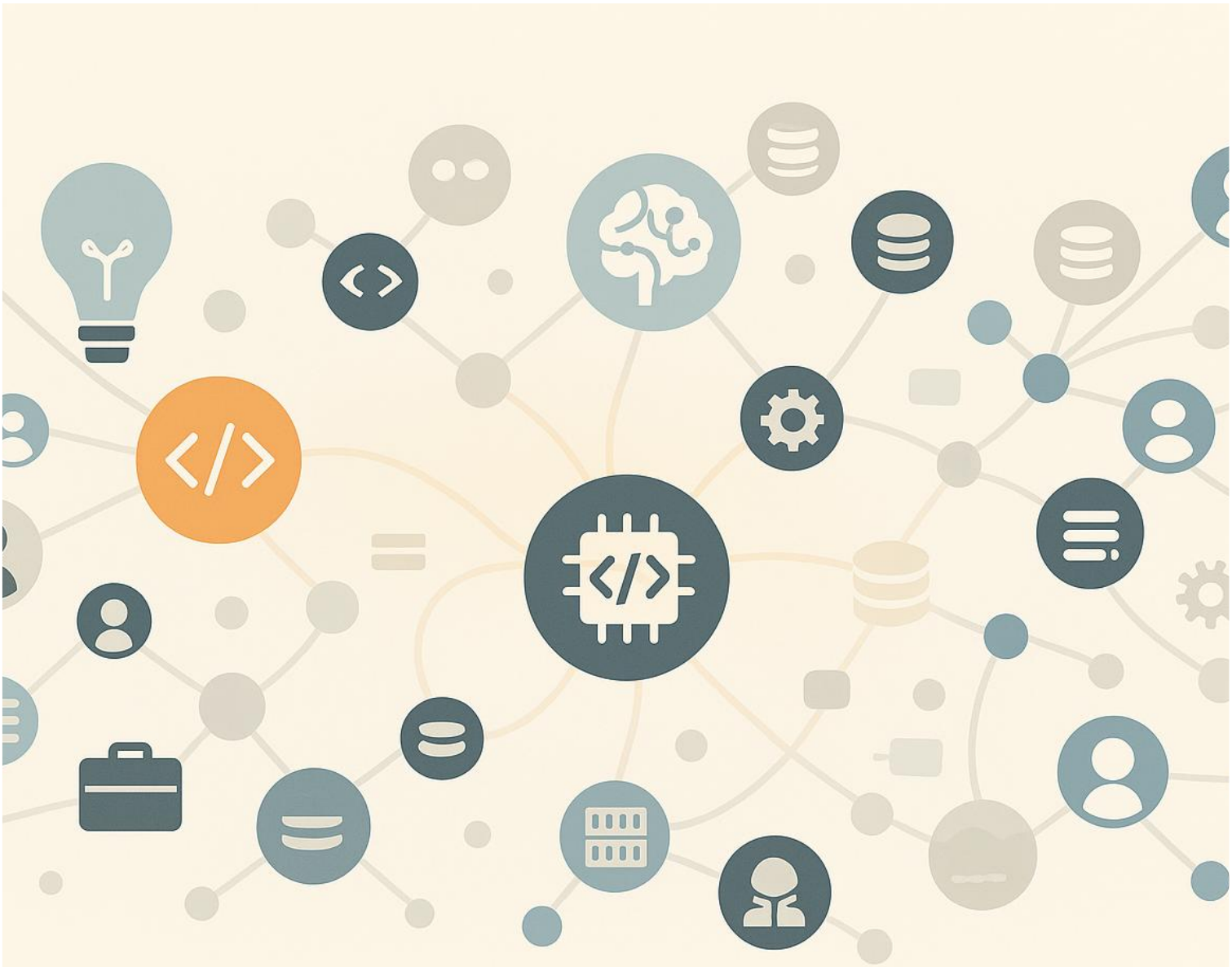


Fra ord til økosystem: Språkmodeller og generativ KI i Norge

Kartlegging av det nasjonale økosystemet for utvikling og bruk av språkmodell-basert generativ kunstig intelligens

2025

DIGITALISERINGS- OG FORVALTNINGSDEPARTEMENTET

AGENDA KAUPANG

OPPDRAGSGIVER:	**Digitaliserings- og forvaltningsdepartementet**
RAPPORTNUMMER:	**R1023358**
RAPPORTENS TITTEL:	**Fra ord til økosystem: Språkmodeller og generativ KI i Norge**
ANSVARLIG KONSULENT:	**Morten Stenstadvold**
KVALITETSSIKRET AV:	**Jonas Rusten Wang**
ILLUSTRASJON PÅ FORSIDEN:	**Generert med DALL·E (OpenAI) gjennom ChatGPT.**
DATO:	**29.09.2025**

Forord

Digitaliserings- og forvaltningsdepartementet har gitt Agenda Kaupang, med bistand fra Simula Research Laboratory, i oppdrag å kartlegge økosystemet for språkmodellbasert kunstig intelligens i Norge. Hensikten med oppdraget har vært å skape oversikt over aktører, roller og samhandlingsmønstre, samt å belyse behov, barrierer og forventninger knyttet til utvikling og bruk av språkmodeller.

Arbeidet bygger på en kombinasjon av dokumentanalyse og semistrukturerte intervjuer med sentrale aktører fra hele verdikjeden for språkmodeller – forskningsmiljøer, utviklere, infrastrukturleverandører, myndigheter og brukere i offentlig og privat sektor. I tillegg har prosjektet trukket veksler på faglige vurderinger fra oppdragsmiljøene.

Vi ønsker å rette en stor takk til Digitaliserings- og forvaltningsdepartementet for et interessant og fremtidsrettet oppdrag. En særlig takk går til alle informanter som har deltatt i intervjuene og delt erfaringer, refleksjoner og vurderinger. Deres bidrag har vært avgjørende for å belyse temaet på en nyansert måte. Ansvar for innhold og konklusjoner ligger hos forfatterne. Vi håper at rapporten vil være et nyttig grunnlag for departementets videre arbeid med utvikling, regulering og strategisk styring av språkmodellbasert kunstig intelligens i Norge.

Rapporten er utarbeidet av Agenda Kaupang og Simula Research Laboratory. Ansvarlig konsulent har vært Morten Stenstadvold fra Agenda Kaupang. Prosektdeltagere er Jonas Rusten Wang og Tom Markussen (Agenda Kaupang) og Omar Richardson og Knut Andre Grytting Prestsveen (Simula Research Laboratory. Jonas Rusten Wang har også vært kvalitetssikrer for rapporten.

Oslo, september 2025

Innhold

Sammendrag

Digitaliserings- og forvaltningsdepartementet har gitt Agenda Kaupang og Simula Research Laboratory i oppdrag å kartlegge økosystemet for språkmodellbasert kunstig intelligens i Norge. Formålet med kartleggingen har vært å identifisere de viktigste aktørene og deres roller, forstå samhandlingsmønstre, kartlegge bruksområder og forretningsmodeller, samt få innspill om behov, barrierer og forventninger til offentlig innsats.

Økosystemet for kunstig intelligens og språkmodeller i Norge er i rask utvikling. Det er et stort og voksende økosystem av norske selskaper som utvikler KI-løsninger, med over 350 identifiserte aktører, hvor mange er nyetablerte. Språkmodellene som brukes er i all hovedsak utenlandske, og infrastrukturen som brukes er store internasjonale skyplattformer. Tilgjengeliggjøring av språk-modeller for bruk i applikasjoner skjer i all hovedsak gjennom API-er levert av internasjonale selskaper som Microsoft og Google.

Kartleggingen viser samtidig at økosystemet for utvikling *norske* språkmodeller er dominert av et fåtall hovedsakelig offentlige og akademiske aktører. Nasjonal treningsinfrastruktur er hovedsakelig knyttet til Sigma2, NTNU og tilgang til LUMI-superdatamaskinen i Finland, og flere aktører etterlyser økt kapasitet. Datagrunnlaget for norske språkmodeller er hentet fra et bredt spekter av kilder, blant annet Nasjonalbiblioteket, mediehus og åpne internettkilder. Treningsdata samles og kurateres i hovedsak gjennom samarbeid mellom offentlige institusjoner som Nasjonalbiblioteket, UiO og NTNU.

Utvikling av store generelle modeller fra bunnen er svært krevende med tanke på både datatilgang, kompetanse og infrastruktur. Norske aktører som utvikler språkmodeller, som Nasjonalbiblioteket, UiO, NTNU og Bineric, fokuserer derfor på finjustering og tilpasning av åpne internasjonale språkmodeller som Mistral og Llama. Bruken av norskutviklede språkmodeller i praktisk bruk ser ut til å være begrenset, med unntak av Whisper-modellen fra Nasjonalbiblioteket.

Barrierene for utvikling av norske språkmodeller er mange og gjensidig forsterkende: mangel på datasett, begrenset tilgang til tungregnekraft, svak infrastruktur, kortsiktige finansierings-mekanismer, uklare ansvarsforhold og mangel på kompetanse. Opphavsrettslige begrensninger og lav datamengde for nynorsk og samiske språk pekes også på som utfordringer. Dette gjør det krevende å utvikle robuste modeller og varige løsninger. Offentlig sektor spiller en avgjørende rolle, både som regulator, premissgiver og tilrettelegger. Samtidig fremstår innsatsen som fragmentert og lite koordinert.

Aktørene etterlyser tydeligere styring, bedre rammeverk for datadeling, langsiktige finansierings-ordninger, test- og sandkassefasiliteter, samt kompetanseheving. Barrierene som er beskrevet kan ikke løses av enkeltaktører alene. Det kreves en helhetlig offentlig politikk og et tett samspill mellom myndigheter, forskningsmiljøer og næringsliv for å utløse potensialet. Etableringen av KI Norge kan bidra til dette.

Agenda Kaupang vurderer at den offentlige ressursinnsatsen, og særlig KI Norges rolle, bør målrettes mot å stimulere til faktisk bruk av kunstig intelligens, som samtidig ivaretar norsk språk og kultur. Ved å prioritere tiltak som øker etterspørselen hos de aktørene som skal ta KI i bruk innen sine ansvarsområder, kan man sikre bedre balanse mellom tilbud og behov. Dette vil bidra til at løsningene som bygges opp i både privat og offentlig sektor blir relevant, nyttig og tett koblet til de konkrete utfordringene den skal løse. Offentlig sektor, og KI Norge, bør i størst mulig grad prioritere områder som det offentlige har særlig gode forutsetninger for å løse. Vi har følgende konkrete anbefalinger:

- **Innføre felles standarder og rammeverk for sikker bruk av KI**
  Vi anbefaler å arbeide med å etablere tydelige mekanismer for å dokumentere at en språkmodell er «god nok» for bestemte typer bruk. Bruk av standardiserte evalueringsrammeverk og godkjenningsordninger kan redusere terskelen for bruk av språkmodeller og gjør det lettere å ta i bruk KI-løsninger utviklet i privat sektor.

- **Tydeliggjøre regelverk og praksis for behandling av sensitive data**
  I offentlig sektor generelt er det uttrykt et sterkt behov for klarhet i hvordan personopplysninger og sensitiv informasjon kan håndteres ved bruk av språkmodeller. Vi anbefaler å kartlegge dagens praksis og tolkninger på dette området, og hvis mulig gi felles føringer for hva som er forsvarlig praksis.

- **Bruke innovasjonsvirkemidler til å stimulere bruk av språkmodellbasert KI**
  Ordninger som finnes i dag bør også kunne bruke midlene til språkmodellbaserte løsninger. Det bør vurderes å styrke disse ordningene for å stimulere til økt eksperimentering og bruk av KI.

- **Finansiere frigjøring og tilgjengeliggjøring av norske data**
  Et gjennomgående hinder for trening av norske språkmodeller er begrenset tilgang til åpne og rettighetsavklarte datasett. Det bør derfor settes av midler til å kuratere, kvalitetssikre, tilpasse og tilgjengeliggjøre datasett med stor samfunnsverdi.

- **Vurdere å finansiere utvikling av domenespesifikke modeller**
  Språkmodeller som er tilpasset fagområder som helse, juss, skole og offentlig forvaltning, krever målrettet arbeid med data og fagspråk. Det bør vurderes sektorvise satsinger, der offentlige fagmyndigheter koordinerer modellutvikling innen sitt felt.

- **Bevilge infrastrukturmidler til tungregning via nasjonale aktører**
  Det bør settes av dedikerte midler til videreutvikling og drift av tungregningsinfrastruktur gjennom nasjonale aktører. Samtidig bør det stilles krav om samarbeid med europeiske initiativer, for å sikre ressursutnyttelse og teknologisk samspill på tvers av land.

- **Sørge for gode rammevilkår for bruk av norske tilbydere av KI-infrastruktur**
  Det bør legges til rette for at norske aktører som tilbyr sikker norsk infrastruktur og språkteknologi har stabile og konkurransedyktige rammevilkår. Strategiske hensyn som nasjonal sikkerhet, teknologisk suverenitet, og hensyn

- **Bygge kapasitet innen anskaffelse av løsninger basert på norske språkmodeller**
  Vi anbefaler egnede tiltak som kan gjøre det lettere å anskaffe forsvarlige KI-løsninger, eksempelvis språkmodeller som speiler norsk språk og verdier. Dette kan senke terskelen for å ta i bruk KI-løsninger, samtidig som det mobiliserer leverandørmarkedet til å bidra med innovasjon og løsninger som fremmer ansvarlig bruk.

- **Bygg et fagmiljø for koordinering av investeringer**
  I dag skjer investeringer, utvikling og innføring av språkmodeller i Norge fragmentert.
  En nasjonal fagmyndighet for språkmodeller, forankret i et uavhengig og tverrsektorielt mandat, bør få ansvar for å kartlegge og holde oversikt, og prioritere satsinger.

- **Bygge opp og styrke offentlige mekanismer for proaktiv veiledning**
  Vi anbefaler å satse på regulatoriske sandkasser og pilotordninger der språkmodeller kan prøves ut.

- **Gjennomføre en systematisk kompetansekartlegging**

  Kompetansekartleggingen bør søke å identifisere dagens kompetansegap innenfor ulike områder, og være bør være bred nok til å favne alle nivåer (ledere, fagpersoner, prosjektledere, utviklere, spesialister mv) og foreslå målrettede tiltak.

- **Legge til rette for aktiv eksperimentering**

  Legge til rette for pilotprosjekter der teknologien kan testes i forsvarlige rammer vurderes derfor som vesentlig.

Vår vurdering av hva KI Norges rolle bør være samsvarer i stor grad med rollen KI Norge allerede er tiltenkt. Vi vurderer videre at KI Norge også kan ta rollen å følge opp mange av tiltakene over:

- Ha oversikt over, vurdere og anbefale felles standarder og rammeverk for sikker bruk av KI
- Identifisere utfordringer forbundet med ulike regelverk og praksis for KI (inkl. behandling av sensitive data) og iverksette egnede tiltak opp mot etater og departement.
- Gjøre faglige vurderinger av hvilke data som bør tilgjengeliggjøres ut fra en helhetlig vurdering, og eventuelt bistå i prosessene med å avklare rettigheter
- Bistå sektormyndigheter som ønsker å utvikle domenespesifikke språkmodeller
- Bistå anskaffelsesmiljøet i DFØ med å legge til rette for sikker og ansvarlig bruk av KI i offentlige anskaffelser
- Gjennomføre samfunnsøkonomiske vurderinger av investeringer knyttet til språkmodeller som berører mange sektorer (frikjøp av innhold, investeringer i felles infrastruktur, utvikling av norske språkmodeller mv.)
- Vurdere samlede kompetansebehov og tiltak.

1 Innledning

Kunstig intelligens (KI) er i ferd med å endre grunnleggende forutsetninger for både offentlig og privat sektor og samfunn i Norge. I nasjonal digitaliseringsstrategi «Fremtidens digitale Norge» er det en ambisjon om at Norge skal utnytte mulighetene som ligger i kunstig intelligens, og fram mot 2030 vil regjeringen arbeide for å få på plass en nasjonal infrastruktur for kunstig intelligens. Nasjonal KI-infrastruktur består av grunn- og språkmodeller trent på norsk og samisk innhold, samt nødvendig tungregningskapasitet levert av nasjonal infrastruktur for tungregning. I tillegg skal KI-infrastrukturen ha en organisatorisk ramme som forvalter og formidler tilgangen til den.

Det er mange aktører på tvers av sektorer og nivåer som engasjerer seg i utvikling og utnyttelse av KI. Digitaliserings- og forvaltningsdepartementet (DFD) ønsker at initiativer, tiltak og utvikling som pågår, eller planlegges startet opp i regi av offentlig sektor, eller som del av et offentlig-privat samarbeid, skal bli best mulig koordinert og trekke i samme retning.

For å få full nytte av en KI-infrastruktur er det behov for å få en oversikt over økosystemet for utvikling og bruk av språkmodell-basert KI i Norge. Dette økosystemet består blant annet av aktører som trener grunnmodeller, driver med fintrening av modeller (for egen bruk eller som tilbud til flere), tilpasninger ved hjelp av RAG-teknologi eller liknende teknologier, og/eller utvikling av applikasjoner, som KI-assistenter, taleroboter og agenter.

Det nasjonale økosystemet vil også omfatte den nasjonale organiseringen for håndheving av den norske KI-loven som vil bygge på EUs KI-forordning: tilsyn, regulatoriske sandkasser og ny nasjonal arena for innovativ og ansvarlig KI i Digitaliseringsdirektoratet (KI Norge), samt relevante forskningsmiljøer innenfor kunstig intelligens.

1.1 Kartlegging av økosystemet

DFD har gitt Agenda Kaupang og Simula Research Laboratory i oppdrag å kartlegge økosystemet for språkmodellbasert kunstig intelligens. Manglende oversikt over aktører og roller i dette økosystemet kan føre til fragmentering, dobbeltarbeid og lav effektivitet. Norge står overfor en digital transformasjon der det er avgjørende at offentlige ressurser investeres smart og strategisk. Forståelsen av økosystemet for språkmodeller er derfor en forutsetning for å realisere regjeringens ambisjoner om en koordinert og ansvarlig utvikling av KI.

Hensikten er å danne et bilde av hvem som gjør hva, hvordan aktører samhandler, og hvilke behov og forventninger som finnes knyttet til KI-infrastruktur og språkmodeller.

Kartleggingen skal bidra til å:

- Identifisere de mest sentrale aktørene og deres rolle i verdikjeden for språkmodeller
- Forstå samhandlingsmønstre og grad av koordinering mellom aktørene i økosystemet
- Kartlegge eksempler på offentlig-private samarbeid, formål og organisering
- Undersøke hvordan språkmodeller brukes i praksis
- Få innspill på behov for regnekraft, teknisk kompetanse og støttefunksjoner
- Kartlegge forretningsmodeller og tilgangsbetingelser
- Peke på utfordringer og barrierer som aktørene opplever

I tillegg til kartleggingen ble vi også bedt om å bidra med våre egne betraktninger og anbefalinger om tiltak for bedre samordning og ressursbruk når det gjelder språkmodellbasert kunstig intelligens.

1.2 Gjennomføringen av kartleggingen

Kartleggingen bygger på to hovedkilder: dokumentanalyse og semi-strukturerte intervjuer med nøkkelaktører i økosystemet for språkmodeller. Litteraturgjennomgangen omfattet offentlige dokumenter, rapporter fra sentrale organisasjoner, forskningsartikler, medieomtale og informasjon fra utviklerplattformer. Intervjuene ble gjennomført med aktører fra hele verdikjeden for språk-modeller; FoU-miljøer, utviklere, infrastrukturtilbydere, store brukere, startups og myndigheter. Denne kartleggingen forsøker ikke å gi et komplett bilde av økosystemet for språkmodeller. Vi har benyttet også av en snøball-metodikk for å identifisere de viktigste informantene. (Snowball-sampling[1]), Hovedfokus har vært mot aktører som utvikler språkmodeller og leverandører av infrastruktur. Omfanget av virksomheter som leverer applikasjoner basert på språkmodeller eller tar i bruk KI i egen virksomhet er svært stort, så her dekker kartleggingen kun en minimal andel av totalen. Utvalgsmetoden kan ha ført til skjevheter i det bildet som dannes, men vi håper vi gir et relativt dekkende bilde av økosystemet.

I arbeidet er KI brukt til gjennomgang av litteratur og til utarbeidelse av intervjureferater. For å sikre kvalitet er alt materiale dobbeltsjekket mot originalkilder og manuelt kvalitetssikret. Bruken av GPT har skjedd i en lukket versjon uten trening på prosjektdata, med nøye utformede prompt for å redusere risiko for feil og sikre pålitelighet. Nærmere beskrivelse av datainnsamling og metode finnes i vedlegg 1.

1.3 Rapportens oppbygning

Rapporten gir en helhetlig framstilling av økosystemet for språkmodellbasert kunstig intelligens i Norge. I kapittel 2 gis en gjennomgang av økosystemet for språkmodellbasert KI, med forklaringer på hva som menes med store språkmodeller, en beskrivelse av verdikjeden, en oversikt over sentrale norske aktører, innsatsområder, bruks- og utviklingsformer, aktuelle forretningsmodeller og nødvendig teknisk infrastruktur. Kapittel 3 tar for seg samarbeid og nettverk, herunder viktige samarbeidspartnere, organisasjons- og samarbeidsformer og etablerte samarbeidsarenaer.
I kapittel 4 beskrives barrierer som hemmer arbeid med språkmodeller. Kapittel 5 belyser offentlig innsats og politikkutforming, med fokus på organiseringen av KI-arbeidet nå og framover og behovene som aktører uttrykker for offentlig støtte og tiltak. Til slutt presenterer kapittel 6 Agenda Kaupangs egne vurderinger og anbefalinger, der det foreslås konkrete regulatoriske, økonomiske, tekniske, organisatoriske og pedagogiske tiltak for å fremme utvikling og bruk av språkmodellbasert KI i Norge.

[1] Snowball sampling er en utvalgsmetode der man rekrutterer nye deltakere gjennom eksisterende deltakere, som henviser til personer i sitt nettverk. I denne sammenheng startet vi et utvalg «selvskrevne» informanter og ba disse komme med innspill på nye informanter.

2 Økosystemet for språkmodellbasert KI

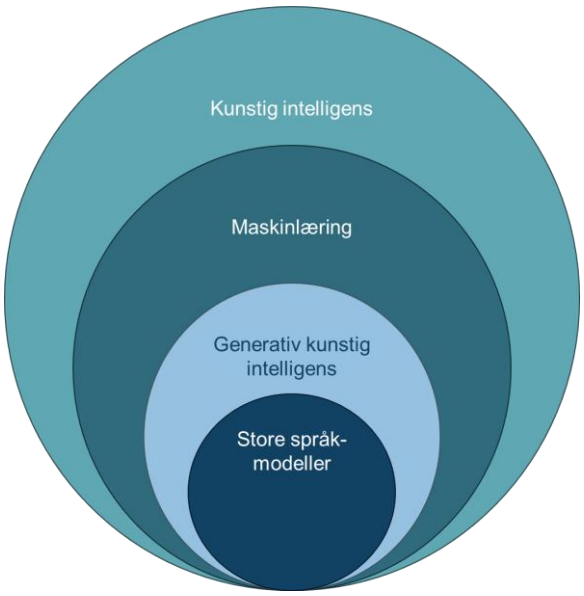
Dette kapittelet gir en helhetlig framstilling av økosystemet for språkmodellbasert kunstig intelligens i Norge. Først forklares hva som menes med store språkmodeller og hvordan de inngår i det bredere KI-landskapet. Deretter presenteres en forenklet modell av verdikjeden for språkmodeller, med beskrivelse av innsatsfaktorer, utviklingsledd og anvendelser.

Med dette som utgangspunkt gir kapitlet en bred kartlegging av det norske økosystemet. Det omfatter en oversikt over sentrale aktører og deres roller, innsatsområder for bruk av språk-modeller, samt hvordan modeller utvikles og anvendes i praksis. Videre belyses hvilke forretnings-modeller som finnes, og hvordan teknisk infrastruktur – som datakraft, datasett og kompetanse – danner grunnlaget for utvikling og drift av språkmodellbaserte løsninger i Norge.

2.1 Hva menes med store språkmodeller?

Denne rapporten er avgrenset til økosystemet for store språkmodeller i Norge (kjent som Large Language Models, LLM). Det betyr at den dekker kun en liten av fagfeltet kunstig intelligens, som spenner mye bredere.

Figur 1 viser en svært forenklet skisse av språkmodeller som en underkategori av generativ KI. Generativ KI er igjen en underkategori av maskinlæring, som igjen er en del av feltet som kalles kunstig intelligens. Det er mange ulike definisjoner av kunstig intelligens, men vi viser her til EUs ekspertgruppe definisjon: *Kunstig intelligente systemer utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål.*[2] Kunstig intelligens kan altså også inkludere tradisjonelle regelbaserte systemer, selv om det i dagligtale ofte viser til de nyeste teknologiene innenfor feltet.



Figur 1: Forholdet mellom kunstig intelligens og store språkmodeller

[2] Independent High Level Expert Group set up by the European Commission (2019): A definition of AI: Main capabilities and diciplines

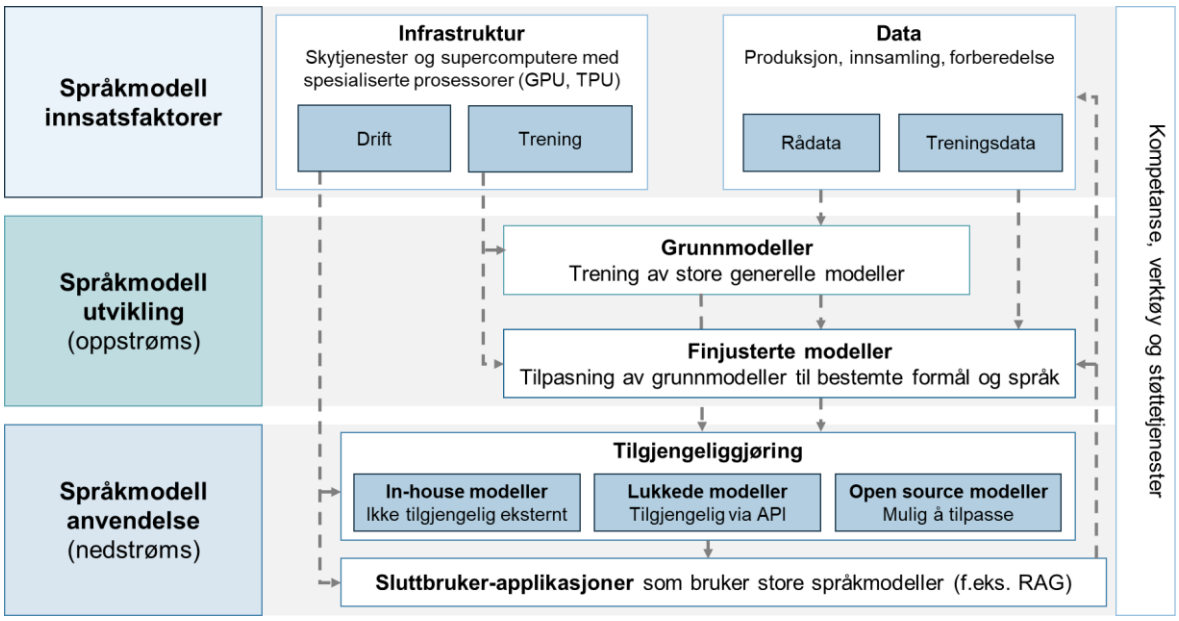
Kunstig intelligens inkluderer mange underfelt som utvikler seg kontinuerlig, og som har vist allsidige bruksområder og samfunnsverdi. Disse underfeltene inkluderer *datasyn* (f.eks. å la droner autonomt oppdage feil i kraftlinjer), *tidsseriefremskrivning* (f.eks. prognoser for energietterspørsel og -tilbud) og *anbefalingssystemer* (f.eks. å identifisere hvilke kurs som bør tilbys en student basert på tidligere interesser).

Maskinlæring er en underkategori av kunstig intelligens og dekker en rekke ulike teknikker, der reglene utledes fra de dataene systemet trenes på, i motsetning til regelbaserte systemer der reglene er gitt av mennesker, ofte basert på eksperterfaring, forretningslogikk eller regelverk. Aktuelle teknikker som har blitt anvendt over lengre tid inkluderer veiledet læring, ikke-veiledet læring, forsterket læring og dyplæring. Generativ KI[3] er en underart av dyplæring.

De siste årene har generativ KI og store språkmodeller fått særlig mye offentlig oppmerksomhet. Store språkmodeller er en underklasse av generativ KI. Store språkmodeller tilbyr ofte tilgang til andre grunnmodeller eller generative kapabiliteter, slik at brukere kan skape egne bilder, oppdage og klassifisere objekter, generere eller tolke tale, og lage videoer fra bilder. Store språkmodeller kan klassifiseres (i stigende rekkefølge av kompleksitet) i modeller som bare kan forstå språk, modeller som i tillegg kan generere språk, og modeller som i tillegg kan svare på instruksjoner.

2.2 Verdikjeden for språkmodellbasert kunstig intelligens

Utvikling og bruk av språkmodellbasert kunstig intelligens skjer i en verdikjede. Figuren under viser en overordnet modell av verdikjeden for språkmodellbasert kunstig intelligens. Denne er videreutviklet av Agenda Kaupang med utgangspunkt i en rapport fra Copenhagen Economics (Abecasis og de Michiels, 2024) som igjen er bygget på en rapport fra britisk konkurransemyndigheter.[4]



Figur 2: Forenklet modell av verdikjeden for store språkmodeller. Kilde: Copenhagen Economics, videreutviklet av Agenda Kaupang

[3] Generativ kunstig intelligens (generativ KI) er en fellesbetegnelse på en type kunstig intelligens som kan lage unikt innhold – både tekst, lyd, bilder og video – ved å få enkle instrukser i naturlig språk. (Ellen Strålberg, Teknologirådet)

[4] AI Foundation Models Update paper (April 2024). Competition and Markets Authority. Update Paper

Modellen viser verdikjeden for store språkmodeller fra og er strukturert i tre hovednivåer:

- Språkmodell-innsatsfaktorer (infrastruktur og data)
- Språkmodell-utvikling (grunnmodell og finjustering)
- Språkmodell-anvendelse (tjenester og applikasjoner)

Kompetanse, verktøy og støttetjenester understøtter alle nivåene.

Innsatsfaktorene for språkmodeller er vist i form av (fysisk) infrastruktur og data. Infrastruktur viser til den fysiske infrastrukturen som brukes enten for å trene språkmodeller eller for å drifte dem. Dette inkluderer både skytjenester eller høyytelsesmaskiner (HPC). Både infrastruktur for drift og trening av språkmodeller baserer seg ofte på såkalte GPUer.[5] Selv om det er noen fellestrekk mellom infrastruktur for trening og drift velger vi å omtale de hver for seg i rapporten, fordi det er ulike krav og ofte ulike aktører.

Den andre innsatsfaktoren i figuren er data. Store språkmodeller er trent på enorme mengder treningsdata. Treningsdata beskriver teksteksemplene modellen har blitt eksponert for. Modeller som har sett mer relevant treningsdata for en gitt oppgave, vil kunne gi mer presise resultater knyttet til den oppgaven; dette gjelder både språk og teksttyper, som for eksempel programmeringsspråk eller juridiske tekster. Rådataene viser til alle former for tekst som brukes, eksempelvis offentlige dokumenter, åpent innhold på internett eller avisartikler. Treningsdata er i denne sammenhengen data som er bearbeidet for å brukes i trening av språkmodeller.

Deretter følger språkmodell-utvikling. Vi skiller mellom trening av grunnmodeller og finjustering. Grunnmodeller er store, generelle modeller som er laget for å dekke mange ulike formål. Dette krever enorme menger data og datakraft, i tillegg til høy kompetanse. Eksempler på grunnmodeller er ChatGPT (OpenAI), Gemini (Google) og Llama (Meta).

Når en modell er trent, kan vektene frigis som åpen kildekode (*open source*) slik at andre kan laste dem ned og bruke dem i applikasjoner, forske på dem, eller bruke dem som utgangspunkt for å trene sin egen modell videre, gjennom en prosess kalt fintrening (*fine-tuning)*. Llama og Mistral er eksempler på grunnmodeller som tilgjengeliggjøres som åpen kildekode. Mange av de mest brukte språkmodellene (som Chat-GPT) er i hovedsak lukkede modeller, som ikke er tilgjengelig for fine-tuning. Disse tilbys direkte via et web- eller chatte-grensesnitt, slik at man kan sende egne data og få et resultat tilbake.

Nederst i illustrasjonen finner vi språkmodell-anvendelse, der modellene brukes i praksis.
Dette kan være in-house-modeller (kun interne), lukkede modeller (tilgjengelig via API) eller åpne modeller (open source, som kan tilpasses).

Sluttbrukertjenester kan være applikasjoner som bygger på store språkmodeller, inkludert løsninger som kombinerer modeller med søk i dokumenter (RAG-løsninger)[6]. Samlet illustrerer modellen hvordan innsatsfaktorer, utvikling og bruk henger sammen i en helhetlig verdikjede for språkmodeller.

Modellen må ses på som en forenkling blant annet fordi det i praksis vil være flere relasjoner og feedback-loops mellom de ulike delene, og fordi den ikke inkluderer relasjoner med andre digitale løsninger slik som interne databaser, systemer og applikasjoner. Den er likevel nyttig for å vise hvor de ulike aktørene har sin plass.

[5] GPU (Graphics Processing Unit) er en spesialisert prosessor utviklet for å håndtere grafikk og parallellberegninger. TPU er en prosessortype spesifikt for maskinlæring.

[6] En RAG-modell (Retrieval-Augmented Generation) er en språkmodell som kombinerer generering av tekst med oppslag i en ekstern kunnskapskilde, slik at svarene bygger både på modellens språkforståelse og oppdatert informasjon.

2.3 Sentrale aktører i det norske språkmodell økosystemet

Våre informanter er relativt konsistente om hvilke aktører som har sentrale roller i det norske økosystemet for språkmodell basert KI. På tvers av intervjuer nevnes spesielt Nasjonalbiblioteket, flere universitetsmiljøer (særlig UiO og NTNU), samt samarbeidsprosjekter som NorwAI og infrastrukturaktører som Sigma2 og LUMI-konsortiet som sentrale aktører.

Nasjonalbiblioteket blir omtalt av mange som den mest sentrale enkeltaktøren i utviklingen av norsktrente språkmodeller. Dette begrunnes med deres rolle som dataleverandør, modelltrener og samarbeidspartner for både offentlige og private aktører. Nasjonalbiblioteket blir beskrevet som den mest etablerte og ressurssterke aktøren for språkmodellutvikling rettet mot norsk språk. De har trent egne modeller, samarbeidet om prosjekter som Mimir[7] og delt datasett med andre miljøer. I litteraturstudien bekreftes dette bildet. Nasjonalbiblioteket oppgis å spille en nøkkelrolle som tilrettelegger for språkteknologisk utvikling i Norge.

Innen akademia fremheves Universitetet i Oslo (UiO) og NTNU som de viktigste forsknings-miljøene, blant annet gjennom miljøene Language Technology Group ved UiO og NorwAI ved NTNU. Disse har samarbeidet tett med Nasjonalbiblioteket, og har ansvar for evaluering og utvikling av norsktrente språkmodeller i flere prosjekter. Informanter beskriver disse miljøene som de mest erfarne og aktive i forskningen på språkmodellteknologi i norsk kontekst.

Sigma2 omtales av flere informanter som en sentral aktør for tilgang til regnekraft i Norge, og som en fasilitator for forskning på språkmodeller gjennom infrastruktur og støtte. De tilbyr regne-ressurser til utviklingsmiljøer som NorwAI og UiO, og samarbeider blant annet om prosjekter som Mimir. Også LUMI-superdatamaskinen i Finland trekkes frem i denne sammenheng, spesielt av akademiske aktører som har trent modeller på storskala infrastruktur. I litteraturstudien nevnes Sigma 2s rolle som leverandør av nasjonal kontrollert regnekraft som en forutsetning for utvikling av språkmodeller. Det påpekes et økende gap mellom behov og kapasitet, og et behov for investeringer i offentlig regneinfrastruktur (Forskningsrådet, 2024).

Innen offentlig forvaltning og regulatorisk tilrettelegging nevnes aktører som Digitaliserings-direktoratet, Helsedirektoratet og Skatteetaten som aktører som er engasjert i bruk språkmodeller offentlig sektor.

På infrastruktursiden er også Norwegian AI Cloud (NAIC) en aktør som har ambisjoner om å bygge nasjonal kapasitet og støtte utvikling av KI løsninger som ivaretar hensynet til suverenitet og demokratisk kontroll.

Flere informanter nevner også NRK som offentlig mediaaktør, og Schibsted som en viktig kommersiell medieaktør. Disse har bidratt med datasett og erfaring fra anvendelse i medie-produksjon, og har vært aktive partnere i utviklingssamarbeid.

I litteraturstudien finner vi også henvisning til andre aktører. Ifølge AI Report Norway 2025 finnes det over 350 norske selskaper som tilbyr KI-verktøy (RankmyAI, 2025). Språkrådet har blant annet bidratt til arbeidet med vurdering av språkmodellbruk i helsesektoren (Helsedirektoratet, 2025). SimulaMet, med tilknytning til OsloMet, oppgis også å ha bidratt med teknisk og vitenskapelig ekspertise knyttet risikovurdering og tilpasning av språkmodeller til helsesektoren (Helsedirektoratet, 2025). Digdir[8] har utviklet og driftet flere KI-løsninger i offentlig sektor bl.a.

[7] Mímir, et prosjekt hvor forskningsmiljøer har trent en rekke nye språkmodeller og vurdert betydningen aviser og bøker under opphavsrett kan ha for denne typen kunstig intelligens.

[8] Digitaliseringsdirektoratet. Oversikt over prosjekter i offentlig sektor (Nettside: https://data.norge.no/kunstig-intelligens)

inkludert Altinn-assistenten – en språkmodell-basert chatbot trent på dokumentasjon fra Altinn-plattformen.

Oppsummert, fremstår det norske økosystemet som dominert av ganske få, i hovedsak offentlige og akademiske aktører.

2.4 Innsats for bruk av språkmodellbasert KI

Informantene beskriver et bredt spekter av aktiviteter knyttet til språkmodellbasert KI. Aktivitetene fordeler seg langs en akse fra forskning og utvikling av egne modeller, til bruk og tilpasning av eksisterende løsninger for spesifikke behov. Også i litteraturstudien nevnes eksempler på aktiviteter som kan knyttes til disse kategoriene. Opplistingen under oppsummerer funn både fra intervjuer og litteraturstudien. Det er viktig å påpeke at noen av aktørene opererer i flere ledd av verdikjeden.

Tilrettelegging og drift av infrastruktur for språkmodellutvikling

Enkelte aktører har som hovedoppgave å legge til rette for at andre kan utvikle eller bruke språk-modeller gjennom tilgang til regnekraft og teknisk støtte.

- Sigma2 tilbyr tungregning og lagring, og støtter forskningsprosjekter med teknisk infrastruktur og veiledning i bruk av superdatamaskiner som LUMI.
- Norwegian AI Cloud (NAIC) er et fellesprosjekt mellom en rekke aktører[9] som har som mål å gi tilgang til infrastruktur for KI og maskinlæring for mindre aktører og forskningsmiljøer.
- Nasjonalbiblioteket leverer infrastruktur i form av treningsdata. Gjennom Språkbanken tilbyr de omfattende mengder norske tekst- og taledata som er kritiske for trening av språkmodeller tilpasset norsk språk og kontekst.
- Telenor AI factory er i en tidlig fase, men satser på å levere norsk-eid og -kontrollert infra-struktur. De leverer per nå mest til oppstartsselskaper og Telenor selv, men ser offentlig kunder med sensitive data som en viktig kundegruppe framover. Telenor forvalter allerede store mengder kritisk infrastruktur og er underlagt sikkerhetsloven.

Utvikling av og forskning på språkmodeller

Flere aktører har som hovedaktivitet å utvikle og trene språkmodeller, ofte med utgangspunkt i åpne modeller som Mistral og LLaMA, kombinert med norske datasett.

- Nasjonalbiblioteket trener også egne modeller. De jobber også med evaluering og distribusjon av modeller for offentlig og akademisk bruk. Nasjonalbiblioteket fungerer dermed både som infrastrukturleverandør og forskningsaktør, blant annet gjennom etableringen av en egen KI-lab og eksperimenter med generativ KI på egne samlinger. (Nasjonalbiblioteket, 2024).
- UiO (Language Technology Group) forsker på effektiv trening, evaluering og tilpasning av språkmodeller for norsk språk. De har koordinert et større partnerskap som har utviklet en samling store språkmodeller for norsk språk kalt NorLLM (også kjent som NORA.LLM).
- NorwAI (NTNU-ledet) arbeider med utvikling og forbedring av norsktrente modeller for generativ tekst, ofte i samarbeid med offentlige og private aktører. NorwAI, nevnes også som en aktør som trener og utvikler språkmodeller, og som har trent modeller tilpasset norske forhold. Dette samarbeidet har også involvert mediehus som Schibsted og NRK, samt datasentre og infrastruktur-partnere som Sigma2 og LUMI.

[9] UiO, NORA, NTNU, UiB, SIMULA, UiA, SINTEF, UiT, SIGMA2 og NORCE Research

- Bineric (privat) er et oppstartsselskap som utvikler og tilbyr KI-løsninger med sterkt fokus på personvern og lokal kontroll. De tilbyr en plattform med API-tilgang for utviklere og chatgrensesnitt for sluttbrukere. Her får man tilgang til en rekke språkmodeller fra OpenAI, Google, Anthropic m.fl., i tillegg til deres egen språkmodell NorskGPT som er optimalisert for nordiske språk.

Implementering og tilpasning av språkmodeller i løsninger

Flere av våre informanter arbeider med å implementere og tilpasse kommersielle språkmodeller til egne eller kunders behov. Eksempler er:

- Intility har utviklet sin egen interne løsning (Intility GPT) for å sikre datasikkerhet og intern effektivisering.
- Chronos har utviklet en juridisk assistent som teknologipartner til selskapet Lawai. Her brukes språkmodeller i kombinasjon med et utvalg av juridisk innhold til å tilby en KI-basert rådgiver innen arbeidsrett. Løsningen er tilgjengelig for kunder i dag og har blant annet vært behandlet i Datatilsynets regulatoriske sandkasse.
- NRK anvender språkmodeller for transkribering og gjenfinning i arkivene, og har bygget interne verktøy for journalistisk bruk basert på åpen teknologi kombinert med egne data.
- Teknologikonsulentselskaper bistår kunder i både privat og offentlig sektor med tilpasning og integrasjon av språkmodeller i eksisterende systemer. Dette inkluderer støtte til fintrening og teknisk rådgivning.
- Helse Vest IKT utvikler Klinisk NorBERT, en variant av den norske modellen NorBERT tilpasset kliniske domener og helsefaglig språkbruk. (Helsedirektoratet, 2025)
- Helsedirektoratet og samarbeidspartnere benytter RAG i kombinasjon med internasjonale språkmodeller f.eks., i en applikasjon som Helsesvar (Helsedirektoratet, 2024a).
- Egde Consulting har utviklet Agder KI, en KI-plattform som kombinerer chatbot-teknologi, generativ KI og språkmodeller for å levere intelligente og nyanserte svar på brukerforespørsler.

Effektivisering og forbedring av tjenester

Våre informanter viser også til eksempler på hvordan språkmodeller brukes i virksomheter for å effektivisere arbeid, forbedre interne prosesser og utvikle nye funksjoner for tjenestetilbydere.

- Helsedirektoratet arbeider med tilrettelegging for trygg bruk av språkmodeller i helsesektoren, særlig i diagnostikk og dokumentasjon.
- NRK bruker språkmodeller for å strukturere, klassifisere og personalisere innhold, og har utviklet interne løsninger for både journalister og publikum.
- Divvun/UiT utvikler KI-verktøy for talegjenkjenning og talesyntese for samiske språk, og integrerer språkmodeller i praktiske tjenester som API-er og stemmegeneratorer.

Kompetansebygging og formidling

Flere av våre informanter har som kjerneoppgave å styrke kunnskap, forståelse og bruk av språk-modeller i samfunnet.

- Nemonoor og Digital Norway fokuserer på kompetanseheving, utvikling av kurs og veiledere, samt støtte til virksomheter som vil ta i bruk språkmodeller. De fremmer forståelse for verdien av å bruke egne data i KI-løsninger.
- NORA arbeider med koordinering og støtte til utdanning, innovasjon og forskning innen KI og språkmodeller.

Fra litteraturstudien ser vi at de samme kategoriene og aktørene nevnes. I tillegg pekes det på anvendelse og kommersialisering av språkmodellbaserte løsninger som viktige aktiviteter for aktører i økosystemet. Omtalen omhandler alle typer språkmodeller, men det er få eksempler på anvendelse og kommersialisering av norske språkmodeller. I kartleggingen av bruk av språk-modeller i offentlig sektor (Rambøll, 2025) påpekes det statlige virksomheter bruker hovedsakelig åpent tilgjengelige og kommersielle språkmodeller som ChatGPT og Microsoft Copilot. Bruken er ofte begrenset til enkeltpersoner eller testgrupper i organisasjonen, og anvendelsen skjer oftest i form av generelle kontorstøtteoppgaver, som tekstproduksjon, språkvask, idéutvikling og søk i dokumenter. En ser også at bruk av KI er på vei pilotprosjekter innen saksbehandling, arkiv og anskaffelser.

2.5 Bruksområder for språkmodeller

Våre informanter beskriver et mangfold av anvendelsessammenhenger for språkmodeller, som spenner over ulike sektorer og deler av verdikjeden – fra datafangst og informasjonsbehandling til sluttbrukerrettede tjenester. Bruken kan kategoriseres i fem overordnede bruksområder: media og innholdsproduksjon, offentlig sektor og forvaltning, helse og omsorg, forskning og språkteknologi-utvikling, samt næringsliv og kundespesifikke løsninger.

Media og innholdsproduksjon

Innen mediebransjen benyttes språkmodeller i stor skala til å håndtere store mengder tekst, lyd og video. Formålet er både automatisering, effektivisering og innovasjon i måten innhold produseres, organiseres og distribueres på.

- NRK bruker språkmodeller til transkribering av sitt arkiv, metadata-generering, stemme-gjenkjenning, og personalisering av innhold. De har også testet løsninger hvor brukere kan «chatte» med arkivet for å finne relevant informasjon.
- En tidligere leder i et stort mediekonsern beskrev bruk av norsktrente språkmodeller i journalistisk produksjon og automatisert tekstgenerering i aviser.
- Nasjonalbiblioteket samarbeider med mediehus om verktøy for transkribering og klassifisering av tekst og tale.

Offentlig forvaltning

Språkmodeller anvendes i offentlig sektor for effektivisering av saksbehandling, publikumsdialog, og internt tekstarbeid.

- Helsedirektoratet vil utforske bruken av språkmodeller i utvikling av beslutningsstøtteverktøy, oppsummering av journaler, og støtte for helsepersonell. Internt benyttes modeller som Copilot til effektivisering av skrivearbeid og dokumentproduksjon.
- Språkmodeller brukes som støtteverktøy i offentlig forvaltning. Aktører som Nemonoor og Digital Norway har bistått med utvikling av veiledere og rammeverk for trygg bruk i statlige virksomheter.
- Skatteetaten har et innovasjonsteam som blant annet har testet konsepter for å strukturere informasjon fra dokumenter, lagd kunnskapsgrafer for å identifisere interne og eksterne lovhenvisninger, og gjort tester av hallusinasjonstendens og bruk av små språkmodeller.
- NAV bruker blant annet språkmodeller for å bistå arbeidsgivere med å formulere bedre stillingsannonser på arbeidsplassen.no[10]

[10] https://arbeidsplassen.nav.no/enklere-a-skrive-gode-kvalifikasjoner

Fra litteraturstudien nevnes bruk som:

- Informasjonstilgang og søk: LLM-basert søk i data.norge.no gjør det mulig å koble bruker-spørsmål i naturlig språk til relevante datasett gjennom en «Orakelkatalogen»-prototype som ble satt i produksjon i 2024 (DataNorge og AI, 2024).
- Borgerservice og saksstøtte: Altinn-assistenten, basert på GPT-4o og RAG-arkitektur, gir teknisk veiledning til utviklere (Digdir, 2023, vedlegg).

Helse og omsorg

Innen helsesektoren anvendes språkmodeller for å støtte klinisk arbeid.

- Bruken inkluderer transkribering av journalsamtaler, oppsummering av medisinsk informasjon, og utvikling av KI-assistenter for helsepersonell.
- Helsesektoren ser også på bruk av språkmodeller for støtte i beslutningsprosesser og for utvikling av sektortilpasset semantikk og begrepsapparat, spesielt i samarbeid med Nasjonal-biblioteket og forskningsmiljøer.

Fra litteraturstudien nevnes:

- Språkrelaterte verktøy: Store språkmodeller kan brukes til automatisk tekstproduksjon, fritekststrukturering i pasientjournaler, helsefaglig koding, klarspråk, oversettelse, talegjenkjenning, triagering av pasienthenvendelser og kliniske samtaleroboter (Helsedirektoratet, 2025).
- Gjennom prosjektet Enklere tilgang til informasjon (ETI), utvikles en generativ KI-løsning som benytter RAG for å hente og generere informasjon basert på offentlige kilder, særlig rettet mot familier med alvorlig syke barn (Rambøll 2025).
- Administrasjon: Bruk av KI i dokumentasjon, planlegging og ressursstyring inngår som del av den felles KI-planen for helsesektoren (Helsedirektoratet, 2024b).

Forskning og språkteknologiutvikling

Mange av aktørene innen akademia og språkteknologi bruker språkmodeller som forskningsobjekt og som basis for nye modeller tilpasset norsk språk og kultur.

- UiO, NTNU/NorwAI og Nasjonalbiblioteket utvikler, trener og evaluerer modeller for bruk i ulike domener, som akademia, offentlig forvaltning og næringsliv.
- Modellene ble også anvendt i Mimir-prosjektet, hvor de ble testet på forskjellige norske datasett for å evaluere semantisk forståelse og språklig presisjon.
- Divvun/UiT har utviklet egne samiske språkmodeller, spesielt for talegjenkjenning og syntese.

Næringsliv og kundespesifikke løsninger

Flere teknologiselskaper og konsulentvirksomheter bruker språkmodeller for å utvikle skredder-sydde løsninger til kunder i ulike sektorer.

- Intility har satt opp Intility-GPT for å øke kontrollen over hvordan data prosesseres i løsningen og tilbyr det til sine kunder.
- Mange konsulentselskaper integrerer språkmodeller i løsninger for alt fra kundedialog og dokumentanalyse til domenespesifikke KI-assistenter, ofte kombinert med andre teknologier som søkemotorer og databaser. Blant annet viser et teknologiselskap vi intervjuet til at de utviklet en juridisk assistent ved å kombinere egne dokumenter, språkmodeller og en rekke andre teknikker for å presisjon og relevans i svarene.

- Enkelte utvikler også egne fintrente modeller på mer domenespesifikke data, særlig i sektorer som krever høy grad av presisjon og tilpasning. Dette omfatter områder som jus, finans, helse og media.
- Kundeservice og salgsstøtte: Generative KI-assistenter brukes i kundedialog og teknisk support (RankmyAI, 2025).
- Produktutvikling: Språkmodeller integreres i bransjespesifikke løsninger innen maritim sektor, finans, media og HR-teknologi (RankmyAI, 2025).

2.6 Bruk og utvikling av språkmodeller

Våre informanter rapporterer et variert bilde når det gjelder hvilke språkmodeller de benytter eller utvikler, og hvor modellene stammer fra. Valg av modelltype – åpen, kommersiell eller egenutviklet – avhenger av formål, ressurser, krav til kontroll og teknisk kompetanse. Mange kombinerer ulike modeller og tilnærminger for å balansere kostnader, fleksibilitet og sikkerhet. I det store bildet er det de lukkede internasjonale modellene, særlig Chat-GPT, som brukes i applikasjoner. Det er kun Whisper-modellen som trekkes fram av informantene som modeller som har blitt brukt for utvikling av løsninger.

Bruk av lukkede modeller

De aller fleste informantene som utvikler sluttbrukerløsninger, baserer disse på kommersielle lukkede språkmodeller. Dette omfatter konsulentselskaper, mediaaktører, men også offentlige virksomheter. GPT-modeller fra OpenAI, Gemini fra Google, og Copilot fra Microsoft brukes både via åpne API-er og gjennom skytjenester som Azure OpenAI Service eller Google Cloud. Bruken av kommersielle modeller begrunnes med høy ytelse, enkel tilgang og bred funksjonalitet, men flere aktører uttrykker bekymring for manglende innsyn, datasikkerhet og kulturell tilpasning.

Bruk av åpne modeller

Flere av våre informanter innen akademia oppgir at de primært benytter eller videreutvikler åpne språkmodeller i sin forskningsvirksomhet. Modeller som Mistral, LLaMA, Whisper og andre brukes som basis for utvikling og tilpasning. Åpne modeller foretrekkes fordi de gir tilgang til vektverdier, treningsteknikk og grunnlagsdata, noe som er avgjørende for å kunne evaluere og justere ytelsen, særlig i norske kontekster. Modellene videreutvikles ofte gjennom fintrening på norske datasett, som aviser, bøker, offentlige dokumenter, lydopptak og taleinnhold fra mediearkiver. Flere miljøer kombinerer ulike åpne flerspråklige modeller og tilpasser dem til norsk språk gjennom videre trening. Åpne modeller er fleksible og gir mulighet for tilpasning. De er også mer transparente og gir større mulighet for innsikt og kontroll.

Bruk av norske språkmodeller

Flere av informantene utvikler eller har utviklet egne språkmodeller fra bunnen av eller ved videre-utvikling av åpne modeller.

Nasjonalbiblioteket har utviklet flere egne norsktrente modeller for generering, transkribering og klassifisering, basert på digitaliserte norske tekster. Flere informanter trekker fram at de har testet eller brukt Whisper-modellene som er laget for talegjenkjenning på norsk. NRK bruker både kommersielle modeller (som Gemini) og spesialtrente versjoner av Whisper, som er videreutviklet i samarbeid med Nasjonalbiblioteket på NRKs egne data.

Divvun/UiT har utviklet egne samiske språkmodeller, spesielt for talegjenkjenning og syntese. De bruker åpne modeller som Whisper som grunnlag, men har også trent egne stemmer og transkripsjonsmodeller basert på samisk data.

Statistikk på nedlastning av norske modeller

Hugging Face regnes som den ledende plattformen for åpne språkmodeller, og fungerer som vertstjeneste for alle typer maskinlæringsmodeller og datasett. Disse er tilgjengelige for nedlasting. De miljøene vi har identifisert som norske språkmodell-utviklere bruker Hugging Face til å gi tilgang til modellene sine. Ut fra brukerprofilene på Hugging Face har NorwAI tilgjengeliggjort 9 modeller, Bineric 7 modeller, og NORA.LLM 5 modeller.[11] Nasjonalbiblioteket har en stor samling av modeller innen både tekst og tale, i tillegg til datasett tilgjengelig for nedlastning. På grunn av hvordan åpen kildekode brukes, er det vanskelig å finne nøyaktige estimater på hvor mye modellene faktisk tas i bruk i ulike applikasjoner.[12] Statistikken vil variere en del etter akkurat når man gjør tellingen, siden mange nedlastninger skjer rett etter at en ny modell utgis.

Tabell 1: Nedlastninger av utvalgte norske språkmodeller i juli 2025. Kilde: Hugging Face

Modellsamling	Utgiver	Nedlastninger juli 2025
NorskGPT	Bineric	Ca. 1 300
nb-gpt	NB	Ca. 500 000
Normistral	NORA	Ca. 1 000
NorwAI-Mixtral	NorwAI	Ca. 5 000
Norbert	UiO	Ca. 24 000
Internasjonale modeller		
Llama	Meta	31 millioner
Mistral	Mistral AI	5 millioner

2.6.1 Betydningen av norske språkmodeller

Informantene ble spesifikt spurt om deres syn på betydningen av norske språkmodeller. Det er bred enighet blant informantene om at norsktrente språkmodeller kan ha stor betydning for både offentlig sektor, medier, næringsliv og forskning.

Mange aktører, som Nasjonalbiblioteket, UiO, NorwAI, NRK og Helsedirektoratet, peker på at språkmodeller som er trent på norsk tekst, gir bedre semantisk forståelse, kontekstilpasning og treffsikkerhet. Flere nevner at modeller trent kun på engelsk eller flerspråklige datasett ofte gir upresise eller kulturelt malplasserte svar, og at dette undergraver tilliten og nytteverdien i norsk kontekst.

Helsedirektoratets rapport om språkmodeller i helse- og omsorgstjenesten peker på at tilpasning til norske forhold er nødvendig for å sikre korrekt terminologi, juridisk presisjon og kulturforståelse i helsesektoren (Helsedirektoratet, 2025).

Mange informanter fremhever at språket vårt er en del av den digitale infrastrukturen, og at det er et offentlig ansvar å sikre at norsk språk, inkludert nynorsk og samiske språk, har en naturlig plass i

[11] Telling av modeller tilgjengelig på Hugging Face.com, juli 2025

[12] Vi bruker Hugging Face sitt mål for «total downloads last month» som en tilnærming for å måle populariteten til modeller, men påpeker at mange modeller har flere versjoner tilgjengelig, at nedlastingsstatistikken nullstilles når de oppdateres, og at én enkelt nedlasting både kan være et første eksperiment med åpne LLM-er eller en implementering i hele en virksomhet.

KI-utviklingen. Dette krever investeringer i åpne modeller, åpne datasett, fellestjenester og evalueringsverktøy, spesielt rettet mot norske språkvarianter og domener.

Teknologirådet anbefaler å tilby norske språkmodeller som en nasjonal fellestjeneste slik at offentlig sektor og næringsliv får tilgang til modeller trent på kvalitetsdata fra Norge, med fokus på datasuverenitet og kvalitet. Videre anbefaler Teknologirådet å stille formelle kvalitetskrav til norske språkmodeller for å sikre pålitelighet og unngå skjevheter (Teknologirådet, 2024).

2.7 Datakilder og treningsgrunnlag

Informantene er gjennomgående opptatt av kvaliteten og sammensetningen av treningsdata, og hvor modellene henter språklig kunnskap fra.

Selv med Språkbanken og andre kilder er tilgangen til tilstrekkelig mengde domenespesifikke norske data begrenset. Opphavsrett og begrenset tilgang til store, høyverdige tekstkorpuser (som aviser, lærebøker og offentlige dokumenter) gjør det vanskelig å bygge modeller med bred språkforståelse. Flere ønsker bedre nasjonale avtaler for deling av data til språkmodelltrening, på tvers av medier, forvaltning og minoritetsspråk.

Respondentene er gjennomgående opptatt av kvaliteten og sammensetningen av treningsdata, og hvor modellene henter språklig kunnskap fra. Tilgang til gode datakilder beskrives av mange som kritisk forutsetning både for trening av språkmodeller, og i utviklingen av applikasjoner som brukes språkmodeller.

Aktører som UiO, NorwAI og Nasjonalbiblioteket bruker datasett som inkluderer bøker, aviser, offentlige dokumenter og Wikipedia, men peker på at tilgang til høyverdige datasett er juridisk krevende på grunn av opphavsrett. Verdien av opphavsrettslig materiale i språkmodellutvikling ble testet i Mimir-prosjektet. En evalueringsrapport fra prosjektet i 2024 gir oversikt over størrelsen på ulike korpus. Det som omtales som *Base* i Tabell 2 er i hovedsak innhold uten opphavsrett, mens extended inneholder alt tilgjengelig innhold i Nasjonalbibliotekets samlinger. Tabell 3 viser ulike delkorpus som kun er opphavsrettslig beskyttet materiale. I prosjektet ble det trent 17 ulike modeller basert på Mistral-arkitekturen, basert på ulike deler av korpuset, som deretter ble testet på en lang rekke faktorer. En av hovedresultatene var at opphavsrettslig beskyttet materiale bidrar til forbedring av modellens ytelse.

Tabell 2: Antall dokumenter og ord per komplett korpus for pretrening. Kilde: Mimir-prosjektet 2024

Korpus	Dokumenter	Ord
Base	60,182,586	40,122,626,817
Extended	125,285,547	82,149,281,266

Tabell 3: Antall dokumenter og ord per delkorpus for finjustering, som kun inneholder opphavsrettslig beskyttet materiale. Kilde: Mimir-prosjektet 2024

Delkorpus	Dokumenter	Ord
Bøker	492,281	18,122,699,498
Aviser	46,764,024	9,001,803,515
bøker + aviser	47,256,305	26,078,915,554
Skjønnlitteratur	117,319	5,287,109,366
Faglitteratur	359,979	12,384,323,012
faglitteratur + aviser	42,083,532	20,340,539,068

norsk litteratur	392,887	13,352,261,605
norsk litteratur + aviser	47,156,911	22,354,065,120
oversatt litteratur	96,258	4,695,814,506

Det har vært et pågående arbeid over lengre tid for å i større grad kunne tilgjengeliggjøre opphavsrettslig beskyttet materiale for trening av norske språkmodeller. Det ble 2. september 2025 offentliggjort at regjeringen planlegger å bevilge 45 mill. kr over statsbudsjettet for 2026 for å kunne inngå en avtale med norske aviser om bruk av beskyttet innhold.[13]

Som tidligere nevnt er tilgang til treningsdata en utfordring for utvikling av gode norske språkmodeller for å håndtere bokmål. Dette gjelder i enda større grad for nynorsk og samiske språk.

Nynorske språkmodeller

Nynorsk trekkes frem som et underrepresentert og oversett språk i de fleste modeller og i utviklingsprosjekter.

Flere informanter påpeker at både treningsdata og evalueringsverktøy i hovedsak er rettet mot bokmål, og at dette fører til lavere presisjon og kvalitet på nynorsk i generative oppgaver. Selv modeller som er trent på store norske tekstkorpuser, skiller ofte ikke mellom bokmål og nynorsk, og at dette kan føre til uønsket språklig blanding eller stilbrudd i output. Enkelte informanter mener at nynorsk krever egne treningsløp, tilpasning av prompt-design, og tydeligere merking og kontroll i både trening og drift for å oppnå tilfredsstillende resultater.

Samiske språkmodeller

Samisk omtales som et særlig utfordrende område av enkelte informanter. Samiske språk har helt egne strukturelle og ressursmessige behov. Kommersielle modeller og datasett i støtter ikke disse språkene i særlig grad. Arbeidet med samisk språkteknologi er drevet av små fagmiljøer, og bygger på åpne modeller, samarbeid med allmennkringkastere i land hvor samisk språk brukes (NRK, YLE, SVT), og datainnsamling fra samiske kilder. Divvun/UiT benytter samiske taleopptak fra NRK Sápmi og finsk allmennkringkasting (YLE), i tillegg til egne opptak fra undervisnings- og mediemateriell.

Utfordringer nevnt inkluderer mangel på tilstrekkelig samiske datasett, behov for spesialiserte annotatorer, samt risiko for at samisk språk og kultur blir marginalisert i KI-utviklingen. Informantene fremholder også at det finnes skepsis i samiske miljøer knyttet til hvem som kontrollerer data og modeller, og at reell samisk medbestemmelse i teknologisk utvikling er en forutsetning for tillit.

2.8 Forretningsmodeller knyttet til bruk av språkmodeller

En verdikjede er avhengig av en inntektsstrøm. Informantene ble spurt om hvilke forretningsmodeller de baserer sin virksomhet på. Intervjuene viser et bredt spekter av forretningsmodeller, avhengig av aktørenes rolle i verdikjeden – fra grunnmodellutvikling og fintrening til integrasjon, drift og rådgivning.

Flere private teknologibedrifter og konsulentselskaper baserer seg på prosjektbasert inntektsmodell, der de utvikler og integrerer språkmodellbaserte løsninger for kunder mot timebasert betaling eller fastprisavtaler. Dette gjelder både for generiske KI-løsninger og spesialtilpassede

[13] Se f.eks. omtale om Nasjonalbibliotekets nettsted: Regjeringen vil styrke norsk kunstig intelligens med 45 millioner | Nasjonalbiblioteket

modeller. Enkelte av disse aktørene tilbyr også driftstjenester for språkmodellapplikasjoner, der løpende vedlikehold, oppdatering og infrastrukturforvaltning inngår i prisingen.

En oppstartsbedrift som utvikler egne modeller oppgir å ha en kombinasjon av API-basert lisensiering og SaaS[14]-abonnement, hvor kunden betaler for tilgang til modellene etter bruk (f.eks. per token eller per forespørsel).

Innen offentlig sektor og forskningsmiljøer er forretningsmodellene mindre inntektsdrevne og mer rettet mot prosjektfinansiering gjennom interne midler, offentlige midler, forskningsprogrammer eller EU-støtte. Her er verdien knyttet til utvikling av fellestjenester, kompetanseheving og sektor-overgripende nytte fremfor direkte kommersielt salg. Nasjonalbiblioteket, NorwAI og NAIC opererer eksempelvis i stor grad innenfor et offentlig finansieringsregime, med mål om å tilgjengeliggjøre modeller og data for hele økosystemet.

Flere informanter, både offentlige og private, peker på at partnerskapsbaserte forretningsmodeller er viktige. Dette kan innebære felles utvikling av modeller der partene bidrar med data, kompetanse eller infrastruktur, og deler på kostnader og gevinster. Eksempler inkluderer samarbeidsprosjekter mellom mediehus, forskningsinstitusjoner og teknologiselskaper for å utvikle norsktrente modeller.

Flere informanter trekker fram at for små og mellomstore aktører er finansiering en sårbar faktor, og de understreker behovet for ordninger som kan dekke oppstartskostnader ved modelltrening og -drift. Slike ordninger kan være avgjørende for at nye løsninger kommer ut i markedet. Samtidig påpeker enkelte informanter på at offentlige kunder ofte uttrykker behov for nasjonale løsninger, men velger internasjonale, kommersielle aktører i praksis – noe som utfordrer bærekraften i nasjonale forretningsmodeller.

2.9 Språkmodellers infrastruktur

I vurderingen av infrastruktur er det relevant å differensiere mellom infrastruktur knyttet til trening (grunnmodeller og fintrening) og drift (inferens) av løsninger.

Trening av grunnmodeller (foundation-models) krever store mengder regnekraft og data (tekst, bilder, mv), og gjøres i en tidsavgrenset periode. Dette skjer ved hjelp av spesialisert infrastruktur.[15] Fintrening (tilpasning til domene eller virksomhet) tar utgangspunkt **i** en eksisterende grunnmodell (open-source) og skreddersyr denne med et mindre datasett, for eksempel interne dokumenter, eller tekster på bestemte fagfelt som juss og helse. Fintrening er typisk gjerne mer begrenset i tid og kapasitet. Det er verdt å understreke at overgangen mellom trening av grunnmodeller og fintrening fra et ressursperspektiv kan være flytende: dersom datamengden og ambisjonsnivået for fintreningen er høyt vil det medføre ressurs- og kompetansebehov som i større grad ligner på det som gjelder for trening av grunnmodeller.

Drift handler om at modellen skal besvare spørsmål ute hos brukerne.[16] Da er det viktig med en infrastruktur som sørger for lav responstid, høy stabilitet og begrenset kostnad per spørring. Trafikken bør derfor kunne skaleres basert på løpende etterspørsel. Energieffektivitet, orkestrering[17] og observabilitet[18] er viktig, fordi systemet må levere svar i løpet av millisekunder

[14] Programvare leveres som en skytjeneste, der brukerne får tilgang via internett uten å måtte installere eller drifte systemet selv

[15] Dette vises til som High Performance Computing, som baserer seg på parallellprosessering. I KI-sammenheng brukes det ofte såkalt GPU-enheter som er spesielt tilpasset dette.

[16] Prosessen der man gir en språkmodell en instruks som lager et svar tilbake vises ofte til som «inferens».

[17] Orkestrering betyr at systemet automatisk styrer hvordan modellressurser aktiveres i forhold til etterspørsel.

[18] Observabilitet handler på sin side om å ha gode instrumenter for å følge med på hvordan systemet faktisk fungerer for å sikre stabil drift og forutsigbar tjenestekvalitet.

døgnet rundt til en forutsigbar pris. For tilstrekkelig små språkmodeller er det imidlertid mulig å kjøre drift lokalt på en sluttbrukerenhet, som en bærbar PC eller en mobiltelefon.

2.9.1 Tilgang til ressurser for trening av språkmodeller

Flere informanter fremhever at tilgang til kraftige beregningsressurser er avgjørende for trening og fintrening av språkmodeller. Olivia (Sigma2) og Idun (NTNU) er norske høyytelsesmaskiner (HPC) brukt blant annet til forskning på og utvikling av norsktrente modeller. LUMI (i Finland) trekkes ofte frem som den viktigste internasjonale ressursen.

Sigma2 koordinerer nasjonal tilgang til regnekraft og lagring, og deres infrastruktur brukes som plattform for eksperimentering og modelltrening av flere forskningsmiljøer. Videre koordinerer Sigma2 den norske deltagelsen og bruken av LUMI. Flere forsknings- og språkmiljøer som bruker Sigma2-maskinene i Norge, kombinert med LUMI-superdatamaskinen i Finland, påpeker at søknads- og køsystemer gir ventetid og lite fleksibilitet. Tilgangskvoter, søknadsprosesser og administrativt oppsett oppleves tidvis som tidkrevende og komplisert, særlig for mindre aktører. Flere trekker også fram at kapasiteten til Sigma2 er en begrensende faktor for trening av språkmodeller.

Trening og tilpasning av store språkmodeller krever tungregning (HPC), GPU-klynger og super-datamaskiner med høy båndbredde mellom noder og datasentre. Flere informanter og kilder peker på at slike ressurser må være tilgjengelige nasjonalt for å sikre datasuverenitet og støtte forsknings- og samfunnsnyttige formål (Forskningsrådet, 2024; Teknologirådet, 2024).

Flere rapporter peker på et markant og økende behov for tungregnekraft i Norge, særlig drevet av veksten i generativ kunstig intelligens og datadrevet forskning. Norges forskningsråd anslår at de årlige investeringene i nasjonal tungregneinfrastruktur bør økes til om lag 250 mill. kr for å komme på nivå med sammenlignbare forskningsnasjoner. Dagens kapasitet vurderes som utilstrekkelig. For å dekke det kjente behovet for tilstrekkelig nasjonal tungregnekapasitet er det estimert et behov for en årlig vekst i kapasitet på 12-15 prosent for CPU, mens behovet for GPU-kapasitet anslås å vokse 40-50 prosent årlig (Forskningsrådet, 2024).

Det store spennet i anslagene reflekterer usikkerhet knyttet til hvor avanserte modeller som skal trenes, og om man skal bygge norske grunnmodeller fra grunnen av basert på norske data, eller basere seg på eksisterende internasjonale modeller. En informant som leverer tungregnekraft rapporterer om en svært rask økning i etterspørselen siden 2022, og forventer ytterligere vekst som følge av multimodale generative modeller. I tillegg til nasjonal infrastruktur trekkes det fram internasjonale kommersielle tilbud som Google TPU Research Cloud. For mindre ressurskrevende fintrening viser en annen aktør at de har fått gratiskvoter fra Amazon, og brukt GPU-kapasitet fra Telenor AI Factory.

Samtidig er det en hurtig utvikling internasjonalt knyttet til etablering av datasentre for trening og drift av KI-løsninger. 31. juli 2025 ble planen om Stargate Norway annonsert, som er et samarbeid mellom Aker, Nscale og OpenAI. Planen er å opprette et datasenter med i første omgang 100.000 Nvidia GPUer for å betjene europeiske kunder. OpenAI oppgir i sin pressemelding at de vil ha «kontakt med regjeringsrepresentanter for å utforske muligheter for samarbeid, inkludert å fremme AI-adopsjon og hjelpe til med å oppnå Norges nasjonale AI-mål til fordel for landets innbyggere.»[19]

[19] Artikkel på OpenAI sitt nettsted: Vi introduserer Stargate Norway | OpenAI

2.9.2 Tilgang til ressurser for drift av språkmodeller

Flere av informantene beskriver at de i dag har fungerende løsninger for drift[20] av språkmodeller. Av løsninger som er drift baseres dette på som regel på kommersielle skyløsninger som Microsoft Azure, Google Cloud eller Amazon Web Services (AWS). Telenor AI-factory er i en oppstartsfase hvor de leverer norskbasert KI-infrastruktur, i første omgang til oppstartsselskaper og for å dekke Telenors egne behov.

Intility har utviklet en intern løsning, «Intility GPT», som bygger på OpenAI-modeller via Microsoft Azure, men som er tilrettelagt for høy datasikkerhet å kjøre på lokale servere i kombinasjon med europeiske datasenter.

Flere fremhever at driftsfasen er teknisk og økonomisk krevende, særlig ved høyt brukervolum. Tekniske behov i driftsfasen omfatter rask responstid, høy oppetid og sikker databehandling. Kostnadene for skybaserte API-kall kan bli betydelige over tid, og dette oppgis som en barriere for bred utrulling av egne modeller. Særlig oppstartsaktører peker på at drift er den mest kostbare fasen i hele verdikjeden, og at det krever skalerbar infrastruktur for å kunne tilby tjenester konkurransedyktig.

Kommersielle skyplattformer trekkes fram som vanlig å bruke for løsninger som baserer seg på språkmodeller. Skyplattformene har ofte gode verktøy for å skalere basert på løpende behov. Skyplattformene brukes gjerne også for andre deler av en virksomhets data, løsninger og prosesser, og derfor er det ofte nærliggende å utvikle KI-løsninger integrert på disse plattformene. Det er en rekke støtteverktøy og løsninger tilgjengelig på plattformene for å understøtte utvikling og drift av KI-løsninger, noe som gjør plattformene svært attraktive. En større offentlig aktør trekker fram at de store internasjonale skyplattformene har også har langt bedre sikkerhet enn det er realistisk å bygge opp internt i egen virksomhet.

Mange informanter viser til viktigheten av at data skal behandles innenfor Norge eller EØS, det etterlyses norske eller nordiske driftstjenester for å møte regulatoriske krav og for å ha kontroll over hvor dataene lagres. Noen trekker også fram fordelen ved at datasenteret er geografisk nært der tjenesten drives fra, i tilfeller der det er applikasjoner som krever kjapp responstid.

Det synes å være ulike oppfatninger blant informantene om hva som ligger i begrepet sikker infrastruktur og datasuverenitet. Noen viser til stedet hvor datasenteret er fysisk plassert og hvorvidt dataene fysisk befinner seg i Norge eller Europa, mens andre legger vekt på nasjonaliteten til eierselskapet. Enkelte understreker at utenlandske-eide datasentre lokalisert i Norge eller Europa ikke nødvendigvis er tilstrekkelig for å sikre nasjonal kontroll over dataene.

Kompetanse trekkes også frem som kritisk i driftsfasen – ikke bare for å håndtere selve modellene, men også for optimalisering av ressursbruk og tilpasning til sluttbrukernes behov.

2.9.3 Betydningen av norsk KI-infrastruktur

På tvers av intervjuene fremheves norsk KI-infrastruktur som strategisk viktig både for teknologisk suverenitet, sikkerhet, og langsiktig verdiskaping. Flere aktører peker på at nasjonal kontroll over kritiske ressurser som regnekraft, lagring, og sikre dataressurser, er avgjørende for å beskytte sensitive data innen sektorer som helse, forsvar og andre deler av offentlig forvaltning.

[20] Ved drift av språkmodeller brukes ofte begrepet «inference» som viser til den konkrete prosessen fra én spørring mot språkmodellen til leveranse av et resultat/output.

Det trekkes frem at egen infrastruktur gir bedre forutsetninger for å overholde norsk og europeisk lovverk, redusere avhengighet av internasjonale kommersielle aktører, og sikre at investeringer i modellutvikling også gir nasjonal nytte.

Samtidig understreker mange at Norge ikke kan bygge alt alene, og at nordisk og europeisk samarbeid (eksempelvis gjennom LUMI-konsortiet) gir tilgang til infrastruktur som det ville være urealistisk å etablere nasjonalt i samme skala.

2.10 Oppsummering av det norske økosystemet

Kartleggingen viser at det norske økosystemet for språkmodellbasert KI er i rask utvikling og består av mange ulike aktører. Merk at den beskrivelsen av økosystemet som gis her ikke er en fullstendig oversikt over alle relevante deltakere i dette økosystemet, men et bilde basert på våre informanters beskrivelser av økosystemet. Vi bruker verdikjedemodellen introdusert i kapittel 2.2 for å strukturere beskrivelsen.

Hoved-kategori	Underkategori	Oppsummering av det norske økosystemet
Språkmodell innsats-faktorer	Infrastruktur for drift	Infrastruktur for drift av språkmodeller domineres i svært stor grad av internasjonale aktører som tilbyr dette som en del av store skyplattformer. Enkelte norske aktører har egne servere med GPU-er som kan drifte språkmodeller, men vi er ikke kjent med at noen av disse er brukt i større skala. Telenor har nylig etablert en tjeneste for sikker og norskbasert drift av (blant annet) språkmodeller, men er foreløpig i en oppstartsfase med et fåtall brukere.
	Infrastruktur for trening	Infrastruktur for trening av norske språkmodeller er i stor grad sentrert rundt infrastrukturen til Sigma2 og NTNU, samt tilgangen til bruk av LUMI i Finland. Mange aktører etterlyser mer kapasitet til dette. Det er også flere eksempler på at det brukes internasjonale tjenester som Googles TPU Research Cloud.
	Rådata	Rådata for trening av norske språkmodeller kommer fra en rekke kilder, herunder det åpne internett, innhold fra mediehus (NRK, Schibsted, Amedia mv), og innhold fra Nasjonalbibliotekets samlinger (både offentlig og opphavsrettslig beskyttet materiale. Opphavsrett er en sentral utfordring knyttet til viktige deler av materialet, slik som bøker og aviser. Mengden tilgjengelige data er en utfordring for nynorsk og samiske språk. For utvikling av norske talegjenkjenningsmodeller beskrives særlig transkripsjoner fra Stortinget og tekstede sendinger fra NRK m.fl. som særlig verdifulle, blant annet fordi de dekker mange ulike dialekter.
	Treningsdata	Arbeidet med å samle og kuratere treningsdata for norske språkmodeller ser ut til å i stor grad å være basert på samarbeid mellom store offentlige aktører som Nasjonalbiblioteket, UiO (særlig Language Technology Group) og NTNU (NorwAI) i samarbeid med andre offentlige og private partnere.

Språkmodell utvikling	Utvikling av grunnmodeller	Utvikling av store generelle modeller fra bunnen av beskrives som svært krevende med tanke på både datatilgang, kompetanse og infrastruktur. Denne aktiviteten domineres derfor av store internasjonale aktører (som OpenAI, Meta og Google), mens norske aktører jobber med å forbedre og tilpasse åpne internasjonale modeller (f.eks. Mistral og Llama) til å bedre speile norsk språk, samfunn og kultur.
	Finjustering av modeller	Arbeidet med finjustering og tilpasning av åpne modeller til norske forhold domineres i stor grad av store offentlige aktører som Nasjonalbiblioteket, UiO og NTNU. I privat sektor tilbyr oppstartsselskapet Bineric tjenester knyttet blant annet til finjustering av norske modeller til bestemte formål.
Språkmodell anvendelse	Tilgjengelig-gjøring av modeller	Når det gjelder tilgjengeliggjøring av modeller for bruk i applikasjoner ser det ut de aller fleste som utvikler applikasjoner basert på språkmodeller, benytter de lukkede modellene som tilgjengeliggjøres via API fra store internasjonale selskaper (Microsoft, Google mv). Noen aktører med høy kompetanse eksperimenter også med å kjøre åpne språkmodeller lokalt på egne servere. Vi finner få tilfeller der norskutviklede språkmodeller brukes i praktiske anvendelser. Unntaket ser ut til å være Nasjonalbibliotekets Whisper-modell, som enkelte oppgir å ha brukt eller testet.
	Sluttbruker-applikasjoner	Det ser ut til å være et svært stort antall selskaper og aktører som tester og utvikler løsninger basert på kunstig intelligens, inkludert språkmodellbasert kunstig intelligens. Rapporten RankMyAI (2025) har identifisert over 350 norske selskaper og løsninger basert på kunstig intelligens, som leverer innenfor et stort antall områder. 30 prosent er etablert i 2022 eller senere. Det er grunn til å tro at antallet virksomheter er langt høyere siden KI i økende grad integreres i alle typer produkter og tjenester.
Kompetanse, verktøy og støttetjenester		Kompetanse trekkes fram som en kritisk faktor av de aller fleste informanter. Av miljøer som trekkes fram som toneangivende i dag trekkes det fram forskningsmiljøene, konsulentselskaper, oppstartsselskaper, i tillegg til store teknologiselskaper/-leverandører. Flere trekker fram organisasjoner som for eksempel Nemonoor, Digital Norway og NORA som nyttige gjennom å bidra med kurs, veiledning og koordinering. Mange er også opptatt av arbeidet som gjøres med å utvikle gode verktøy for å benchmarke og evaluere KI-løsninger. Når det gjelder utvikling av KI-løsninger peker flere på at de store skyplattformene har et stort utvalg av relevante verktøy for å utvikle og produksjonssette løsningene. Dette gjør bruk av skyplattformer til et attraktivt valg for mange utviklere.

3 Samarbeid og nettverk i språkmodell-økosystemet

Dette kapittelet undersøker hvordan samarbeid og nettverk bidrar til utvikling og bruk av språkmodellbasert KI i Norge. Kapittelet gir først en oversikt over sentrale samarbeidspartnere – fra offentlige institusjoner og forskningsmiljøer til næringslivsaktører og internasjonale partnere. Deretter beskrives ulike former for samarbeid og organisering, som prosjektbaserte initiativer, formelle partnerskap og mer uformelle nettverk. Til slutt omtales etablerte samarbeidsarenaer der aktørene møtes for å utveksle kunnskap, dele ressurser og utvikle felles løsninger.

Samarbeid og nettverksbygging er en sentral del av arbeidet med språkmodeller i Norge. Aktørene opererer i et økosystem som inkluderer offentlig sektor, privat næringsliv, akademia og internasjonale partnere. Samarbeidene varierer i form og omfang – fra formelle, langsiktige avtaler og forskningskonsortier til uformelle nettverk og praksisnære testarenaer. Det deles data, infrastruktur og kompetanse på tvers av sektorer, og både nasjonale og internasjonale forbindelser spiller en avgjørende rolle for utvikling, testing og bruk av språkmodellteknologi. Samlet sett er det likevel relativt få deltakere i det norske økosystemet.

3.1 Sentrale samarbeidspartnere

Samarbeidet om språkmodeller og KI i Norge omfatter et bredt spekter av aktører, fra offentlige institusjoner og forskningsmiljøer til mediehus, teknologiselskaper og mindre oppstartsbedrifter. Alle de sentrale aktørene som ble presentert i kapittel 2 deltar i forskjellige samarbeid. Dette avsnittet gir en samlet oversikt over de mest sentrale nasjonale og internasjonale samarbeidene som våre informanter deltar i eller kjenner til.

Nasjonale samarbeid

- Nasjonalbiblioteket samarbeider tett med akademia, mediehus, teknologiselskaper og offentlige organer om utvikling av norsktrente modeller, tilgang til datasett, evaluering og testing.
- Universiteter og forskningsinstitusjoner som UiO, NTNU, UiT, UiB og USN trekkes frem som sentrale samarbeidspartnere i forsknings- og utviklingsprosjekter. Bidragene spenner fra metodeutvikling og domeneekspertise til deltagelse i fellesprosjekter og undervisning.
- Mediehus som NRK og Schibsted bidrar både med datasett for utvikling og testing av språkmodeller og deltar i prosjekter. Samarbeidet som ble etablert i forbindelse med Mimir-prosjektet nevnes som særlig fruktbart.
- Teknologi- og infrastrukturleverandøren Sigma2 tilbyr regnekraft, datasettforvaltning og nettverkstilgang. Sigma2 er bl.a. samarbeidspart for små og mellomstore bedrifter som mangler egen tungregningskapasitet.
- Digital Norway og Nemonoor fremstår som viktige tilretteleggere for spredning av kunnskap og erfaringer på tvers av sektorer.
- Offentlige organer som Helsedirektoratet deltar i samarbeidsprosjekter for utvikling av helserettede språkmodeller. De har også kommunisert regulatoriske krav til leverandører gjennom nettverket Norway Health Tech.
- Agder AI trekkes frem som eksempel på et velfungerende samarbeid der kommuner, næringsliv og akademia utvikler og tester KI-løsninger i fellesskap. Modellen bygger på åpen deling av erfaringer, felles risikovurderinger og høy grad av tillit.

Internasjonale samarbeid

- Europeiske prosjekter og konsortier som f.eks. LUMI-superdatamaskinen i Finland brukes av en rekke norske aktører, særlig til modelltrening som krever mer kapasitet enn nasjonale løsninger kan tilby.
- Horizon Europe, Open Euro og Hyperformance Language Technologies er arenaer for samarbeid om språkmodellutvikling og evaluering på tvers av europeiske språk.
- Microsoft, Amazon, Google, Meta, Anthropic og HP samarbeider med norske aktører både som leverandører av modeller og infrastruktur og som rådgivere. Eksempler inkluderer gratis GPU-kapasitet fra Telenor AI-fabrikken i samarbeid med HP, kreditt fra Amazon Web Services, og direkte faglig støtte fra Meta og Anthropic. Mange av disse relasjonene har oppstått gjennom egeninitiert kontakt fra norske aktører.
- Internasjonale nettverk som Digital Innovation Hubs (DIH) knytter norske aktører til over 200 organisasjoner i Europa, mens EBU (European Broadcasting Union) brukes av NRK og andre medieaktører til å teste språkmodeller på tvers av språk og kontekster. AI Sweden og OECD-fora gir norske aktører tilgang til erfaringer, standarder og regulatoriske diskusjoner.

3.2 Former for samarbeid og organisering

Samarbeidene i språkmodell-økosystemet organiseres på ulike måter, avhengig av aktørenes størrelse, rolle og ressurser. Her beskrives hovedtypene av samarbeidsformer slik de er omtalt av informantene, fra store prosjektbaserte initiativer til mer fleksible og uformelle nettverk.

Prosjektbaserte initiativer

Den mest utbredte samarbeidsformen er tidsavgrensede prosjekter med tydelige mål, rollefordeling og tverrsektorielt innhold. Mimir-prosjektet, ledet av Nasjonalbiblioteket, er et eksempel som samler akademia, mediehus, offentlige organer og private teknologibedrifter om utvikling og evaluering av norsktrente språkmodeller.

Lignende strukturer finnes i sektorielle initiativer, som helsesamarbeid mellom Helsedirektoratet, helseforetak og Nasjonalbiblioteket, i MediaFutures for medieproduksjon, og i nordiske prosjekter for samisk språkteknologi der Divvun ved UiT samarbeider med NRK, SVT og YLE gjennom faste arbeidsprosesser.

Formelle partnerskap og strategiske avtaler

Langsiktige samarbeidsavtaler brukes der kontinuerlig koordinering og ressursdeling er avgjørende. Nasjonalbiblioteket har avtaler med mediehus og offentlige etater om datautveksling og felles utvikling.

Helsesektoren kombinerer styringsstrukturer og regulatoriske nettverk (som KI-rådet og Datatilsynets sandkasse) med teknologileverandører for å utvikle sektorielle løsninger.

Forskningsklynger og teknologibedrifter har avtaler med universiteter og mediehus som regulerer ansvarsforhold i modelltrening, mens konsulentselskaper bygger partnerskap med Microsoft, Google og Amazon for tilgang til modeller, infrastruktur og felles innovasjonsprosjekter.

Uformelle og hybride samarbeid

For mange mindre aktører oppstår samarbeid gjennom direkte kontakt og relasjonsbygging heller enn via formaliserte programmer, eksempelvis Agder AI som samler kommuner, akademia og næringsliv i åpne delingsarenaer og felles styringsmodeller uten juridisk binding.

Innen samisk språkteknologi beskrives et nordisk samarbeid som bygger på tillit og pragmatisk problemløsning fremfor formelle avtaler.

Flere miljøer kombinerer likevel formelle og uformelle elementer, som i LUMI-konsortiet, der definerte roller suppleres med fleksibel tilpasning til partnernes behov, eller i nasjonale innovasjonsmiljøer som deltar i EU-prosjekter parallelt med uformelle opplærings- og veilednings-samarbeid.

3.3 Spesifikke samarbeidsarenaer

Utvikling og utprøving av språkmodeller skjer ofte ii dedikerte arenaer som gir struktur for samarbeid. I dette avsnittet oppsummeres eksempler på formelle og uformelle arenaer, både nasjonale, regionale og internasjonale som informantene opplyste om.

Sandkasser og testarenaer med formell struktur

Flere informanter har deltatt i nasjonale og/eller sektorielle sandkasser som kombinerer praktisk utprøving med regulatorisk veiledning. Datatilsynets regulatoriske sandkasse trekkes frem som verdifull for å avklare juridiske og etiske problemstillinger. Et eksempel er utviklingen av en digital assistent for arbeidsrett (LawAI) som gikk gjennom sandkassen. Her fikk deltakerne både direkte råd om personvern og mulighet til å teste løsninger i et trygt miljø. På helsesiden brukes nasjonale sandkasser og pilotprosjekter til å teste språkmodellbaserte løsninger i samarbeid mellom myndigheter, helseforetak og teknologileverandører, ofte koblet til nasjonale KI-råd.

Regionale samarbeid- og testarenaer

Agder AI organiserer lokale sandkasser der kommuner, akademia og næringsliv prøver ut KI i konkrete tjenester i kommunene. Arbeidet kjennetegnes av felles risikovurderinger, deling av verktøy og åpen erfaringsutveksling, noe som har bidratt til å bygge tillit og senke terskelen for å teste ny teknologi i offentlig sektor.

Internasjonale testarenaer og konsortier

Norske aktører deltar også i internasjonale test- og utviklingsmiljøer, som EBU (European Broadcasting Union) for mediehus, og Digital Innovation Hubs (DIH) for bredere teknologiutvikling. I tillegg brukes infrastrukturprosjekter som LUMI i Finland som testarena for språkmodeller som krever mer regnekraft enn nasjonale løsninger kan tilby. Slike arenaer gir tilgang til et bredere erfaringsgrunnlag, samtidig som de bidrar til kompetanseutvikling og nettverksbygging på tvers av landegrenser.

Prosjekter som tester ut løsninger

En del utviklingsløp fungerer som testarenaer uten å være formelt definert som det. Dette kan være kundeprosjekter der språkmodellbaserte løsninger prøves ut i produksjonsmiljøer, eller samarbeid med kultur- og helseinstitusjoner om prototyper. For oppstartsbedrifter er slike prosjekter ofte den mest realistiske måten å teste løsninger på, siden de gir direkte brukererfaring uten krav til deltakelse i ressurskrevende, formelle programmer.

3.3.1 Offentlig-private samarbeid

Mange av informantene har erfaring fra samarbeid mellom offentlig og privat sektor, og beskriver dette som en viktig drivkraft for utvikling og anvendelse av språkmodeller. De fleste informantene beskriver offentlig-private samarbeid som både nødvendige og verdifulle for utvikling og bruk av språkmodeller. Samarbeidene gir tilgang til data, modeller, infrastruktur, kompetanse og test-brukere som enkeltaktører ellers ville hatt vanskelig for å skaffe. Eksempler er Nasjonalbibliotekets samarbeid med Schibsted og NRK i Mimir-prosjektet, som ga tilgang til store datasett for norsk-trening, og helseprosjekter der Helsedirektoratet og private leverandører utvikler helsetilpassede språkmodeller i fellesskap.

Selv om samarbeidene stort sett vurderes positivt, opplever mange aktører strukturelle hindringer, særlig i offentlig sektor. Det nevnes lang saksbehandlingstid, rigide anskaffelsesprosesser, begrensede budsjetter og lav risikovilje. Kompetansegap hos offentlige bestillere gjør at samarbeidet ofte blir skjevt, og enkelte mindre aktører peker på at det er liten vilje til å ta i bruk norskutviklede løsninger. Tilgang til kritiske datakilder, som lovdata.no, trekkes også frem som en barriere for utvikling.

Aktører som rapporterer om gode erfaringer peker på felles kjennetegn: tett og åpen dialog mellom partene, delingskultur og tillit, tydelige rammer for data- og kompetansedeling, samt fleksibilitet i gjennomføringen. Uformelle og praksisnære samarbeid trekkes frem som spesielt effektive, som i Agder AI eller i Divvuns samarbeid med små private utviklere av språkopplæringsverktøy. Offentlige støtteordninger og regulatoriske sandkasser, som hos Datatilsynet, fremheves som nyttige verktøy for trygg eksperimentering og avklaring av regelverk.

3.3.2 Internasjonalt samarbeid og finansiering

Flere aktører deltar i internasjonale prosjekter, konsortier og nettverk, og trekker frem betydningen av tilgang til infrastruktur, kompetanse og finansiering fra utlandet. Dette avsnittet beskriver hvordan norske miljøer kobler seg på internasjonale satsinger og hvilke ressurser dette gir.

Deltakelse i forsknings- og utviklingsprosjekter

Flere norske aktører er aktive i europeiske forsknings- og innovasjonsprosjekter, ofte med finansiering fra EU eller som del av store konsortier. UiO / Language Technology Group deltar i prosjekter som *Hyperformance Language Technologies* og *OpenEuroLLM*, begge rettet mot utvikling og evaluering av språkmodeller for europeiske språk. NorwAI bidrar i europeiske initiativer med fokus på germanske språk, og Sigma2 er en del av LUMI-konsortiet for superdatakapasitet, i tillegg til prosjekter under Horizon Europe. Nemonoor inngår i EU-nettverket *Digital Innovation Hubs* (DIH) med over 200 deltakende organisasjoner, og deltar i internasjonale samarbeids-prosjekter rettet mot små og mellomstore bedrifter.

Bruk av og samarbeid om internasjonal infrastruktur

LUMI-superdatamaskinen i Finland nevnes som et kritisk verktøy for mange norske aktører når nasjonal kapasitet ikke strekker til. Bruken spenner fra modelltrening og evaluering til eksperimenter med språkmodeller for minoritetsspråk. Divvun/UiT har samarbeid med språkbanker og tekniske universiteter i Finland, samt prosjekter med nordiske kringkastingshus (NRK, SVT, YLE). De nevner også erfaringsutveksling med urfolksmiljøer i Australia, Afrika og Sør-Amerika om språkteknologiutvikling.

Partnerskap med globale teknologiselskaper

Mange kommersielle og enkelte offentlige aktører samarbeider med teknologiselskaper som Microsoft, Google, Amazon, Meta og Anthropic. Samarbeidene omfatter tilgang til modeller som GPT, Gemini og Llama, samt infrastruktur, teknisk rådgivning og opplæring.

Internasjonal finansiering og strategisk innsikt

Flere akademiske miljøer og nettverk, som NORA og UiO, mottar finansiering fra EU og Forskningsrådet til kompetanseutvikling og prosjekter. Oppstartsbedrifter nevner støtte fra nasjonale innovasjonsordninger, men få av disse har direkte internasjonal finansiering. Enkelte offentlige organer, som Helsedirektoratet, følger tett med på internasjonale regulatoriske prosesser i EU, OECD og andre fora, uten alltid selv å delta som aktive partnere.

3.4 Oppsummering

Kartleggingen viser at samarbeid og nettverk er en grunnleggende forutsetning for utvikling og bruk av språkmodellbasert KI i Norge. Både intervjuer og litteratur peker på at økosystemet er preget av noen få sentrale noder, særlig Nasjonalbiblioteket, universitetsmiljøene og utvalgte medie- og teknologiorganisasjoner, som fungerer som brobyggere mot både nasjonale og internasjonale ressurser. Kombinasjonen av nasjonale ressurser og internasjonal kapasitet fremstår som viktig for norsk språkmodellutvikling.

Samarbeidsformene varierer betydelig. Strukturerte og tidsavgrensede prosjekter dominerer, men det er også mere uformelle og fleksible nettverk. Dette kan åpne for innovasjon og rask tilpasning. Litteraturen gir konkrete eksempler på sektorvise initiativer, internasjonale prosjekter og koblinger til globale teknologileverandører, noe som understøtter at økosystemet utvikler seg i flere parallelle spor.

Formelle arenaer bidrar med struktur, legitimitet og regulatorisk trygghet, mens regionale og uformelle arenaer fremmer fleksibilitet, hurtighet og praktiske løsninger. Til sammen skaper dette et mangfold av samarbeidsarenaer som akselererer utvikling, styrker kvalitet og utvider nettverk på tvers av sektorer og landegrenser.

Offentlig-private samarbeid fremstår som en bærebjelke i utviklingen. Disse fungerer best når stabile strukturer kombineres med fleksibilitet og gjensidig ressursdeling. Samtidig peker litteraturen på utfordringer knyttet til blant annet anskaffelsesprosesser og begrenset tilgang til data, som kan hemme samarbeidet. Uformelle arenaer, hvor aktører bygger tillit og deler erfaringer på tvers, ser ofte ut til å være like effektive som mer formelle ordninger.

Internasjonalt samarbeid er særlig viktig for å gi norske aktører tilgang til regnekraft, data og spesialisert kompetanse. Dette skjer gjennom EU-prosjekter, nordiske initiativer og partnerskap med globale teknologiselskaper, samt via nettverk som Digital Innovation Hubs og OECD-fora. Samtidig er deltakelsen asymmetrisk: store forskningsmiljøer og institusjoner har en sterk formell tilstedeværelse, mens mindre bedrifter i større grad må være opportunistiske og finne fleksible veier inn i samarbeidet gjennom nettverk og direkte kontakt.

4 Barrierer i arbeidet med språk-modeller

Dette kapittelet tar for seg de viktigste barrierene som hemmer utvikling og bruk av språkmodell-basert KI i Norge. Basert på både litteratur og intervjuer identifiseres utfordringer knyttet til tilgang på data, mangel på tungregnekraft og infrastruktur, utilstrekkelig finansiering, fravær av standardisering og tydelige ansvarsforhold, samt manglende tverrfaglig kompetanse. Kapittelet viser hvordan disse barrierene kan føre til fragmentering, redusere innovasjonstakten og gjør det vanskeligere å kommersialisere løsninger.

Kartleggingen viser at viser at en møter en rekke utfordringer i arbeidet med språkmodeller og KI. Utfordringene er av teknisk, juridisk, organisatorisk og markedsmessig karakter. Mange av barrierene går igjen på tvers av sektorer, men påvirker aktørene ulikt avhengig av størrelse, rolle og tilgang på ressurser. Tilgang til data og infrastruktur trekkes særlig frem som kritiske forutsetninger for utvikling og bruk, mens regulering, finansiering og samhandlingsstrukturer nevnes som vesentlige hindringer for implementering og skalering. Mange av disse utfordringen er allerede nevnt, men vi samler dem i framstillingen under.

4.1 Tekniske og økonomiske barrierer

Intervjuene viser at mange aktører opplever kombinasjonen av begrenset nasjonal regnekraft, høye driftskostnader og kortsiktig finansiering som sentrale hindringer.

Behovet for tungregnekapasitet i Norge beskrives som stort på tvers av sektorer. Flere aktører peker på at kapasiteten er for lav og prosessene for tilgang både kompliserte og lite fleksible, og at dagens tilgang til ressurser som Olivia og LUMI ikke dekker framtidige behov. For norske private aktører er det i liten grad tilgang til disse ressursene i det hele tatt.

Norges kapasitet for tungregning vurderes som utilstrekkelig til å møte behovene i forskning og næringsliv, særlig for språkmodellprosjekter, der behovet vokser raskere enn utbyggingen og forskere møter kompliserte søknadsprosesser og prioriteringsregimer. I rapporten Behov for tungregnekraft (Forskningsrådet 2024) vises det til estimater på behov fra ledende KI-miljøer i forskningssektoren i Norge som er i størrelsesorden 10–100 ganger kapasiteten som er tilgjengelig i dag. Det store spennet er blant annet knyttet til hvor avanserte modeller som skal trenes og om man skal bygge norske grunnmodeller fra grunnen av basert på norske data, eller bare bygge videre på eksisterende modeller (Forskningsrådet, 2024).

Finansiering gjennom enkeltprosjekter skaper heller ikke stabile rammer for utvikling og drift av språkmodeller.

Kostnader til drift trekkes av mange fram som en utfordring forbundet med å sette løsninger innen språkmodellbasert kunstig intelligens i produksjon. Skyleverandørene som tilbyr dette tar gjerne betalt per spørring og størrelsen på denne basert på antall ord.[21] Selv om prisene for dette stadig faller, vil det uansett utgjøre en kostnadsdriver som er høyere og mer uforutsigbar enn tradisjonell dataprosessering. Driftsløsninger som tilbys av internasjonale skyleverandører beskrives likevel som velfungerende, blant annet fordi de har en lang rekke verktøy og tjenester som gjør det enklere drifte løsningen.

[21] Prismodellene er som regel basert på bruk av såkalte «tokens». Tokens viser til ord eller bestanddeler av ord.

Flere problematiserer at driften skjer hos utenlandske leverandører, hvor man har mindre grad av kontroll knyttet til hvordan data prosesseres. Flere fremhever viktigheten av at man har nasjonal kontroll på drift av KI-løsninger, blant annet i lys av geopolitisk uro. Enkelte påpeker at det finnes sikre nasjonale alternativer for drift i dag (som Telenor), men at virksomheter som skal ta i bruk KI ikke etterspør det i særlig grad når det kommer til faktiske investeringer og avtaler.

Flere av informantene peker på at norske språkmodeller, selv når de holder høy kvalitet, ofte bare tilbys som nedlastbare filer uten driftede API-tjenester, noe som begrenser bruken i praksis.[22]

4.2 Juridiske og regulatoriske barrierer

Intervjuene peker på at tilgang til treningsdata er en av de største barrierene for utvikling av språk-modeller i Norge. Den mest gjennomgående utfordringen gjelder rettigheter og juridisk usikkerhet: Mange datasett, særlig innen nyheter, bøker og fagområder som juss og helse, er opphavsrettslig beskyttet, og det finnes få klare rammer for hvordan de kan brukes i modelltrening. Dette gjør det vanskelig, tidkrevende og i mange tilfeller umulig å bruke data kommersielt eller på tvers av sektorer.

Et eksempel er tilgangen til innhold fra lovdata.no, som trekkes fram som en utfordring både i enkelte av intervjuene, og i rapporten fra Teknologirådet (2024). Lovdata inneholder blant annet norske rettsavgjørelser, som de har enerett på. Brukeravtalen med Lovdata presiserer at imidlertid innholdet ikke skal brukes i språkmodeller. Det trekkes både fram at disse dataene kan være en svært verdifull kilde for å trene opp norske språkmodeller i å forstå norsk jus. Innholdet vil også være svært nyttig for selskap som utvikler juridiske KI-assistenter, ifølge et teknologiselskap som arbeider med dette.

Rapporten fra Teknologirådet viser også til potensialet knyttet til bruk av Arkivverkets data, som er krevende fordi de inneholder mye personopplysninger.

I helsesektoren gjør strenge krav til behandling av sensitive data testing og validering vanskelig, men litteraturen peker på behov for sertifiseringsordninger og tydelige retningslinjer for å gi forutsigbarhet (Helsedirektoratet, 2024a). I tillegg begrenser fraværet av nasjonale og nordiske standardiserte rammer for datadeling ressursutnyttelse og samarbeid (Teknologirådet, 2024).

Flere aktører beskriver mangel på datasett innen smale fagområder og minoritetsspråk, særlig for samisk og andre språk med lite digitalt materiale. Her er både mengde og kvalitet en utfordring. Samtidig opplever mange små og mellomstore aktører at de ikke får tilgang til offentlige data, eller at det mangler verktøy og prosesser for sikker og effektiv datadeling.

Et annet gjennomgående funn er at internasjonale teknologiselskaper har langt bedre tilgang på data enn norske aktører, noe som skaper en skjev konkurransesituasjon. Norske bedrifter mangler både datamengde og økonomisk evne til å bygge modeller i samme skala.

Det er behov for å etablere nasjonale fellestjenester og rammeverk for datadeling, inkludert åpne datasett, juridiske retningslinjer og infrastruktur for sikker lagring og bruk. Uten slike tiltak vil norske språkmodeller fortsatt ha begrenset kvalitet og rekkevidde.

Til sist bidrar også kompetansemangel og organisatorisk usikkerhet til at mange virksomheter vegrer seg for å dele eller bruke data, blant annet på grunn av usikkerhet om lovverk og manglende interne rutiner (Vestlandsforsking, 2022).

[22] Unntaket er NorskGPT som tilbys som en chattjeneste og via API

4.3 Kompetansemangel

Intervjuene viser at mangel på kompetanse er en av de største barrierene for utvikling og bruk av språkmodeller i Norge. Det er få fagmiljøer med avansert teknisk ekspertise til å trene og evaluere modeller, og det mangler særlig tverrfaglig kompetanse som kombinerer språk, teknologi og domeneinnsikt (f.eks. helse, juss, samisk).

Enkelte peker på at offentlig sektor har generelt lav teknologiforståelse og mangler bestillerkompetanse, noe som fører til avhengighet av eksterne leverandører og manglende evne til å utvikle eller ta i bruk norsktrente modeller. Kompetanse om livsløp, forvaltning og sikker drift av språkmodeller er også svak, både blant utviklere og brukere.

Det er bred enighet om behovet for nasjonale satsinger på kompetansebygging, særlig gjennom styrking av forskning, utdanning og praktisk opplæring i offentlig sektor. Mangelen på kompetanse begrenser i dag både utvikling, bruk og strategisk kontroll over språkmodeller i Norge.

Flere etterlyser et felles nasjonalt økosystem for språkmodeller, inspirert av plattformer som Hugging Face, der modeller, datasett og kompetanse kan deles på tvers av offentlig, privat og akademisk sektor. Dette ses som et viktig virkemiddel for å oppnå både strategiske mål og økt verdiskaping, ved å skape synergier mellom sektorer og redusere dobbeltarbeid. I litteratur-gjennomgangen ser vi også mangel på kompetanse som en sentral barriere. I offentlig sektor er det særlig i kompetansemangel i kommunesektoren. Mange små og mellomstore kommuner har ikke tilgang til nødvendig teknisk kompetanse for å utvikle eller implementere KI-løsninger, og er derfor avhengige av ekstern bistand (Vestlandsforskning, 2022). Selv i større kommuner er det ofte bare én eller få ansatte med relevant kompetanse, noe som skaper sårbarhet og avhengighet av eksterne aktører.

Manglende kompetanse gjelder både teknisk spisskompetanse (f.eks. innen maskinlæring og dataforvaltning), og tverrfaglig kompetanse som trengs for å forstå etiske, juridiske og samfunns-messige implikasjoner ved bruk av KI. Dette gjør det vanskelig å vurdere risiko, implementere forsvarlige løsninger og sikre etterlevelse av regelverk. Det en erkjennelse av at Norge, som lite land, ikke kan bygge opp spisskompetanse på tvers av hele KI-feltet. Det pekes derfor på behovet for å konsentrere innsatsen og styrke kompetanse innen utvalgte områder hvor Norge har fortrinn (DFD, 2020).

I tillegg vises det til behovet for kvalitetskriterier og evalueringsevne i møte med ulike språk-modeller. Uten tilstrekkelig faglig forståelse i forvaltningen og akademia kan det bli vanskelig å gjøre informerte valg om bruk av slike modeller (Teknologirådet, 2024).

Til sammen danner både funn fra intervjuer og litteraturgjennomgangen et bilde av en sektor og et landskap der kompetansemangel hemmer både utvikling, anvendelse og regulering av KI-teknologi.

4.4 Svakheter i dagens marked eller infrastruktur?

Et gjennomgående funn er at Norge mangler helhetlige og skalerbare løsninger for drift og tilgjengeliggjøring av språkmodeller.

Oppstartsbedrifter beskriver samtidig at de kan utvikle gode modeller, men at det er vanskelig å etablere bærekraftig drift når internasjonale aktører tilbyr rimeligere og mer tilgjengelige løsninger. For store virksomheter, som et offentlig mediehus, betyr dette at man i stor grad er avhengig av amerikanske modeller som Gemini og OpenAI for avanserte oppgaver, ettersom norske alternativer foreløpig ikke vurderes som tilstrekkelig effektive i komplekse produksjonsmiljøer.
I tillegg trekker Nemonoor fram at mange norske virksomheter har lav modenhet og uklare

verdiforslag knyttet til KI, noe som bremser etterspørselen etter nasjonale løsninger og dermed svekker utviklingen av et konkurransedyktig norsk økosystem.

Litteraturgjennomgangen bekrefter i stor grad dette inntrykket. Mangel på driftede, skalerbare løsninger og sterk konkurranse fra globale plattformer svekker markedsposisjonen for norske språkmodeller. (Teknologirådet, 2024).

4.5 Hull i verdikjeden eller kompetansen?

Informantene peker særlig på mangelen på gode datasett, svak overføring fra forskning til drift og behovet for tverrfaglig kompetanse.

Informanter fra forskningsmiljøer fremhever at utviklingen av konkurransedyktige språkmodeller krever større, bedre kuraterte og rettighetsavklarte datasett for norsk og samiske språk. For oppstartsbedrifter er tilgang til høykvalitets data den største barrieren for fintrening, der samarbeid med datasetteiere ofte blir både ressurs- og tidkrevende. Offentlige mediehus og infrastruktur-aktører peker samtidig på at mange forskningsprosjekter stopper ved prototypestadiet fordi det mangler finansiering og støtte til å ta løsningene videre i produksjon. Konsulentselskaper og innovasjonsmiljøer understreker også at fraværet av tverrfaglig kompetanse – som kombinerer teknologisk forståelse med domeneinnsikt og forretningsforståelse – fører til at prosjekter utvikles saktere og gir mindre effekt enn de ellers kunne hatt.

Litteraturstudien viser i stor grad det samme bildet. Utviklingen av norske språkmodeller hemmes av mangelen på store, kuraterte og rettighetsavklarte datasett for norske og samiske språk, noe som forsterkes av fraværet av nasjonale standarder og delingsmekanismer (Teknologirådet, 2024). Mange prosjekter stopper dessuten på prototype-stadiet, der uklare ansvarsforhold og manglende finansieringsmekanismer trekkes fram som sentrale årsaker (Forskningsrådet, 2024). I tillegg mangler det tverrfaglig kompetanse, både kunnskap om språkmodeller (og kombinasjonen av teknologi-, domene- og forretningsforståelse, som blir en barriere for effektiv implementering og skalering (Rambøll, 2025). Samarbeid om data med eksterne eiere er særlig krevende for mindre aktører, og litteraturen peker på at fraværet av standardiserte avtaler utgjør en viktig barriere (Teknologirådet, 2024.)

Intervjuer og dokumenter peker på de samme hullene, men litteraturen tilfører et tydelig system-nivåperspektiv med fokus på standardisering, finansiering og ansvarsdeling, mens intervjuene gir mer praksisnære eksempler.

4.6 Oppsummering

Kartleggingen viser at arbeidet med språkmodeller i Norge møter en rekke barrierer som hemmer utvikling, implementering og skalering. Disse barrierene finnes både på teknisk, organisatorisk og strukturelt nivå, og de rammer ulike aktører i ulik grad.

En viktig barriere er mangelen på store, kuraterte og rettighetsavklarte datasett for norsk og samiske språk. Dette begrenser muligheten til å utvikle modeller med høy presisjon og språklig bredde. Tilgangen til tungregnekraft og infrastruktur trekkes også fram som en flaskehals, spesielt for mindre aktører som ikke har ressurser til å bruke internasjonale løsninger i stor skala. Samtidig fremheves fraværet av nasjonale standarder og delingsmekanismer for data som en viktig barriere, særlig når prosjekter skal skaleres utover prototypestadiet.

Finansieringsmekanismer og ansvarsdeling beskrives som uklare. Flere aktører peker på at kort-siktige prosjektmidler ikke er tilstrekkelige for å bygge opp langsiktige løsninger og infrastruktur. I tillegg er mangel på tverrfaglig kompetanse et gjennomgående hinder. Det gjelder både dyp teknisk

kunnskap om språkmodeller og evnen til å kombinere teknologi med domeneinnsikt og forretnings-forståelse.

Intervjuer og litteratur peker samlet sett på et mønster der barrierene forsterker hverandre: manglende data gjør det vanskelig å utvikle gode modeller, svak infrastruktur gjør det krevende å trene og drifte dem, og utilstrekkelige finansieringsordninger gjør det vanskelig å sikre kontinuitet.

5 Offentlig innsats og politikk-utforming

Dette kapittelet tar for seg hvordan offentlig sektor bidrar til utvikling og rammer for språkmodell-basert KI i Norge. Først beskrives organiseringen av det offentlige KI-arbeidet, inkludert roller og ansvar som ulike statlige etater har fått gjennom nasjonale strategier og som følge av EUs KI-forordning. Deretter presenteres funn fra kartleggingen om aktørenes behov og forventninger til offentlig innsats knyttet til regulering, datadeling, finansiering, kompetansebygging og etablering av test- og sandkasseordninger.

5.1 Offentlig organisering av KI-arbeidet nå og framover

Utviklingen og bruken av kunstig intelligens, inkludert språkmodeller, krever et klart rammeverk for ansvar, regulering og koordinering i offentlig sektor.

5.1.1 Formell ansvarsdeling for det offentlige KI-arbeidet

Det er flere offentlige aktører som har formelle roller knyttet til utvikling og bruk av språkmodeller i Norge. Under redegjøres det for de sentrale aktørene.

Digitaliseringsdirektoratet (Digdir)

Utnyttelse av muligheter knyttet til datadeling og kunstig intelligens er viktige prioriteringer i Digitaliseringsstrategien. Ifølge tildelingsbrevet for 2025 har Digdir har ei pådriver- og veilednings-rolle for ansvarlig utvikling og bruk av KI, og har et ansvar for å videreutvikle virkemidler for å nå målene i strategien. I tildelingsbrevet til Digitaliseringsdirektoratet for 2025 vises det også til at arbeidet med tilsyn og veiledning på det digitale området i praksis deles mellom Digdir, Nkom og Datatilsynet, og at de tre etatene skal styrke dialogen, og i fellesskap utarbeide et forslag til hvordan samarbeidet skal foregå (Tildelingsbrev 2025 for Digitaliseringsdirektoratet).

KI-forordningens kapittel 6 fastslår at det skal opprettes minst én regulatorisk sandkasse for kunstig intelligens. Supplerende tildelingsbrev slår fast at den regulatoriske KI-sandkassen etter KI-forordningen være plassert i Digdir i KI Norge. KI Norge skal være den nasjonale arenaen for ansvarlig innovasjon, utvikling og bruk av kunstig intelligens i offentlig og privat sektor. Den regulatoriske KI-sandkassen skal etableres, driftes og utvikles innenfor et formelt samarbeid mellom Digdir/KI Norge, Nkom og Datatilsynet.

Videre står det at norske virksomheter i sandkassen skal kunne eksperimentere, utvikle og trene innovative KI-systemer innenfor trygge og kontrollerte rammer, før løsningen settes i produksjon eller markedsføres. Hensikten er at myndighetene i sandkassen skal overvåke og veilede virksomhetene for å identifisere og håndtere risiko, særlig knyttet til grunnleggende rettigheter, helse og sikkerhet, og bidra til at KI-systemene oppfyller kravene i KI-forordningen og annen relevant lovgivning.

Særlig oppstartsselskaper og små og mellomstore bedrifter (SMBer) skal prioriteres i sandkassen. Det er et mål at sandkassen skal bidra til kunnskapsbasert regulatorisk læring, og støtte deling av beste praksis gjennom samarbeid mellom myndighetene som er involvert.

Nasjonal kommunikasjonsmyndighet (Nkom)

Nasjonal kommunikasjonsmyndighet (Nkom) er den norske tilsyns- og reguleringsmyndigheten for elektronisk kommunikasjon. De har ansvar for å sikre at Norge har velfungerende, sikre og fremtidsrettede kommunikasjonsnett og -tjenester. Nkom jobber også med beredskap og sikkerhet i kritisk digital infrastruktur.

I supplerende tildelingsbrev nr. 2 til Nkom for 2025 vises det til at Digitaliserings- og forvaltnings-departementet skal gjennomføre EUs forordning om kunstig intelligens (KI-forordningen) i norsk rett med mål om å fremme en lovproposisjon for Stortinget våren 2026. som del av implementeringen skal det utpekes og etableres en nasjonal forvaltningsstruktur for håndheving av KI-forordningen. Det er besluttet at Nasjonal kommunikasjonsmyndighet (Nkom) skal være nasjonal koordinerende markedstilsynsmyndighet og nasjonalt kontaktpunkt overfor EU.

Som nasjonalt kontaktpunkt vises det til at Nkom skal representere Norge i relevante fora, samt koordinere og formidle informasjon mellom norske myndigheter og EU-organer. Rollen som koordinerende tilsynsmyndighet innebærer blant annet å (sitert fra supplerende tildelingsbrev for 2025 nr. 2 til Nkom):

- Bidra til en enhetlig forståelse og praktisering av KI-regelverket blant nasjonale myndigheter,
- Bistå øvrige markedstilsynsmyndigheter med avklaringer av særlig komplekse tekniske og/eller juridiske problemstillinger,
- Gi veiledning om innretning og gjennomføring av KI-tilsyn, og ved behov stille fagkompetanse til rådighet for andre myndigheter.
- Fordele tilsynsansvar i saker som ikke naturlig faller inn under eksisterende myndigheters ansvarsområder, eller der det er uklarhet om sektortilhørighet – og ved behov selv føre tilsyn i slike saker.

Datatilsynet

Datatilsynet etablerte sin regulatoriske sandkasse i 2020. Sandkassen for personvernvennlig innovasjon og digitalisering legger til rette for utvikling av innovative løsninger som bruker person-opplysninger på en trygg og lovlig måte. I sandkassen får offentlige og private virksomheter tilpasset og dialogbasert veiledning slik at de kan utvikle og ta i bruk kunstig intelligens (og annen innovativ teknologi) på en personvernvennlig måte. I 2024 var tre av fire gjennomførte prosjekter knyttet til generativ kunstig intelligens.[23]

Som nevnt ovenfor har Digdir, Nkom og Datatilsynet fått et felles oppdrag om å etablere samarbeid framover, og samarbeide om en om etablering, utvikling og drift av den regulatoriske KI-sand-kassen som vil være organisatorisk plassert i KI Norge. Det er foreløpig ikke avklart hvordan Datatilsynets regulatoriske skal fungere opp mot den regulatoriske KI-sandkassen.

Nasjonalbiblioteket

Nasjonalbiblioteket besitter store samlinger av norsk språk med høy kvalitet, og har siden 2010 hatt ansvar for å tilby grunnlagsressurser til utvikling av språkteknologi på bokmål, nynorsk og norske dialekter, til bruk bl.a. i retteprogrammer, oversettingsverktøy og talestyring.

I forbindelse med Statsbudsjettet for 2025 fikk Nasjonalbiblioteket i oppdrag å trene og tilgjengelig-gjøre norske og samiske språkmodeller. Nasjonalbiblioteket fikk en bevilgning på 15 mill. kr for å etablere en enhet for trening, oppdatering og tilgjengeliggjøring av norske og samiske, og 5 mill. kr til lokal infrastruktur for trening av språkmodeller i Nasjonalbiblioteket. Formålet er å legge til rette

[23] Prosjektene kom fra NTNU, Helsedirektoratet og firmaet Lawai/Juridisk ABC. Kilde: Tid for generativ KI i sandkassa | Datatilsynet

for at samfunnet kan ta i bruk teknologi av høy kvalitet som er tilpasset norsk og samiske språk og norske og samiske samfunnsforhold (Prop. 1S Kulturdepartementet, 2024-2025).

Sigma2

Sigma2 AS er et statlig eid aksjeselskap, hvor eierskapet forvaltes av Sikt – kunnskapssektorens tjenesteleverandør. Sigma2 tilbyr blant annet tungregning (High performance computing) både gjennom nasjonale superdatamaskiner og gjennom å ivareta Norges deltagelse i LUMI-konsortiet, som gir tilgang til Europas kraftigste superdatamaskiner.

Sigma2 fungerer som sekretariat og administrativ støtte for Ressursfordelingskomitéen (RFK) som er ansvarlig for å evaluere og tildele beregnings- og lagringsressurser (inkludert avansert bruker-støtte) på den nasjonale e-infrastrukturen som driftes av Sigma2. Komiteen består av ledende norske forskere fra relevante brukergrupper og oppnevnes av Sigma2s styre. Tildelinger skjer to ganger i året, og søknader vurderes etter vitenskapelig fremragende arbeid og tydelig dokumentert behov.[24]

I tildelingsbrevet til Sikt for 2025 legges det til grunn at «regjeringen vil frem mot 2030 etablere en nasjonal infrastruktur for kunstig intelligens (KI), hvor superdatamaskiner (tungregning) og norsk-utviklede språkmodeller utgjør en viktig del, jf. Nasjonal digitaliseringsstrategi 2024–2030». Det legges opp til at Sigma2 AS står for investering og drift av de største nasjonale superdata-maskinene, mens Nasjonalbiblioteket skal stå for utvikling og drift av språkmodellene. Bevilgningen til Sigma2 er økt med 20 mill. kr i 2025 for å finansiere trening av Nasjonalbibliotekets modeller. Arbeidet til Sikt og Sigma2 skal sees i sammenheng med oppdraget Forskningsrådet har fått om å utrede det nasjonale behovet for tungeregnekraft til formålene forskning, offentlig forvaltning og utvikling og bruk av KI, herunder hvordan en nasjonal infrastruktur for tungregning bør organiseres (Sikt tildelingsbrev for 2025).

5.1.2 KI Norge

KI Norge skal etableres som en del av Digdir og være operativ fra august 2026[25]. Formålet er å utvikle en nasjonal arena for ansvarlig innovasjon, utvikling og bruk av KI i både offentlig og privat sektor. Som arena skal KI Norge fungere både som et spissmiljø og et bindeledd mellom ulike aktører, inkludert offentlige virksomheter, næringslivet, forskningsmiljøer og akademia.

Pådriver og veileder

KI Norge skal det være en pådriver for ansvarlig KI-utvikling, med utadrettet aktivitet og samarbeid som bidrar til kunnskapsdeling og erfaringsutveksling på tvers av sektorer. KI Norge skal bidra til å skape et helhetlig og samordnet nasjonalt rammeverk for veiledning om KI for å gi virksomheter støtte i hvordan kunstig intelligens kan utvikles og tas i bruk innenfor trygge og ansvarlige rammer. Sammen med Nkom og Datatilsynet skal det utarbeides klare rolle- og ansvarsfordelinger i veiledningen på KI. Dette skal sikre tydelig kommunikasjon, unngå overlapp og styrke koordineringen i myndighetenes samlede innsats.

Regulatorisk sandkasse

En sentral oppgave for KI Norge blir å drifte en regulatorisk KI-sandkassen, som etter EUs KI-forordning skal være et virkemiddel for innovasjon og økt konkurranseevne. Sandkassen vil gi norske virksomheter, særlig oppstartsbedrifter og små og mellomstore bedrifter, mulighet til å teste og utvikle KI-løsninger under kontrollerte forhold. Myndighetene vil i denne sammenhengen

[24] Se Ressursfordelingskomité (RFK) | Sigma2

[25] https://www.regjeringen.no/no/aktuelt/gjor-norge-klar-for-trygg-og-innovativ-ki-bruk/id3093081/

overvåke og veilede virksomhetene for å identifisere og håndtere risiko, og for å bidra til at løsningene er i samsvar med KI-forordningen og øvrig relevant regelverk.

Status for KI Norge

Vi har intervjuet representanter for etater som deltar i KI Norge om status for arbeidet med etableringen. Endelig innhold og form er ennå ikke fastsatt, men informantene har synspunkter på hva KI Norge skal bidra til.

KI Norge skal fungere som en pådriver, rådgiver og arena for praktisk utprøving av KI, men samtidig være en institusjon som evner å balansere innovasjon og regulering. Digdir ser etableringen av KI Norge som en naturlig videreføring av deres tidligere oppgaver med premiss-giving, veiledning og utvikling av fellesløsninger innen digitalisering. Digdir har historisk ikke hatt et formelt KI-mandat, men teknologiens strategiske betydning gjør det uunngåelig å engasjere seg. Informantene ser for seg at KI Norge kan vokse til å bli en slags «KI-direktorat». Enheten skal rette seg både mot offentlig sektor og mot små og mellomstore bedrifter i privat sektor. Nkom beskriver sin rolle som markedstilsynsmyndighet og «vaktbikkje», med ansvar for å håndheve AI-for-ordningen nasjonalt. Deres mandat innebærer å sikre at KI-utvikling skjer innenfor trygge rammer, men samtidig bidra til å legge til rette for innovasjon. KI Norge skal bidra med veiledning, kunnskapsbygging og tillitsarbeid.

Et sentralt virkemiddel i KI Norge blir den regulatoriske sandkassen. Sandkassen beskrives som en arena hvor aktører kan få konkret veiledning om hvordan regelverket slår ut i praksis. Små og mellomstore bedrifter fremheves som den viktigste målgruppen, fordi de ofte mangler ressurser til å håndtere komplekse reguleringer på egen hånd. Informantene ser for seg at sandkassen i første omgang vil måtte etableres på et minimumsnivå i tråd med EU-kravene, men at det på lengre sikt kan være behov for å utvikle flere sektorspesifikke og mer tekniske sandkasser.

Samordning og rolleavklaringer mellom myndighetene fremstår som en nøkkelfaktor for å lykkes. Det er viktig med samordnede og tydelige oppdrag til etatene. Det er viktig at KI Norge kan fungere som et tydelig kontaktpunkt som samler og formidler helhetlige råd.

Offentlig–privat samarbeid trekkes også frem som nødvendig. Informantene legger vekt på at det må etableres partnerskap med private aktører som sitter på mye av innovasjonskraften. Myndighetenes rolle først og fremst er å legge til rette for gode rammer, tilgjengelige data og tydelig veiledning, mens det er de private som skal drive utviklingen av produkter og tjenester.

Informantene knytter suksesskriteriene for KI Norge både til innovasjon og tillit. Informantene understreker at KI Norge må fremstå som noe mer enn «nok et koordineringstiltak» og faktisk bidra med reell merverdi. Det er avgjørende at næringsliv, akademia og offentlig sektor opplever KI Norge som relevant, nyttig og godt samordnet med øvrige myndigheter.

Informantene mener også at nasjonal politikk bør forankres i et europeisk og internasjonalt samarbeid, og at Norge i mange tilfeller må bygge videre på løsninger utviklet i EU eller globalt.

Informantene peker også på behovet for å se KI i et helhetlig perspektiv. Selv om språkmodeller får mye oppmerksomhet i dag, finnes det mange andre bruksområder for KI som kan få stor betydning fremover. En viktig oppgave for KI Norge blir derfor å bidra til at det bygges et helhetlig kunnskapsgrunnlag som dekker hele økosystemet for KI.

Kritiske perspektiver på KI Norge

Flere av våre informanter fra økosystemet uttrykker bekymring for at KI Norge kan bli en struktur som først og fremst gagner store, etablerte aktører med kapasitet til å delta i konsortier og nasjonale satsinger. En utvikler fra en mindre teknologibedrift beskriver at bedrifter som han

representerer ofte havner «mellom to stoler», de er verken oppstartsbedrift i tidligfase eller tunge akademiske institusjoner, og dermed faller de utenfor mange av de virkemidlene som finnes i dag. Det pekes på at et KI Norge som bygger videre på eksisterende strukturer uten å inkludere nye, mindre aktører, vil kunne forsterke denne skjevheten.

Flere informanter stiller spørsmål ved om KI Norge vil løse reelle problemer, eller bli et toppstyrt initiativ uten praktisk gjennomslagskraft. De uttrykker bekymring for at en tungrodd og ensrettet struktur kan hemme utviklingen ved å standardisere for mye og ikke fange opp ulike sektorbehov. Samtidig etterlyses større tydelighet i rollefordelingen mellom infrastruktur, kompetanse, regulering og innovasjon, med advarsel mot å «gjøre alt». Klare prioriteringer og forankring i faktiske brukerbehov trekkes fram som avgjørende for at KI Norge skal lykkes.

5.2 Uttrykte behov for offentlig innsats

Med bakgrunn i etableringen av KI Norge KI ble informantene i økosystemet spurt om hvilke generelle behov for offentlig innsats de kunne ønske seg og evt. hvilken rolle KI Norge kunne ha i dette. En rekke ønsker, og behov ble fremmet:

Infrastruktur, data og nasjonale fellestjenester

Et sentralt behov som tas opp av mange aktører, er etablering av nasjonal infrastruktur for utvikling, drift og distribusjon av språkmodellbasert KI. Det gjelder særlig behovet for regnekraft, lagringskapasitet og stabile plattformer for drift. Flere informanter peker på at dagens avhengighet av utenlandske aktører er problematisk, særlig innenfor helse, offentlig forvaltning og andre områder med sensitive data. Når det gjelder drift av KI-løsninger er enkelte skeptiske til at dette skal bygges opp i offentlig regi, siden det finnes aktører i markedet for dette, og at det er vanskelig å konkurrere mot internasjonale skyplattformer på sikkerhet, kostnader og funksjonalitet. Nasjonal tungregnekraft og fellestjenester for trening og drift av språkmodeller er kritisk for å sikre utvikling og bruk av KI i Norge

Tilgang til treningsdata trekkes også frem som et sentralt hinder. Både forskningsmiljøer, medie-aktører og utviklingsselskaper påpeker at rettighetsproblematikk knyttet til tekst og språkdata gjør det krevende å trene gode norsktrente modeller. Flere ønsker derfor at KI Norge bør bidra med både juridiske rammer og tekniske infrastrukturer for ansvarlig deling og forvaltning av datasett.

Kompetanse og anvendelsesstøtte

Mangel på kompetanse går igjen som et hovedproblem. Aktører fra hele verdikjeden etterlyser tiltak som kan bygge kapasitet, både i forsknings- og utviklermiljøer og ute hos brukere. Her ser mange for seg at KI Norge kan spille en nøkkelrolle som en nasjonal kunnskapsplattform som tilbyr veiledning, kompetanseheving, opplæring og eksempler på god praksis.

Et annet moment er at KI Norge kan ha en rolle som støtte til utvikling av bestillerkompetanse og veiledning i anskaffelse og bruk av KI-løsninger i offentlig sektor. KI Norge bør hjelpe til med å gjøre kunstig intelligens tilgjengelig og anvendbar i praksis, særlig for aktører uten egen KI-kompetanse.

Tilgjengeliggjøring og drift – behov for API-er og brukerstøtte

Selv blant aktører som utvikler/trener egne modeller, er det et klart uttrykt behov for bedre støtte til drift, inferens og produksjonssetting. En informant peker på at det finnes mange lovende norsk-trente modeller, men at få av dem er tilgjengelige via robuste API-er eller plattformer som gjør dem brukervennlige for tredjepartsaktører. En annen informant viser til at de har tilgjengeliggjort sin norske språkmodell som en tjeneste. KI Norgekunne få en rolle i å bidra til å tilgjengeliggjøre språkmodeller på en sikker og stabil måte, med støtte for bruk i ulike sektorer og domener.

Regulering, rammeverk og standardisering

Et aspekt som flere informanter også fremhever er behovet for regulatoriske og etiske rammer som gjør det trygt og legitimt å bruke språkmodellbasert KI. Offentlige brukere etterlyser særlig dette. De ønsker at KI Norge bør tilby regulatorisk veiledning, standarder og rammeverk for kvalitetssikring og etterlevelse. Offentlige brukere nevner også behovet for felles rammeverk for risikovurdering og deling av erfaringer.

Flere informanter fremhever også behovet for sandkasser eller testarenaer hvor ny teknologi og regelverk kan prøves ut i kontrollerte omgivelser. Erfaringene med Datatilsynets sandkasse trekkes frem som svært nyttige, og det foreslås at KI Norge kan tilby tilsvarende nasjonalt.

I tillegg etterlyser konsulentselskaper og teknologipartnere tekniske standarder for modellutvikling, dokumentasjon og evalueringspraksis. Dette er nødvendig for å sikre sammenlignbarhet, transparens og ansvarlig bruk på tvers av aktører og sektorer.

Samarbeid og økosystembygging

På tvers av sektorer uttrykkes en forventning om at KI Norge skal fungere som samhandlingsplattform, ikke bare tilby teknologisk infrastruktur. Flere peker på verdien av offentlig-private samarbeid i tidligere prosjekter, men også på manglende struktur og langsiktighet.

Fra andre ønskes det at KI Norge bidrar til å etablere mekanismer for langsiktig samhandling, prosjektfinansiering, og ressursdeling – ikke minst for små og mellomstore offentlige eller private aktører som er for små til å delta i de store nasjonale satsingene.

Offentlig innsats som understøtter, ikke konkurrere med, private aktører

Informantene fra økosystemet peker mot et tydelig skille mellom offentlig og privat innsats: Offentlige aktører bør først og fremst være muliggjørere, ikke konkurrenter. Rollen bør være å bygge fundamentet som private aktører kan utvikle videre. Konkretiseringer av dette skillet som nevnes er f.eks. at myndighetenes rolle bør først og fremst være å tilby fellesgoder og grunninfrastruktur i form av nasjonale fellestjenester som datasett, tungregningskapasitet og evalueringsverktøy. Slik sikres det at offentlige investeringer kommer hele økosystemet til gode, uten at staten bygger konkurrerende kommersielle produkter.

Offentlige midler bør i tillegg rettes mot partnerskap der privat og offentlig sektor utfyller hverandre: det offentlige kan bidra med data, reguleringsforståelse og infrastruktur, mens private aktører driver innovasjon, kommersialisering og skalering.

For å fremme rettferdige konkurransevilkår må anskaffelser være teknologinøytrale og åpne for små og mellomstore bedrifter, og løsninger som utvikles bør i størst mulig grad være gjenbrukbare og tilgjengelige slik at verdien sprer seg utover enkeltprosjekter.

5.3 Oppsummering

Kapittelet viser at offentlig sektor har en viktig rolle i utviklingen av språkmodellbasert KI i Norge, både som regulator, premissgiver og tilrettelegger. Samtidig kan innsatsen oppleves fragmentert, med uklare ansvarsforhold og varierende grad av koordinering mellom departementer, direktorater og tilknyttede virksomheter.

På nasjonalt nivå har strategier og handlingsplaner lagt et grunnlag for arbeid med KI, men de adresserer språkmodeller kun indirekte. Implementeringen av EUs KI-forordning trekkes fram som et område som vil ha stor betydning fremover, både når det gjelder regulering av risikobruk og etablering av systemer for etterlevelse. Flere aktører etterlyser tydeligere nasjonal styring og mer langsiktige rammer for prioriteringer og investeringer.

Kartleggingen viser også at offentlige virkemidler i begrenset grad er tilpasset utvikling av språkmodeller. Eksisterende støtteordninger er ofte prosjektbaserte og kortsiktige, noe som gir utfordringer for aktører som ønsker å bygge varige løsninger og robust infrastruktur. Samtidig fremheves behovet for bedre ordninger for deling av data, særlig i offentlig sektor, der store mengder språkdata i dag ikke er tilgjengelig for utvikling eller testing. Aktørene har en rekke forslag til områder hvor offentlig innsats er ønsket. Det ønskes en nasjonal satsing på infrastruktur og treningsdata og offentlige fellestjenester for språkmodellbasert kunstig intelligens. Manglende kompetanse er et gjennomgående hinder, og KI Norge foreslås som en kunnskaps- og veiledningsplattform som kan bygge kompetanse. Det etterlyses også regulatoriske rammer, standarder, testarenaer og tekniske kvalitetskrav for å sikre ansvarlig bruk. Samarbeid og økosystembygging er også en oppgave det offentlige kan bidra til.

6 Vurderinger og anbefalinger

Dette kapittelet inneholder Agenda Kaupangs egne vurderinger og anbefalinger til hvilke tiltak departementet bør prioritere framover. Disse er delt inn i regulatoriske, økonomiske, tekniske (infrastruktur), organisatoriske og pedagogiske tiltak. Målet er å legge grunnlaget for et mer helhetlig og koordinert økosystem som både ivaretar nasjonale behov og legger til rette for innovasjon, ansvarlig bruk og verdiskaping.

Når vi vurderer virkemidler, bygger vi på beskrivelsene av det norske økosystemet som er lagt fram tidligere i rapporten. Vi tar også utgangspunkt i temaer som har vist seg særlig viktige, blant annet koordinering, infrastruktur og kompetanse. Samtidig har vi vurdert hvilken rolle staten bør ha, blant annet gjennom initiativet KI Norge, og denne rollen skisseres avslutningsvis i kapittelet. Vi har gjort en prinsipiell vurdering av hvor offentlig ressursbruk eller reguleringer er mest nødvendig ut fra ulike former for markedssvikt, hensyn til fordeling, eller hensyn til sikkerhet og nasjonal suverenitet.

6.1 Overordnet vurdering

Dagens økosystem for språkmodeller i Norge preges av stort engasjement, men lav modenhet. Mange aktører, både offentlige og private, har vist interesse for å ta i bruk eller utvikle språkmodell-basert KI, og det finnes et bredt spekter av behov og forventninger i ulike sektorer. Det er et stort antall selskaper som på ulike måter bruker, eller vil bruke kunstig intelligens som del av tjenestene eller produktene de tilbyr. Antallet selskaper og virksomheter som tester eller tilbyr produkter basert på kunstig intelligens er raskt økende.

Forskning på og utvikling av språkmodeller i Norge domineres i stor grad av offentlige institusjoner som Nasjonalbiblioteket, universitetene og enkelte forskningskonsortier. Kommersielle norske aktører ser i liten grad ut til å ha tatt steget til å drifte eller tilby norske språkmodeller som produkter i stor skala. Det er mange pågående prosjekter og initiativer, men samarbeidsformene er ofte prosjektbaserte og kortvarige. Utfordringen for Norge er ikke mangel på engasjement eller initiativ, men mangel på koordinering, felles retning og infrastruktur. Der andre land har satt av betydelige ressurser og etablert nasjonale strategier for språkmodeller, forblir det norske økosystemet fragmentert. Konsekvensen kan være lavere konkurransekraft, svakere modeller for norsk språk og større avhengighet til internasjonale aktører.

Per nå ser det ut til å være liten reell etterspørsel etter å bruke norsktrente språkmodeller i sluttbrukerapplikasjoner, eller norskbasert infrastruktur for drift av språkmodeller. Lukkede modeller som tilbys av de store utenlandske selskapene dominerer. De fleste virksomheter som utvikler applikasjoner basert på språkmodeller bruker internasjonale modeller som ChatGPT. Mange kjenner til de norskutviklede språkmodellene, men de er i liten grad tilgjengelige for bruk, og de vurderes ikke som konkurransedyktige opp mot internasjonale modeller.

Dermed blir utviklingsaktiviteten i det norske økosystemet i stor grad forskningsdrevet, og verdi-kjeden stopper ofte før løsningene settes i ordinær drift. Selv om det er store forventninger knyttet til generativ kunstig intelligens, er det foreløpig begrensede investeringer i bruk av norskbaserte språkmodeller og norskbasert infrastruktur.

Vi vurderer at den offentlige ressursinnsatsen, og særlig KI Norges rolle, bør målrettes mot å stimulere til faktisk bruk av kunstig intelligens, som samtidig ivaretar norsk språk og kultur. Ved å prioritere tiltak som øker etterspørselen hos de aktørene som skal ta KI i bruk i sine ansvars-områder, kan man sikre bedre balanse mellom tilbud og behov. Dette vil bidra til at infrastrukturen som bygges opp i både privat og offentlig sektor blir relevant, nyttig og tett koblet til de konkrete utfordringene den skal løse.

Offentlig sektor, og KI Norge, bør i størst mulig grad prioritere områder som det offentlige har særlig gode forutsetninger for å løse. De neste delkapitlene kommer med noen konkrete anbefalinger.

6.2 Regulatoriske tiltak

Det regulatoriske landskapet for språkmodeller i Norge fremstår for mange som krevende og til dels uoversiktlig. Flere informanter peker på utfordringer knyttet til juridisk usikkerhet, spesielt i møte mellom EUs AI Act, GDPR, sektorlovgivning og nasjonale føringer. Dette gjelder særlig når språkmodeller skal brukes på sensitive data eller i sektorer med høye krav til sikkerhet og etterprøvbarhet, slik som helse, justis og offentlig forvaltning. Mange uttrykker bekymring for at regelverket er for generelt eller for uklart, og at det legger en dempende effekt på innovasjon og etterspørsel. Dette gjelder særlig i offentlig sektor, hvor det ofte er lav risikovilje og man er avhengig av tydelig juridisk forankring før teknologi tas i bruk i ordinær drift.

Informantene etterlyser både mer tilpasset lovverk og mer veiledning for å navigere i lovverket. Det etterlyses praktiske tiltak som veiledning, felles verktøy, evalueringsrammeverk og ordninger som gjør det mulig å teste ut språkmodeller i trygge og kontrollerte omgivelser.

Enkelte peker også på faren for at offentlige krav og anskaffelsesregimer kan bli for rigide, og dermed indirekte favorisere utenlandske løsninger som allerede er tatt i bruk, framfor norske alternativer som fortsatt er under utvikling.

Anbefalte tiltak:

- **Innføre felles standarder og rammeverk for sikker bruk av KI**
  Vi anbefaler å arbeide med å etablere tydelige mekanismer for å dokumentere at en språkmodell er «god nok» for bestemte typer bruk. Mange virksomheter gjør nå mange av de samme vurderingene knyttet til svarkvalitet, språk, skjevheter osv. Bruk av standardiserte evalueringsrammeverk og godkjenningsordninger kan redusere terskelen for bruk av språkmodeller og gjør det lettere å ta i bruk KI-løsninger utviklet i privat sektor.
- **Tydeliggjøre regelverk og praksis for behandling av sensitive data.**
  I offentlig sektor generelt er det uttrykt et sterkt behov for klarhet i hvordan personopplysninger og sensitiv informasjon kan håndteres ved bruk av språkmodeller. Det ser ut til å være store ulikheter rundt hva ulike aktører vurderer som forsvarlig infrastruktur for behandling av sensitive data i språkmodeller. Vi anbefaler å kartlegge dagens praksis og tolkninger på dette området, og hvis mulig gi felles føringer for hva som er forsvarlig praksis.

6.3 Økonomiske tiltak

Økonomiske virkemidler kan spille en viktig rolle for å stimulere til bruk og utvikling av språk-modeller. I en tidlig fase bør støtte målrettes dit hvor de har størst mulig nytteverdi, enten det gjelder å fjerne konkrete barrierer for bruk, å styrke tilgang til kritiske ressurser, eller å fremme innovasjon i samfunnskritiske sektorer.

For en del aktører vil investeringer i språkmodellbasert kunstig intelligens kunne oppleves som risikabelt og eksperimentelt. Det finnes allerede en rekke virkemidler og ordninger for å stimulere innovasjon slik som Stimulab, Medfinansieringsordningen og Skattefunn. Disse kan potensielt styrkes og i større grad prioritere bruk av språkmodellbasert kunstig intelligens, både i offentlig sektor og næringslivet. Dette vil kunne senke terskelen for å eksperimentere med nye løsninger. Det bør også vurderes å gjeninnføre StartOff, eller en tilsvarende ordning, for å stimulere innovative anskaffelser i offentlig sektor og koble behov med leverandørkompetanse på en strukturert måte.

Offentlige midler bør hovedsakelig brukes på områder som ikke prioriteres av markedet. Åpent tilgjengelige treningsdata og språkmodeller vil være fellesgoder som gir gevinster for hele bransjer, sektorer og målgrupper. Et eksempel kan utvikling av spesialiserte norske språkmodeller knyttet til bestemte domener (som juss og helse) hvor det er nødvendig at det offentlige koordinerer og (del)finansierer.

Anbefalte virkemidler:

- **Bruke innovasjonsvirkemidler til å stimulere bruk av språkmodellbasert KI**

  Ordninger som finnes i dag, bør også kunne bruke midlene til språkmodellbaserte løsninger. Det bør vurderes å styrke disse ordningene for å stimulere til økt eksperimentering og bruk av KI.

- **Finansiere frigjøring og tilgjengeliggjøring av norske data**

  Et gjennomgående hinder for trening av norske språkmodeller er begrenset tilgang til åpne og rettighetsavklarte datasett. Det bør derfor settes av midler til å kuratere, kvalitetssikre, tilpasse og tilgjengeliggjøre datasett med stor samfunnsverdi. Dette kan eksempelvis være lovtekster, anonymiserte helsedata, og digitalisert tekst fra offentlig sektor. Slike data bør så langt det er mulig også gjøres tilgjengelig for privat sektor siden data er en svært viktig innsatsfaktor i dette økosystemet.

- **Vurdere å finansiere utvikling av domenespesifikke modeller.**

  Språkmodeller som er tilpasset fagområder som helse, juss, skole og offentlig forvaltning, krever målrettet arbeid med data og fagspråk. Hvis det offentlige skal finansiere språkmodellutvikling, kan det vurderes sektorvise satsinger, der offentlige fagmyndigheter koordinerer modellutvikling innen sitt felt. Slik kan man sikre både kvalitet og relevans i modellene, og unngå dobbeltarbeid og fragmentering. Dette er nærliggende å følge opp i hver sektor, mens KI Norge potensielt kan ha en rolle i å tilrettelegge og gi veiledning til dette arbeidet.

6.4 Tilgang til infrastruktur

Tilgang til teknisk infrastruktur, for tungregning, datalagring og drift, fremstår som en grunnleggende forutsetning for utvikling og bruk av språkmodeller. Flere informanter peker på at dette ikke bare er en praktisk nødvendighet, men også et spørsmål om sikkerhetspolitikk og teknologisk suverenitet. For språkmodeller som skal brukes i offentlig sektor eller andre områder med sensitive data er det avgjørende at data og modeller kan behandles under norsk eller europeisk regelverk.

Samtidig bør staten unngå å tre for tungt inn i et marked som i stor grad fungerer. Det finnes både kommersielle og offentlige aktører som tilbyr relevant infrastruktur. Teknisk infrastruktur må forstås som en strategisk ressurs i språkmodelløkosystemet, og tiltak på dette området bør sikre *både* nasjonal kontroll og effektiv ressursutnyttelse, uten å svekke dynamikken i et allerede eksisterende og kompetent marked. I vurderinger av offentlig KI-infrastruktur er det også nyttig å differensiere mellom trening og drift av språkmodeller.

Staten bør først og fremst prioritere områder hvor markedet ikke selv løser behovene, eksempelvis knyttet til trening av norske modeller, støtte til forskning og lagring av nasjonale datasett. Det vises i denne sammenheng til Forskningsrådet konseptvalgutredning (KVU) for tungregning, som peker på sterkt økende behov for tungregneinfrastruktur. Flere informanter viser til veletablerte internasjonale samarbeid (f.eks. LUMI) som det vil være fornuftig å vurdere i sammenheng med investeringer som gjøres i Norge. Gitt dagens geopolitiske situasjon kan nordiske eller europeiske samarbeid synes mest aktuelle.

Når det gjelder infrastruktur for drift av språkmodeller er det ikke åpenbart at dette bør etableres i nasjonal statlig regi eller være offentlig subsidiert tjeneste. Det er et stort og hurtig voksende marked for slike tjenester internasjonalt, med også norske og europeiske aktører som bygger opp relevante tjenester. Internasjonale selskaper bygger også opp slik driftsinfrastruktur i Norge, eksemplifisert gjennom kunngjøringen av Stargate Norway, hvor Aker, Nscale og OpenAI planlegger å bygge et enormt datasenter for KI i Narvik. Offentlig infrastruktur som tilbyr drift av KI-løsninger i en viss skala vil i praksis konkurrere både med store internasjonale skyplattformer, og med mindre aktører i Norge og Europa som bygger opp infrastruktur som er innrettet mot økt nasjonal kontroll.

Videre vil det være naturlig å vurdere drift av KI-løsninger i sammenheng med skytjenester mer generelt. Nasjonal sikkerhetsmyndighet gjennomførte i 2022-23 en konseptvalgutredning om nasjonal skytjeneste, der ulike konsepter ble vurdert. Her vises det til at hensynet til nasjonal kontroll på den ene siden må avveies mot ønske om fleksibilitet og funksjonalitet på den andre. Hvordan disse hensynene veies opp mot hverandre er mer et spørsmål politisk prioritering heller enn en samfunnsøkonomisk vurdering (Nasjonal sikkerhetsmyndighet, 2023). Mange offentlige og private aktører, inklusiv mange av de mulige brukerne av språkmodellbasert KI i Norge, behandler allerede store mengder data i skyen. Det er derfor nærliggende å vurdere behovet for nasjonal digital infrastruktur samlet, heller enn å vurdere drift av generativ kunstig intelligens isolert.

Siden det allerede finnes private aktører som søker å etablere alternativer som gir høy grad av kontroll bør det vurderes om slike hensyn i større grad kan etterspørres i offentlige anskaffelser. Tiltak for å understøtte dette kan eksempelvis være veiledning eller maler/standardavtaler.

Anbefalte tiltak:

- **Bevilge infrastrukturmidler til tungregning via nasjonale aktører**
  Det bør settes av dedikerte midler til videreutvikling og drift av tungregningsinfrastruktur gjennom nasjonale aktører som Sigma2. Midlene bør kanaliseres med et tydelig formål om å understøtte språksuverenitet og utvikling av fellesskapsorienterte språkmodeller. Samtidig bør det stilles krav om samarbeid med europeiske initiativer som LUMI, for å sikre ressurs-utnyttelse og teknologisk samspill på tvers av land.
- **Sørge for gode rammevilkår for bruk av norske tilbydere av KI-infrastruktur**
  Det bør legges til rette for at norske aktører som tilbyr sikker norsk infrastruktur og språk-teknologi har stabile og konkurransedyktige rammevilkår. Det bør legges til rette for at strategiske hensyn som nasjonal sikkerhet, teknologisk suverenitet, og hensyn til norsk språk og kultur i større grad etterspørres i offentlige anskaffelser.

6.5 Organisatoriske tiltak

Vår vurdering er at teknologiske investeringer alene ikke er tilstrekkelige for å styrke det norske økosystemet. Det trengs målrettede organisatoriske tiltak som kobler infrastruktur, kompetanse, finansiering og samarbeid på tvers av sektorer. Investeringer i kunstig intelligens kan bli svært kostbare, og det er derfor avgjørende å bygge kapasiteten til å prioritere de riktige initiativene.

Mange offentlige aktører mangler kompetanse til å vurdere kvalitet, risiko og gevinster ved språkmodellprosjekter, og er derfor avhengige av leverandørenes vurderinger. Samtidig viser kartleggingen at offentlig etterspørsel etter generativ kunstig intelligens generelt, og norske løsninger spesielt, fortsatt er begrenset. Terskelen for å ta i bruk KI-løsninger kan senkes dersom det utarbeides tydeligere og mer enhetlige krav til hva KI-løsninger skal levere, enten det gjelder sikker infrastruktur, kvalitet, personvern eller språkmodeller som speiler norsk språk og kultur.

Direktoratet for forvaltning og økonomistyring (DFØ) har allerede etablerte et miljø innen offentlige anskaffelser, som blant annet tilbyr en markedsplass for skytjenester, og et sett med avtalemaler eksempelvis innen skyavtaler og innovative anskaffelser. Ved bruk av lignende standardiserte beskrivelser hva KI-løsninger skal levere, enten det gjelder sikker infrastruktur, kvalitet, personvern eller språkmodeller som speiler norsk språk og kultur kan dette redusere barrierer mot bruk av KI i det offentlige.

Anbefalte tiltak:

- **Bygge kapasitet innen anskaffelse av løsninger basert på norske språkmodeller**
  Vi anbefaler å bruke anskaffelsesmiljøet i DFØ til å finne egnede tiltak som kan gjøre det lettere å anskaffe forsvarlige KI-løsninger, eksempelvis språkmodeller som speiler norsk språk og verdier. Dette kan senke terskelen for å ta i bruk KI-løsninger, samtidig som det mobiliserer leverandørmarkedet til å bidra med innovasjon og løsninger som fremmer ansvarlig bruk.
- **Bygg et fagmiljø for koordinering av investeringer**
  I dag skjer investeringer, utvikling og innføring av språkmodeller i Norge fragmentert, gjennom enkeltprosjekter i offentlige etater, forskningsmiljøer og private virksomheter, ofte uten tilstrekkelig koordinering. Dette skaper risiko for overlappende satsinger, manglende standardisering og ineffektiv bruk av ressurser. En nasjonal fagmyndighet for språkmodeller, forankret i et uavhengig og tverrsektorielt mandat, bør få ansvar for å kartlegge og holde oversikt, og prioritere satsinger basert på faglige kriterier og identifiserte behov for sikker bruk og verdiskaping.

6.6 Pedagogiske tiltak

Kompetanse beskrives gjennomgående som en av de mest kritiske ressursene for utvikling, tilpasning og bruk av språkmodeller. Mange aktører beskriver mangel på spesialisert KI-kompetanse i Norge, både innen modelltrening og fintrening, drift og evaluering. Mange peker også på manglende kompetanse i anvendelse av KI-baserte språkmodeller. Verdiskapende anvendelse krever tverrfaglig kompetanse som kombinerer teknologisk forståelse med domeneinnsikt, for eksempel innen helse, juss eller språk.

Det finnes et stort antall tilgjengelige kurs og opplæringsmateriell i dag både i Norge og internasjonalt. Kompetansebehovene gjelder på alle nivåer og i et stort antall ulike sektorer og fagfelt. Det vil derfor være kritisk at eventuelle opplæringstiltak er fundert i reelle behov og tilstrekkelig spisset.

Videre er det naturlig å gå ut fra at en stor del av kompetanseoppbyggingen skjer gjennom igangsetting av konkrete prosjekter og tiltak der kunstig intelligens brukes på eget fag- eller ansvarsområde. En viktig del av kompetansehevingen vil derfor være å sørge for at det settes i gang tilstrekkelig med aktivitet på området.

Anbefalte tiltak:

- **Bygge opp og styrke offentlige mekanismer for proaktiv veiledning**
  Vi anbefaler å satse på regulatoriske sandkasser og pilotordninger der språkmodeller kan prøves ut under realistiske forhold, men med støtte fra relevante myndigheter. Tilgang til veiledning og testarenaer kan senke terskelen for bruk, særlig i offentlig sektor, og gjør det lettere å gå fra pilot til produksjon.
- **Gjennomføre en systematisk kompetansekartlegging**
  Kompetansekartleggingen bør søke å identifisere dagens kompetansegap innenfor ulike områder, og være bør være bred nok til å favne alle nivåer (ledere, fagpersoner, prosjekt-ledere, utviklere, spesialister mv) og foreslå målrettede tiltak. Det bør vurderes et bredt sett av

virkemidler, slik som opplæringsprogrammer, incentivordninger og regler for å få inn spesialisert arbeidskraft fra utlandet.26

- **Legge til rette for aktiv eksperimentering**

  Den største kompetansehevingen kommer trolig når generativ kunstig intelligens brukes i praksis på egne fagområder. Å legge til rette for pilotprosjekter der teknologien kan testes i forsvarlige rammer vurderes derfor som vesentlig. Mange av de øvrige tiltakene nevnt over understøtter nettopp dette.

6.7 Anbefalt rolle for KI Norge

Vår vurdering av hva KI Norges rolle bør være samsvarer i stor grad med rollen de allerede er tiltenkt, se omtale i kapittel 5.1. Dette gjelder særlig KI Norges tiltenkte rolle som pådriver for ansvarlig bruk av kunstig intelligens, og strategien om å legge til rette for eksperimentering under trygge rammer gjennom regulatoriske sandkasser. Proaktiv veiledning rundt et sammensatt regelverk er viktig og etterspurt, og det tar utgangspunkt i en spisskompetanse som de tre virksomhetene i KI Norge allerede besitter.

I tillegg til dette vurderer vi at KI Norge kan ta rollen som en form for offentlig fagmyndighet for kunstig intelligens i Norge. I praksis vil dette bety å ha oppdatert og relevant faginnsikt til å kunne gjøre løpende gode vurdering av hvilke tiltak som er nødvendig i økosystemet. Dette forutsetter tett dialog med miljøer på tvers av offentlig og privat sektor. Mange av de foreslått tiltakene over vil naturlig være en del av KI Norges oppgaveportefølje. Dette inkluderer:

- Ha oversikt over, vurdere og anbefale felles standarder og rammeverk for sikker bruk av KI
- Identifisere utfordringer forbundet med ulike regelverk og praksis for KI (inkl. behandling av sensitive data) og iverksette egnede tiltak opp mot etater og departement.
- Gjøre faglige vurderinger av hvilke data som bør tilgjengeliggjøres ut fra en helhetlig vurdering, og eventuelt bistå i prosessene med å avklare rettigheter
- Bistå sektormyndigheter som ønsker å utvikle domenespesifikke språkmodeller
- Bistå anskaffelsesmiljøet i DFØ med å legge til rette for sikker og ansvarlig bruk av KI i offentlige anskaffelser
- Gjennomføre samfunnsøkonomiske vurderinger av investeringer knyttet til språkmodeller som berører mange sektorer (frikjøp av innhold, investeringer i felles infrastruktur, utvikling av norske språkmodeller mv)
- Vurdere samlede kompetansebehov og tiltak.

[26] En norsk variant av “tech-visum” eller talentbaserte oppholdstillatelser kombinert med økonomiske incentiver som gjør det mer attraktivt for eksperter. Norge kunne lage bilaterale avtaler med land som har kompetanse. gjensidig utveksling av teknologer og forskere. Styrkede forskningsstipender og talentprogrammer innen AI

Vedlegg 1 Datainnsamling og metode

Kartleggingen er basert på to typer datainnsamling: dokumentanalyse og semi-strukturerte intervjuer.

Litteraturgjennomgang

Vi har gjennomført en litteraturgjennomgang for å kartlegge aktører, prosjekter, roller og teknologiske tilnærminger når det gjelder språkmodeller. I vår søkestrategi har hovedvekten vært på norske bidrag, men vi har også gjennomgått internasjonale publikasjoner om emnet.

Dokumenter og litteratur vi har sett på omfatter:

- Offentlige styringsdokumenter, stortingsmeldinger og strategier relatert til digitalisering, KI og forskning.
- Rapporter og prosjektbeskrivelser fra sentrale organisasjoner som Nasjonalbiblioteket, Forskningsrådet, NORA.AI, og Sigma2.
- Faglige artikler og medieomtale som belyser sentrale aktører og utviklingstrekk.
- Åpent tilgjengelig informasjon fra bedrifter og forskningsmiljøer (nettsider, pressemeldinger).
- Utviklerportaler som GitHub, Hugging Face mv. som gir strukturert informasjon om hvilke modeller som har blitt utviklet, og ulike iterasjoner med videreutvikling (gjør det mulig å se relasjoner mellom data, språkmodeller og rammeverk).

Semi-strukturerte intervjuer

Den andre datakilden vi har brukt er semistrukturerte intervjuer med nøkkelaktører i økosystemet rundt språkmodeller. Målet er å få frem praktiske erfaringer, innsikt i barrierer og behov, samt syn på samspill og koordinering. Intervjuutvalget har dekket flere ledd i verdikjeden, inkludert:

Verdikjedeledd	Informanter
Språkmodell infrastruktur	• Intility: utvikler intern løsning («Intility GPT») for sikrere bruk av språkmodeller. • Norwegian AI Cloud (NAIC): nasjonal skyinfrastruktur for AI, med fokus på suverenitet og støtte til mindre språkmodeller • Sigma2: nasjonal leverandør av tungregnings- og lagringsinfrastruktur for forskning • Telenor AI factory: tilbyr infrastruktur for drift av språkmodeller
Språkmodell utvikling (Grunnmodeller og fintrening)	• Bineric: oppstartsbedrift innen språkmodellbasert KI, fintrener åpne LLM-er (LLaMA, Mistral, Qwen) for skandinaviske språk • Divvun/UiT: utvikler språk- og taleteknologi for samiske språk, inkludert egne talegjenkjenningsmodeller • Nasjonalbiblioteket: leverer både datasett og infrastruktur for språkmodellutvikling • NTNU (NorwAI): utvikler og trener norsktrente språkmodeller basert på åpne modeller som Lama og Mistral • UiO (Language Technology Group): utvikler åpne norske språkmodeller og verktøy, både fra bunnen og via varmstart

Språkmodell anvendelse	• Agder AI/Egde Consulting: anvender og utvikler løsninger for offentlig sektor, basert på kommersielle modeller • CapGemini: konsulentselskap som tilbyr rådgiving og integrerer språkmodeller i kundeløsninger • Chronos: teknologiselskap som utvikler løsninger innen kunstig intelligens, blant annet en juridisk KI-assistent • NAV: bruker språkmodeller eksternt mot brukere og internt for effektivisering av saksbehandling. • Nemonoor/Digital Norway: bistår virksomheter med bruk og tilpasning av språkmodeller, særlig gjennom opplæring • NORA: nasjonalt forskningskonsortium for kunstig intelligens som fremmer forskning, utdanning, innovasjon og samarbeid på tvers av akademia, næringsliv og offentlig sektor. • NRK: bruker språkmodeller for transkribering, metadata-generering og personalisering av innhold. Bidrar også inn i store prosjekter for å utvikle språkmodeller. • Skatteetaten: har et innovasjonsteam som utforsker og tester mulige bruksområder blant annet for kunstig intelligens • Sopra Steria: konsulentselskap som tilbyr rådgiving og integrerer språkmodeller i kundeløsninger • Vend/Schibsted: bruker eksisterende modeller i medieproduksjon, deltar i utviklingsprosjekter med NB
Offentlige myndigheter	• Digdir: Utvikler og forvalte felles retningslinjer, metoder og verktøy for ansvarlig bruk av KI i offentlig sektor • Forskningsrådet: Finansierer KI – forskning • Helsedirektoratet: Arbeider med å tilrettelegge for trygg bruk av KI i helsesektoren • NKOM: Tilsyns- og koordineringsmyndighet for kunstig intelligens i Norge.

Det er til sammen gjennomført 23 intervjuer. Aktørene er satt sammen for å dekke en bredde når det gjelder ulike roller i økosystemet (infrastruktur, språkmodell utvikling mv), og på tvers av offentlig og privat sektor.

I gjennomføring av intervjuing har vi benyttet oss av en *snowball sampling*-tilnærming for å identifisere relevante aktører: Hver intervjuet aktør ble bedt om å peke på andre sentrale aktører, initiativer eller miljøer de samarbeider med eller kjenner til. På denne måten håpet vi å fange opp mindre synlige eller nye aktører, og avdekker nettverk og systemiske sammenhenger i økosystemet. Metoden kan føre til skjevheter i utvalget, men vi mener likevel vi har grunnlag for en rimelig dekkende beskrivelse av økosystemet.

Bruk av KI i datainnsamling

Vi har benyttet oss av KI i litteraturgjennomgangen og til oppsummering av intervjuer.
I litteraturgjennomgangen er artikler gjennomgått. I tillegg har vi brukt KI for å identifisere og informasjon fra litteraturen som er relevant for prosjektets problemstillinger.

Intervjuene er gjennomført i Teams og vi har benyttet oss av transkripsjonsfunksjonen i Teams. Transkripsjonen er lastet opp i en GPT sammen med intervjuguide for å generere gode intervjureferater. Informantene er informert om transkriberingen og har gitt sitt samtykke.

GPT-en vi har brukt, er en versjon hvor data vi legger inn ikke brukes til trening av modellen. For å sikre kvalitet datainnsamlingen har vi gjennomført en kvalitetssikringsprosess i alle faser av arbeidet. Etter at intervjureferatene var generert ved hjelp av GPT ble de gjennomlest og for å kontrollere at innhold ble korrekt gjengitt i forhold til vår opplevelse fra intervjuet. Eventuelle feil, uklarheter eller mangler ble rettet opp manuelt. Tilsvarende ble informasjon hentet fra litteratur-gjennomgangen kontrollert mot de opprinnelige kildene for å sikre at sitater, begrepsbruk og konklusjoner var riktig gjengitt. Denne dobbeltkontrollen, både av intervjureferater og litteratur-grunnlag, mener vi redusere risikoen for feil og sikret at den endelige fremstillingen bygger på et pålitelig og godt dokumentert datagrunnlag. Promptene som har vært brukt har blitt utformet med bakgrunn i forskningsanbefalinger for å redusere risiko for hallusinasjoner, overgeneralisering og feilinformasjon.

Litteraturliste

Abecasis og De Michiels, Innsiktsrapport fra Copenhagen Economics. Generative AI Partnerships: Separating Good from Bad (CPI Antitrust Chronicle, juli 2024).*AI Foundation Models Update paper (April 2024). Competition and Markets Authority*

AI Index (2025). *Artificial Intelligence Index Report 2025*. Stanford: Stanford HAI.

Buono, D., Felecan, M. & Tessitore, C. (2024). *An introduction to Large Language Models and their relevance for statistical offices*. Luxembourg: Publications Office of the European Union.

DataNorge og AI (2024). *Norway's National Data Portal – improving data discovery using AI*. Digitaliseringsdirektoratet.

DFD (2020). *Nasjonal strategi for kunstig intelligens*. Digitaliseringsdepartementet.

Digitaliseringsdirektoratet. Oversikt over prosjekter i offentlig sektor (Nettside: https://data.norge.no/kunstig-intelligens)

Forskningsrådet (2024). *Behov for tungregnekraft for forskning og kunstig intelligens*. Oslo: Norges forskningsråd.

Helsedirektoratet (2024a) Store språkmodellar i helse- og omsorgstenesta – eit kunnskapsgrunnlag.

Helsedirektoratet (2024b). *Felles KI-plan for trygg og effektiv bruk av KI i helse- og omsorgstjenesten 2024–2025*. Oslo: Helsedirektoratet.

Helsedirektoratet (2025). *Rapport om store språkmodeller i helse- og omsorgstjenesten: risikoer og tilpasninger til norske forhold*. Oslo: Helsedirektoratet.

Independent High Level Expert Group set up by the European Commission (2019): A definition of AI: Main capabilities and disciplines.

Kultur- og likestillingsdepartementet. Prop 1S for budsjettåret 2025 under Kultur- og likestillingsdepartementet (2024-2025) Statsbudsjettet.

Kunnskapsdepartementet (2025) Sikt tildelingsbrev for 2025.

Mimir-prosjektet (2024). Evaluering av virkningen av opphavsrettsbeskyttet materiale på generative store språkmodeller for norske språk. Teknisk rapport. Tilgjengelig fra: 20240807_The Mímir Project - Technical Report [Short] (nob)

OECD Digital Economy Papers (2023) AI Language Models: Technological, Socio-Economic and Policy Considerations. Utarbeidet av OECD-sekretariatet i samarbeid med AIGO og CDEP

Nasjonal sikkerhetsmyndighet (2023) Nasjonal skytjeneste: Konseptvalgutredning. Oslo.

Nasjonalbiblioteket (2024). *KI i Nasjonalbiblioteket – Prosjekt Mímir*. Oslo: Nasjonalbiblioteket.(Powerpoint presentasjon)

Rambøll (2025) Generativ KI i statlige virksomheter Oslo: Rambøll/Comte Bureau

RankmyAI (2025). *AI Report Norway 2025*. Bergen: NHH Norwegian School of Economics & RankmyAI.

Teknologirådet (2024). *Generativ kunstig intelligens i Norge: Teknologi, betydning og tiltak*. Oslo: Teknologirådet.

Digitaliserings- og forvaltningsdepartementet (2025) Tildelingsbrev 2025 for Datatilsynet.

Digitaliserings- og forvaltningsdepartementet (2025) Tildelingsbrev 2025 for Digitaliseringsdirektoratet.

Digitaliserings- og forvaltningsdepartementet (2025) Supplerende tildelingsbrev nr. 2 til Nkom for 2025

Vestlandsforsking og Rambøll (2022) Bruk av kunstig intelligens i offentlig sektor og risiko for diskriminering.

AGENDA
KAUPANG

Agenda Kaupang bidrar til omstilling og utvikling av offentlig sektor. Vi bistår ledere og medarbeidere med faktabaserte beslutningsgrunnlag og effektivisering av prosesser. Agenda Kaupang gjennomfører analyser og rådgiving innen ledelsesutvikling, styring, økonomi, organisasjonsutvikling og digitalisering.