegg er det altså svært mange og til dels betydelige problemer i materialet som ikke er registrert som feil, jf. figur 1.

3.2 Vurdering av samlet prestasjon

Feilene og problemene setter et sterkt preg på de tekstene vi har gransket. Tekstene kan derfor vanskelig karakteriseres som gode.

Det er to hovedfunn:

- Alle praterobotene gjør så mange feil at tekstene vil trenge grundig språkvask og korrektur før eventuell publisering.
- Alle praterobotene mestrer bokmål langt bedre enn nynorsk. Forskjellene mellom skriftspråkene hos hver praterobot er signifikante (på 0,01-nivået). Vi kommer tilbake til dette funnet i avsnitt 3.5 nedenfor.

Mange vil etterlyse et tredje funn: Hvilken praterobot er det som gjør færrest feil? Men undersøkelsen gir ikke grunnlag for en kåring av beste praterobot.

Figur 2 gir et grovt inntrykk av prestasjonene: På bokmål er Copilot og NorMistral best og jevngode, nokså tett fulgt av ChatGPT, med Le Chat et godt stykke bak. På nynorsk er NorMistral best, tett fulgt av Copilot, dernest av ChatGPT, igjen med Le Chat et godt stykke bak. Men det er bare noen av forskjellene som er signifikante og kan brukes til å peke (forsiktig) ut over det materialet vi har undersøkt. Målt mot gruppegjennomsnittet på samme skriftspråk skiller disse kombinasjonene seg signifikant ut:

- positivt: Copilot bokmål, Copilot nynorsk og NorMistral nynorsk
- negativt: Le Chat bokmål og Le Chat nynorsk

Vi har også sammenlignet de enkelte kombinasjonene med hverandre. Da fant vi bare én signifikant forskjell:

- Le Chat bokmål (høyest feilprosent) mot Copilot bokmål

Noen forskjeller er nesten signifikante:

- Le Chat bokmål (høyest feilprosent) mot NorMistral bokmål og ChatGPT bokmål
- Le Chat nynorsk (høyest feilprosent) mot NorMistral nynorsk og Copilot nynorsk

ChatGPT kommer litt dårligere ut i vårt materiale enn Copilot og NorMistral, men forskjellen er ikke signifikant når vi måler robotene mot hverandre direkte. Vi har altså ikke godt nok grunnlag for å rangere de tre praterobotene NorMistral, Copilot og ChatGPT.

Det vi kan slå fast, er at alle de fire robotene gjør mange feil av mange slag, og at Le Chat gjør flest.

Merk at søylene for nynorsk i figur 2 ikke bare representerer nynorskfeil, men feil generelt i tekster på nynorsk. Forskjellen mellom nynorskkompetansen og bokmålskompetansen til robotene er med andre ord større enn søylene gir inntrykk av, men akkurat hvor mye større er usikkert, for vi har ikke sammenlignet nynorskspesifikke og bokmålsspesifikke feil direkte. En sammenligning av resultatene i feilkategorien «grammatiske feil» i figur 3 nedenfor forteller likevel en god del om forholdet, siden disse feilene er de mest bokmåls- og nynorskspesifikke.

3.3 Vurdering av feil i ulike kategorier

Er det mulig å si noe om sterke og svake sider ved praterobotene?

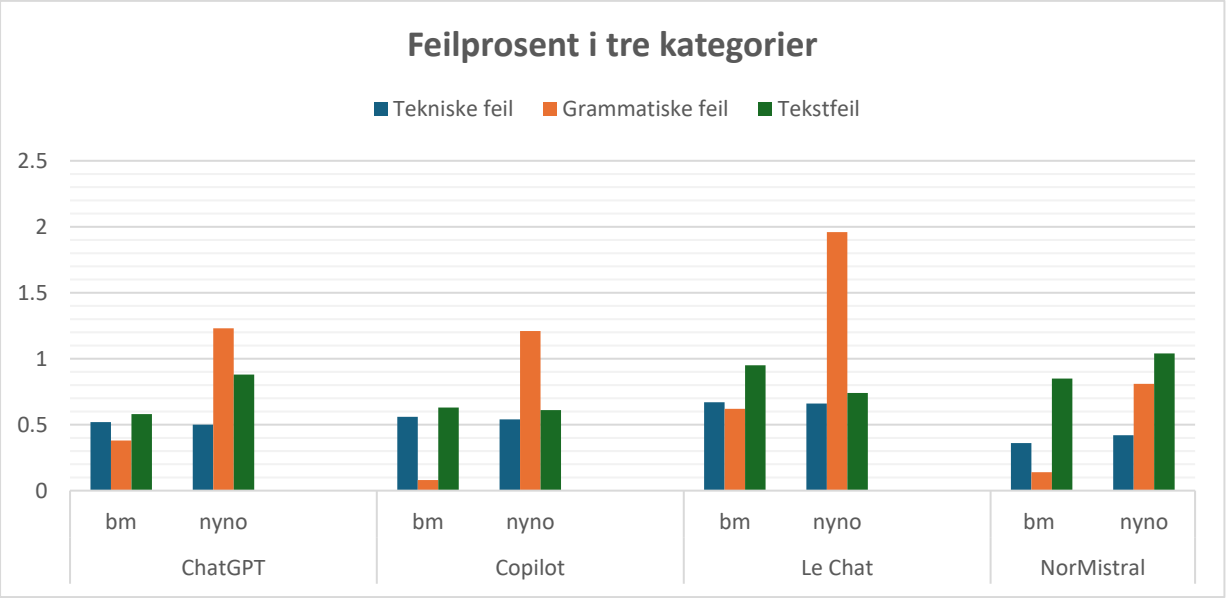
Vi kan ikke si noe særlig om helt spesifikke feiltyper. Til det er tallene for små, og de tre feilkategoriene er nokså grove. Men feilene i de to første kategoriene (tekniske feil og grammatiske feil) er tross alt ganske sammenlignbare. Tabellen nedenfor viser både samlet feilprosent og feilprosent innenfor hver feilkategori. De klareste avvikene fra gruppegjennomsnittet for alle robotene er markert med rødt (for relativt dårlig resultat) og blått (for relativt positivt resultat), og de er signifikante.

Feilprosent i tre kategorier

Praterobot og skriftspråk	Alle feiltyper	Tekniske feil	Grammatiske feil	Tekstfeil
ChatGPT bokmål	1,48	0,52	**0,38**	**0,58**
ChatGPT nynorsk	2,61	0,50	**1,23**	0,88
Copilot bokmål	**1,28**	0,56	**0,08**	0,63
Copilot nynorsk	**2,35**	0,54	**1,21**	0,61
Le Chat bokmål	**2,23**	0,67	0,62	**0,95**
Le Chat nynorsk	**3,36**	0,66	**1,96**	0,74
NorMistral bokmål	1,35	**0,36**	**0,14**	0,85
NorMistral nynorsk	**2,27**	0,42	0,81	**1,04**

Tabell 2: Feilprosenter hos praterobotene, samlet (alle feiltyper) og i underkategorier. Tallene som skiller seg klarest ut, er markert med farger: blått for positive avvik og rødt for negative.

Her er det samme resultatet framstilt med en figur:



Figur 3: Feilprosent fordelt på kategorier etter praterobot og skriftspråk.

Observasjoner

- Tekniske feil: NorMistral skiller seg positivt ut, særlig på bokmål.
- Grammatikk, bokmål: NorMistral og særlig Copilot skiller seg positivt ut. ChatGPT gjør det bra målt mot gjennomsnittet for robotene, men ikke mot NorMistral og Copilot særskilt.
- Grammatikk, nynorsk: NorMistral skiller seg positivt ut. De andre tre gjør det svakt.
- Tekstfeil, bokmål: ChatGPT skiller seg positivt ut, mens Le Chat skiller seg mest negativt ut.
- Tekstfeil, nynorsk: NorMistral skiller seg negativt ut, mens Copilot har den laveste feilprosenten.

NorMistral og Copilot gjør det altså forholdsvis skarpt på bokmål på grunn av det relativt lave feiltallet innenfor grammatikk (skrivemåte og bøyning), og NorMistral har dessuten færre «tekniske feil» (tegnsettingsfeil o.l.) enn de andre robotene. Videre ser NorMistral ut til å mestre nynorsk rettskrivning best. Men det forspranget NorMistral får på de to første områdene, forsvinner når vi tar tekstfeil med i regningen; NorMistral har mange slike.

3.4 Særskilt om anglisismer

Vi har notert engelske innslag og oversettelseslån[4] fra engelsk for seg. ChatGPT har 15 anglisismer til sammen (11 av dem oversettelseslån), Copilot har 10 (alle oversettelseslån), Le Chat har 11 (alle oversettelseslån). NorMistral har bare 4 (3 oversettelseslån), men tekstomfanget er mindre. Dette svarer til en «anglisismepromille» på litt under 0,6 for Le Chat og Copilot, omtrent 0,8 for ChatGPT og 0,3 for NorMistral. På grunn av de små tallene er ingen av forskjellene signifikante. Tallene er egentlig større, for mange oversettelseslån er bare notert som problemer, og de inngår altså ikke i tallmaterialet som danner grunnlag for den kvantitative framstillingen. De opptalte anglisismene utgjør mindre enn én per side, selv hos ChatGPT, men til gjengjeld er mange av dem ganske grove:

- *en kopp [kaffe] som kunne tilby henne en ny start* (ChatGPT)
- *plutselig fanget øyet hans noe* (ChatGPT)
- *ho tok eit djupt pust* (Copilot)
- *for no var han tilfreds med å vere i byen* (Copilot)
- *å informere fiskeripolitikken* (Copilot)
- *begge kjente den [skogen] som sine egne lommer* (Le Chat)
- *sa hun med en stolt stemme* (Le Chat)
- *de ulike aktørene involvert* (Le Chat)
- *en kveld som hadde gått fullstendig ut av kontroll* (NorMistral)

3.5 Forskjellsbehandling av bokmål og nynorsk

Alle praterobotene presterer dårligere i tekster på nynorsk enn i tekster på bokmål. Hovedtallene i figur 2 kan lett gi inntrykk av at kvalitetsforskjellen på bokmålstekst og nynorsktekst er omtrent den samme hos alle praterobotene. Hvis vi trekker feilprosenten på nynorsk fra feilprosenten på bokmål, får vi nesten den samme forskjellen mellom skriftspråkene:

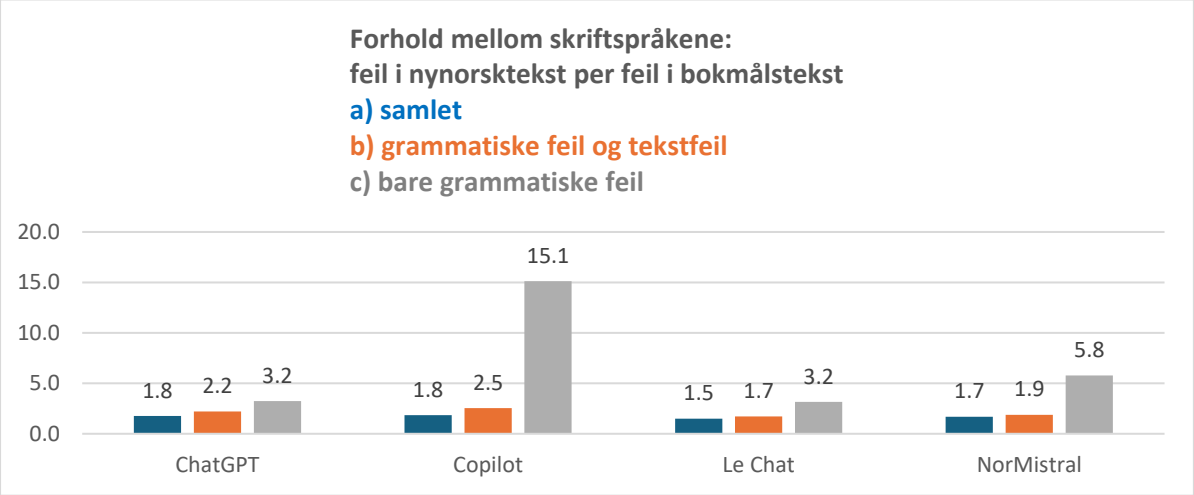
- ChatGPT, Copilot og Le Chat: 1,1 prosentpoeng forskjell mellom feil i bokmålstekst og feil i nynorsktekst
- NorMistral: i underkant av 1,0 prosentpoeng forskjell mellom feil i bokmålstekst og feil i nynorsktekst

Bildet forandrer seg litt hvis vi i stedet ser på forholdet mellom feilratene:

- Chat GPT og Copilot: 1,8 feil i nynorsktekst per feil i bokmålstekst
- NorMistral: 1,7 feil i nynorsktekst per feil i bokmålstekst
- Le Chat: 1,5 feil i nynorsktekst per feil i bokmålstekst

[4] Oversettelseslån er ord eller uttrykk som er lånt inn gjennom direkte oversettelse ledd for ledd, for eksempel *gjøre en forskjell* (av engelsk *make a difference*).

Vi kan skjerpe motsetningen mellom skriftspråkene ved å trekke fra de feilkategoriene der vi for det meste finner feil som er felles for nynorsk og bokmål (det vil si tekniske feil og tekstfeil). Da får vi disse forholdene:



Figur 4: Forholdet mellom antall feil i nynorsktekst og antall feil i bokmålstekst samlet og i ulike feilkategorier for hver av robotene.

Når vi går mot høyre i hver søylegruppe (til den oransje og den grå søylen), får vi et skarpere – men ikke nødvendigvis rimelig – bilde av forholdet mellom skriftspråkene. Det er på et vis riktig at Copilot diskriminerer nynorsk mest, og at Le Chat diskriminerer minst. Men spør vi i stedet hvilken praterobot som produserer de beste og dårligste tekstene på nynorsk, er jo svaret det stikk motsatte (se hovedtallene i figur 2). En praterobot kan altså behandle et skriftspråk bedre enn andre roboter gjør, men samtidig diskriminere det, altså forskjellsbehandle det negativt. Eller omvendt. Når Copilot og NorMistral kommer så dårlig ut på figurene ovenfor, skyldes det snarere at de er relativt gode i bokmål, enn at de er spesielt dårlige i nynorsk. Det er også flere andre svakheter ved sammenligningen som gjør at vi ikke bør legge for mye vekt på disse forholdstallene.[5]

4 Problemet er ikke bare formelle feil

Tallene i tabellene og figurene er altså utelukkende basert på det vi har regnet som ganske klare feil, små eller store. Vi har notert en lang rekke problemer i tillegg til disse feilene, men vi har ikke talt dem. De fleste problemene gjelder tekstbinding, logikk og uidiomatiske

[5] Vi nøyer oss med å nevne to av dem: For det første inneholder kategorien «grammatiske feil» en god del konsekvensfeil (vingling mellom tillatte former), og det er ikke nødvendigvis alvorlige feil. Det at nynorsk har mye valgfrihet som faktisk blir brukt, kan bidra til å blåse opp feiltallene. På den andre siden er en del av de nynorskspesifikke syntaksproblemene å finne i kategorien «tekstproblemer» (og disse er dermed ikke talt).

formuleringer. Stikkprøver i sakprosatekstene viser at problemtallet kan være minst dobbelt så stort som feiltallet. I noen av tekstene er det 6–7 problemer (i tillegg til feil) per 100 ord hos de fleste praterobotene (litt færre for Copilot bokmål og litt flere for Le Chat nynorsk).

Med andre ord kan det godt være over 30 små og store feil og problemer på en A4-side. I vedlegg 1 sist i rapporten har vi tatt med smakebiter både på feil og andre problemer.

5 Noen problemer med bruk av praterobotter til tekstarbeid

Tekster med 30 feil og problemer per A4-side krever mye etterarbeid om de skal bli presentable. Det gjør også tekster med 20 eller 10 feil og problemer. Tekstene trenger grundig gjennomgang uansett. Alle KI-tekstene vi har generert, inneholder for mange feil og problemer av ulike slag til at de kan fanges opp ved hjelp av en enkel sjekkliste eller ved hjelp av retteprogrammer. Det er verdt å merke seg at påfallende og grove feil ofte kan rettes opp fort og automatisk, mens det kan ta lang tid å rydde opp i små, spredte og diffuse problemer.

Det hevdes ofte at praterobotter egner seg godt til tekstforbedring og språkvask. Vi har ikke testet disse egenskapene spesifikt, men undersøkelsen viser tydelig at det praterobotene selv genererer, krever språkvask og korrektur. Det å erstatte menneskelig sluttkontroll med kunstig intelligens er altså risikabelt. En undersøkelse gjort av byrået Totaltekst i 2025 (Totaltekst 2025) viser at ChatGPT ikke er til å stole på som språkvasker.

To andre anbefalte bruksområder for praterobotter er generering av tekstutkast og generering av sammendrag. Det sier seg selv at utkast og sammendrag trenger språklig bearbeiding på linje med annen KI-generert tekst. Man bør heller ikke glemme at nettopp det å skrive utkast og sammendrag kan være nyttig for å holde både skrivekompetanse og fagkompetanse ved lag hos mennesker.

Det kan tenkes at noen vil bruke KI *både* til å lage tekstutkast og til å bearbeide og kontrollere tekster. Det er en ekstra risikabel metode. Foreløpig er det umulig å produsere tekster med akseptabel språklig kvalitet på denne måten, enten man bruker samme praterobot til begge oppgaver, eller setter én praterobot til å bearbeide og kontrollere et utkast generert av en annen.

Et tredje anbefalt bruksområde for KI i skriveprosessen er drøfting av ideer. Et påfallende trekk ved sakprosatekstene vi har generert, er at de mangler kildehenvisninger. Ønsker man et solid grunnlag for drøfting av ideer, må man altså huske å be praterobotene om å oppgi kilder. Trolig bør man også kontrollere dem nøye. Kildekritikk og kildekunnskap er viktigere enn noen gang, både i skolen og forvaltningen.

Til slutt må vi nevne risikoen for at KI-generert tekst kommer til å endre måten vi skriver på. De tekstene vi har generert i denne undersøkelsen, er ikke bare preget av feil, men også av svevende, generelle og uidiomatiske formuleringer (se vedlegg 1).

Oppsummeringsformuleringene på slutten av hver sakprosatekst (se eksempel 7, 8 og 9 i vedlegget) er blant de mest uidiomatiske i materialet vårt. Det vil være særdeles uheldig om den stilen de representerer, sprer seg i norsk forvaltningsprosa, for eksempel gjennom KI-genererte sammendrag.

6 Råd om forsvarlig bruk av prateroboter til tekstarbeid i offentlige organer

Språkloven paragraf 9 («klarspråksparagrafen») (Lovdata 2021) slår fast at offentlige organ skal bruke et klart og korrekt språk som er tilpasset målgruppa. Hvis et offentlig organ vil bruke de praterobotene vi har testet, til å produsere tekst fra grunnen av, må de derfor

- vurdere bruksområdet nøye
- sette av nok ressurser til etterarbeid, inkludert kontroll av språk, fakta og kilder
- sette av ekstra ressurser til kontroll av og etterarbeid med tekster på nynorsk
- ha kompetente tekstarbeidere i staben eller hyre profesjonelle språkkonsulenter
- ha den nødvendige språkkompetansen til å vurdere når det er verdt ressursbruken og risikoen å overlate skrivearbeid til prateroboter
- sørge for at den menneskelige skrivekompetansen (og dermed tekstbehandlings- og tekstvurderingskompetansen) i statsorganet ikke forvitrer på lengre sikt

Vi gir ikke her råd om hvordan problemene vi har beskrevet, bør håndteres i skolen og den høyere utdanningen. Men det er viktig at problemene er kjent og tas på alvor.

Konklusjonene om forsvarlig bruk av KI til tekstproduksjon gjelder for alle praterobotene i undersøkelsen og er ikke avhengige av de nøyaktige feiltallene i denne eller andre undersøkelser.

Vedlegg 1: Eksempler på feil og andre problemer i kategorien «tekstfeil», hentet fra tre sakprosatekster på bokmål

1) «Det er nettopp i skjæringspunktet mellom disse ulike målene at landbrukspolitikken blir så kompleks og viktig, og at det er nødvendig med en kontinuerlig politisk debatt om prioriteringer og veivalg for fremtiden.» (ChatGPT om landbrukspolitikk)

   Kommentar: Samordningen av de to at-setningene med *og* er notert som en uidiomatisk formulering på grensa til syntaktisk brist, og som en feil. Den merkelige framstillingen ellers – at landbrukspolitikken «blir» så kompleks og viktig i skjæringspunktet mellom målene – er ikke registrert, men kunne godt ha vært notert som et problem.

2) «Det norske landbruket kjennetegnes av små og mellomstore familiebruk, der produksjonen er fordelt utover hele landet med betydelige variasjoner i klima, jordbruksstruktur og produksjonsforhold.» (ChatGPT om landbrukspolitikk)

   Kommentar: *Der* har ikke et skikkelig korrelat. Viser det til familiebrukene, er det saklig galt, viser det til landbruket, er det syntaktisk galt. Notert som feil.

3) «Dette førte til at Norge inntok en forsiktig og balansert linje; landet ønsket både å sikre sin egen sikkerhet gjennom avskrekking innenfor NATO, samtidig som det ønsket å unngå en overdreven provokasjon av Sovjetunionen.» (ChatGPT om Norges forhold til Nato)

   Kommentar: «*Både* + *samtidig*» er notert som feil, likeledes kommaet foran *samtidig* og «avskrekking innenfor NATO». Den merkelige substantiviske formuleringen «unngå en overdreven provokasjon av» er notert som et problem, sammen med det lite elegante «sikre sin egen sikkerhet».

4) «Fiskeripolitikk er et komplekst og viktig område som omhandler regulering og forvaltning av fiskeressurser.» (Copilot om fiskeripolitikk)

   Kommentar: At *området* «omhandler» noe, er notert som feil. «Komplekst og viktig» og lignende er standardfraser som går igjen i de fleste tekstene. De er notert som et problem der de ikke tilfører noen innsikt.

5) «Miljøhensyn er en annen viktig komponent i fiskeripolitikk. Bærekraftig fiskeriforvaltning innebærer å minimere miljøpåvirkningen fra fiskeriaktivitetene. Dette kan inkludere tiltak for å beskytte sårbare habitater, redusere forurensning og bevare biologisk mangfold.» (Copilot om fiskeripolitikk)

Kommentar: «Viktig komponent», «innebærer» og «dette kan inkludere» er standardformuleringer som brukes i de fleste tekstene uten hensyn til den språklige sammenhengen. Notert som problem noen steder. Kan kalles typisk KI-stil.

6) «Fiskeripolitikk er et dynamisk felt som krever kontinuerlig tilpasning og innovasjon. Nye utfordringer og muligheter oppstår, og det er viktig å være fleksibel og åpen for endringer.» (Copilot om fiskeripolitikk)

   Kommentar: Første setning bærer preg av ordgyting. Det at de samme formuleringene går igjen i flere tekster uavhengig av emne, forsterker det inntrykket. Det er likevel ikke notert som et problem for seg her. Men den andre setningen henger i lufta, og det er registrert som en tekstbindingsfeil.

7) «Den norske NATO-politikken er også sterkt preget av Norges bidrag til kollektiv sikkerhet gjennom alliansens felles kapasiteter. […] Alt i alt er Norges forhold til NATO dypt forankret i landets historie, politikk og strategiske kultur.» (ChatGPT om Norges forhold til Nato)

   Kommentar: Oppsummeringsformuleringen «alt i alt … forankret» er registrert som feil. I tillegg er det registrert et par problemer med ullen framstilling.

8) «Samlet sett er fiskeripolitikk en balansegang mellom ulike hensyn og interesser.» (Copilot om fiskeripolitikk)

   Kommentar: Oppsummeringsformuleringen «samlet sett … balansegang» er notert som en tekstfeil (jf. «alt i alt … forankret» hos ChatGPT ovenfor).

9) «For å møte fremtidige utfordringer innen landbrukspolitikk, er det nødvendig med en fleksibel og tilpasningsdyktig tilnærming. […] I sammendrag er landbrukspolitikk et komplekst felt som krever et balansert syn på økologiske, økonomiske, sosiale og kulturelle aspekter.» (Le Chat om landbrukspolitikk)

   Kommentar: I første setning er standardformuleringen «fleksibel og tilpasningsdyktig tilnærming» notert som et ordgytingsproblem, og det samme gjelder «aspektene» og «kompleksiteten» i siste setning, som går igjen i mange tekster. Kommaet foran *er* er notert som en feil. Endelig er oppsummeringsformuleringen «[er] i sammendrag et komplekst felt» notert som feil (jf. ChatGPTs «alt i alt» og Copilots «samlet sett»).

10) a) «Fiskerinæringen er ikke bare en økonomisk aktivitet; den er også en kulturell bærebjelke langs kysten, bidrar til sysselsetting og levende lokalsamfunn.» (NorMistral om fiskeripolitikk)

b) «Norge har bidratt med styrker til flere NATO-operasjoner gjennom årene, både i fredsbevarende, fredsopprettende og opplæringsoppdrag.» (NorMistral om Norges forhold til Nato)

Kommentar: Her har vi to ulike eksempler på gal bruk av oppramsingsregelen «1, 2 og 3». Feiltypen er vanlig hos alle praterobotene.

Vedlegg 2: Referanser

Lovdata (2021): Lov om språk (språklova). <https://lovdata.no/dokument/NL/lov/2021-05-21-42/%C2%A79#%C2%A79> (oktober 2025).

Samuel, David et al. (2025): *Small Languages, Big Models: A Study of Continual Training on Languages of Norway*. < https://aclanthology.org/2025.nodalida-1.61.pdf> (oktober 2025).

Språkrådet (2025a): KI-språkets fallgruver. <https://sprakradet.no/aktuelt/ki-sprakets-fallgruver/> (oktober 2025).

Språkrådet (2025b): Språkrådets test av skriveprogrammer på norsk. <https://sprakradet.no/wp-content/uploads/rapport_fra_sluttkorrektur_29.9.2025.pdf> (oktober 2025).

Totaltekst (2025): *Vi tester KI-språkvask*. <https://www.totaltekst.no/blogg/vi-tester-ki-sprakvask> (oktober 2025).