Helsedirektoratet

Store språkmodellar i helse- og omsorgstenesta – eit kunnskapsgrunnlag

Først publisert: 06. mai 2024
Siste faglige endring: 06. mai 2024

Innhold

1. Forord 3
2. Samandrag 4
3. Introduksjon 5
4. Om store språkmodellar 7
5. Treningsdata 18
6. Bruksområde for språkmodellar 23
7. Utfordringar og risikoar 30
8. Moglege tiltak 31
9. Engelsk-norsk ordliste 34
10. Vedlegg A: Døme på risikoar ved bruk av store språkmodellar 35

Forord

Dette dokumentet skal gje eit statusbilete av modenskapen til språkmodellar i helse- og omsorgssektoren. Dokumentet skal også peike framover på moglege bruksområde og risikoar, i tillegg til overordna vegval og tiltak.

Innhaldet i dokumentet er basert på artiklar, forskingslitteratur, presentasjonar og intervju med fleire aktørar i og utanfor helsesektoren:

- Geir Thore Berge og Tor Oddbjørn Tveit (Sørlandet sjukehus)
- Christian Autenried (Helse vest IKT)
- Yngvil Beyer, Per-Erik Solberg, Magnus Brede Birkenes, Per-Erik Kummervold og Svein-Arne Brygfjeld (Nasjonalbiblioteket)
- Gina Helstad (Helsearkivet)
- Kristine Eide, Ann-Helen Langaker og Matias Jentoft (Språkrådet)
- Jens Andresen Osberg og Alexandra Schultz (Digitaliseringsdirektoratet)
- Pål Brekke (DIPS AS)
- Jon Atle Gulla (Noregs teknisk-naturvitskaplege universitet (NTNU))
- Alexander Lundervold (Høgskulen på Vestlandet)
- Arvid Lundervold (Universitetet i Bergen)
- Hercules Dalianis (Nasjonalt senter for e-helseforsking)
- Finn-Henry Hansen (Helse nord RHF)
- Karl Øyvind Mikalsen (Senter for pasientnær kunstig intelligens (SPKI)) og Geir Villy Isaksen (Helse nord IKT)
- Lilja Øvrelid (Universitetet i Oslo)
- Fredrik Andreas Dahl og Pierre Lison (Norsk Regnesentral (NR))
- Håkon Haaheim (Helseplattforma AS)
- Elin Thygesen (Universitetet i Agder)

Vi vil rette ein særskilt takk til Alexander Lundervold (Høgskulen på Vestlandet) for tolmod og viktige bidrag i løpet av fleire samtalar til den tekniske delen av dokumentet.

Samandrag

Dokumentet gir eit oversyn over bruk og utvikling av store språkmodellar i helse- og omsorgssektoren i Norge. Det beskriv korleis slike modellar kan bidra til forbetra helsetenester, inkludert potensielle fordeler og utfordringar.

Hovudpunkter:

1. **Introduksjon:** Prosjektet "Bedre bruk av kunstig intelligens" vart starta i 2019 som ein del av nasjonal helse- og sjukehusplan (NHSP 2019-2023), med Helsedirektoratet i leiinga.
2. **Store språkmodellar:** Store språkmodellar er statistiske modellar som kan analysere og generere tekst basert på naturleg språk. Dei kan vere generelle eller tilpassa spesifikke fagområde som helse.
3. **Ettertrening og helsefagleg rettleiing:** Store språkmodellar kan tilpassast helsefag gjennom ettertrening og rettleiing. Finjustering, instruksjonsjustering, og etisk justering er blant metodane.
4. **Strategiar for ettertrening og rettleiing:** Forskjellige tilnærmingar til utvikling av helsefaglege språkmodellar er presenterte, inkludert alternativ som berre inneber helsefagleg rettleiing.
5. **Forvaltning av språkmodellar:** Diskusjonen om korleis språkmodellar bør forvaltast i helse- og omsorgssektoren, enten på nasjonalt nivå, regionalt nivå eller gjennom kommersielle aktørar.
6. **Treningsdata:** Viktigheten av tilgang til høgkvalitets treningsdata for å sikre gode språkmodellar blir understreka. Dette inkluderer ulike typar helsefaglege tekstar.
7. **Bruksområde:** Rapporten skisserer ulike måtar språkmodellar kan bidra til i helse- og omsorgssektoren, som for eksempel i strukturering av journaldata, avgjerdsstøtte, og som utdanningsverktøy.
8. **Utfordringar og risikoar:** Potensielle risikoar som personvernbrot, feilinformasjon, og mangel på transparens blir diskutert.
9. **Moglege tiltak:** Tiltak for å møte utfordringane og styrke bruken av språkmodellar i sektoren blir føreslått, inkludert kapasitetsbygging og risikoanalyse.

Rapporten avsluttar med å understreke behovet for vidare forsking og utvikling, samt nødvendige vurderingar rundt forvaltning og bruk av språkmodellar i helse- og omsorgssektoren.

(Samandraget er skrive av ChatGPT 4 og er uredigert)

Introduksjon

Det nasjonale koordineringsprosjektet «Bedre bruk av kunstig intelligens» (koordineringsprosjektet for KI) starta opp i siste halvdel av 2019 som ein del av arbeidet med nasjonal helse- og sjukehusplan (NHSP 2019-2023). I prosjektet samarbeidde Helsedirektoratet, Direktoratet for e-helse, Statens legemiddelverk, dei regionale helseføretaka, Kommunesektorens organisasjon (KS), Helsetilsynet og Folkehelseinstituttet. Kompetansenettverket Kunstig intelligens i norsk helseteneste (KIN) har vore observatør. Helsedirektoratet har leia arbeidet. Arbeidet blir vidareført i 2024 gjennom eit nytt prosjektet som skal leggje til rette for trygg innføring av KI i helse- og omsorgstenesta. Denne rapporten om store språkmodellar i helse- og omsorgstenesta er ein leveranse frå disse prosjekta.

Rapporten er eit kunnskapsgrunnlag som har som mål å gje eit augeblinksbilete over landskapet for store språkmodellar på helsefeltet i Noreg pr. mars 2024. Det blir forventa at nokre område vil utvikle seg raskt dei komande åra, og det er vanskeleg å fange opp alle aktivitetar. Biletet vil derfor ikkje være komplett. Rapporten skal nyttast av Helsedirektoratet i det vidare arbeidet på KI-området, spesielt i arbeidet med felles KI-plan.[1]

Bakgrunn

Rapporten «Tilgang til data til kunstig intelligens i helse- og omsorgstjenesten» peikar på at løysingar som kan bidra til kategorisering, klassifisering og strukturering av data på ein enklare måte for helsepersonell, vil kunne gje betre data til KI og redusere meirbelastinga på klinisk personell. KI kan også nyttast til struktureringsarbeidet.[2] Koordineringsprosjektet tilrådde at Direktoratet for e-helse utarbeider eit kunnskapsgrunnlag som summerer opp metodar og verktøy som kan strukturere helseinformasjon på ein føremålstenleg måte. Det skulle blant anna gjerast greie for bruk av naturleg språkprosessering for å effektivisere arbeidet med datafangst, klassifisering og koding, og forenkle søk ved datauttrekk.

I løpet av hausten 2022 fekk store (språk)modellar og generative modellar stor utbreiing. Spesielt gjeld dette KI-verktøy som behandlar tekst, som samtalerobotar (chatbot). Språkmodellar har vist seg å gi oppsiktsvekkjande gode svar på spørsmål. Ein måte å illustrere forbetringane av tolking av tekstar og kor rask utviklinga har vore, er å sjå på korleis språkmodellar svarar på eksamensspørsmål for medisinstudentar. I november 2022 svarte den beste språkmodellen DRAGON berre 47,5 prosent rett. I mars 2023 oppnådde Med-PALM-2 frå Google DeepMind korrekt svarprosent på 86.5 prosent. Den ville bestått medisineksamen med god margin.

Utviklinga av store språkmodellar og generative modellar har styrkt moglegheita for at KI-verktøy kan nyttast for fleire område i helse- og omsorgstenesta. Det er mogleg at fleire manuelle oppgåver med tekst i helsetenesta kan effektiviserast, som til dømes å sjekke at kode og journalnotat stemmer, foreslå kodar, summere opp journalnotat og lage utkast til epikrise. Samstundes føreset det grundige vurderingar av t.d. risikoar som brot på personvern, hallusinasjonar (oppdikta innhald) og bias, mangel på transparens og så bortetter.

Koordineringsprosjektet for KI oversende eit notat til Helse- og omsorgsdepartementet 2. juni 2023 som skildra oppgåver der KI har eit forventa potensial som personellsparande teknologi i løpet av den neste fireårsperioden. Eit av områda var å effektivisere skriveoppgåver.[3]

Bruk av store språkmodellar i helse- og omsorgstenesta vil likevel ha som føresetnad at språkmodellane og/eller applikasjonane som blir nytta, er godt tilpassa språkbruken og kulturen i den norske helsetenesta.

Det finst fleire sentrale initiativ knytt til språkmodellar i helse- og omsorgssektoren i Noreg. Innsikt i desse har vore viktige i dette kunnskapsgrunnlaget:

- Nasjonalt senter for e-helseforsking, DIPS, Universitetssjukehuset i Nord-Noreg: ClinCode-prosjektet
- Helse vest IKT, saman med DIPS: Klinisk NorBERT
- Helsearkivet
- Helsedirektoratet
- Sørlandet sjukehus (no avslutta)
- Universitetet i Agder

Denne rapporten skal gje eit innblikk i kva språkmodellar er og ulike måtar å utvikle og finjustere dei på. Dokumentet skal også vurdere potensielle bruksområde for språkmodellar i helsetenesta, og gje ein kort status over aktuelt arbeid med språkmodellar i helsetenesta i Noreg. Kunnskapsgrunnlaget skal også skissere hindringar og risikoar ved utvikling, finjustering, bruk og forvalting av språkmodellar i tillegg til å peike på framtidige vegval og moglege tiltak på området.

[1] TB2024-34: Kunstig intelligens i helse- og omsorgstjenesten: https://www.regjeringen.no/contentassets/d8f63d7d01d64def982cb7c8ce1eeb64/tildelingsbrev-til-helsedirektora

[2] https://www.ehelse.no/publikasjoner/tilgang-til-data-til-kunstig-intelligens-i-helse-og-omsorgstjenesten

[3] Notatet er internt, men beskrivinga av dei moglege arbeidsoppgåvene finst i kapittel 3: https://www.helsedirektoratet.no/rapporter/status-og-forslag-til-videre-arbeid-med-kunstig-intelligens-ki-i-helse-o

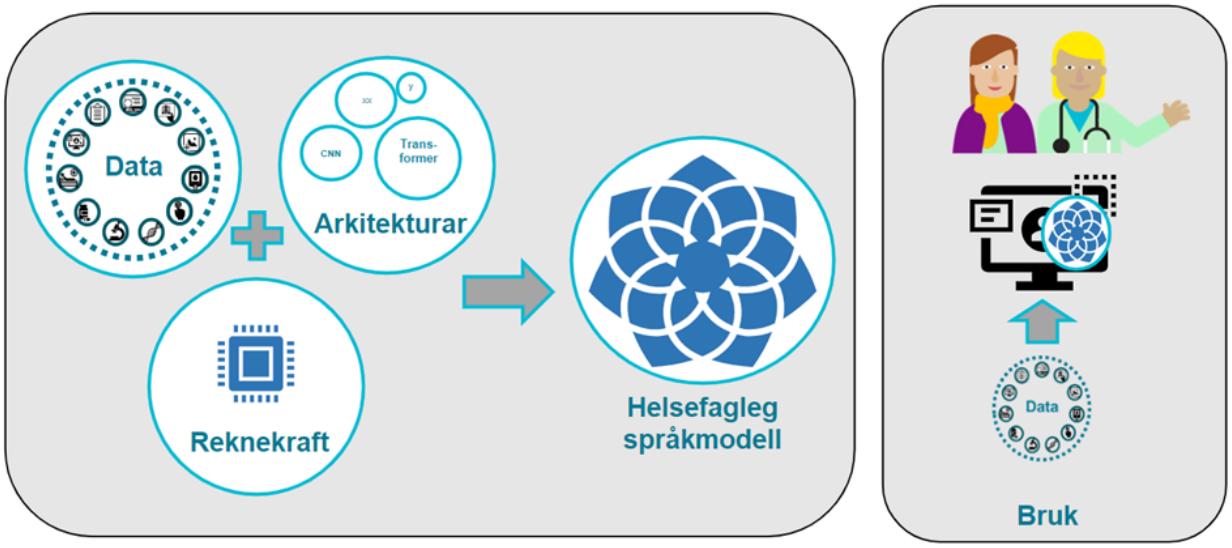
Om store språkmodellar

Kva er ein stor språkmodell?

Ein stor språkmodell (*large language model* (LLM)) er ein statistisk modell av eit språk som nyttar seg av ulike teknikkar innan fagområda KI og språkteknologi. Modellen er ei sannsynsfordeling over sekvensar av ord, og han kan brukast til å analysere og generere tekst basert på naturleg språk[4]. Ein stor språkmodell blir trena på enorme mengder tekstdata som innputt (vanlegvis frå internett, bøker og artiklar) som blir prosessert ved hjelp av reknekraft og maskinlæringsarkitektur[5], sjå figur 1. Språkmodellar er i stand til å kjenne att mønster i språket og dimed tolke, skape og føreseie nytt innhald. Omgrepet "stor" viser hovudsakleg til talet på parameterar og omfanget av data han er trena på. Parameterar i denne samanhengen er dei interne variablane i modellen som blir justerte gjennom treningsprosessen for å læra seg mønster i språkdata han er eksponert for.

Ein trena språkmodell blir brukt som del av ein applikasjon eller ei nettside, og kan brukast for eksempel på ein PC eller i ein mobiltelefon. I applikasjonen/nettsida blir data frå brukssituasjonen brukt som innputt.

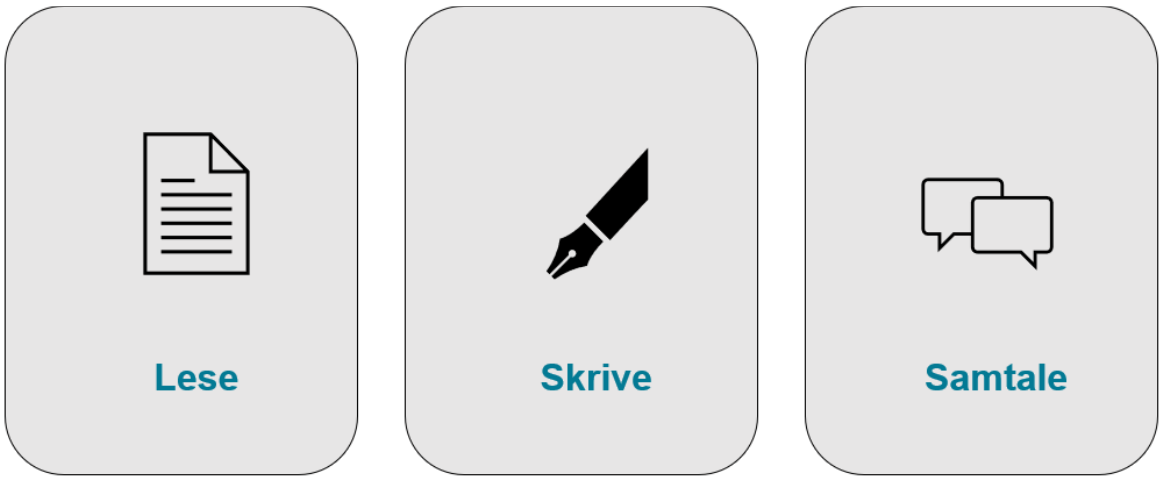
Slike store språkmodellar er generelle, dvs. dei er ikkje knytt til eit bestemt fag- eller bruksområde. Dei blir difor ofte kalla for grunnmodellar (*foundation models*).



Figur 1: Utvikling og trening av språkmodellar

Den såkalla transformer-teknologien[6] er ein type arkitektur som har ført til eit paradigmeskifte for språkmodellar og kunstig intelligens, særleg i kjølvatnet av den publiserte artikkelen «Attention Is All You Need» av Vaswani o.a. i 2017. Språkmodellar blei etter dette langt meir effektive og dei kunne nyttast til fleire oppgåver.

Ein kan skilje mellom generative og ikkje-generative språkmodellar. Ein generativ språkmodell kan i tillegg til å lese tekst, produserer *ny tekst*. Det det mest kjende dømet er GPT-modellen som ligg til grunn for ChatGPT. Andre store, internasjonale generative språkmodellar er franske Mistral, tyske Aleph Alpha og Gemini og LLaMA frå amerikanske Google og Meta.



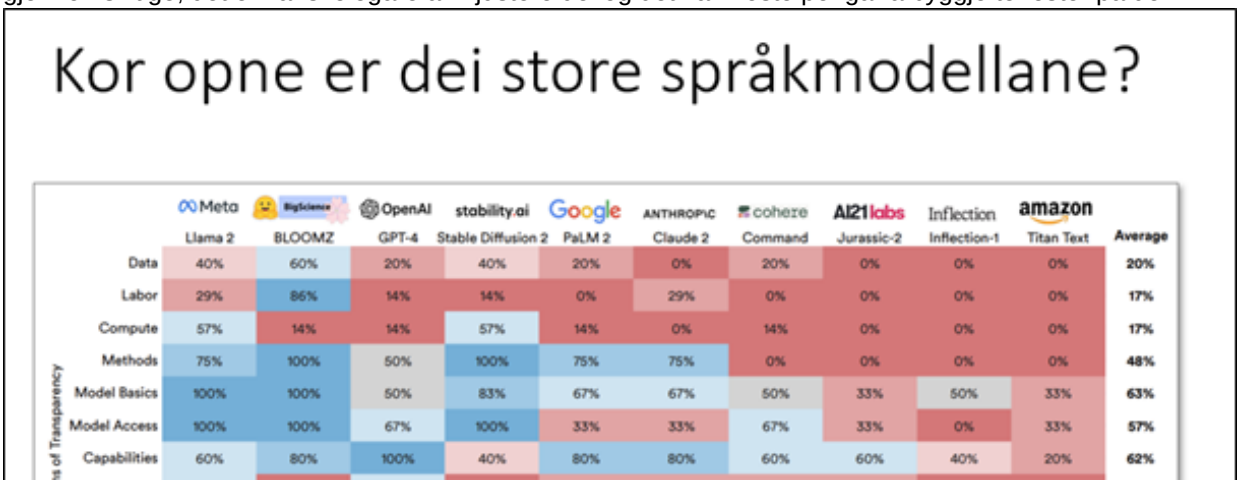
Figur 2: Brukseigenskapar til generative språkmodellar

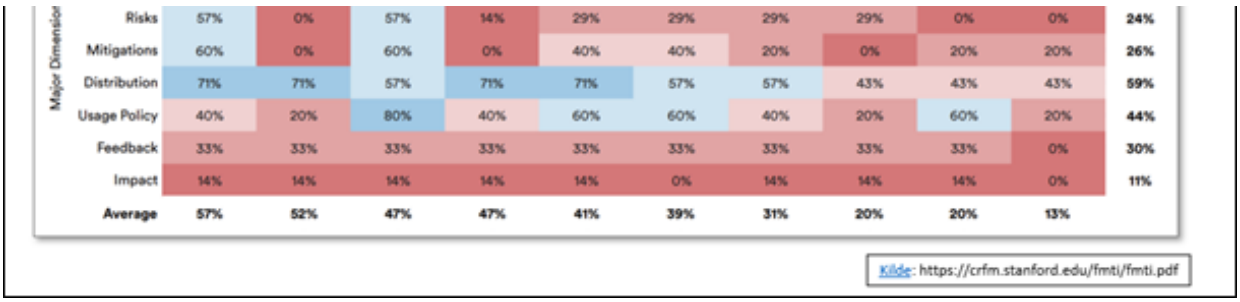
Ikkje-generative språkmodellar er annleis ved at dei leser og analyserer innputtdata til t.d. kategorisering, men dei produserer ikkje ny tekst. Dei såkalla BERT-modellane er eit døme på det. Ein generativ språkmodell vil krevje langt større mengder innputtdata og reknekraft enn ein ikkje-generativ språkmodell. I tillegg finst det ein tredje type språkmodell som kombinerer desse to modellane: den såkalla enkodar-dekodar-modellen (t.d. T5-modellen). Han kombinerer generative og ikkje-generative eigenskapar og kan brukast til dømes til kategoriseringsarbeid og maskinomsetjing.

Det finst i dag fleire ferdigutvikla store språkmodellar basert på BERT-teknologien[7] i Noreg. I tillegg finst det to initiativ frå etablerte fagmiljø til å etablere ein stor norsk generativ språkmodell: Norwegian Artificial Intelligence Research Consortium (NORA) og Norwegian Research Center for AI Innovation (NorwAI). NorwAI testar ut to ulike tilnærmingar, både gjennom å byggje opp ein norsk språkmodell frå botn og gjennom å vidareutvikle internasjonale modellar (både amerikansk og europeisk arkitektur). NORA har søkt Forskningsrådet om å setje i gang med å byggje ein norsk språkmodell frå botn frå 2025. Begge initiativa legg opp til finjustering for helseføremål (sjå under). I tillegg finst det private initiativ, som t.d. Norsk GPT[8].

Språkmodellar kan vere unimodale (dvs. trena på éin type data, t.d. tekst) eller multimodale (dvs. trena på fleire typar data, t.d. tekst, tale, bilete og genomdata). Multimodale modellar har fått større merksemd i det siste, også på helsefeltet[9]. Det gjer det mogleg å nytte ulike typar data for t.d. avgjerdsstøtte (beslutningsstøtte), jf. kap. 3. NORA-initiativet har som mål å undersøkje multimodalitet i ein mindre skala.

Språkmodellar kan vere opne eller lukka. Opne språkmodellar har dokumentert og gjort programmeringskodar tilgjengelege. Dei er enklare å finjustere utan avtalar med eigarane av modellane og det kostar ofte ingenting å byggje tenester på dei. Lukka språkmodellar er det er derimot lite gjennomsiktige, det er vanskelegare å finjustere dei og det kan koste pengar å byggje tenester på dei.





Figur 3: Grad av openheit i dei store, internasjonale språkmodellane[10]

[4] https://snl.no/spr%C3%A5kmodell

[5] Slike maskinlæringsmetodar omfattar til dømes maskering/skjuling av ord i setningar (*masked language modelling*) eller prediksjon av neste setning (*next sentence prediction*).

[6] https://medium.com/@amanatulla1606/transformer-architecture-explained-2c49e2257b4c

[7] Språkteknologigruppa (LTG) ved Universitetet i Oslo har utvikla NorBERT-modellen som kjem i fleire storleikar og som er open tilgjengeleg, i tillegg til ein norsk T5-modell. Vidare har AI-labben ved Nasjonalbiblioteket utvikla NB-BERT, som også er open tilgjengeleg.

[8] https://norskgpt.com/

[9] https://www.nature.com/articles/s41586-023-05881-4

[10] https://crfm.stanford.edu/fmti/

Språk og kultur i språkmodellar

For at ein språkmodell skal fungere i eit språksamfunn, må han vere trena på store tekstmengder på det aktuelle språket. Sjølv om innhaldet i ein internasjonal språkmodell automatisk blir omsett til norsk, vil det kunne oppstå framande ord og språklege vendingar, jamvel feilomsetjingar. Stil, tone og høflegheit varierer mellom ulike språk og kulturar, og det kan slå ut på uønskt måte ved bruk av éin internasjonal språkmodellar. I tillegg er det grunn til å tru at omsetjingar av medisinsk tekst vil by på utfordringar på grunn av svært domenespesifikt og spesialisert språk. Det gjeld især fagterminologi, sjargong, forkortingar og andre kjenneteikn ved språkbruk i t.d. journalføring.

Trening på store mengder norskspråklege tekstar vil i tillegg gje fag- og kulturrelevant innhald til språkmodellane, t.d. nasjonale behandlingsplanar på helsefeltet. Erfaringane frå uttesting av KI-modellen IBM Watson i Danmark for ein del år sidan synte problem med at modellen gav råd basert på amerikanske kreftbehandlingsplanar, som var annleis enn dei danske[11].

Sjølv om internasjonale språkmodellar som t.d. Med-PaLM og ChatGPT gjer det bra i internasjonale valideringar (med t.d. spørsmål og svar-sett), er det ingen garanti for at modellen fungerer i norsk helsefagleg kontekst.

Alternativet mellom å byggje ein norsk språkmodell frå botn eller å bruke ein internasjonal grunnmodell blir for tida testa ut og vurdert i NorwAI-satsinga (NTNU) og ved språkteknologigruppa (Language Technology Goup (LTG) ved UiO som pilot for NORA-satsinga. Resultatet vil vonleg vere klart i løpet av våren 2024.

[11] Barua, Ishita (2023) *Kunstig intelligens redder liv*. Cappelen Damm.

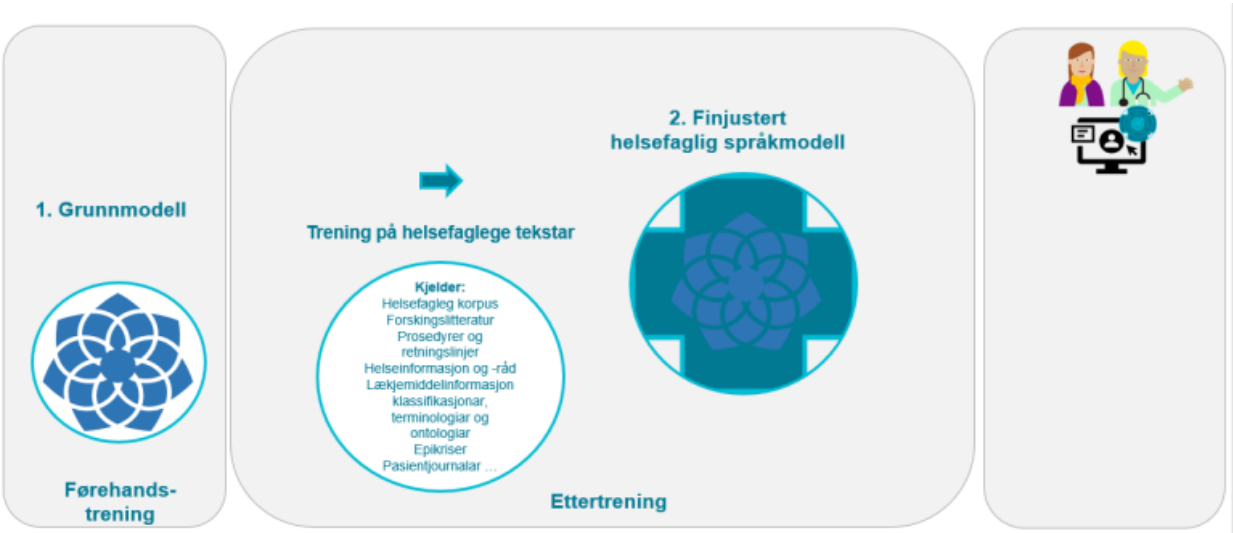
Ettertrening og helsefagleg rettleiing av språkmodellar

For at ein stor språkmodell kan nyttast i helse- og omsorgssektoren, kan det vere at han må tilpassast helseområdet. Ettertrening (*post-training*) kan t.d. gjerast ved å (fin)justere ein grunnmodell etter at den er førehandstrent, og/eller ved å rettleie ein modell mot ønska oppførsel. Vi går gjennom nokre tilnærmingar for å ettertrene og rettleie store språkmodellar på helseområdet i dei påfølgjande delkapitla.

1.3.1 Finjustering med helsefagleg innhald

Ein førehandstrena grunnmodell kan omfatte for mykje allmennspråkleg innhald. Språkmodellen kan difor med fordel tilpassast eit spesifikt fagområde, til dømes helsefag. Slik tilpassing vert kalla for finjustering (*fine tuning*). Det skjer ved at språkmodellen blir tilført nye *helsefaglege tekstar* som ny innputt og ein ny prosess med maskinlæring blir gjennomført, jf. figuren under. Heile eller delar av grunnmodellen vil då bli trena på helsefagleg terminologi, forkortingar, sjargong og så bortetter. for å kunne vere i stand til å handsame tekstar frå spesielle bruksområde, t.d. kliniske tekstar. Ein finjustert helsefagleg språkmodell vil difor vere i stand til å «forstå» fagspråket til helsepersonell betre.

I Noreg er Helse Vest IKT i ferd med å utvikle ein finjustert helsefagleg språkmodell basert på BERT-teknologien. Modellen skal etter planane vere ferdig i løpet av 2024.



Figur 4: Finjustering med helsefagleg innhald

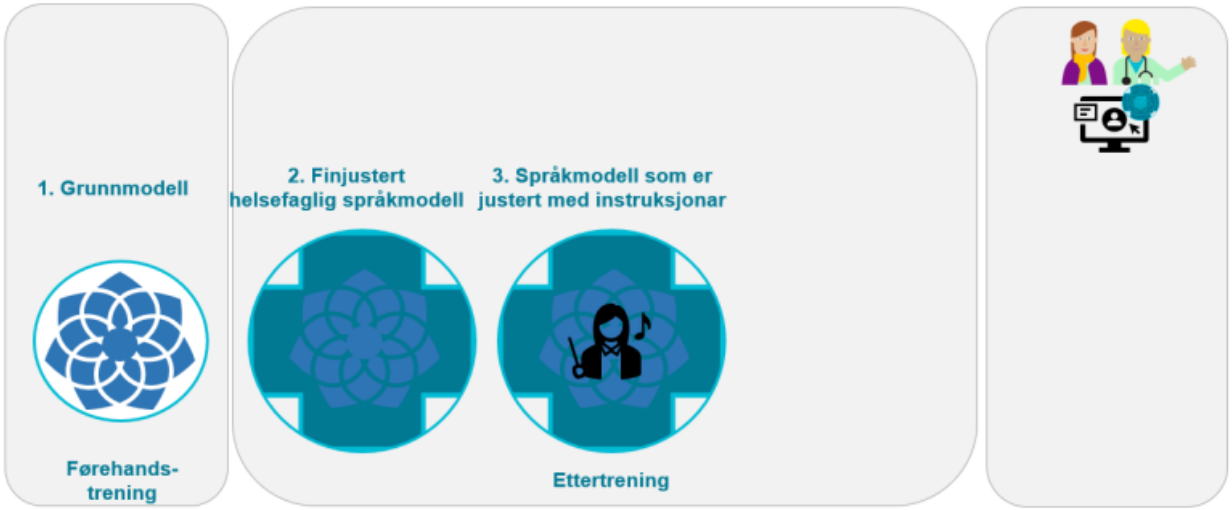
Eit nytt områda i KI er finjustering av små språkmodellar (*small language models* (SLM)), t.d. Microsofts Phi-2, Googles Gemini Nano. Desse har langt færre parameterar enn dei store språkmodellane og er difor langt mindre ressurskrevjande. Dei kan jamvel nyttast på mobiltelefonar utan å vere kopla til internett. Små språkmodellar kan bli finjusterte med helsefaglege tekstar, men då til ganske spesifikke bruksområde pga. den relativt avgrensa modellen.[12][13]

1.3.2 Instruksjonsjustering mot helsefaglege oppgåver

Ein språkmodell som er finjustert mot helsefag er enno ikkje nødvendigvis tilpassa ulike bruksføremål, altså til å utføre spesifikke oppgåver. Slik tilpassing må gjerast med meir trening knytt til bruksføremålet.

Instruksjonsjustering (*instruction fine tuning*) av ein KI-modell er ein prosess der ein tilpassar ein allereie trena modell til å utføre nye, spesifikke oppgåver eller forstå nye instruksjonar betre. Dette blir gjort ved å finjustere modellen med eit mindre datasett som er direkte relevant for dei nye oppgåvene eller instruksjonane.

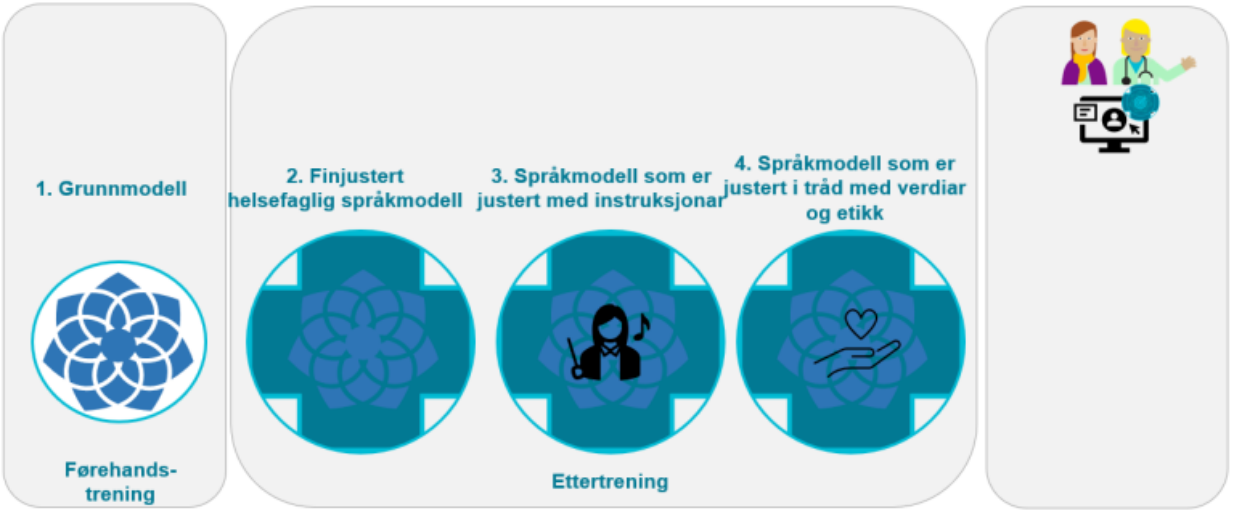
Dette kan gjerast ved å samle inn ei stor mengd innputtdata og utputtdata, og deretter påskjøne (belønne) modellen for å generere ønska utputt. I denne påskjøningsprosessen blir parametrane i modellen endra. Til dømes kan dette vere pasientjournalnotat som innputtdata og kodar, t.d. ICD-10-kodar, som utputtdata.



Figur 5: Instruksjonsjustering mot helsefaglege oppgåver

1.3.3 Justering i tråd med verdiar og etikk

Ein språkmodellen kan også bli justert i tråd med menneskelege verdiar, etiske retningslinjer, prinsipp og/eller målsettingar (*alignment*). Til dømes kan ein samtalerobot bli utvikla til å vere hjelpsam, ærleg og harmlaus for helseføremål. Ein modell kan til dømes lære menneskelege verdiar ved å trene på datasett som reflekterer desse verdiane. Slik justering og forbetring av åtferda til modellen kan også skje gjennom interaksjon med menneske.



Figur 6: Justering i tråd med verdiar og etikk

1.3.4 Rettleiing med helsefagleg kunnskap

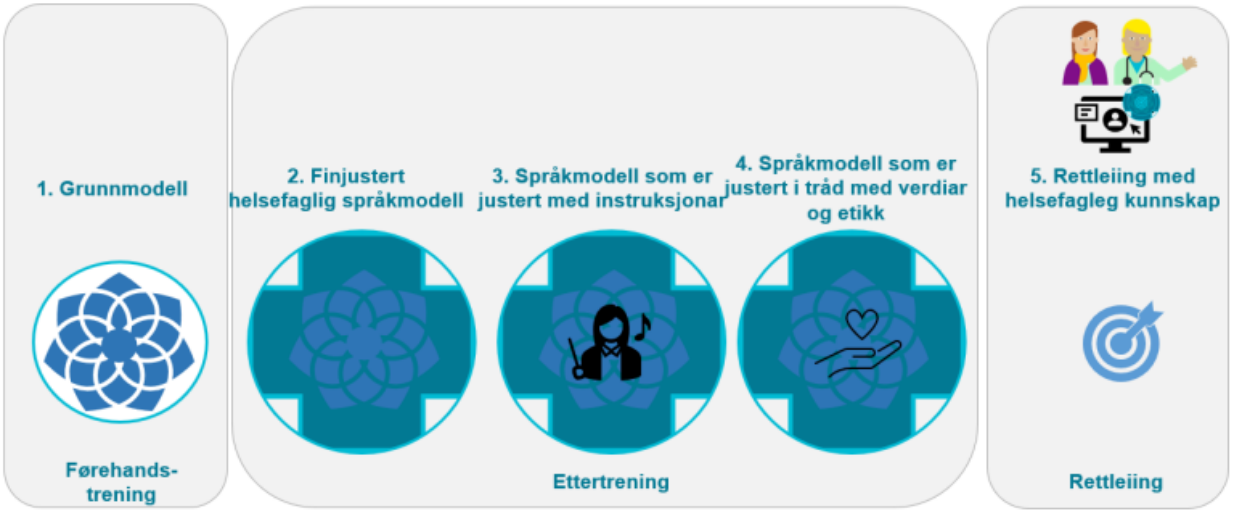
Utvikling av ein språkmodell til eit ønska bruksføremål i helse- og omsorgstenesta kan også skje i den konkrete applikasjonen. Dette steget gjev ei spesifikk rettleiing av språkmodellen slik at han utfører oppgåver på ein tilfredsstillande måte basert på helsefagleg kunnskap.

Denne tilnærminga endrar ikkje på sjølve språkmodellen, men gir heller ei slags rettleiing for korleis reisa gjennom det nevrale nettverket skal gå. Det betyr at teknikken kan brukast på både lukka modellar (som t.d. GPT-4) og opne modellar. Dette skil seg frå dei ulike stega i ettertreninga, der ein gjer faktiske endringar i KI-modellen, noko som er enklast å få til på opne modellar.

Rettleiing kan gjerast anten ved til dømes instruksjonsutforming (*prompt engineering*) eller kopling til ekstern, relevant kunnskap (*grounding*).

Instruksjonsutforming kan skje gjennom ulike formar for instruksjonar, t.d. direkte instruksar (*zero-shot prompting*), instruksar med eitt døme (*one-shot prompting*) eller fleire døme (*few-shot prompting*).[14] Gjennom nøye utforma spørsmål eller kommandoar kan ein språkmodell rettleiast til å generere den ønska typen respons. Dette kan skje manuelt eller gjennom bruk av ferdige sett av instruksjonsdata som effektiviserer arbeidet.

Rettleiing kan også skje ved kopling til ekstern, relevant kunnskap om til dømes medisin og helse i form av filer, lenkjer, gjennom API-ar (programmeringsgrensesnitt) og liknande innputt. I slik rettleiing kan ein leggje til rette for såkalla søkjeforsterka generering (*Retrieval-Augmented Generation* (RAG)) for å gje eit meir treffsikkert resultat[15]. Med andre ord skjer det ei slags kunnskapskuratering der ein vel ut nokre tekstar som er ekstra kvalitetssikra og autoritative som språkmodellen skal nytte. Desse tekstane blir indekserte og tilpassa ved hjelp av eigne verktøy. Slik sikrar ein at bruken av språkmodellen er i tråd med den etablerte kunnskapen på bruksområdet. Bruk av RAG kan på denne måten løyse problem som t.d. hallusinasjon ved bruk av språkmodellar eller risiko for utdatert kunnskap.



Figur 7: Rettleiing med helsefagleg kunnskap

[12] https://medicalfuturist.com/using-chatgpt-offline-the-emergence-of-small-language-models/

[13] https://medium.com/@soulawalid/how-to-train-small-language-model-for-healthcare-a-step-by-step-guide-c

[14] https://www.promptingguide.ai/techniques/zeroshot

[15] https://www.promptingguide.ai/techniques/rag

Strategiar for ettertrening og rettleiing av språkmodellar

Utviklinga av ein stor språkmodell kan skje på ulike måtar, ikkje nødvendigvis steg for steg frå ein grunnmodell. Ein kan til dømes hoppe over ettertreninga og berre gjere den helsefaglege rettleiinga rett på ein grunnmodell.

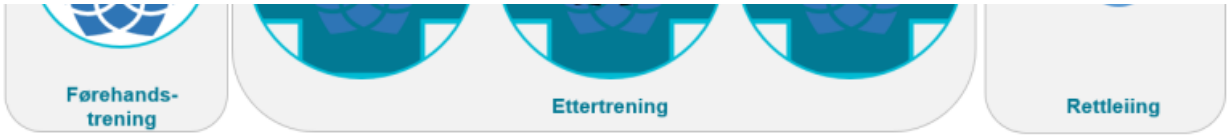
KI-teknologien er ny, og ein veit enno ikkje kva for ein strategi som er mest føremålstenleg for å utvikle ein språkmodell og nytte han på forsvarleg vis i den norske helse- og omsorgstenesta.

Under skisserer vi moglege strategiar for korleis ein førehandstrena grunnmodell kan tilpassast best mogleg bruksområdet og applikasjonen i helse- og omsorgssektoren. Vi skisserer fire moglege steg (alternativ 1, 2, 3 og 4). Vi går igjennom dei ulike alternativa i dei påfølgjande underkapitla med ei oppsummering i form av ein tabell til slutt.

1.4.1 Alternativ 1: Finjustering, instruksjonsjustering, justering i tråd med verdiar og etikk og helsefagleg rettleiing av ein språkmodell

Alternativ 1 inneber at språkmodellen blir utvikla ved å gå gjennom alle stega som er beskrive i kapitelet ovanfor.

Strategien er potensielt meir treffsikker enn dei andre alternativa sidan han er meir finjustert på ulike måtar og rettleia med helsefagleg innhald. Samstundes er han den mest omstendelege og ressurskrevjande av alle og vil truleg ta lengst tid å utvikle. Behovet for treningsdata er ganske omfattande. Han vil også truleg ha det høgaste miljøavtrykket av alle alternativa.

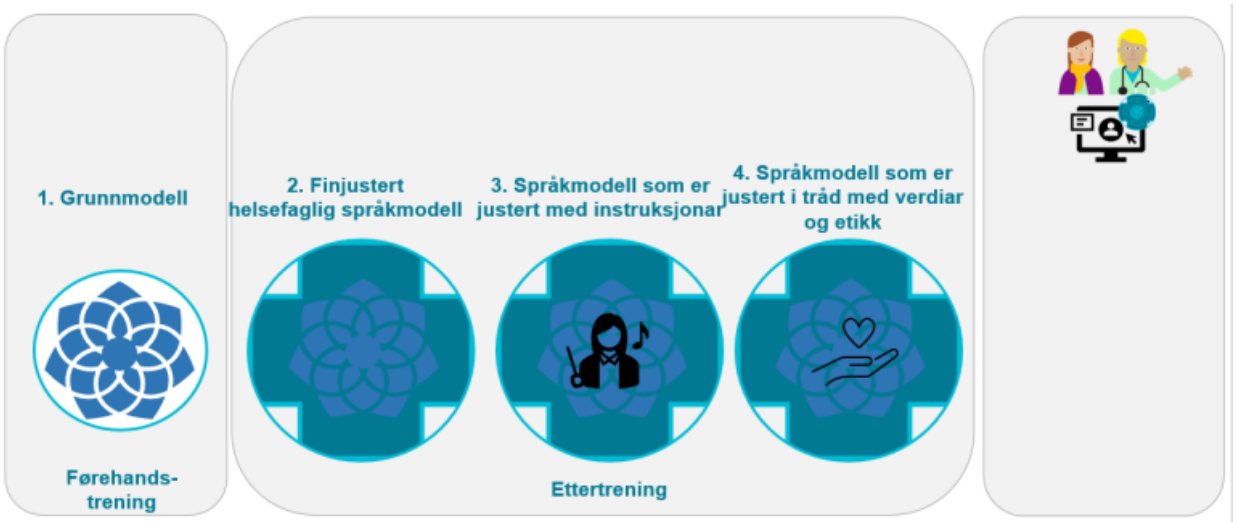


Figur 8: Alternativ 1: Finjustering, instruksjustering, justering i tråd med verdiar og etikk i tillegg til rettleiing med helsefagleg kunnskap av ein språkmodell

1.4.2 Alternativ 2: Finjustering, instruksjonsjustering, justering i tråd med verdiar og etikk av ein språkmodell

Alternativ 2 inneber at språkmodellen blir finjustert, instruksjonsjustert og justert i tråd med verdiar og etikk, men ikkje rettleia. Det kan skuldast at det til dømes ikkje er behov eller moglegheit for å leggje til spesifikke helsefaglege kunnskapsressursar.

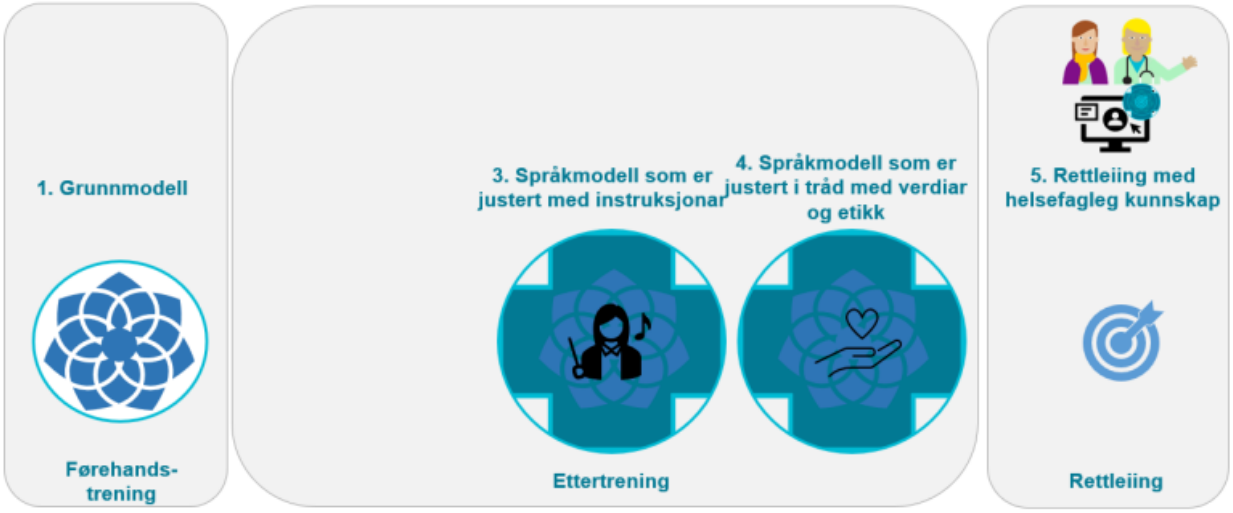
Alternativet vil potensielt ha god treffsikkerheit og vil moglegvis vere lettare å ta i bruk enn alternative under, men vil framleis ha ein del ulemper knytt til ressursbruk, behov for treningsdata og miljøavtrykk. Det er også mogleg at modellen kan innehalde utdatert kunnskap sidan det siste steget manglar.



Figur 9: Alternativ 2: Finjustering, instruksjustering og justering i tråd med verdiar og etikk av ein språkmodell

1.4.3 Alternativ 3: Instruksjonsjustering, forankring og rettleiing av ein språkmodell

Alternativ 3 inneber at ein hoppar over finjusteringa av ein helsefagleg språkmodell. Utviklinga skjer då ved instruksjustering, justering i tråd med verdiar og etikk og helsefagleg rettleiing. Ein slik strategi er teknologisk ressurseffektiv fordi ein hoppar over finjusteringa. Ein treng difor mindre treningsdata til utviklinga av modellen, noko som inneber kortare tid å utvikle og mindre behov for å prosessere data. Den helsefaglege rettleiinga i form av t.d. RAG, sikrar oppdatert kunnskap i modellen. Samstundes flyttar ein utviklingsbyrden over på den helsefaglege rettleiinga, som dimed inneber større grad av helsefagleg kompetanse og kapasitet i tillegg til KI-kompetanse.

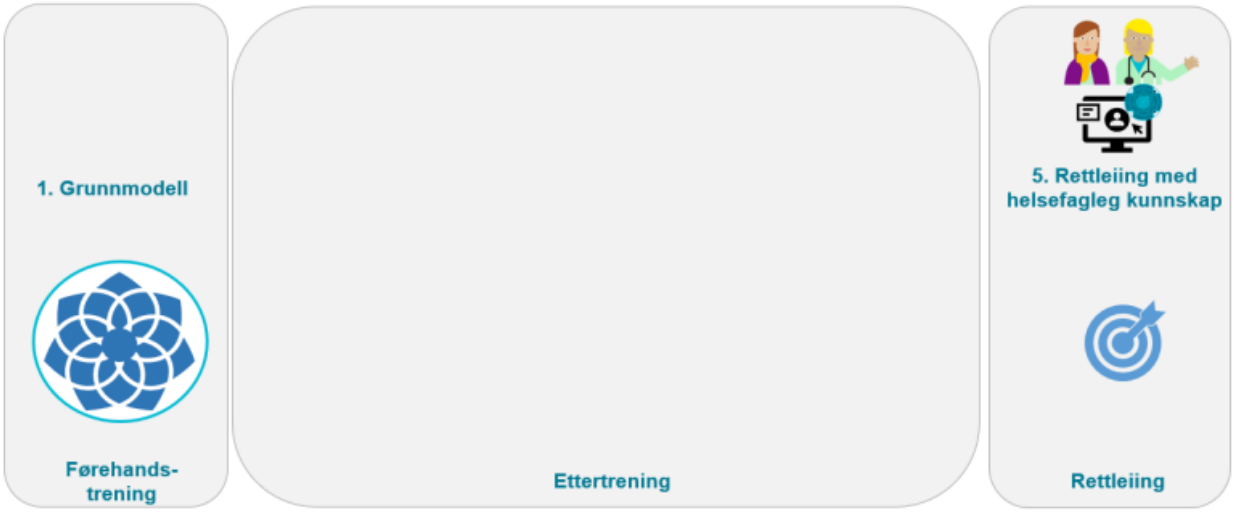


Figur 10: Alternativ 3: Instruksjustering, justering i tråd med verdiar og etikk og rettleiing av ein språkmodell

1.4.4 Alternativ 4: Berre helsefagleg rettleiing av ein grunnmodell

Alternativ 4 er den enklaste, raskaste og meste kostnadseffektive strategien for å ta i bruk ein språkmodell. I dette tilfellet vil ein ta utgangspunkt i ein førehandstrena grunnmodell og tilpassar han til bruksføremålet gjennom rettleiing. Det kan skje ved at ein nyttar ein stor, open grunnmodell og legg til t.d. kuratert fagkunnskap som blir indeksert og nytta i den konkrete applikasjonen (RAG).

Fleire kjelder peiker på at dette alternativet vil vere kostnad- og tidssparande samstundes som at miljøavtrykket vil vere mindre. Ein bør difor vurdere om dette alternativet er godt nok til bruksføremålet før ein set i gang med ulike former for ettertrening.



Figur 11: Alternativ 4: Berre helsefagleg rettleiing av ein grunnmodell

1.4.5 Oppsummering av strategiane

Strategiane ovanfor kan oppsummerast slik:

	Fordelar	Ulemper
Alternativ 1: Finjustering, instruksjustering, justering i tråd med verdiar og etikk og helsefagleg rettleiing av ein språkmodell	Potensielt betre treffsikkerheit og betre tilpassa helsetenesta enn andre alternativ	- Teknologsk ressurskrevjande - Treng mykje treningsdata - Tidkrevjande å utvikle - Høgast miljøfotavtrykk
Alternativ 2: Finjustering, instruksjustering og justering i tråd med verdiar og etikk av ein språkmodell	-Potensielt god treffsikkerheit -Lettare å ta i bruk	- Teknologisk ressurskrevjande - Treng mykje treningsdata - Tidkrevjande å utvikle - Høgt Miljøfotavtrykk -Potensielt utdatert kunnskap
Alternativ 3: Instruksjustering, justering i tråd med verdiar og etikk og helsefagleg rettleiing av ein språkmodell	- Teknologsk ressurseffektiv - Treng lite treningsdata - Rask å utvikle - Mindre miljøfotavtrykk -Potensielt god treffsikkerheit	-Krev meir kompetanse i KI og helsefag enn andre alternativ
Alternativ 4: Berre helsefagleg rettleiing av ein språkmodell	**-** Teknologsk svært ressurseffektiv - Rask å utvikle - Mindre miljøfotavtrykk -Potensielt god treffsikkerheit	• Avgrensa bruksområde • Krev KI og helsefagleg kompetanse • Krev kvalitetssikring av treningstekt

Alternative forvaltningsregime av store språkmodellar

Ein god stor språkmodell krev mykje data og reknekraft, og er difor dyr å både utvikle og vedlikehalde. Samstundes kan dei store kostnadane ved ein grunnmodell rettferdiggjerast sidan han kan brukast til fleire føremål.

I ei tidleg utforskings- og testfase er språkmodellane nokså umodne. I Noreg skjer testing, utvikling, ettertrening og rettleiing lokalt ved dei ulike sjukehusa eller helseføretaka – i tillegg til forskingsinstitusjonar.

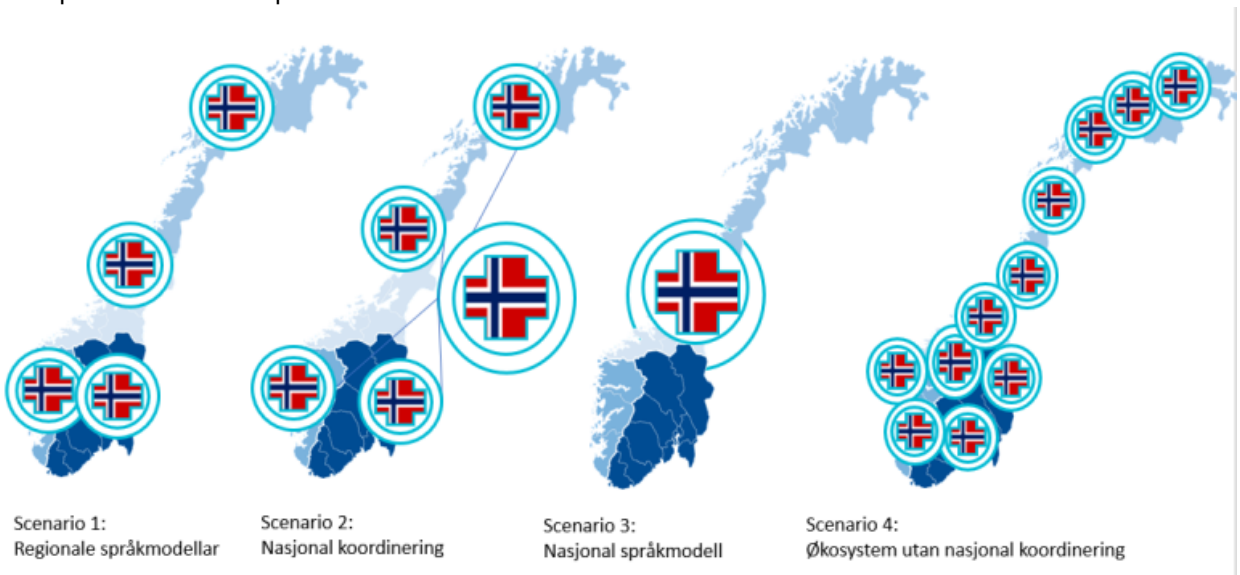
På lenger sikt når språkmodellane når eit visst modenskap, trengst ein diskusjon om korleis dei skal forvaltast og vidareutviklast: Skal forvaltinga eller vidareutviklinga skje av ein eller fleire kommersielle aktørar som tilbyr slike språkmodellar? Eller skal språkmodellar koordinerast på regionalt eller nasjonalt nivå i offentleg sektor? Kven skal i så fall forvalte språkmodellane? Eit nasjonalt organ eller eit helseføretak på vegner av heile sektoren? Eller skal ikkje forvaltinga koordinerast i det heile? Blir det opp til kvar einskild organisasjon å utvikle og forvalte sin eigen bruk av språkmodellar med utgangspunkt i eigne behov og ressursar?

Ein kan sjå for seg ulike scenario for dette, sjå figur under.

1. ***Regionale språkmodellar:*** Kvar helseregion forvaltar sin eller sine språkmodellar.

2. ***Nasjonal koordinering:*** Trening og (fin)justering av språkmodellar blir koordinert på nasjonalt nivå gjennom t.d. ei tilnærming som føderert læring. Det kan bli særleg aktuelt dersom personvern gjer at treningsdata ikkje kan overførast mellom ulike datasystem, institusjonar og nivå i helsetenesta.[16].
3. ***Nasjonal språkmodell:*** Helse- og omsorgssektoren forvaltar éin felles, nasjonal språkmodell.
4. ***Økosystem utan nasjonal koordinering:*** Ei rekkje ulike språkmodellar blir implementert og forvalta på ulike bruksområde i eit økosystem utan koordinering på nasjonalt eller regionalt nivå.

Korleis språkmodellar skal forvaltast i primærhelsetenesta står også att som eit viktig spørsmål. Det store talet på aktørar kan skape meir innfløkte scenario.



Figur 13: Ulike scenario for forvalting av språkmodellar

Det trengst nærmare kunnskap om moglege scenario og vurdering av fordelar og ulemper ved dei. Dette vil bli vurdert nærmare i arbeidet med felles KI-plan for helse- og omsorgstenesta.[17]

[16]
https://medium.com/@TheMAGIC/openfedllm-training-large-language-models-on-decentralized-private-data

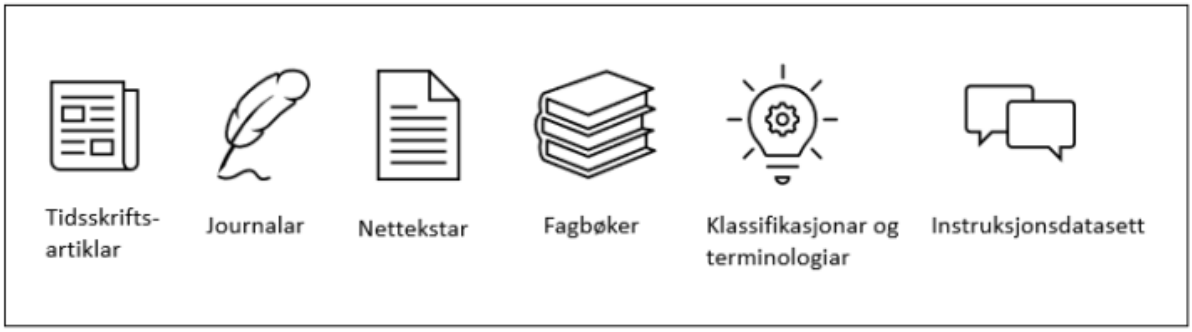
[17] TB2024-34: Kunstig intelligens i helse- og omsorgstjenesten:
https://www.regjeringen.no/contentassets/d8f63d7d01d64def982cb7c8ce1eeb64/tildelingsbrev-til-helsedirek

Treningsdata

Tilgang til store mengder språkdata til trening og finjustering av språkmodellar er avgjerande for at dei skal fungere godt. I ein ny studie av språkmodellar for elektroniske pasientjournalar vert det hevda at "*[o]ne primary limiting factor for obtaining high quality predictions is limited data*"[1]. Like viktig som kvantitet er kvalitet: Språkdata må vere aktuelle (*timeliness*) og komplette (*completeness*), dvs. representative og balanserte. Eit anna viktig moment er opphavet til tekstane: Kjem tekstane frå autoritative, verifiserte kjelder eller er det snakk om automatisk omsette nettsider, til tider av låg kvalitet?

I tillegg til autentiske tekstar trengst det også eigne instruksjonsdata og valideringsdata for å trene og validere språkmodellar.

Det finst ulike typar tekstar som blir brukt eller kan brukast som treningsdata til helsefaglege språkmodellar, sjå figur under.



Figur 14: Typar tekst til trening av helsefaglege språkmodellar
[18] https://www.sciencedirect.com/science/article/pii/S1532046420302653

Helsefaglege tidsskriftsartiklar

> Internasjonale helsefaglege språkmodellar er ofte trena på vitskapelege, biomedisinske tekstar, som PubMed. PubMed er eksempelvis ein internasjonal portal med ein søkjemotor som indekserer samandrag av biomedisinske artiklar. Mange av samandraga lenkjar vidare til fulltekstartiklar, som kan vere fritt tilgjengelege eller krevje abonnement. 16 millionar abstrakt og 5 millionar heile artiklar er tilgjengelege i eit eige datasett for trening av store språkmodellar[19].
>
> I Noreg blir norskspråklege medisinske artiklar publisert i tidsskrift som t.d. Tidsskrift for Den norske legeforening, Kirurgen, Onkonytt, Sykepleien o.a.. Truleg vil også oppgåver og avhandlingar på norsk frå Universitetet i Oslo (digitale utgjevingar ved UiO (DUO)), Universitetet i Bergen (Bergen Open Access Archive (BORA)) og andre universitet vere moglege kjelder. Det er avgjerande at spørsmålet om opphavsrettar og godkjenning av bruk er avklart for alle desse kjeldene.
>
> [19] https://en.wikipedia.org/wiki/The_Pile_(dataset)

Journaldokument

Nokre internasjonale språkmodellar er trena på journalnotat, både strukturerte og ustrukturerte. Dei er som regel er trena på eit allment tilgjengeleg datasett frå Beth Israel Deaconess Medical Center i Boston, USA (MIMIC-III-datasettet). Dette innhaldet baserer seg på anonymiserte data frå ca. 38 500 pasientar før år 2012 i radiologirapportar og epikriser. Datasettet representerer difor eit utsnitt av faktisk helsefagleg språkbruk. Nokre andre helsefaglege språkmodellar hentar data frå Massachusetts Institute of Technology (eICU-datasettet, frå ca. 139 000 pasientar) og Clinical Practice Research Datalink (longitudinelle data (data over ei lengre periode) frå 7 % av alle pasientar i Storbritannia).

I Noreg har ei lovendring opna for at Helsedirektoratet kan gje dispensasjon frå teieplikta (tausheitsplikten) for å tilgjengeleggjere helseopplysingar frå pasientjournalar når føremålet er utvikling av avgjerdsstøtteverktøy (beslutningsstøtteverktøy) basert på kunstig intelligens, og til å ta slike verktøy i bruk i klinikk, jf. helsepersonellova § 29.

Generelt kan ein seie at det må gjerast ei konkret vurdering av kvart enkelt datasett. Dette må gjerast av dataansvarleg før tilgjengeleggjering.

Det er etablert ei tverretatleg rettleiingsteneste med juristar frå Helsedirektoratet, Direktoratet for medisinske produkt (DMP) og Helsetilsynet slik at prosjekt som har spørsmål knytt til kunstig intelligens kan få svar på spørsmål etter fleire regelverk samstundes: https://www.helsedirektoratet.no/tema/kunstig-intelligens/tverretatlig-veiledningstjeneste

Døme er Helse vest IKT som nyttar journaldokument frå Helse vest HF i trening av sin språkmodell. Journaldokument har også blitt brukt i ClinCode-prosjektet hos Nasjonalt senter for e-helseforsking, Universitetssjukehuset i Nord-Norge (UNN) og DIPS AS.

Spørsmål og svar-tekstar

Autentiske tekstar frå spørsmål og svar-tenester kan også vere nyttige treningsdata for språkmodellar, særleg om språkmodellen skal nyttast til slike tenester. Det kan t.d. dreie seg om innsende spørsmål til fagstyresmakter med påfølgjande svar på helserelaterte spørsmål. I tiltaket Enklare tilgang til informasjon (ETI)[20] blir spørsmål og svar-tekstar nytta som treningsdata (sjå lenger nede under bruksområde).

Slike typar treningsdata må anonymiserast for å kunne brukast til dette føremålet.

[20] https://alvorligsyktbarn.no/enklere-tilgang-til-informasjon

Helsefaglege nett-tekstar

Ei kjelde til treningsdata er helsefaglege autoritative nett-tekstar av høg kvalitet. Det kan eksempelvis dreie seg om retningslinjer og rettleiarar frå Helsedirektoratet, ulike metodebøker frå ulike spesialitetar, t.d. metodebok i handkirurgi frå Helse nord[21] og metodebok for beinbrot frå Oslo skadelegevakt[22] og felles nettløysing for spesialisthelsetenesta[23].

[21] https://www.unn.no/497fd8/siteassets/documents/metodeboker/metodebok-ved-handskader-og-handleddss

[22] https://skadelegevakten.no/bruddskader

[23] https://www.fnsp.no/

Digitaliserte helsefagbøker

Store mengder helsefaglege bøker er digitaliserte hos Nasjonalbiblioteket. Det kan dreie seg om alt frå pensumbøker i medisin og sjukepleie til formidlingsbøker til allmennheita. Innhaldet i bøkene er av høg kvalitet og vil potensielt kunne nyttast til å trene helsefaglege språkmodellar.

Kunnskapsdepartementet har gitt Nasjonalbiblioteket i oppdrag å vurdere verdien av materiale med opphavsrett ved trening av språkmodellar.[24] Arbeidet vil skje med i samarbeid med UiO og NorwAI, og eit forslag er venta i løpet av året.

[24] https://www.regjeringen.no/no/aktuelt/skal-undersoke-bruk-av-opphavsrettslig-beskyttet-materiale-i-trening-

Klassifikasjonar og terminologiar som språkdata

Helsefaglege klassifikasjonar (t.d. ICD-11) og terminologiar (t.d. SNOMED CT) er kunnskapsmodellar der omgrepa er strukturerte i hierarki med relasjonar som definerer innhaldet. Desse er i utgangspunktet språklege kjelder av høg kvalitet som ein kunne tenkje seg å nytte som datainnputt i språkmodellar. Det er likevel ei vesentleg innvending: Språkmodellar byggjer på konteksten i setningar. Denne logikken bryt med kunnskapsmodellar, som inneheld listar av termar som ikkje følgjer setningsstruktur, men eit semantisk hierarki. Ei liste over termar (utan setningar) kan difor skape støy og redusere kvaliteten i ein språkmodell.

Det er likevel mogleg at helsefaglege klassifikasjonar og terminologiar kan inngå som ein del av treninga og finjusteringa av språkmodellar, t.d. ved at relasjonen i hierarkiet vert uttrykt ved hjelp av naturleg språk (t.d. «kronisk lungebetennelse er ein type lungebetennelse», eller «kuldeurtikaria er ein type urtikaria forårsaka av låg temperatur» frå terminologien SNOMED CT).

Tekstar der klassifikasjonar og terminologiar blir nytta i naturlege setningar, t.d. tekstdefinisjonar, vil uansett kunne bli viktige. Ei anna kjelde kan vere kodingsreglar (for ICD-10/ICD-11). Dei vil innehalde fagtermar i naturleg språkleg kontekst.

Sjølv om slike ressursar kan vere vanskelege å bruke i treninga av ein grunnmodell eller finjusteringa av han, peiker nokre studiar på at kunnskapsmodellar (*knowledge graphs*) kan komplementere,og bli integrert i, språkmodellar. Difor bør klassifikasjonar og terminologiar vurderast som ein viktig ressurs.[25]

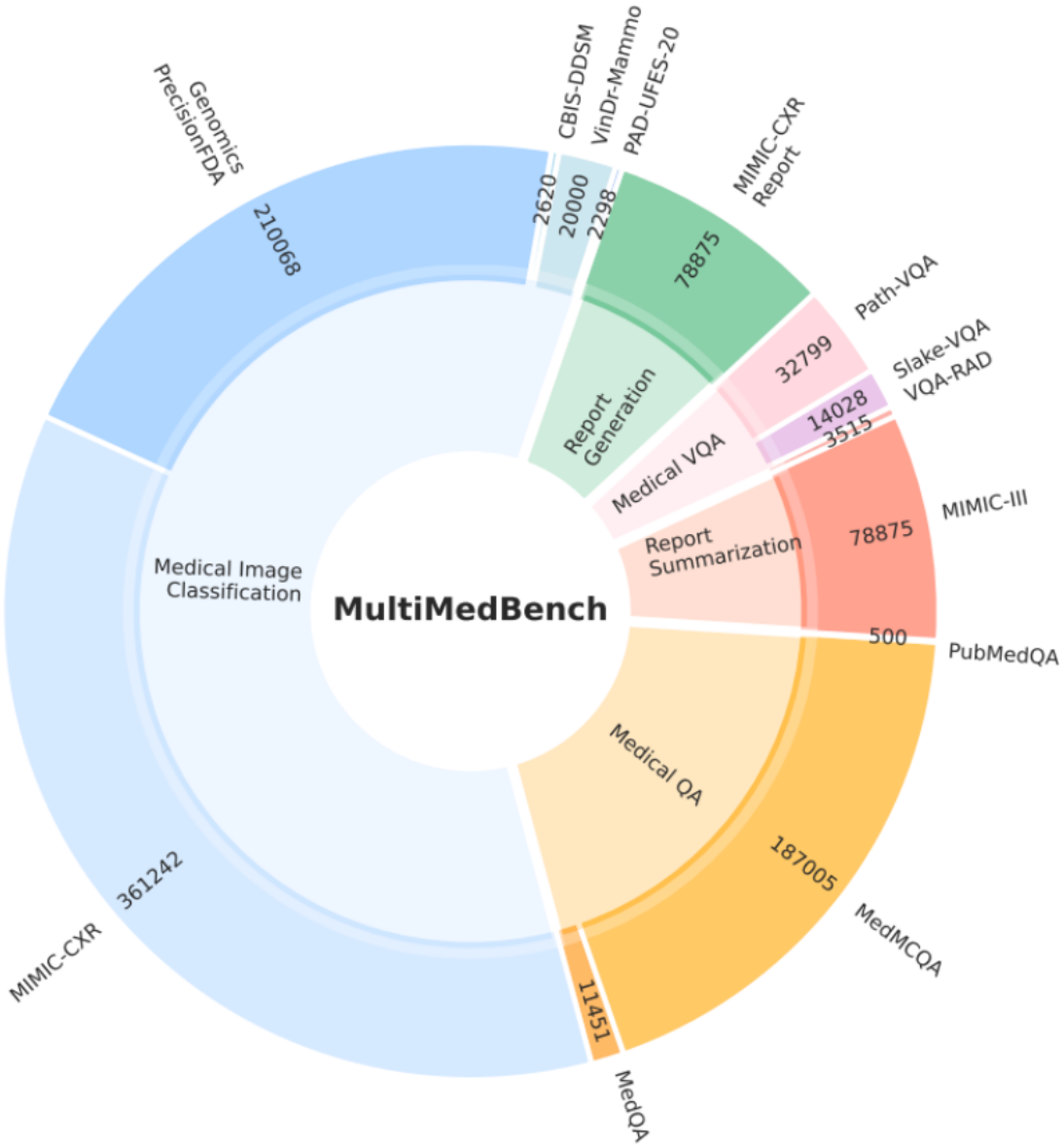
[25] https://www.nature.com/articles/s41586-023-05881-4

Instruksjonsdata

Ein særskilt type treningsdata er instruksjonsdata. Slike datasett blir brukt når ein trenar maskina til å gje forventa resultat basert på ein type innputt, t.d. spørsmål og svar ("Kva er hovudstaden i Frankrike"/"Paris" eller opne spørsmål med tilhøyrande svar). Bruken av instruksjonsdata skjer i steget som er kalla instruksjonsjustering (*instruction fine tuning*), sjå kap 1.2 ovanfor). Instruksjonsdata er difor datasett som inneheld både innputt og utputt.[26]

Det finst opne, internasjonale instruksjonsdatasett som t.d. Databricks Dollt 15K, som inneheld 15 000 menneskelaga instruksjonar (*promtps*) med tilhøyrande svar.

For bruk av språkmodellar i helse- og omsorgssektoren vil det truleg vere behov for instruksjonsdata som er spesielt utvikla for sektoren. Eit døme kan vere spørsmål og svar basert på medisineksamen eller tilpassa den arbeidsoppgåva helsepersonellet skal bruke språkmodellen i. Det finst fleire internasjonale instruksjonsdatasett for helsefag, t.d. MultiMedBench[27] og MultiMedQA. MultiMedBench inneheld 14 ulike oppgåvesett frå sju ulike helsespesialitetar, inkludert genomikk. Oppgåvesetta omfattar meir enn éin million døme, sjå figur 18.



Figur 15: Samansetjing av instruksjonsdatasettet MultiMedBench [28]

MultiMedQA er eit instruksjonsdatasett som er samansett av seks ulike spørsmål- og svar-datasett med medisin, forsking og innbyggjarførespurnadar samt medisinske spørsmål henta frå nettsøk.[29]

Å utvikle instruksjonsdatasett for norsk helsefagleg bruk vil krevje spesialiserte menneskelege ressursar. Ei mogleg tilnærming vil vere å omsetje og tilpasse internasjonale instruksjonsdatasett til norske tilhøve, men det er usikkert om det er tilstrekkeleg.

[26] https://medium.com/@kshitiz.sahay26/how-i-created-an-instruction-dataset-using-gpt-3-5-to-fine-tune-llama

[27] https://arxiv.org/abs/2307.14334

[28] https://the-decoder.com/med-palm-m-is-google-deepminds-all-purpose-ai-medical-expert/

[29] https://www.nature.com/articles/s41586-023-06291-2

Valideringsdata

Valideringsdata blir brukt til å validere språkmodellen med omsyn til m.a. yting og kvalitet.

Validering kan skje på ulike måtar. Intern validering målar ytinga til modellen på ein tilfeldig del av det opphavlege datasettet, som er trekt ut på førehand og halde unna modelltreninga. Ekstern validering målar ytinga til modellen på eit uavhengig datasett, og blir utført av ein uavhengig instans.

Som ved instruksjonsdatasett, så er ein avhengig av datasett som er tilpassa norsk helsefagleg praksis.

Bruksområde for språkmodellar

Bruk av språkmodellar på helseområdet er eit relativt umodent område. Forventingane er store, men det står att erfaringar og utprøvingar, og det manglar betre tilpassa språkmodellar og kompetanse om bruk.

Samtalerobotar som brukar store språkmodellar har både styrker og veikskapar, som blir peikt på i denne artikkelen:

> *The application of greatest potential and concern is the use of chatbots to make diagnoses or recommend treatment. A user without clinical experience could have trouble differentiating fact from fiction*".[30]

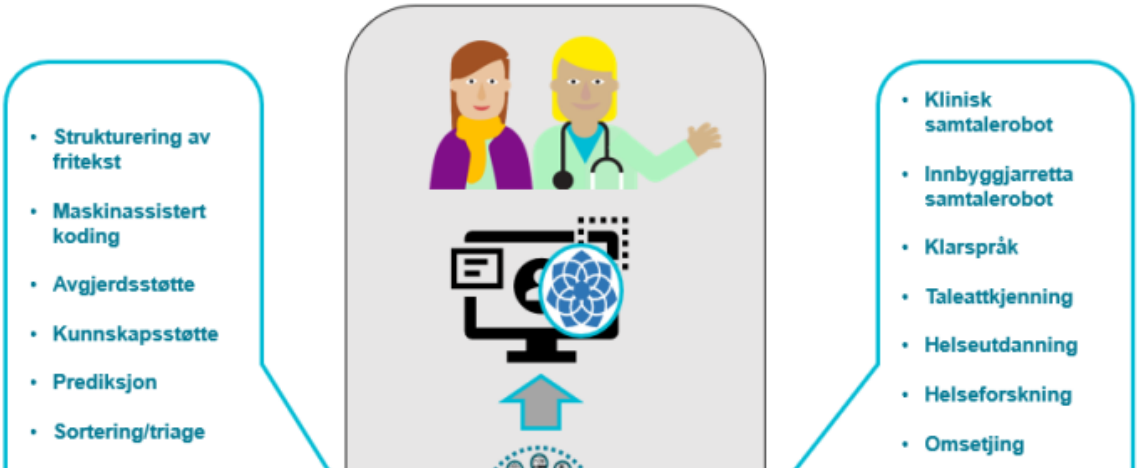
Forfattarane konkluderer likevel med at samtalerobotar vil bli viktige verktøy i medisinsk praksis:

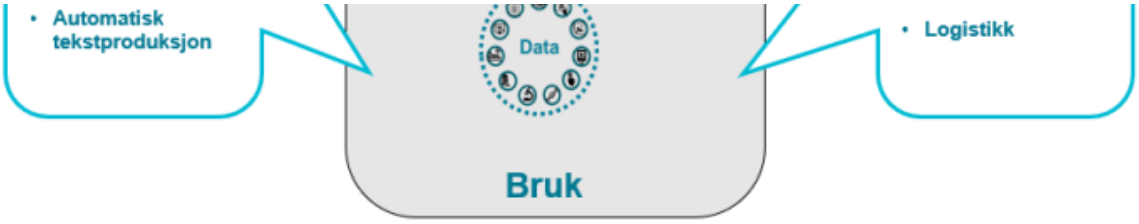
> *Like any good tool, they can help us do our job better, but if not used properly, they have the potential to do damage. Since the tools are new and hard to test with the use of the traditional methods noted above, the medical community will be learning how to use them, but learn we must. There is no question that the chatbots will also learn from their users. Thus, we anticipate a period of adaptation by both the user and the tool.*

Vi kan likevel allereie no sjå konturane av kva slags område språkmodellar kan vere nyttige, frå å effektivisere tekstoppgåver til hjelp i det kliniske arbeidet.

Kjeldene peiker på fleire bruksområde som kan effektivisere eksisterande oppgåver eller bidra til nye og betre arbeidsprosessar og dimed heve kvaliteten på helsetenester. Brukt i administrative oppgåver kan KI frigjere tid som kan nyttast til pasientbehandling. I tillegg kan kunstig intelligens vere ein viktig ressurs i behandlinga. Samstundes meiner vi det er viktig å sjå på kunstig intelligens som ein partnar og støtte, og ikkje ei erstatning for helsepersonell og mellommenneskeleg kontakt. Bruk av KI er også avhengig av godkjenningar på same nivå som annan teknologi i helsetenesta.

Under følgjer ein gjennomgang som baserer seg på intervju og litteratur. Gjennomgangen er ikkje komplett, og dei ulike bruksområda syner ulik grad av modenskap. Til dømes er tekstproduserande bruksområde frå generative språkmodellar meir usikre på grunn av faren for hallusinasjonar. Bruksområda som er nemnde her, er ikkje ei endeleg vurdering av kva område og KI-løysingar som bør takast i bruk, og rapporten må ikkje sjåast på som ei vurdering av forsvarlegheit, tilråding eller godkjenning av Helsedirektoratet. Det er også mogleg at fleire av bruksområda kan løysast av vidareutvikling av regelstyrt maskinlæring i staden for KI.[31]





Figur 16: Moglege bruksområde for språkmodellar

[30] https://www.nejm.org/doi/full/10.1056/NEJMra2302038

[31] Roboten Nora Nord har effektivisert arbeidet med å automatisere manuelle rutiner, t.d. hente data frå ulike kjelder til helsepersonell, sende ut epikriser og så bortetter ved Nordlandssjukehuset

Strukturering av fritekst i journalar

> Det er behov for å strukturere pasientinformasjon for dokumentasjon, samhandling og gjenbruk i pasientforløpet eller til sekundærbruk (statistikk, finansiering, innrapportering, analysar og maskinlæring). Mykje av pasientinformasjonen finst i fritekst, men kan potensielt strukturerast ved hjelp av språkmodellar. Til dømes kan beskrivingar av kliniske symptom hos pasienten ved hjelp av eit standardisert språk brukast om att på tvers av ulike helseinstitusjonar og spesialitetar (semantisk samhandling). Helsepersonell kan difor raskt hente informasjon som er registrert tidlegare i pasientforløpet. Minst like viktig er bruk av strukturert pasientinformasjon til sekundærbruk som t.d. rapportering til ulike helsefaglege kvalitetsregister og til maskinlæring.
>
> Helsearkivregisteret arbeider med å strukturere fritekst av historiske pasientjournalar ved hjelp av KI og språkmodellar. Slik strukturering gjer det mogleg med semantiske søk for forsking.

Maskinassistert helsefagleg koding

> Helsepersonell er pålagt å setje helsefaglege kodar i journalsystemet, t.d. ein diagnosekode for ein pasient eller ein utført prosedyre. Denne koden er henta frå ein klassifikasjon, t.d. diagnosekodar frå ICD-10 eller ICPC-2 og prosedyrekodar frå Norsk prosedyrekodeverk, og kodinga dannar grunnlaget for rapportering til statistikk og finansiering. I tillegg blir terminologien SNOMED CT brukt i delar av helse- og omsorgssektoren for dokumentasjon og samhandling.
>
> Det blir forska på korleis språkmodellar kan kome med automatiske forslag til helsefaglege kodar. Prosjektet ClinCode ved Nasjonalt senter for e-helseforsking er eit døme på det. I prosjektet blir det forska på korleis fritekst og strukturert tekst kan bli analysert ved hjelp av ein språkmodell for å generere ein ICD-10 kode.[32]
>
> Den same teknologien kan også nyttast til å retrospektivt analysere og kvalitetssikre ein kode som helsepersonell allereie har sett.[33]
>
> Maskinassistert helsefagleg koding kan bidra til å lette klinisk byrde for helsepersonell.

[32] og [33]
https://ehealthresearch.no/en/projects/clincode-computer-assisted-clinical-icd-10-coding-for-improving-effici

Avgjerdsstøtte (beslutningsstøtte) i behandling

Språkmodellar kan danne grunnlaget for å utvikle avgjerdsstøtteverktøy (beslutningsstøtteverktøy) for å vurdere alternative behandlingsmetodar.

Det blir peikt på moglegheiter for støtte i form av automatisk oppslag i metodebøker, retningslinjer og liknande ressursar. Beskrivingar av symptom og diagnose hos den einskilde pasienten kan ved hjelp av språkmodellar gje helsepersonell rask og god tilgang til behandlingsforslag.

Språkmodellar kan gjere det lettare å oppdage allergiar og andre tilstandar i journaltekstar, slik at ein kan unngå allergiske reaksjonar i pasientbehandlinga. Dette har tidlegare blitt testa ut med hell ved Sørlandet sjukehus (Informasjonssystem for klinisk konseptbasert søk (IKKS-prosjektet)).

I Danmark er det utvikla ein språkmodell (Den Intelligente Patientjournal) som raskare finne kritisk informasjon i pasientjournalar, t.d. tidlegare blødingsepisodar som kan gje risiko for ny bløding.[34]

Ei form for avgjerdsstøtte der det har blitt forska på bruk av språkmodellar, er forslag til differensialdiagnosar[35]. Differensialdiagnostikk er det å vurdere kva for ein av fleire moglege sjukdomar som ligg føre, ved å samanlikne symptom, sjukdomshistorie, funn ved undersøking av pasienten, laboratorieverdiar og i mange tilfelle også røntgenfunn.[36] Gjennom analysar av kliniske funn, symptom og sjukdomshistorie kan språkmodellar hjelpe helsepersonell med å liste opp moglege sjukdomar hos pasienten.[37] Dette bruksområde må framleis reknast som nokså umodent.

[34] https://www.sdu.dk/en/forskning/sdurobotics/researchprojects/ipj

[35] https://arxiv.org/abs/2312.00164

[36] https://sml.snl.no/differensialdiagnostikk

[37] https://www.nature.com/articles/s41746-024-01010-1

Kunnskapsstøtte

Kunnskapsstøtte er IT-verktøy som kan gje helsepersonell tilgang til forskingsbasert kunnskap i arbeidskvardagen. Ei utfordring for helsepersonell er å orientere seg i den store mengda fagartiklar og andre forskingsbaserte publikasjonar. Språkmodellar kan bidra til kunnskapsstøtte for klinikarar slik at dei kan hente oppdatert informasjon frå ei eller mange kjelder, prosessere han og lage ei oppsummering.

Dette gjeld ikkje minst fag- og forskingsartiklar, som blir publisert i høgt tempo internasjonalt. Språkmodellar kan vere eit verkemiddel for å halde seg fagleg oppdatert på forskingsfronten innan ulike helsefag. Folkehelseinstituttet (FHI) har synt til mogleg stor tidssparing ved å bruke KI til kunnskapsoppsummeringar.

Prediksjon

Forsking peiker på at språkmodellar kan nyttast til ulike former for prediksjon, t.d. risikoprediksjon og mortalitetsprediksjon. Forsking har sett på korleis språkmodellar kan nyttast til å føreseie til dømes lengde på innlegging, risiko for gjeninnlegging (innan 30 dagar etter utskriving) og risiko for dødsfall under sjukehusinnlegging.[38]

Eit mogleg døme er bruk av språkmodellar til analysar av samtalar i psykoterapi. Høg grad av dropout og manglande gjennomføring av terapiane er ei utfordring i behandlinga. Ein språkmodell kan analysere og identifisere markørar i samtalen som indikerer om det er høg risiko for at pasient kjem til å forlate behandling. Denne kunnskapen kan gjere det mogleg å kjenne att når det er fare for dropout og intervenere på føremålstenleg måte.[39]

[38] https://www.nature.com/articles/s41586-023-06160-y

[39] Frostad, Stein. E-post 05.04.2024

Sortering og triage

Det har blitt forska på bruk av språkmodellar for å raskt og effektiv sortere alvorsgrada av skadar og sjukdom hos pasientar, t.d. hos lækjarvakta. Teknologien kan vurdere om pasienten treng akutt hjelp eller ikkje.[40] Til dømes er verktøyet TriageGO utvikla for slik bruk blant sjukepleiarar i USA. Verktøyet nyttar KI for å analysere ulike data (t.d. vitale teikn, kliniske funn, diagnosar og anamnese) for å triagere pasientar med grunngjeving.

Andre automatiske sorteringsfunksjonar kan vere knytt til tilvisingar: Teknologien kan fordele pasientar til rett lege automatisk basert på analyser av tilvisingsbrev.[41]

[40] https://medium.com/columbia-journal-of-science-tech-ethics-and-policy/harnessing-the-power-of-ai-in-emer

[41] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7679210/

Automatisk tekstproduksjon

Språkmodellar kan bidra til automatisk generering av utkast og samandrag til klinisk bruk, t.d. epikrisar i spesialisthelsetenesta og tilvisingsrapportar frå fastlegar[42]. Det finst allereie løysingar i Noreg i dag som i tillegg genererer slike samandrag basert på taleteknologi[43]. Løysinga vil kome med eit forslag til ein epikrise som helsepersonell kan kvalitetssikre og eventuelt justere før utsending. Dette bruksområdet kan lette den kliniske byrden for helsepersonell.

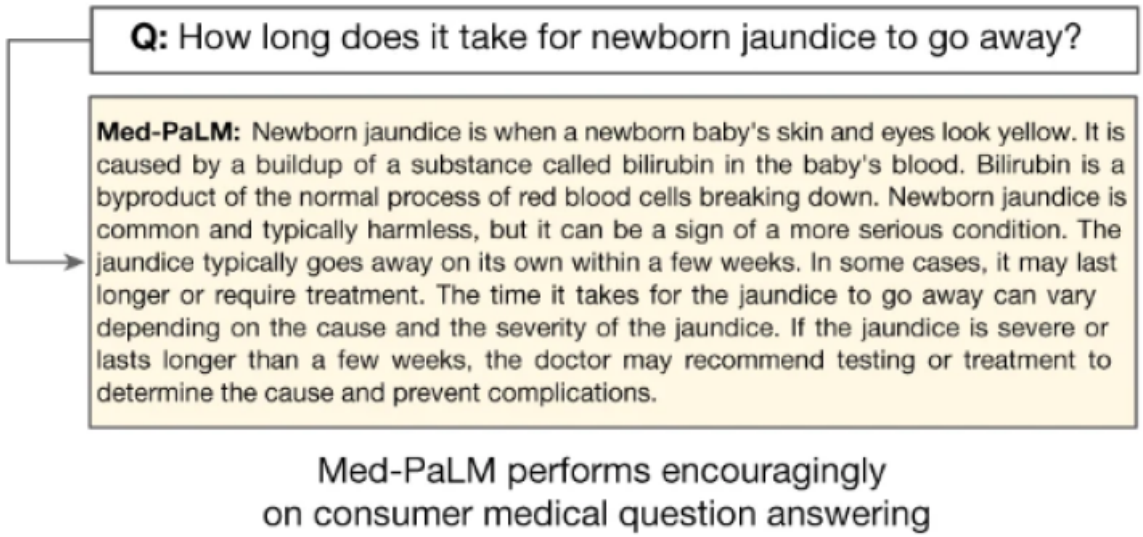
Epikrisar blir også lesne av pasientar, som kan ha vanskar med å forstå den medisinske terminologien. Bruk av språkmodellar kan nyttast til å lage ein (parallell) versjon av epikrisa med eit språk som er lettare forståeleg for lekfolk (sjå punkt 3.10).

[42] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10635391/

[43] Presentasjon av Omilon på EHiN 2022 8. november 2023

Klinisk samtalerobot (spørsmål og svar)

Generative språkmodellar kan fungere som ein samtalerobot som gjev svar på kliniske spørsmål, og såleis fungere som eit avgjerdsstøtteverktøy (beslutningsstøttesystem), jf. figur 17. Språkmodellen Med-PaLM har blitt løfta fram som eit døme på kor raskt utviklinga av språkteknologi går. Med-PaLM var i stand til å svare rett på 67,6 % av spørsmåla (m.a. frå medisineksamen i USA) i desember 2022. I mars 2023 kom den korrekte svarprosenten på 86,5 %.



Figur 17: Døme på spørsmål og svar i språkmodellen Med-PaLM[44]

[44] https://www.nature.com/articles/s41586-023-06291-2/figures/1

Innbyggjarretta samtalerobot (spørsmål og svar)

Generative språkmodellar kan fungere som ein samtalerobot som gjev svar på spørsmål som innbyggjarar måtte ha om symptom, sjukdomar, behandling, førebygging, medisinar og så bortetter. Truleg blir eksisterande generative språkmodellar som t.d. ChatGPT nytta til slikt allereie.

I utvikling av innbyggjarretta samtalerobotar er det viktig at modellen blir brukstilpassa (*alignment*) slik at han er i stand til å ha ein god dialog med menneske.

Helsedirektoratet ønskjer å utforske om generativ KI kan brukast til å gjere informasjon frå helsevesenet meir tilpassa og forståeleg for forskjellige brukargrupper, blant anna barn og unge.[45] Prosjektet undersøker korleis m.a. ungdom skal få betre svar på spørsmål ved å forankre ein språkmodell på utvalde relevante tekstar og kunnskapskjelder (*grounding)*.

Tiltaket Enklare tilgang til informasjon (ETI), som høyrer til porteføljen Alvorleg sjukt barn, har som mål å utvikle ein informasjonsassistent som legg til rette for innsamling og presentasjon av relevant informasjon

frå eit stort antal offentlege kjelder.[46] Her blir språkmodellar brukt til både å førehandsgenerere spørsmål og svar og å utvikle ein interaktiv samtalerobot.

[45] https://www.datatilsynet.no/aktuelt/aktuelle-nyheter-2023/tid-for-sprakmodeller-i-sandkassa/

[46] https://alvorligsyktbarn.no/enklere-tilgang-til-informasjon

Klarspråk

Radiologiske rapportar er utforma av spesialistar og inneheld difor fagspråk som kan vere vanskeleg å forstå både for pasientar og anna helsepersonell enn radiologar, som til dømes allmennlegar. Språkmodellar kan nyttast til å omforme tekst og tilpasse han til ulike målgrupper.

Mohn-senteret for medisinsk biletdiagnostikk og visualisering (MMIV) i samarbeid med Høgskulen på Vestlandet er i startgropa med eit prosjekt som skal prøve ut og evaluere om store språkmodellar kan brukast til å "omsette" spesialisert fagspråk til meir forståeleg språk, ev. til andre språk. Prosjektet vil i første omgang teste språkmodellar på radiologirapportar, men språkmodellar vil truleg også kunne «omsetje» andre rapportar utarbeida av helsepersonell, t.d. epikrisar og labsvar til eit meir forståeleg språk for pasientar og evt. allmennlegar.

I pasient- og brukarrettslova står det i § 3-5 at informasjon frå helse- og omsorgssektoren skal "… være tilpasset mottakerens individuelle forutsetninger, som alder, modenhet, erfaring og kultur- og språkbakgrunn". Prosjektet kan difor bidra til å oppfylle dette lovpålagde ansvaret.

Taleattkjenning

Taleteknologi har eksistert lenge, t.d. i tale-til-tekst i helsesektoren. Med språkmodellar har nye verktøy som t.d. Whisper[47] heva kvaliteten på slik språkteknologi. Nasjonalbiblioteket har gjort ein norsk versjon av Whisper, som er trena på transkribert norsk tale, tilgjengeleg. Godt fungerande taleattkjenning basert på språkmodellar kan redusere arbeidsbyrda for helsepersonell ved at tale gjennom ein diktafon blir transkribert og lagra rett stad i journalen. Bruk av språkmodellar gjer det også mogleg å automatisk generere samandrag av den innlesne informasjonen (sjå 4.7).

[47] https://en.wikipedia.org/wiki/Whisper_(speech_recognition_system)

Helseutdanning

Det har også blitt påpeikt at språkmodellar kan nyttast i medisin-, sjukepleie- og farmasiutdanninga. Studentar kan nytte språkmodellar til å trene seg på å setje diagnosar i ei interaktiv løysing som gjev tilbakemelding med det same.[48] [49]

[48] Presentasjon på KIN-møte #3 2023

[49] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10463084/

Helseforsking

> Språkmodellar kan også nyttast til forsking på helseområdet, gjennom t.d. å oppdage mønster i store mengder data, programmering i t.d. bioinformatikk og hjelp til akademisk skriving.[50]
>
> [1] https://www.nature.com/articles/s43856-023-00370-1 The future landscape of large language models in medicine | Communications Medicine (nature.com)

Omsetjing

> Språkmodellar har allereie i fleire år blitt brukt til automatisk omsetjing mellom ulike språk. Gjennom helsefaglege språkmodellar kan ein automatisk omsetje tekstar med klinisk fagspråk til eller frå norsk. I helse- og omsorgssektoren kan det vere nyttig når språklege barrierar oppstår mellom helsepersonell og pasientar og tolketenester ikkje er tilgjengelege.

Logistikk

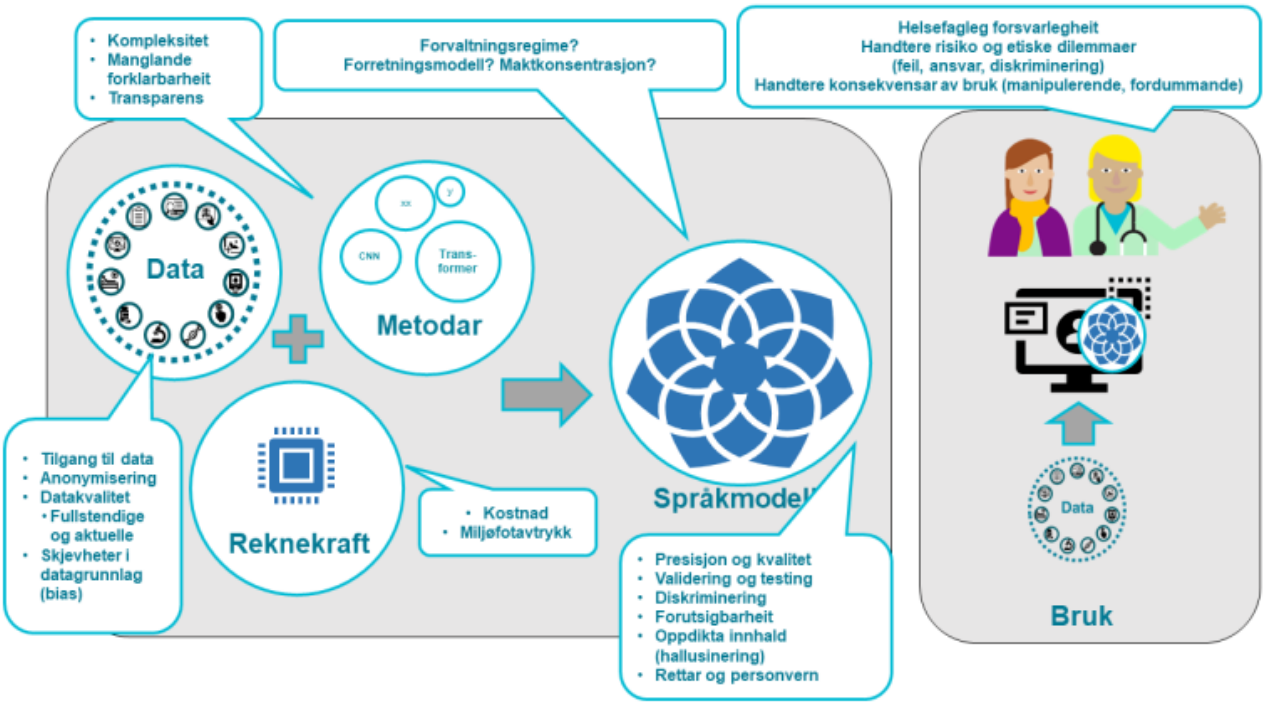
> Språkmodellar kan nyttast til å effektivisere logistikkoppgåver som t.d. å føreseie lagerbehov, automatisering av bestilling og monitorering av bruk av medisinsk utstyr.[51]
>
> [15] https://srinath-sridharan.medium.com/enhancing-healthcare-efficiency-the-role-of-large-language-models-in

Utfordringar og risikoar

Sjølv med betydelege framsteg i ytinga til store språkmodellar, kan dei ikkje brukast utan vidare i den norske helse- og omsorgstenesta. Problem som hallusinasjonar, innebygde fordommar og utfordringar med personvernproduksjon av tilsynelatande truverdig feilinformasjon, krev merksemd. Det blir spesielt viktig å forbetre og tilpasse KI-modeller til kulturen og språket som blir brukt i norsk helseteneste, å redusere skjeivheiter/bias i treningsdata og å utvikle robuste metodar for å validere innhald generert av KI, slik at språkmodellar kan brukast trygt og effektivt i helsesektoren.

Kjeldene peiker på fleire spørsmål, utfordringar og bekymringar knytt til språkmodellar, mellom anna:

1. **Utvikling og finjustering av språkmodeller**
   Utfordringar knyter seg til tilgang til og kvalitet på treningsdata, datakraft, miljøavtrykk, kostnad, nytte og risiko og validering.
2. **Bruk av språkmodellar**
   Det er framleis risiko for at generative språkmodellar hallusinerer og skaper uriktig informasjon. Andre bekymringar knyter seg til kvalitet, skeivheiter (bias), etikk, gjennomsiktigheit, ansvar og kompetanse.
3. **Forvaltning av språkmodellar**
   Det er opne spørsmål knytte til korleis språkmodellar forvaltast, om vi treng nasjonal forvaltning, og om det er risiko for maktkonsentrasjon.



Figur 18: Nokre utfordringar og bekymringar ved språkmodellar

Desse problemstillingane er i større grad konkretiserte i vedlegg A.

Moglege tiltak

Dette kunnskapsgrunnlaget har gitt et augeblinksbilete av språkteknologilandskapet på helsefeltet i Noreg pr. mars 2024. Det er forventa store og raske teknologiske endringar i den nærmaste framtida, og det opnar for både moglegheiter, men også for risikoar og utfordringar som må løysast.

Nasjonal helse- og samhandlingsplanen for 2024-2027 (NHSaP 2024-2027) slår fast at det er eit stort potensial for å etablere ei meir føremålstenleg oppgåvedeling og god organisering av arbeidsprosessar i vår felles helseteneste. Vidare peiker planen på at det er eit behov for kontinuerleg vurdering om oppgåver kan løysast på heilt andre og personellsparande måtar, t.d. ved bruk av kunstig intelligens, ved å automatisere arbeidsoppgåver og endre arbeidsprosessar som tidlegare blei gjort manuelt. Regjeringa forventar at kunstig intelligens og annan personellsparande teknologi vil kunne bidra til at vi kan oppretthalde kvaliteten i tenesta i åra framover og reduserer ventetidene på dei stadene der KI allereie er tatt i bruk i dag. [52]

Store språkmodellar og generativ KI er delar av KI-feltet som nettopp blir sett på som personellsparande teknologi. Samstundes er bruken av KI i helse- og omsorgstenester heilt i startgropa. Det blir viktig å vinne erfaringar og det står att fleire utfordringar som må løysast med bruk av KI. For å ta ut dei moglege gevinstane som dei store språkmodellane gir og for å møte utfordingane og den teknologiske endringa på området, skisserer vi nokre moglege tiltak nedanfor. Konkretisering og prioritering av tiltaka vil gjerast i arbeidet med felles KI-plan for helse- og omsorgstenesta.[53]

[52] Meld. St. 7 (2019–2020). Nasjonal helse- og sykehusplan 2020–2023. Helse- og omsorgsdepartementet. https://www.regjeringen.no/no/dokumenter/meld.-st.-7-20192020/id2678667

[53] TB2024-34: Kunstig intelligens i helse- og omsorgstjenesten: https://www.regjeringen.no/contentassets/d8f63d7d01d64def982cb7c8ce1eeb64/tildelingsbrev-til-helsedirektora

Vurdere risiko

> Bruk av språkmodellar, særleg generative språkmodellar, inneber mange moglegheiter, men også usikkerheit og risikoar. Det er behov for å gjere systematiske risikoanalysar av bruk av språkmodellar i sektoren ved å ta omsyn til m.a. bruksområde og teknisk arkitektur.
>
> Helse- og omsorgsdepartementet skriv i sitt tildelingsbrev til Helsedirektoratet at "etatene [skal] også vurdere hvilke risikoer store språkmodeller introduserer".
>
> Moglege tiltak:
>
> - systematisk analyse av risikoar for aktuelle bruksområde og ulike typar språkmodellar
> - identifisere risikodempande tiltak

Tilgang på språkmodellar tilpassa norske tilhøve

Det er usikkert om tilgjengelege språkmodeller tek omsyn til norsk språk, kultur og praksis i helse-omsorgssektoren i tilstrekkeleg grad. Språkmodellar i sektoren må vere trena på tilstrekkelege norskspråklege treningsdata til bruksføremålet.

I tildelingsbrev frå helse- og omsorgsdepartementet står det at Helsedirektoratet skal "identifisere tiltak for å sikre at helse- og omsorgstjenesten har tilgang på språkmodell(er) som er tilpasset norske forhold."

Moglege tiltak:

- analysere ev. behov for nasjonal helsefagleg språkmodell på bakgrunn av resultat frå pågåande utviklingsprosjekt som t.d. Klinisk NorBERT, NorwAI og NORA.
- følgje med og delta i nasjonale prosjekt som utviklar store norske språkmodellar som kan nyttast i helse- og omsorgssektoren
- analysere kor føremålstenlege dei ulike utviklingsstega er for tilpassing av språkmodell til ulike bruksføremål i helse- og omsorgssektoren
- analysere moglege forvaltningsregime for ev. ein eller fleire nasjonale helsefaglege språkmodellar
- vurdere kva slags norske helsefaglege treningsdata som er relevante og i kva grad dei er tilgjengelege for utvikling, trening og finjustering av språkmodellar, inkludert opphavsrett og personvern
- vurdere behovet for og mogleg organisering av ein nasjonal infrastruktur for å tilgjengeleggjere norske treningsdata

Fremje strategisk viktige bruksområde

Det er framleis usikkert kva bruksområde som gjev høgast gevinst i helsetenesta. Nokre bruksområde er lovande, men treng meir utprøving og analyser. Analysane kan dreie seg om arkitektur og tekniske behov, kostnadar og gevinstar, risikoar, hindringar og forvaltning. Det er særskilt viktig å starte med lågthengande frukter.

Moglege tiltak:

- følgje med på nasjonale, nordiske og internasjonale initiativ og spreie kunnskap og erfaringar
- løfte fram lovande norske initiativ og leggje til rette for nasjonal bruk
- vurdere å starte med lågthengande frukter
- vere pådrivar for at nye, store forskingssenter også omhandlar KI i helse- og omsorgstenesta
- vurdere å etablere ei finansieringsordning for dei mest lovande bruksområda

Styrke kapasitet og kompetanse

Det er mangel på tilstrekkeleg kompetanse om bruk, utvikling og finjustering av språkmodellar i helse- og omsorgssektoren.

Kompetansen bør difor styrkast både internt i Helsedirektoratet og i heile helse- og omsorgssektoren. Framtidige strategiske avgjerder vil i større grad vere knytt til digital transformasjon og teknologisk utvikling. Difor er det viktig med grunnleggjande kompetanse også blant leiarar knytt til risikoar og gevinstar ved å bruke språkmodellar i helse- og omsorgstenesta.

Moglege tiltak:

- initiere eit fagråd eller nettverk med ekspertar i språkmodellar og helsefag frå helse- og omsorgssektoren og universitets- og høgskulesektoren
- innhente erfaringar frå nordiske og andre land
- arrangere seminarserie om temaet saman med sektor
- arrangere internt arbeidsseminar (workshop) i Helsedirektoratet
- identifisere tiltak for å styrke leiarkompetanse i sektoren
- løfte fram viktige føregangsprosjekt ved hjelp av ein årleg KI-pris for sektoren
- teste ut helsefagleg rettleiing (til dømes RAG) på eit klinisk bruksområde for å skaffe kunnskap om utvikling og bruk av språkmodeller og synleggjere potensielle bruksområde
- basert på erfaringar: utarbeide rettleiingar for finjustering, RAG og andre sider ved utvikling av språkmodellar for sektor
- utvikle og forvalte ei dynamisk oversikt på ei nettside over språkmodell-initiativ i sektoren i Noreg

Engelsk-norsk ordliste

Alignment	justering i tråd med verdiar og etikk
Chatbot	samtalerobot
federated learning	føderert læring
few-shot prompting	instruksar med fleire døme
fine tuning	finjustering
foundation model	grunnmodell
grounding	rettleiing
large language model	stor språkmodell
one-shot prompting	instruksar med eitt døme
zero-shot prompting	direkte instruksar

Vedlegg A: Døme på risikoar ved bruk av store språkmodellar

Dette vedlegget listar opp utfordringane og risikoane som har kome fram i arbeidet med dette kunnskapsgrunnlaget. Utfordringane og risikoane vil bli tydeleggjort i det vidare arbeidet med store språkmodellar i etatane, som felles KI-plan for helse- og omsorgstenesta.[54]

A.1 Utvikling og finjustering av språkmodellar

- Tilgang til og kvalitet på treningsdata:
  - Tilstrekkeleg mengder treningsdata
  - Treningsdata som har høg nok kvalitet
  - Treningsdata som er oppdaterte og komplette
  - Treningsdata på norsk, både bokmål og nynorsk, og andre skandinaviske språk i tillegg til nord-samisk.
  - Treningsdata med rettsleg grunnlag og lovleg tilgang
  - Treningsdata som ikkje strir mot opphavsrettar (IPR)
  - Treningsdata i tråd med personvern (der sensitive data er nytta for trening), i kva grad kan den trena modellen lekke sensitive treningsdata
  - Treningsdata som ikkje er manipulert av aktørar som ynskjer å påverke åtferda til ein språkmodell (forgiftingsåtak (*poisoning attacks*)
- Datakraft
  - Tilgang til nok maskinkraft til å trene eller finjustere språkmodellar
  - Økologisk avtrykk frå svært høgt energiforbruk ved dataprosessering i utvikling og finjustering av språkmodellar
- Kostnad, nytte og risiko
  - Kva kostnadar vil utvikling av språkmodellar medføre?
  - Kven skal bære kostnadane?
  - Kva er forretningsmodellen til leverandørar av «gratis» verktøy?
- Validering
  - Mangel på metode for evaluering og testing av språkmodellar og tilhøyrande teknologi som t.d. RAG
  - Tilgang til evalueringssett i norsk helsefagleg språk. Slik kan ein seie noko om ytinga til modellane. Det er mogleg at slike evalueringssett kan bli tilgjengelege ved tilpassing av medisinske eksamenar for engelsk til norsk.
  - Korleis gjere CE-merking av språkmodellar
- Oversikt
  - Mangel på oversikt over lokale og regionale initiativ
- Teknologisk kompleksitet
  - Trening av språkmodellar er så kompleks at det blir vanskeleg å forstå eller gjere greie for avgjerdsgrunnlaget til språkmodellen
  - Teknologien endrar seg raskt, og det er vanskeleg å planleggje på lengre sikt

A.2. Bruk av språkmodellar

- Kvalitet
  - Kva er faren for at generative språkmodellar hallusinerer?

Det er framleis fare for at generative språkmodellar hallusinerer og skaper uriktig informasjon. Språkmodellane er framleis på teststadiet og enno ikkje klare for innføring.

Dersom ein generativ språkmodell hallusinerer, kva er mogelegheita for systemet og/eller brukarane av systemet til å oppdage dette så ikkje feilen forplantar seg vidare?

    - I kva grad gjev generative språkmodellar *relevante* og *spesifikke nok* svar?
    - I kva grad vil ny, oppdatert kunnskap bli vekta høgt nok samanlikna med utdatert kunnskap (*static knowledge* vs. *dynamic knowledge*)
    - Kva konsekvensar har det at ein og same språkmodell gjev ulike svar for den same instruksen, t.d. spørsmål (*reproducability*)?[55]
- Yting/infrastruktur
    - Vil ein kunne køyre språkmodellane med akseptabel hastigheit?
    - Vil ein kunne tilpasse språkmodellane svingingar blant mange samtidige brukarar?
- Kultur
    - Er helsefaglege språkmodellar tilstrekkeleg tilpassa norsk helsefagleg kultur og praksis?
- Bias
    - Gjev språkmodellar like gode tenester/svar til ulike delar av samfunnet (kjønn, etnisitet, klasse)? Vil språkmodellar diskriminere?
    - Vil innføring av språkmodellar føre til mindre likebehandling? Eller meir?
    - Reproduserer språkmodellar stereotypar, t.d. innan psykisk helse?
- Etikk
    - Kva etiske og moralske problemstillingar vil språkmodellar medføre?
- Transparens
    - I kva grad er treningsdata og treningsmetodar for ein språkmodell gjort opne?
    - I kva grad er programmeringskodene for utviklinga av modellen, opne?
    - I kva grad er det mogleg å forstår *kvifor* ein språkmodell gjev eit gitt resultat (svart boks-problemet)
    - Kva konsekvensar får dei nye EU-krava knytt til transparens i språkmodellar, især modellar med stor innverknad (*high impact foundation models*)?
- Ansvar
    - Vil innføring av språkmodellar endre ansvarleggjeringa av helsepersonell? Kven har ansvaret dersom språkmodellar fører til feil?
    - Vil bruken føre til ein slags kunnskapseksternalisering slik at ein på sikt mister evner til å resonnere, tenkje nytt, kreativt eller kritisk?
    - I kva grad vil pasientane vite at avgjerder er blitt tekne på grunnlag av språkmodellar?
- Kompetanse
    - Har helsepersonell tilstrekkeleg kompetanse til å sjå moglegheitene og risikoane ved innføring av språkmodellar?
    - Har helsepersonell tilstrekkeleg kompetanse til å nytte tenester basert på språkmodellar?
- Aksept
    - Kva haldningar finst det til språkmodellar blant helsepersonell?
    - Kva haldningar finst det til språkmodellar blant innbyggjarar?

- Gevinstrealisering
    - Korleis kan ein måle kost-nytte og gevinstrealisering ved språkmodellar?
- Rett bruk og misbruk:
    - Er bruken av språkmodellen i tråd med det den er trent til?
    - Er det tenkt gjennom korleis språkmodellen kan misbrukast (kan den til dømes lurast til å gi ut informasjon det ikkje er ynskjeleg at den gir ut, slik som sensitiv informasjon, informasjon som kan misbrukast og informasjon om modellen og instruksjonar (*prompt injection attacks*)? Finst det tiltak som hindrar at språkmodellen blir misbrukt, og har det blitt vurdert kor lett det er å omgå desse restriksjonane?

    - Er språkmodellen sårbar for innputt som det krev mykje ressursar å prosessere, og som dermed kan hindre god yting i systemet?
- Kontroll på informasjon:
    - Ved bruk av språkmodellen, vil ein måtte dele mykje informasjon (typisk gjennom innputt til modellen) med ein eller fleire andre aktørar? Ynskjer ein å dele denne type informasjon, og kan den overførast på ein trygg måte?

        Dersom ein nyttar kunnskapsbasering (RAG), har ein kontroll på dei eksterne kjeldene som blir brukt, at dei er av god kvalitet og ikkje kan endrast av aktørar som ynskjer å påverke språkmodellen (indirect prompt injection attack).

A.3 Forvaltning av språkmodellar

- Forvaltningsregime
    - Bør språkmodellar forvaltast sentralt, eller i form av ein distributiv modell, sjå kapittel 1.5?
    - Treng vi nasjonal forvaltning til nokre typar bruk?
- Maktkonsentrasjon
    - Vil innføring og bruk av språkmodellar føre til større maktkonsentrasjon blant teknologiske leverandørar (*vendor lock-in*) og monopolliknande tilstandar? Kva med kunnskapskonsentrasjon, dvs. at nokre få aktørar styrer tilgang til helsefagleg kunnskap ?
- Kompetanse
    - I kva grad har styresmakter tilstrekkeleg kompetanse til å regulere språkmodellar?
- Svart boks og risiko

Det finst mange generelle språkmodellar som kan nyttast som grunnmodell og tilpassast til bruk i helse- og omsorgssektoren. Men det kan vere vanskeleg å få tak i kunnskap om kvalitet og risiko i dei ulike grunnmodellane, og kunne samanlikne dei ulike alternativa. Bruk av ein ope tilgjengeleg språkmodell som grunnmodell kan gjere systemet meir utsett for åtak sidan åtakarar kan nytte den opne modellen for å finne effektive åtak, og så overføre desse åtaka til den tilpassa modellen

[54] TB2024-34: Kunstig intelligens i helse- og omsorgstjenesten: https://www.regjeringen.no/contentassets/d8f63d7d01d64def982cb7c8ce1eeb64/tildelingsbrev-til-helsedirektora

[55] The future landscape of large language models in medicine | Communications Medicine (nature.com)