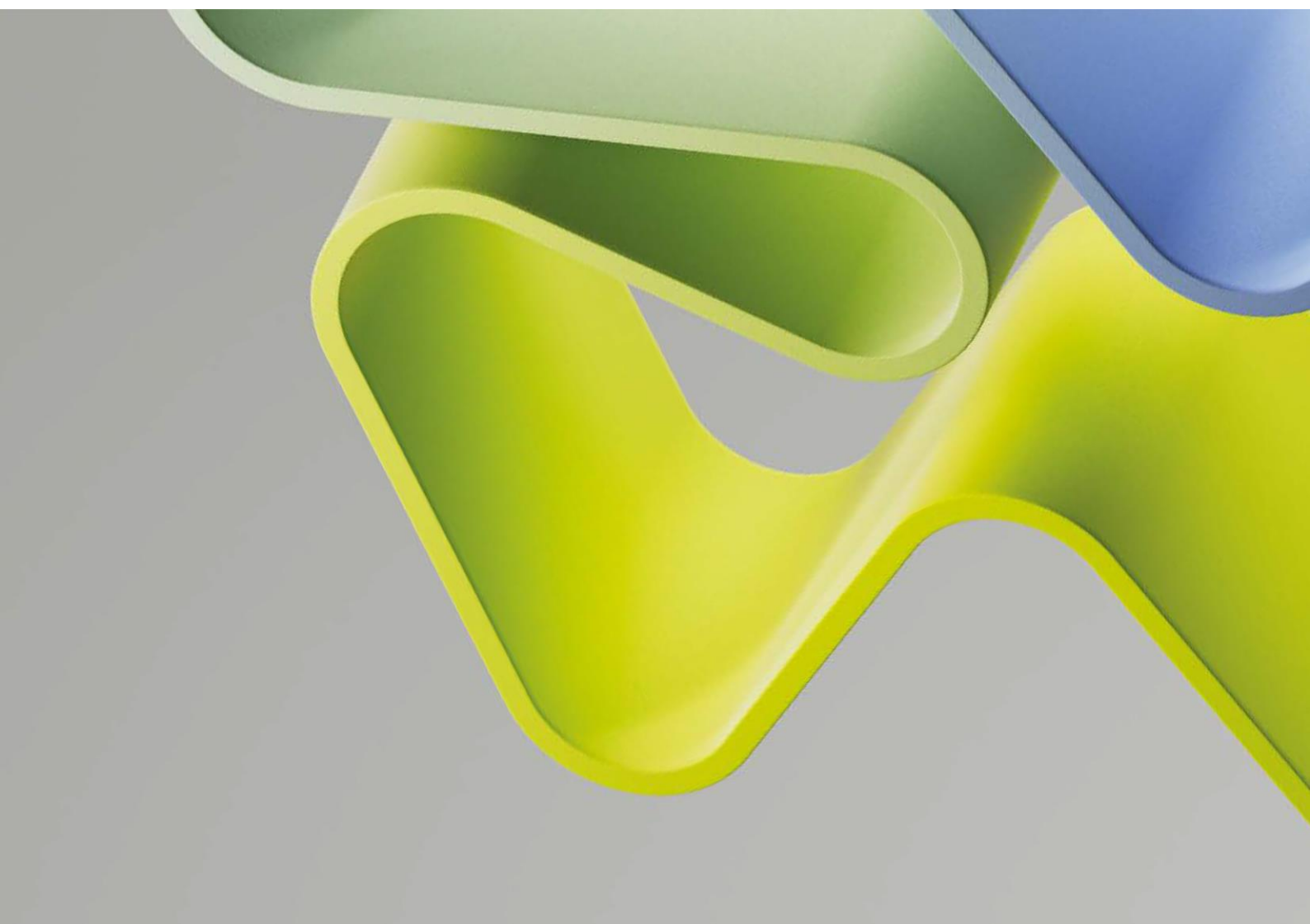


Rapport til Kunnskapsdepartementet fra Norges Forskningsråd

Behov for tungregnekraft for forskning og kunstig intelligens



Behov for tungregnekraft for forskning og kunstig intelligens	2
Sammendrag	4
Oppsummering av funn og Forskningsrådets anbefalinger	5
Innledning	8
Oppdraget	9
Avgrensing og gjennomføring	9
Dagens omfang, sammensetning og forventet utvikling i et femårsperspektiv	11
Sigma2	11
Forskningsinstitusjonene	12
Samfunnsnyttige tjenester basert på tungregning	14
Tungregnekraft til språkmodeller	16
Kvanteteknologi	16
Næringsliv	17
Tilgang til tungregnerressurser på ulike graderingsnivåer	18
Behov for tungregnerressurser lokalisert i Norge	19
Funn og anbefalinger	20
Ambisjonsnivået i våre naboland	23
Funn og anbefalinger	24
Deltagelse og muligheter i EuroHPC	25
Norsk deltagelse og innretning i dag	25
Norsk deltagelse og innretning fremover	27
Funn og anbefalinger	29
Kostnadsanslag i et femårsperspektiv	30
Investeringer	30
Drift	30
Totalkostnad i en femårsperiode	30
Funn og anbefalinger	31
Vedleggsoversikt	32

Sammendrag

Det økende volumet og mangfoldet av data kombinert med nye og mer avanserte algoritmer gir store muligheter for forsknings- og innovasjonsmiljøer til å utvikle ny kunnskap, produkter og tjenester. Mulighetene for hva som kan modelleres har økt enormt, men det krever mye regnekapasitet. Generativ KI har bidratt til en eksplosiv økning i behov for tungregnekraft.

5. April 2024 mottok Forskningsrådet et oppdrag fra Kunnskapsdepartementet (KD) om å utrede behov for tungregnekraft for forskning og kunstig intelligens. Med utgangspunkt i kjente behov i forskningssektoren, offentlig sektor og næringslivet, i tillegg til vurderinger knyttet til fremtidige ambisjoner, ønsker departementet at utredningen svarer på to grupper av spørsmål:

- 1) Omfang og sammensetning av tungregneressursene i dag
- 2) Forventet utvikling i et femårsperspektiv

Vi er glade for at KD ønsker å gjennomføre en slik utredning siden dette er viktig infrastruktur for Norge. Underveis har vi hatt løpende dialog med KD for avklaringer om gjennomføring og avgrensninger. Vi har gjennom arbeidet fått belyst betydningen og behov for tungregning, og bygger opp under tidligere utredninger. Denne rapporten vil være et viktig kunnskapsgrunnlag for videre arbeidet med oppfølging av investeringer og innretning av tungregning for forskning, forvaltning og næringsliv i Norge.

I arbeidet med utredningen har Forskningsrådet vært i kontakt med utvalgte relevante offentlige og private aktører innenfor og utenfor forskningssystemet, inkludert leverandører av skjermingsverdig og gradert FoU. Vi har også innhentet informasjon fra våre nordiske naboland og relevante EU-land.

Nåværende kapasitet innen tungregnekraft er ikke dekkende for dagens kjente behov. Å holde investeringene på samme nivå som i dag, tilsier at gapet mellom behov og kapasitet vil fortsette å øke. Det er derfor avgjørende at investeringene i nasjonale tungregneressurser økes i årene framover for å opprettholde forskning av høy kvalitet, sikre samfunnskritiske tjenester, bygge kompetanse og ekspertise, og styrke norsk konkurranseevne og teknologisk suverenitet.

Mange forskere og forskningsmiljøer kan ikke kjøpe regnekraft fra kommersielle tilbydere fordi sammensetningen av regnekraft og tjenester fra de kommersielle ikke er tilpasset forskningens behov. EU har nylig konkludert med at kommersielle tjenestetilbydere ikke vil kunne gi forskning, offentlig og deler av privat sektor tilstrekkelig tilgang til tungregnekraft for utvikling av KI-modeller, og at det derfor bør være en offentlig oppgave.

Næringsliv og offentlig sektor er avhengig av kompetanse og kapasitet for å utvikle og ta i bruk tjenester og produkter som samfunnet har behov for. For å nyttiggjøre oss mulighetene teknologien og det store tilfanget av data gir må vi bygge kompetanse, og forskningsmiljøer i verdensklasse. Da trengs tilgang til tilstrekkelig tungregningskapasitet, samt infrastruktur og støttetjenester for håndtering av data. Investering i tungregning er nøkkelen til å bygge kompetanse i norske forskningsmiljøer og utvikle produkter og tjenester som er viktig for Norge. Et eksempel på dette er utvikling av norsk språkmodell.

Det er knyttet stor usikkerhet til fremtidig behov, dette gjelder tungregneressurser både for tradisjonelle tungregneoppgaver og spesielt for regnessurser optimalisert for kunstig intelligens (KI). Selv om det er stor usikkerhet har alle aktørene vi har hentet informasjon fra pekt på at behovet er stort og raskt voksende. Alle som vil utvikle, tilpasse og drifte store språkmodeller eller som vil bygge nye generative KI-tjenester på toppen av disse, trenger tilgang på regnekraft. Kunstig intelligens blir nå relevant for stadig flere forskningsfelt og offentlige og private aktører. Dermed vil også etterspørselen etter regnekraft øke kraftig. Disse vurderingene deles også i våre naboland.

Våre nordiske naboer og relevante EU-land ligger på et betydelig høyere nivå enn Norge, både når det gjelder investeringer og tilgang til i nasjonale ressurser, samt investeringer gjennom EuroHPC. De fleste landene vi har undersøkt planlegger store investeringer framover, særlig gjennom EuroHPC.

Vi foreslår en opptrappingsplan for investeringer i nasjonal tungregneinfrastruktur, samt tiltak for å styrke samarbeidet med EuroHPC.

Oppsummering av funn og Forskningsrådets anbefalinger

Her følger en oppsummering av våre viktigste funn og anbefalinger sortert etter spørsmålene i oppdraget.

1a) Hva slags norskeid offentlig og privat tungregnekraft er i dag tilgjengelig i Norge til tradisjonelle tungregneoppgaver og kunstig intelligens?

- **Sigma2 er den nasjonale tilbyderen av tungregnetjenester for forskning.** De fire største universitetene er de viktigste brukerne, men svært mange av de forspurte forskningsinstitusjonene benytter seg av Sigma2s tjenester. Også offentlig sektor og private aktører kan få tilgang, men kapasiteten er begrenset.
- **Noen institusjoner har egne lokale ressurser** mens andre er avhengige av ressurser gjennom Sigma2, LUMI, EuroHPC eller andre samarbeidspartnere. De institusjonene som har mest egne ressurser, er også de største brukerne av Sigma2, LUMI og EuroHPC (og/eller andre internasjonale ressurser).
- **Private norske leverandører av tungregnekraft benyttes ikke/i svært liten grad av forskningsmiljøene**, bortsett fra i prosjektsamarbeid med norske industriaktører som har tilgang til egne regneressurser. Kun noen få institusjoner svarer at de utelukkende benytter kommersielle skytjenester, og da de internasjonale.
- **Nåværende kapasitet innen tungregnekraft er ikke dekkende for dagens kjente behov.**
- **De deler av næringslivet som bruker tungregneressurser i dag, bruker i hovedsak egne ressurser eller kommersielle tilbydere.** Dette gjelder etter det vi kjenner til også for brukere av regneressurser i deler av offentlig sektor. Forskningsmiljøenes ressurser brukes i noen tilfeller der det er samarbeid.

1b) Hvilken tilgang har norske forsknings- og innovasjonsmiljøer, inkludert næringslivet, i dag til forskjellige typer tungregnekraft gjennom internasjonale forsknings- og innovasjonsprogrammer?

- **Norske aktører har dedikert tilgang til nasjonal kvote på LUMI-maskinen** i Finland, i tillegg til muligheter for å søke tilgang til EuroHPC-anlegg basert på kvalitet i prosjektet.

1c) Hvilken tilgang har forskningssektoren, offentlig sektor og næringslivet til forskjellige typer tungregnekraft på ulike graderingsnivåer?

- FFI og NSM har tilgang til tungregneressurser på gradert nivå. Ytterligere informasjon om dette er gradert, det er derfor ikke omtalt i mer detalj i denne rapporten.

2a) Hvor mye og hvilken sammensetning av tungregnekraft vil Norge trenge, gitt ambisjonsnivået om å være på nivå med våre nordiske naboer og relevante EU-land?

- Det er økende behov fremover både for tradisjonell tungregning og for KI. Behovet er økende i alle sektorer. Utviklingen dreier mot mer bruk av GPU (ikke bare til KI). 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**. Det er knyttet størst usikkerhet knyttet til utviklingen av GPU-behov.
- Å holde investeringene på samme nivå som i dag, tilsier at **gapet mellom behov og kapasitet vil fortsette å øke.**
- Alle andre sammenlignbare land investerer betydelige offentlige midler i nasjonale tungregneressurser for forskning og forvaltning.
- Alle andre sammenlignbare land i Europa investerer betydelige offentlige midler i EuroHPC.
- For tilgang til særlig store superdatamaskiner bør Norge samarbeide internasjonalt, fortrinnsvis i Norden og Europa (EuroHPC).

2b) Hvordan skal man innrette seg for å få mest mulig utbytte av investeringen og deltakelsen i det internasjonale partnerskapet EuroHPC og internasjonale samarbeidsprogrammer som f.eks. Horisont Europa og DIGITAL?

- Utviklingen i EuroHPC og beslutninger i styringsorganet skjer raskt. For å sikre norske aktører de beste muligheter i partnerskapet, **anbefaler vi at det settes av en nasjonal EuroHPC-pott, som raskt kan mobiliseres og settes i arbeid.**
- Retur-EU ordningen må opprettholdes og være forutsigbar.
- **Vi anbefaler at det etableres en medfinansieringsordning** som kompenser for norske forskningsinstitusjonenes, særlig instituttenes, manglende fullfinansiering av faktiske kostnader i prosjektene fra DIGITAL. Dette for å sikre forutsigbarhet og konkurransedyktighet hos norske miljøer, i tråd med strategiske prioriteringer i den norske DIGITAL-strategien.
- **Vi anbefaler videreføring for nasjonalt kompetansesenter for HPC som tilbyr brukerstøtte og opplæring for offentlig sektor og næringsliv.** Senteret må tilpasse aktivitetene for å utnytte nye muligheter i EuroHPC.
- Vi anbefaler at det settes av **tilstrekkelige ressurser til administrativ oppfølging av EuroHPC.** Dette for å sikre god og tilstrekkelig koordinering med relevante aktører og oppfølging av det administrative arbeidet knyttet til aktiviteter i EuroHPC.
- Det kan være muligheter for å være vertskap for et EuroHPC-anlegg på norsk jord, delfinansiert av EuroHPC og et konsortium. Vi anbefaler at det gjøres en kost-nytte-analyse samt en juridisk vurdering av EØS-lands muligheter for å være fysisk vertskap for en forskningsinfrastruktur som delvis er eid av EuroHPC.
- **Det trengs gode støttetjenester som er tilpasset ulike brukergruppers behov** (forskningsmiljøene, offentlig sektor og næringsliv) for bedre utnyttelse av regneressursene både nasjonalt, men spesielt ressursene vi får tilgang til gjennom EuroHPC.

2c) Hvor stor andel av de fremtidige tungregnerressursene bør lokaliseres i Norge ut fra nasjonale behov for sikkerhet og beredskap?

- **Det er viktig å ha tilgjengelige norskeide ressurser både lokalt, nasjonalt og internasjonalt** for å ivareta ulike behov.
- **Vi anbefaler at det fortsatt investeres betydelige beløp i tungregneinfrastruktur lokalisert i Norge.** Det er to hovedgrunner til dette:
 - **Sikkerhet og beredskap:** risiko for at Norges tilgang til internasjonale tungregnekapasitet blir nedprioritert ved usikkerhet og kriser.
 - **Bygge kompetanse og ekspertise:** norske miljøer må ha enkel tilgang til lokale og nasjonale tungregnerressurser som treningsarenaer.
- Det er **viktig å ha tilstrekkelig nasjonal kontroll over tungregnerressurser**, spesielt for oppgaver som er relatert til forsvar, klima, helse, geofarer og matsikkerhet. Ressurser som eventuelt ikke er plassert i Norge bør ligge i land vi har sikkerhetspolitisk samarbeid med.
- **Nærhet mellom bruker og tungregnerressurser er viktig**, særlig for mindre erfarne brukere. Dette er viktig for å bygge kompetanse både i forskningsmiljøene og i offentlig sektor og næringsliv og enklere kunne realisere samarbeid mellom ulike fagmiljøer fra ulike sektorer (forskning, offentlig sektor og næringsliv).

2d) Hvordan kan man dele på tungregnerressurser på ulike graderingsnivåer i et nasjonalt perspektiv?

- **Det vil være et økende behov for tungregnekapasitet i forsvarssektoren**, både egne graderte systemer som blant annet blir brukt til å vurdere ytelse og trusler, men også ugraderte systemer.
- **Skjerming av informasjon og FoU er faglig krevende og kostbart.** Riktig bruk av graderingsnivå for tungregnerressurser kan bidra til økt sikkerhet og er potensielt kostnadsbesparende.
- **Det bør derfor være en ambisjon å gjenbruke åpen/felles infrastruktur for tungregning innen forsvarssektoren så langt som mulig**, og vurdere behov for å bruke graderte systemer nøye. Det er en forutsetning at det er høyt kunnskapsnivå på å vurdere nødvendigheten av bruk av graderte vs. åpen infrastruktur for tungregning.

2e) Hva er kostnadsanslagene forbundet med punktene over, fordelt på kostnader til investering og drift?

- **Vi anbefaler å lage en opptrappingsplan for investeringer og drift på minst 2 600 millioner kroner for den neste femårsperioden for å dekke behovene innen forskning og utvikling.** Dette inkluderer ikke kostnader tungregning på gradert nivå eller spesielle tilpasninger for personsensitive data.
- **Vi foreslår en samlet investering til den nye KI-relaterte LUMI-maskinen som utgjør 10-20 prosent av den foreslåtte investeringssummen til nasjonal tungregning (Sigma2), over perioden 2026-2027.**
- **Opptrappingsplanen bør revideres årlig.** Det vurderes at estimatet vil endre seg mye både på grunn av usikkerhet i utviklingen i faktisk behov, teknologiutvikling, samt usikre priser på hardware.

Andre funn og anbefalinger

- Flere samfunnsnyttige og samfunnskritiske tjenester utvikles ved bruk av tungregning og samarbeid med forskningsmiljøer.
- **Samfunnskritiske tjenester har behov for regneressurser med rask og forutsigbar tilgang og lite nedetid.**
- **Det må allokeres midler til regnekraft og drift for opprettholdelse og utvikling av samfunnsnyttige/samfunnskritiske tjenester.** Dette inkluderer trening og operativ drift av modeller for andre brukere enn forskning, slik som værvarslingstjenester og språkmodeller.
- **Norskeide regneressurser bør samordnes bedre** for å kunne utnytte kapasitet og kompetanse i kriser.
- Vi anbefaler at det utvikles en egen **nasjonal strategi for kvanteteknologi.**
- De viktigste utfordringene for å ta i bruk av data, KI-verktøy, inkludert tungregning i næringslivet er:
 - manglende kompetanse og innsikt,
 - bekymringer rundt etikk og lovverk (rettigheter til bruk av data og personvern),
 - tilgang til data av god kvalitet.

Vi har stor grunn til å tro at dette også gjelder store deler av offentlig sektor.

- I næringslivet, spesielt SMBer og startup-bedrifter, er det **behov for spisskompetanse, opplæring og brukerstøtte for bruk av data, KI og tungregning.** Det vil være nødvendig med tettere koblinger til forskningssystemet. Dette vil bidra til å ta forskningen og kompetansen i forskningsinstitusjonene raskere i bruk.
- Det er også **behov for infrastruktur og verktøy for datahåndtering** koplet til tungregneressursene.
- Det er **behov for kompetanse og kapasitet for å gjøre data tilgjengelig for bruk.**
- Det er **behov for kompetanse på juridiske og regulatoriske bestemmelser for bruk av data** (inkl rettigheter, personvern mm.).
- Det er **behov for godt samarbeid mellom lokale, nasjonale og internasjonale tungregneressurser**, for bedre utnyttelse av kompetanse og kapasitet.
- Investeringene i tungregning må være **så klimavennlige som mulig**

Innledning

Tungregning, også kjent som High Performance Computing (HPC), er bruk av kraftige datamaskiner og systemer for å utføre komplekse beregninger og analysere på store datamengder som ikke er mulig på vanlige datamaskiner. Disse beregningene krever spesialiserte superdatamaskiner eller beregningsklynger. Gjennom å investere i tungregning, kan Norge sikre at landet er rustet til å møte fremtidens utfordringer og muligheter, både innen forskning, offentlig forvaltning, næringsliv og samfunnsutvikling.

Det økende volumet og mangfoldet av data kombinert med nye og mer avanserte algoritmer gir store muligheter for forsknings- og innovasjonsmiljøer til å utvikle ny kunnskap, produkter og tjenester. Mulighetene for hva som kan modelleres har økt enormt, men det krever mye regnekapasitet.

Tilstrekkelig tilgang til regnekraft er en absolutt nødvendighet for å drive forskning av høy kvalitet, men også for å drive effektiv og god forvaltning på områder som f.eks. helse, klima, miljø, havbruk, landbruk og beredskap. Bruk av superdatamaskiner bidrar til å løse komplekse problemstillinger og utvikle samfunnsnyttige tjenester, som å utvikle nye behandlinger basert på personlig medisin, eller bedre forutsi og håndtere effekten av naturkatastrofer gjennom bruk av avanserte datasimuleringer.

Tilgang på regnekraft og grunnmodeller for kunstig intelligens er med på å åpne nye kommersielle muligheter i privat sektor, og er dermed avgjørende for at Norge skal lykkes med forskning, innovasjon og verdiskaping med KI. Norske grunnmodeller kan videreutvikles av næringsliv eller offentlig sektor for å lage f.eks. språkmodell-baserte KI-verktøy for mer effektive tjenester og produksjon.

Mange forskere og forskningsmiljøer kan ikke kjøpe regnekraft fra kommersielle tilbydere fordi sammensetningen av regnekraft og tjenester fra de kommersielle ikke er tilpasset forskningens behov. EU har nylig konkludert med at kommersielle tjenestetilbydere ikke vil kunne gi forskning, offentlig og deler av privat sektor tilstrekkelig tilgang til tungregnekraft for utvikling av KI-modeller, og at det derfor bør være en offentlig oppgave.

Nasjonal infrastruktur som muliggjør eksperimentell tungeregning- og KI-forskning er avgjørende for å utvikle og opprettholde den spisskompetansen på teknologi og metoder innen tungregning, som blant annet næringslivet etterspør. Å bygge gode forskningsmiljøer er nødvendig for å gi næringsliv og offentlig sektor de tjenestene vi har behov for i Norge. Norge har særlige behov som kun kan ivaretas ved å videreutvikle norske forskningsmiljøer innen tungregning. Dette gjelder særlig utvikling av norsk språkmodell, men også for å utvikle andre tjenester for det norske samfunnet. Dette er en grunn til at andre land investerer tungt

Våre nordiske naboer og relevante EU-land ligger på et jevnt høyere nivå enn Norge, både når det gjelder investeringer i nasjonale ressurser, samt investeringer i internasjonale partnerskap gjennom EuroHPC.

Stortinget har i Innst. 170 S (2022–2023)¹ sluttet seg til regjeringens ambisjon om å sørge for tilstrekkelig nasjonal tungregnekapasitet til å dekke nåværende og fremtidige behov i forskning og forvaltning, samt at forskere ved norske universiteter og institutter skal ha tilgang til datainfrastrukturer som muliggjør forskning og utdanning i verdensklasse, jf. Meld. St. 5 (2022–2023) *Langtidsplan for forskning og høyere utdanning*².

Nåværende kapasitet i Norge innen tungregning for forskning og KI er ikke engang dekkende for dagens kjente behov. Å holde investeringene på samme nivå som i dag, tilsier at gapet mellom behov og kapasitet vil fortsette å øke og Norge vil sakke enda mer akterut når vi sammenligner oss med nordiske og relevante EU-land. Det er derfor avgjørende for å opprettholde forskning av høy kvalitet, samt sikre samfunnskritiske tjenester at investeringene i tungregnekraft økes i årene framover.

¹ [Innstilling 170 S \(2022–2023\) \(stortinget.no\)](#)

² [Meld. St. 5 \(2022–2023\) \(regjeringen.no\)](#)

Oppdraget

Denne rapporten er utarbeidet på bestilling fra Kunnskapsdepartementet (KD) av 05.04.24. 'Behov for tungregnekraft for forskning og kunstig intelligens'. I bestillingen ber KD om svar på følgende spørsmål:

1. Omfang og sammensetning av tungregneressursene i dag

- Hva slags norske offentlig og privat tungregnekraft er i dag tilgjengelig i Norge til tradisjonelle tungregneoppgaver og kunstig intelligens?
- Hvilken tilgang har norske forsknings- og innovasjonsmiljøer, inkludert næringslivet, i dag til forskjellige typer tungregnekraft gjennom internasjonale forsknings- og innovasjonsprogrammer?
- Hvilken tilgang har forskningssektoren, offentlig sektor og næringslivet til forskjellige typer tungregnekraft på ulike graderingsnivåer?

2. Forventet utvikling i et femårsperspektiv

- Hvor mye og hvilken sammensetning av tungregnekraft vil Norge trenge, gitt ambisjonsnivået om å være på nivå med våre nordiske naboer og relevante EU-land?
- Hvordan skal man innrette seg for å få mest mulig utbytte av investeringen og deltakelsen i det internasjonale partnerskapet EuroHPC og internasjonale samarbeidsprogrammer som f.eks. Horisont Europa og DIGITAL?
- Hvor stor andel av de fremtidige tungregneressursene bør lokaliseres i Norge ut fra nasjonale behov for sikkerhet og beredskap?
- Hvordan kan man dele på tungregneressurser på ulike graderingsnivåer i et nasjonalt perspektiv?
- Hva er kostnadsanslagene forbundet med punktene over, fordelt på kostnader til investering og drift?

Analysen av fremtidige behov for tungregnekraft skal ta utgangspunkt i kjente behov i forskningssektoren, offentlig sektor og næringslivet, i tillegg til vurderinger knyttet til fremtidige ambisjoner. Det vil være naturlig at Norge har et ambisjonsnivå på linje med våre nordiske naboer og relevante EU-land.

Analysen skal ta høyde for behov knyttet til skjermingsverdig og gradert FoU på ulike nivåer.

Analysen skal vurdere i hvilken grad behov kan dekkes av Norges deltakelse i internasjonale satsinger, herunder EuroHPC³.

Avgrensing og gjennomføring

I oppdraget etterspørres beskrivelse av hvilken sammensetning av tungregnekraft som er tilgjengelig og som det vil være behov for i forskningssektoren, offentlig sektor og næringslivet.

Med sammensetning menes her sammensetning av tungregneressurser for henholdsvis tradisjonelle tungregneoppgaver (for eksempel numeriske beregninger og modelleringer) og kunstig intelligens (og maskinlæring). Hvilken sammensetning av tungregnekraft vi vil trenge for å dekke behovene bestemmes av hvilke på ulike egenskaper (formål) regnekraften skal dekke. Vi har litt forenklet valgt å beskrive tradisjonell eller generisk tungregnekraft som CPU-ressurser⁴, og tungregning optimalisert for KI som GPU-ressurser⁵. Vi har avgrenset beskrivelsen av tilgjengelig ressurser til regneanlegg over en viss størrelse⁶

³ Kort om EuroHPC

⁴ Central Processing Unit er hovedregne-/prosesseringsenheten i en datamaskin, også kjent som prosesser eller hovedprosessor.

⁵ Graphics Processing Unit er en spesialisert mikroprosessor som brukes til grafikkprosessering og høyttelsesberegninger.

⁶ Større og nyere enn (>V100) 16 GPUer eller 3000 CPUer.

Vi har vært i dialog med KD underveis i arbeidet med oppdraget for å avklare forventninger og avgrensninger.

Innenfor tidsrammen for oppdraget har vi ikke kunnet gjennomføre omfattende analyser av tilgang og behov i offentlig sektor og næringsliv. Detaljerte beskrivelser av tilgjengelig tungregneressurser og fremtidig behov er derfor avgrenset til å gjelde tungregnekraft for forskningsinstitusjonene og deler av offentlig forvaltning. Behovet for tungregning, utfordringer og mulighetene som ligger i utvikling og bruk av KI, vil variere innenfor ulike fagområdene, men innenfor tidsrammen har vi ikke hatt mulighet til å gjøre analyser av domenespesifikke behov og utfordringer.

I arbeidet med utredningen har vi hatt kontakt med relevante offentlige og private aktører innenfor og utenfor forskningssystemet.

For å kartlegge dagens omfang og sammensetning av tungregneressurser og fremtidige behov i forskningssektoren har vi hatt tett dialog med Sigma2 og vi har sendt ut en spørreundersøkelse til det statlige universitetene og 22 utvalgte institutter. 8 av 10 universiteter har respondert, og alle de 22 forespurte forskningsinstituttene har respondert. Se vedlegg 2 for respondenter og spørsmål i undersøkelsen.

Vi har i tillegg hatt oppfølgende dialog med enkelte institutter både i form av skriftlige oppfølgingsspørsmål, samt utdypende møter med Statistisk sentralbyrå (SSB), Folkehelseinstituttet (FHI), og Meteorologisk institutt (MET).

I spørreundersøkelsen har vi også stilt spørsmål om utvikling i prosjekter institusjonene har hatt for eller i samarbeid med andre offentlige aktører og næringslivet. Men innenfor den tidsrammen vi har hatt til rådighet har vi ikke kunnet gjennomføre omfattende kartlegginger og analyser av behov og ambisjonsnivå i større deler av offentlig sektor og næringslivet. Vi viser i denne sammenheng til rapportene "Kartlegging og vurdering av stordata i offentlig sektor"⁷ og "Kunstig intelligens i Norge – nytte, muligheter og barrierer"⁸ som utdyper behov og ambisjonsnivå for offentlig sektor og næringsliv.

For noe mer utfyllende informasjon om offentlig sektor har vi hatt dialog med helsedirektoratet, Digitaliseringsdirektoratet og Nasjonalbiblioteket (om utvikling av norsk språkmodell).

For informasjon om og fra næringslivet har vi hatt møter med Abelia, Telenor, IKT-Norge og DigiCat (Norsk Katapultsenter). Det ble i tillegg lagt inn spørsmål om bruk og fremtidig behov for tungregning i NHOs medlemsundersøkelse i mai. Se vedlegg 2 for respondenter og spørsmål i undersøkelsen.

For behov for graderte tungregneressurser har vi innhentet informasjon fra Forsvarets forskningsinstitutt (FFI) og Nasjonal sikkerhetsmyndighet (NSM).

For informasjon om Nordiske og europeiske aktører som det er relevant å sammenligne oss med har vi hatt skriftlig dialog med DeIC (Danish e-infrastructure Consortium), og Uddannelses- og Forskningsministeriet i Danmark, CSC (IT Center for Science) og Akademi of Finland i Finland, NAISS (National Academic Infrastructure for Super-computing), Vinnova (Sweden's innovation agency) og Vetenskapsrådet i Sverige, Federal Ministry of Education and Research i Tyskland, Ministry of Economic and Climate Affairs i Nederland, og Direktoratet for forskning og innovasjon i Frankrike (Dgri).

Tallgrunnlag for kostnadsestimater i et femårsperspektiv med utgangspunkt i ambisjonsnivået i våre nordiske naboland er hentet fra Sigma2, Nasjonalbibliotekets beregninger av kostnader for norsk språkmodell, de nordiske aktørene nevnt ovenfor, samt informasjon i søknader til Forskningsrådet.

Vi har i tillegg hentet informasjon fra åpne kilder som strategier, rapporter og notater, samt tidligere kunnskapsgrunnlag Forskningsrådet har utarbeidet.

Oppdraget omfatter ikke regnekraft for kvanteberegninger, men flere av aktørene vi har vært i kontakt med mener at i tillegg til en veldig utvikling i behov for regnekraft optimalisert for KI, er det stor sannsynlighet for at behovet for regnekraft for kvanteberegninger vil øke betydelig i løpet av de neste fem årene. Vi vil derfor gi en kort oppsummering av status per i dag, og anbefaler at behov for tungeregnekraft for kvanteberegning bør inkluderes i fremtidige revisjoner av opptrappingsplanen for investeringer i regnekraft.

⁷ [Kartlegging og vurdering av stordata i offentlig sektor \(regjeringen.no\)](https://regjeringen.no)

⁸ [Rapport 35-2023 \(nho.no\)](https://nho.no)

Dagens omfang, sammensetning og forventet utvikling i et femårsperspektiv

Vi vil i denne delen beskrive dagens omfang og sammensetning av nasjonale tungregneressurser, hvilke regneressurser forskningsinstitusjonene i dag benytter, samt beskrive forventet utvikling i behov for ulike tungregneressurser i et femårsperspektiv. Behovet er beskrevet på et overordnet nivå, da det innenfor oppdragets tidsramme ikke har vært mulig å gjøre en omfattende kartlegging av de ulike fagområdenes særskilte behov, utfordringer og muligheter.

Kapittelet er delt inn etter aktører og tema som er relevant for det nasjonale perspektivet. Dagens omfang og framtidig behov beskrives både for tradisjonell tungregnekraft, samt regnekraft optimalisert for KI.

Mange av aktørene, både i forskningssektoren, offentlig forvaltning og næringslivet, som behandler store datamengder vil kunne bruke lokale ressurser eller kjøpe kapasitet hos store kommersielle skytjenesteleverandører for noen typer beregninger. Dersom man skal behandle større datamengder, eller data som krever store parallelle beregninger, er imidlertid ikke dette tilstrekkelig. Da trenger man større regneressurser, som det er mer kostnadseffektivt å etablere på nasjonalt eller internasjonalt nivå. For enkelte forskningsområder er datamengdene så store og kravet til hastighet i beregningene så høye at de ressursene vi har nasjonalt ikke har tilstrekkelig kapasitet. Da er vi avhengig av samarbeid og av å kjøpe kapasitet i internasjonalt samarbeidende konsortier som EuroHPC. EuroHPC er nærmere beskrevet i kapittelet 'Kort om EuroHPC'.

Sigma2

Sigma2 er en nasjonal tilbyder av tungregnetjenester for forskning. Sigma2 er et norsk statlig selskap eid av Sikt – Kunnskapssektorens tjenesteleverandør⁹ og har ansvar for anskaffelser og drift av tungregneressurser og datalagring for norske forskningsmiljøer. Tjenesten leveres i samarbeid med de fire universitetene NTNU, UiO, UiB og UiT. Tungregnetjenesten gir tilgang til nasjonale databehandlingsanlegg og programvare som har langt større kapasitet enn forskningsinstitusjonene normalt har tilgjengelig. Gjennom Sigma2 har norske FoUol-miljøer tilgang til tre nasjonale superdatamaskinene (Betzy, Fram og Saga), i tillegg til den europeiske superdatamaskinen LUMI (som er en del av EuroHPC og i dag Europas kraftigste superdatamaskin). Norske muligheter deltakelse i LUMI og EuroHPC gir er beskrevet i kapittelet 'Kort om EuroHPC'.

Betzy er plassert ved NTNU i Trondheim og tilbyr hovedsakelig CPU-ressurser og noe GPU-ressurser. Fram er plassert ved UiT og tilbyr CPU-ressurser. Saga som er plassert ved NTNU tilbyr GPU- og KI-ressurser.

Den historiske økningen når det gjelder bruk av nasjonale tungregneressurser hos Sigma2 er på 12-13 prosent årlig. For tjenester knyttet til lagring, som er en essensiell del av e-infrastruktur som tilbyr regneressurser, er den historiske økningen i bruk ca. 15 prosent årlig. Sigma2 ser også en dreining mot bruk av, og økt etterspørsel etter, GPU-basert beregningsressurser der hvor prosjekter tidligere brukte mer tradisjonelle CPU-baserte tungregneressurser. Det anslås at kapasiteten i dag fordeler seg på 90 prosent CPU og 10 prosent GPU.

Sigma2 beskriver at utviklingen innen generativ KI for språkmodeller er interessant for å vise hvor raskt etterspørsel etter tungregnekapasitet kan endre seg. I november 2022 var det en moderat etterspørsel, til februar 2023 der hele den totale etterspørsel som Sigma2 kan tilby på LUMI ble tatt i bruk. Dette er med på å gi et godt grunnlag for å kvantifisere behov på kortere og lengre sikt. Det

⁹ [Sikt – Kunnskapssektorens tjenesteleverandør](#)

kan forventes at trening av språkmodeller vil øke i omfang, både i størrelse og frekvens (se også eget delkapittel om 'Tungregnekraft til språkmodeller').

Foreløpige estimater fra de ledende KI-miljøer i forskningssektoren i Norge er sammenfallende, og er i størrelsesorden 10-100 ganger kapasiteten som er tilgjengelig i dag. Det store spennet indikerer stor usikkerhet, 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 etter-trene eksisterende modeller.

Sigma2 trekker fram et annet moment som må tas med i vurderingen er den raske utviklingen innen multimodal generativ KI, der KI brukes sammen med språkmodeller til å genere ikke-tekstlige resultat, som f.eks. bilder og film. For å kunne trene modeller for dette trengs det kapasitet, både innen lagring og regnekraft, som er flere magnituder større enn dagens kapasitet. Parallelt med dette arbeides det med å gjøre KI mindre regneavhengig, slik at veksten i behov for regnekraft på sikt kan flate noe ut.

Forskningsinstitusjonene

Dagens regnkraftressurser og bruk

Åtte universiteter har svart på spørreundersøkelsen, hvor fem av disse har egne lokale tungregnerressurser. De fleste har ressurser som er tilgjengelig for interne brukere, mens under halvparten har ressurser som er tilgjengelig for eksterne brukere. For disse ressursene er det en stor overvekt som brukes til tradisjonell tungeregning (70-90 prosent).

Alle universitetene har tilgang til de nasjonale ressurser hos Sigma2, og selv om de største brukerne er fra de fires største universitetene, svarer alle respondenter at de i større eller mindre grad benytter Sigma2s tjenester. Flere av universitetene benytter felleseuropeiske ressurser som LUMI (EuroHPC) for både tradisjonelle beregningsoppgaver og KI og noen oppgir at de får tilgang til tjenester gjennom samarbeid med universiteter i andre land. Kun ett av universitetene oppgir at de tilbyr bruk av kommersielle skytjenester til sine brukere.

Av totalt 22 institutter som er forespurt og som har svart på undersøkelsen svarer 8 at de har egne tungregnerressurser. Flertallet av disse er kun for interne brukere. Over halvparten bruker Sigma2 sine ressurser. 6 av instituttene svarer at de bruker EuroHPC-ressurser og noen har tilgang til regnekraft gjennom annet nordisk eller europeisk samarbeid. Noen av instituttene benytter også kommersielle skytjenester, men gjerne i tillegg til Sigma2s tjenester. Instituttene opplyser at de i hovedsak benytter tungregnekraft til tradisjonelle tungregneoppgaver. 12 av instituttene bruker noe eller en liten andel til kunstig intelligens.

Et mindretall av universitetene rapporterer at de har en liten andel brukere fra næringslivet. Alle med interne ressurser rapporterer at de har brukere fra offentlig sektor, men at andelen offentlige eksterne brukere er liten, og at det i all hovedsak dreier seg om tradisjonell tungregnekraft. I arbeidet med utvikling av norske språkmodeller har Nasjonalbiblioteket hatt tilgang universitetenes ressurser for trening av modeller. I perioder har all kapasitet til NTNUs anlegg, IDUN, vært allokert til språkmodelltrening. Flertallet av instituttene opplyser at det har vært en økning i bruk av tungregnekraft i prosjekter for/med næringsliv og offentlig sektor. Prosjektene er større, datamengdene er større, og dette fører til behov for mer regnekraft og gode datalagringsløsninger.

Behov

Flere universiteter rapporterer at de i dag har begrensede ressurser både tradisjonelle beregningsoppgaver og beregninger som krever GPU-ressurser. Det er derfor et stort gap mellom hva som er reelt behov og hva forskningsgruppene har finansiering til å bruke.

Tradisjonelle beregningsoppgaver blir brukt av blant annet innen astrofysikk. Der vil nye sensorer, satellitter og observatorier produsere mange ganger så mye data om noen år som de gjør i dag, og det antas derfor at både behov for lagring og CPU, og etter hvert GPU, vil fortsette å øke markant ettersom instrumentene fornyes.

Universitetene støtter Sigma2s vurderinger av fremtidige behov, og anslår en gjennomsnittlig økning på minst 12-15 prosent i forhold til dagens kapasitet, mens andre anslår en tredobling i fremtidige behov i forhold til dagens kapasitet til tradisjonelle beregningsoppgaver.

Når det gjelder behov for tradisjonell tungregning versus behov for tungregnekraft til KI beskriver flere at behovet for tradisjonell tungregning har opprettholdt et stabilt nivå, mens andre beskriver en økning på dette området. Tungregnekraft for KI har vært svært marginalt, men flere ser en vridning mot KI de siste årene, og enkelte beskriver en stor økning ilt. de siste 3 årene, hvor de har gått fra et begrenset GPU behov til et veldig stort behov som er vanskelig å dekke med dagens løsninger.

Selv om KI er en enorm pådriver for teknologiutviklingen, vil det alltid være en svært stor komponent av tradisjonell tungregning - men i stadig økende grad tilpasset nye maskinarkitekturer. Det forventes at behovet for tungregning flytter seg fra systemisk enklere prosesser (enkeltprosesser), til arbeidsflytprogrammer som er knyttet til å løse enten samfunnsutfordringer eller sammensatte problemstillinger i industrien som kobler ulike prosesser sammen.

Flertallet av instituttene beskriver at behovet for tungregnekraft til tradisjonelle tungregneoppgaver vil øke betydelig i den kommende perioden, med estimater som varierer fra en økning på 10 til 100 prosent de neste årene. Dette inkluderer en økning i bruk av både for å forbedre beregninger og håndtere større datamengder i forskjellige sektorer som helseforskning, trafikksimulering og klimaprojekter. Økningen i behov gjelder både CPUer, men også økende behov for bruk av GPUer for å forbedre visse aspekter av beregningene gjennom GPU-akselerasjon.

Instituttene beskriver også stor usikkerhet i sine anslag om fremtidig tungregnekraft generert av KI-behov. Noen institutter ser for seg en eksponentiell økning i behovet for spesialiserte KI-maskiner eller GPUer, og flere anslår en vekst i behovet opp mot en tidobling av dagens behov. Andre peker på det er behovet for kunnskap og kompetanse innen KI som er det mest prekære.

Flere av universitetene peker på at ved etablering av seks strategiske sentere for forskning på KI («KI-milliarden»), vil Norge få et godt fundament innen KI-forskning. Økt kapasitet for tungeregning er en forutsetning for å sikre god effekt av denne investeringen.

Tilgang til tilstrekkelig KI-tilpasset regnekraft vil avgjøre om norske forskere, forvaltning og næringsliv vil kunne bygge nye KI-modeller fra bunnen av basert på norske data, eller bare etter-trene eksisterende modeller. Behovet for regnekraft optimalisert for KI gjelder ikke bare for språkmodeller, men også innen andre anvendelser, som for eksempel billedbehandling og helsedata. Innen helse er det viktig at Norge legger til rette for utnyttelse i Norges forsprang i helsedata og helseregister også innen kunstig intelligens. Dette er særlig relevant for persontilpasset medisin.

Et viktig poeng er at KI/GPU vil komme i tillegg til og ikke i stedet for klassisk tungregning de neste 3- 5 årene. Universitetene anslår behovet for KI-relaterte beregningsressurser til å være kraftig økende fremover, men flere påpeker at det er vanskelig å gi eksakte tall i et KI-landskap raskt i endring, og med tanke på at det ikke er historiske tall som kan brukes til å beregne fremtidige behov. Flere av universitetene påpeker at KI-behovet ikke bare gjelder kapasitet knyttet til KI-maskinvare men vel så viktig er det å ha oppsett og arbeidsflyter som er tilpasset brukerbehov, samt programvare og brukerstøtte for KI-beregninger. I dette ligger også utdanning og forskning for trygg, pålitelig og rettferdig KI og en KI-kompetent arbeidsstyrke.

Også Teknologirådet¹⁰ anbefaler en kraftig oppskalering av norsk regnekraft for å dekke fremtidige behov som skapes av generativ KI.

Gå til oppsummering av Tilgjengelige nasjonale ressurser og behov

¹⁰ [Generativ-KI-i-Norge digital.pdf \(wpd.digital\)](#)

Samfunnsnyttige tjenester basert på tungregning

Flere institutter og forskningsmiljøer utvikler og leverer tjenester av samfunnskritisk betydning som krever bruk av tungregneressurser, som ulike helse-, -varslings og beredskapstjenester.

Økende tilfang av data, inkl. sanntidsdata, kombinert med ny teknologi som KI, gir muligheter for utvikling av flere og bedre tjenester basert på tungregnekraft.

Det trengs betydelig tungregnekapasitet både for å utvikle modeller og tjenester, men også til drift og vedlikehold av disse.

I sin rapport om Teknologirådet anbefaler i sin rapport om «Generativ Kunstig Intelligens i Norge» at drift, forvaltning og tilgang til utvalgte norske språkmodeller bør sikres ved å definere disse som en nasjonal fellestjeneste. Det vil kreve betydelig regnekraft.

Under Korona-pandemien benyttet FHI ressurser i Norge primært ved UiO og NTNU og Sigma2. Det var imidlertid krevende å få gjennomført ønskede analyser på grunn av kapasitetsbegrensninger. Det er behov for en langt større kapasitet før neste store helsekrise. FHI opplyser at de trenger miljøer som kan håndtere personsensitive helsedata med garantert kontinuerlig opptid under helsekriser.

Helsedirektoratet jobber med ulike tiltak for trygg og effektiv bruk av KI i helse og omsorgstjenesten. På oppdrag fra Helse- og omsorgsdepartementet leder Helsedirektoratet nå derfor arbeidet med en felles plan for trygg og sikker innføring av KI i helse- og omsorgssektoren¹¹. Som en del av dette arbeidet inngår vurderinger behovet for regnekraft og hva som bør gjøres nasjonalt og gjennom europeisk samarbeid.

Havforskningsinstitutt (HI) har ansvar for *Trafikklyssystemet for bærekraftig akvakultur*¹² og mer enn 90 prosent av beregningsoppgavene som HI utfører er knyttet til prosjekter under dette oppdraget finansiert av Nærings- og fiskeridepartementet (NFD). Omfanget er omtrent doblet de siste 3-5 årene. Til dette anvendes tjenestene fra Sigma2.

Norges Geotekniske Institutt (NGI) beskriver at det er et uutnyttet potensial for tungregning som kan gi bedre tjenester for eksempel knyttet til håndtering av geofarer der beregninger må gjennomføres raskt, enten når en hendelse oppstår, eller er varslet. De er ikke kjent med tungregneressurser i Norge som kan benyttes til å beregne akutsituasjoner knyttet til f.eks. geofarer, eller allmenn tilgjengelige tungregnesystemer internasjonalt som kan operere opp mot akutsituasjoner. Dette er både på grunn av at et slikt system har behov for integrasjon av data som kobles til infrastruktur sanntid, og på grunn av ressurser for håndtering av slike situasjoner krever at det alltid er tilgjengelig kapasitet, med andre ord at ressursen er prioritert for dette formålet.

NORSAR beskriver at instituttsektoren sitter på store mengder data fra mange ulike sensorer som f.eks fiber optiske kabler. Dette er verdifulle data som ved hjelp av tungregning kan brukes til mange samfunnsnyttige formål som overvåking av klima-endringer, samfunnsberedskap, og innen forsvar.

¹¹ Felles KI-plan for trygg og effektiv bruk av KI i helse og omsorgstjenesten 2024-2025

¹² <https://www.hi.no/hi/temasider/akvakultur/trafikklyssystemet-hi-sin-kunnskap>

Meteorologisk institutt (MET)

MET er en særlig interessant virksomhet i denne sammenheng, da de er en av de største brukerne av tungregning i offentlig sektor. MET bruker tungregning både til forskning og forvaltning, og tungregning er nødvendig for METs samfunnskritiske tjenester.

Drift og forskning/videreutvikling av operasjonell vær- og havvarsling er en del av METs statsoppdrag og gjøres med finansiering dels fra bevilgning over statsbudsjettet og dels med finansiering fra eksterne prosjekter (hovedsakelig bidrag). Klimamodellering gjøres i hovedsak med finansiering fra eksterne prosjekter (bidrag), men er en del av samfunnsoppdraget.

MET har utstrakt europeisk samarbeid med andre meteorologiske institutt, og har et eget tungregneanlegg i samarbeid med det svenske meteorologiske og hydrologiske institutt (SMHI). Anlegget står i Linköping i Sverige. Det er tett faglig samarbeid med Sverige, Finland, Latvia og Estland, og de kjører felles operasjonelle værvarslingsmodeller fordelt på anlegg i Linköping, Norrköping og Finlands eget anlegg.

For forskning, utvikling og produksjon som ikke er tidskritisk, benytter MET både egne tungregneanlegg i Norge og Sverige, nasjonale kvoter på tungregneanlegget til det europeiske senteret ECMWF, og tildelinger på anleggene til Sigma2 og EuroHPC. Klimamodellering med den norske jordsystemmodellen NorESM gjøres i stor grad på anleggene til Sigma2 og EuroHPC, og NIRD benyttes for lagring og deling av data fra NorESM.

Fram til nå har MET budsjettert 11 millioner kroner per år til investering og drift av tungregneanlegget i Linköping. Det er planlagt å investere 10 millioner kroner utover dette i 2025. Økningen har bakgrunn i nye investeringer i anlegget i Linköping, samt økte kostnader til drift.

MET beskriver at de er avhengig av nær 100 prosent oppetid for operasjonelle kjøring. Kapasiteten for dette formål er kritisk for norsk sikkerhet og beredskap og de vil ikke kunne levere sine tjenester med lengre perioder med nedetid. I tillegg til gjenbruk av algoritmer, er krav til oppetid en sentral årsak til at MET ser det som mest hensiktsmessig å samarbeide med andre meteorologiske institutt om dedikerte regneanlegg. Til sammenligning har ikke forskningsmiljøene på universitetene samme krav til oppetid.

MET legger også vekt på at det er svært nyttig med samarbeid om kunstig intelligens (KI) og maskinlæring også med forskere utenfor MET. For å få til gode synergier er det hensiktsmessig med et felles rammeverk for maskinlæring og tilhørende metodeutvikling. Det krever dedikert innsats og investeringer fra forskerne for å kunne effektivt bruke dataene. Til gjengjeld er dataene gode, og det er mulig å få svært gode synergieffekter gjennom et felles rammeverk.

MET beskriver at behovet for tungregnekraft er stadig økende. Tidligere ble det økte behovet i stor grad dekket inn ved at tungregneressursene ble stadig kraftigere mens prisen gikk ned. Denne utviklingen har stoppet, og både operasjonell produksjon og videreutvikling begrenses av tilgangen på tungregnekraft. Det er spesielt krevende å skaffe tilstrekkelig tungregnekraft til klimamodellering, da de eksterne prosjektene ikke har økonomi til å finansiere behovet.

Historisk sett har bruk av METs interne regneressurser økt med rundt 30 prosent per år de siste årene. MET vil framover fortsatt ha behov for tungregnekraft til tradisjonelle (fysikkbaserte) beregningsmodeller. De datadrevne modellene krever langt mindre ressurser enn fysikkbaserte modeller i daglig produksjon, men treningen krever stor tilgang til GPUer. MET beskriver derfor at i forbindelse med utvikling og bruk av KI i forskning og forvaltning, vil de få et økt behov for GPUer, spesielt til trening av datadrevne modeller. Forskning og utvikling av fysikkbaserte modeller som kan utnytte GPUer i beregningene, kan også føre til økt behov for GPUer i daglig produksjon. MET vil også ha stor nytte av en norsk språkmodell, som kan erstatte mye manuelt arbeid og gi effektiviseringsgevinster.

Tungregnekraft til språkmodeller

Teknologirådets rapport «Generativ kunstig intelligens i Norge»¹³ beskriver at alle som vil utvikle, tilpasse og drifte store språkmodeller eller som vil bygge nye generative KI-tjenester på toppen av disse, trenger tilgang på regnekraft.

I dag foregår trening av grunnmodeller i hovedsak på IDUN, en plattform for tungregning ved NTNU, og LUMI. I forhold til dagens krav til regnekraft for trening av store språkmodeller har IDUN begrensede ressurser. I tillegg har Nasjonalbiblioteket benyttet tjenester fra Google's tilbud til forskere (Google TPU Research Cloud). Men tilgangen er uforutsigbar både i tid, kapasitet og stabilitet/tilgjengelighet.

Praksis viser at det er vanskelig å beregne behov og krav til regnekraft for å trene språkmodeller, og enda vanskeligere for bruk av språkmodellene (inferens). I Nasjonalbibliotekets beregninger er det tatt utgangspunkt i at det i en nasjonal kontekst er ønskelig å se behov for regnekraft både for forskning og utvikling i sammenheng. Behovet for investeringer er avhengig av variabler som variasjon i størrelse på modellene, der valg av størrelse avhenger gjerne av sammenhengen der språkmodellene skal brukes, antall treninger av modellene, og antall GPU-timer det er behov for i treningen. Dette vil igjen variere avhengig av maskinvare, programvare og treningsdata. Den siste variabelen er pris per GPU-time, som vil være avhengig av om det bygges egne anlegg i Norge, gjennom EuroHPC-samarbeidet, eller om man benytter kommersielle tilbydere.

Helsedirektoratet beskriver også tilstrekkelig regnekraft som en viktig risiko for bruk av språkmodeller i helsevesenet.¹⁴

Gå til oppsummering av funn og anbefalinger for Samfunnsnyttige tjenester

Kvanteteknologi

En kartlegging Forskningsrådet gjorde på oppdrag fra KD i 2022 (Gjennomgang av norsk forskning på kvanteteknologi¹⁵ av 14.10.22) konkluderte vi med er voksende kvanteteknologisk miljø, til stede på internasjonale arenaer, med aktivitet særlig innen kvanteberegning, - simulering og sansing. Den samlede norske forskningen på kvanteteknologi framstår i de senere årene som mer målrettet. Nå dannes det norske nettverk, samt det er utviklet nye studietilbud på master- og doktorgradsnivå. De norske miljøene jobber aktivt med vekst og deltagelse i internasjonale samarbeid, men har ikke kvanteteknologisk infrastruktur eller finansiering på linje med forskningsmiljøer i Sverige, Danmark og Finland.

Forskningsrådet har ingen utlysninger spesifikt rettet mot kvanteteknologi, og Norge mangler også langsiktige industrielle investorer på feltet, som bidrar sterkt til finansiering i våre naboland.

I motsetning til våre naboland, mangler altså Norge en spesifikk satsing på kvanteteknologier. Miljøene som jobber med kvanteberegning har tilgang til tungregningsklynger for å simulere kvanteberegning, men tilgangen til faktiske kvantemaskiner er i stor grad begrenset til skyressurser fra kommersielle tilbydere som IBM og D-Wave. Effektiv tilgang til store ressurser fra kommersielle tilbydere er i dag dyrere enn det som realistisk kan finansieres innenfor et forskningsprosjektbudsjett.

¹³ [Generativ kunstig intelligens i Norge - Teknologirådet \(teknologiradet.no\)](https://www.teknologiradet.no/Generativ-kunstig-intelligens-i-Norge)

¹⁴ [Utfordringer og risikoer - Helsedirektoratet](#)

¹⁵ Definisjonen for kvanteteknologi er hentet fra EU-kommisjonens strategiske forskningsagenda for kvanteteknologi. Denne handler om teknologier som kan utnytte og kontrollere kvanteeffekter og manipulere enkelte kvanteobjekter som fotoner eller elektroner for utvikling av enheter og systemer. Definisjonen er strukturert rundt fire hovedområder for anvendt kvanteteknologi: kvanteberegning, -kommunikasjon, -simulering, og - sensorer/metrologi.

Fagmiljøene etterspør en nasjonal strategi og satsing på kvanteteknologi. Særlig med tanke på eksperimentell aktivitet har flere miljøer ønske om styrket tilgang på kvantedatamaskiner.

Våren 2024 holdt EuroHPC en anbudsprosess for den nye kvantemaskinen, LUMI-Q, som skal stå i Tsjekkia. Norge via Sigma2 er partner i konsortiet sammen med SINTEF og NORCE. Sigma2 har til nå gått inn med 2 millioner kroner i konsortiet.

Gå til oppsummering av funn og anbefalinger for Kvanteteknologi

Næringsliv

Næringslivet er opptatt av mulighetene for å ta KI i bruk og å bygge KI inn som funksjonalitet i sine produkter og tjenester. Det er en allmenn oppfatning at KI representerer både en mulighet for å vinne nye kunder, men også risiko for å tape konkurransen dersom de ikke lykkes med å ta i bruk KI på riktig måte. Kompetanse og erfaring med superdatamaskiner en forutsetning for å bygge opp og skalere en konkurransedyktig norsk KI-industri.

Dette samsvarer med de funn som beskrives i en rapport¹⁶ utarbeidet i 2023 av Samfunnsøkonomisk analyse på oppdrag fra NHO som omhandler nytten av kunstig intelligens, med særlig vekt på verdiskapingen som kan oppnås dersom virksomheter i offentlig og privat sektor nyttiggjør seg mulighetene. Rapporten konkluderer med at de viktigste utfordringene for å ta i bruk KI-verktøy er knyttet til manglende kompetanse og innsikt på hvordan KI kan nyttiggjøres i deres virksomhet. Noen av virksomhetene intervjuet av Samfunnsøkonomisk analyse, uttrykker et ønske om å ansette kandidater med doktorgrad innen fagfelt som kunstig intelligens, fysikk og matematikk. Det er også noe bekymring knyttet til etikk og lovverk som eierskap til data og personvern. For virksomheter som har kommet i gang med KI, er tilgang til og kvaliteten på data også en utfordring som løftes frem.

Vårt inntrykk, basert på informasjon fra de aktører vi har vært i kontakt med, er at disse utfordringene også gjelder for deler av offentlig sektor.

Gjennom et samarbeid med Abelia har de spurt deres medlemmer om bruk av tungregning i dag og i fremtiden. Hovedkonklusjonen er at mange bedrifter er usikre på behovene fremover, men at de som allerede i dag benytter seg av tungregning mener at behovet for kapasitet vil øke kraftig i årene fremover. Særlig IKT-virksomheter og større virksomheter svarer at de i dag benytter tungregning. Selv om det er få som rapporterer at de bruker tungregning i dag, svarer over ni av ti at behovet for tungregnekapasitet vil øke i virksomheten de neste fem årene. Over halvparten tror behovet vil øke kraftig.

Samtaler med enkeltbedrifter og organisasjoner som IKT Norge, Digital Norway og andre tyder på at næringslivet i dag har gode tilbud fra de kommersielle tilbyderne av regnekapasitet og det er bare i spesielle tilfeller de ser behov for et eget offentlig tilbud. Tilbakemeldingene vi har fått er at de kommersielle tilbyderne har et bedre tjenestetilbud og god brukervennlighet for de mest vanlige behovene. Enkelte bedrifter har investert i egne tungregneressurser for å dekke særlige behov. Flere legger vekt på at det særlig for start-ups er et behov for spisskompetanse og tilgang på arbeidskraft og infrastruktur, og det vil være nødvendig med tettere koblinger til forskningssystemet.

De bedriftene som benytter seg av Sigma2 og EuroHPC springer ofte ut av forskningsmiljøer eller er forskningsintensive. Det er et mål i næringspolitikken å utvikle flere slike «deep-tech» bedrifter. Disse vil i større grad være avhengig av de forskningsnære infrastrukturene, og dersom det satses mer på slik infrastruktur for norske miljøer vil det øke sannsynligheten for at det etableres flere slike virksomheter. Dette forutsetter ikke bare fysisk utstyr, men også tjenestetilbud og brukervennlighet som bidrar til konkurranseevne og veksthastighet.

Flere understreker betydningen av at forsknings- og utdanningsinstitusjonene har god tilgang til tungregnekraft og store superdatamaskiner for å utvikle kompetanse og ekspertise som både næringsliv og offentlig sektor trenger og at Norge er godt koplet på det som skjer på europeisk nivå, både når det gjelder tilgang til superdatamaskiner, utvikling av teknologi, lover og reguleringer.

Gå til oppsummering av funn og anbefalinger for Næringsliv

¹⁶ Rapport 35-2023 Kunstig intelligens i Norge - nytte muligheter og barrierer (nho.no)

Tilgang til tungregnerressurser på ulike graderingsnivåer

I regjeringens Langtidsplan for forsvarssektoren 2025-2036¹⁷ beskrives at regjeringen vil utnytte potensialet i de store datamengdene som produseres til å videreutvikle forsvarssektoren og utvikle forsvarsevnen. Forsvaret skal utnytte potensialet i data for operative formål, men er i for liten grad datadrevet i dag. Regjeringen vil satse på forskning og utvikling innenfor banebrytende teknologier som kvanteteknologi og kunstig intelligens for å bidra til å bygge nasjonal kompetanse og kapasitet. Denne satsingen på et datadrevet forsvar og banebrytende teknologier vil også medføre et økt behov for tungregnekapasitet, både i form av graderte systemer, men også for ugraderte systemer.

Informasjon om tilgang til og behov for forskjellige typer tungregnekraft på ulike graderingsnivåer er hentet inn fra Forsvarets forskningsinstitutt (FFI) og Nasjonal Sikkerhetsmyndighet (NSM). Vi tar utgangspunkt i definisjonene av ulike graderingsnivåer¹⁸ iht. rapporten «Et felles innspill til hvordan forskningssystemet må innrettes for også å håndtere skjermingsverdig og gradert forsknings- og teknologi-samarbeid»¹⁹.

FFI benytter egne driftede systemer til å dekke dagens ugraderte behov. På grunn av endringer i organisering av både IT-organisasjon, og tjenestene som leveres internt av egen IT-organisasjon, jobber FFI med å vurdere hvordan slike tjenester kan skaffes fremover. Tidligere har det vært begrensninger blant annet til flytting av store datamengder, og utfordringer knyttet til prioritet på kjøretid. Ambisjonen er å bruke fellesløsninger så langt det lar seg gjøre for ugradert informasjon.

Ugradert tungregning ved FFI benyttes blant annet til å gjøre avanserte beregninger og analyser, og inngår som en naturlig del av FFIs satsning innen et Datadrevet forsvar. Beregningene blir brukt blant annet til å vurdere ytelse og trusler. I de tilfellene hvor fenomenet eller geometrien på det som skal vurderes er graderte, eller simuleringsresultatene er graderte, er det helt nødvendig med graderte tungregningsressurser. Dette kan FFI ikke beskrive ugradert i mer detalj.

NSM vurderer at behovet for tungregnekraft innenfor forsvarssektoren for en stor del kan dekkes gjennom ugradert tungregning, slik at fellesløsninger i stor grad kan benyttes. Dette gjøres ved å simulere modeller med åpne data, hvor resultatene tas inn i graderte systemer ved behov. Dette vil være kostnadsbesparende sammenlignet med å utføre hele simuleringen på graderte tungregnerressurser. Det vurderes at en stor andel av beregningene som FFI og NSM har behov for kan gjøres i ugraderte systemer. Stabil og sikker tilgang til åpne tungregnerressurser er en forutsetning for å bruke fellesløsninger for å dekke behovene FFI og NSM har.

For å kunne benytte åpen infrastruktur så langt som mulig er det en forutsetning at det er høyt kunnskapsnivå på å vurdere nødvendigheten av bruk av graderte vs. åpen infrastruktur for tungregning. Vi viser i denne sammenheng til anbefaling om at det etableres en systematikk for å identifisere flerbruksmuligheter (dual-use) og skjermet forskning.²⁰

Vi har ikke kjennskap til at det utføres gradert tungregning i næringslivet. Vi er heller ikke kjent med at det utføres gradert tungregning i offentlig sektor utover forsvarssektoren. Når det gjelder skjermet informasjon er det behov for systemer som ivaretar dette, for eksempel i helsesektoren. Skjerming av informasjon og FoU er faglig krevende og dyrt. I hovedsak er det som driver kostnadene opp at gradert arbeid på nivå BEGRENSET krever autorisering, og på nivå

¹⁷ Kapittel 10 i [Prop 87 S \(2023-2024\) \(regjeringen.no\)](#)

¹⁸ Skjermet forskning eller informasjon er forskning eller informasjon som vil kunne skade offentlige interesser, en bedrift, en institusjon eller en enkeltperson om forskningen/informasjonen blir kjent for uvedkommende, eller er skjermet av andre hensyn, eksempelvis på grunn av eksportforbud. Det er også informasjon med skadepotensiale for nasjonale sikkerhetsinteresser knyttet til tilgjengelighets- og/eller integritetsbehov, men hvor det ikke er behov for å beskytte informasjonens konfidensialitet i henhold til sikkerhetsloven. Gradert forskning eller informasjon er forskning eller informasjon som kan skade nasjonale sikkerhetsinteresser om den blir kjent for uvedkommende. For å gradere informasjonen benyttes det fire ulike sikkerhetsgraderinger: begrenset, konfidensielt, hemmelig og strengt hemmelig.

¹⁹ [Sluttrapport - Helhetlig forskningssystem \(forskningsradet.no\)](#)

²⁰ [Sluttrapport - Helhetlig forskningssystem \(forskningsradet.no\)](#)

KONFIDENSIELT eller høyere kreves både sikkerhetsklarering og autorisering. Videre kreves etablering av en sikkerhetsorganisasjon, opplæring av autorisert personell i håndtering av gradert materiale og en endret sikkerhetskultur fra åpenhet og publisering til sikkerhetsmessig forsvarlig håndtering av informasjon og utstyr, i tillegg til digital og fysisk sikker infrastruktur²¹. Vi viser i denne sammenheng til at som en tommelfingerregel kan si at man må beregne et påslag fra ugradert digital infrastruktur på minst 20 prosent for skjermingsnivå BEGRENSET, minst 40 prosent for skjermingsnivå KONFIDENSIELT og minst 60 prosent på skjermingsnivå HEMMELIG for digital infrastruktur.²² Det er med andre ord store økonomiske konsekvenser knyttet til hvilket skjermingsnivå infrastrukturen tilpasses.

Gå til funn og anbefalinger for Tungregnerressurser på ulike graderingsnivåer

Behov for tungregnerressurser lokalisert i Norge

Selv om det er ulike oppfatninger i ulike miljøene hvor stor andel det er hensiktsmessig å ha nasjonalt vs. internasjonalt, er det enighet om at det bør være tilgjengelige norske ressurer både lokalt, nasjonalt og internasjonalt for å ivareta ulike behov. Alle land vi har hentet informasjon om investerer betydelig i nasjonale tungregnerressurser.

For god utnyttelse av kapasitet og kompetanse bør det være en god kopling både mellom lokale og nasjonale ressurer og mellom de nasjonale og de europeiske ressursene vi investerer i. I tråd med anbefalingene i rapporten "Forslag til rammeverk for rask oppskalering av analysekapasitet og -infrastruktur i kriser"²³ bør norske regnerressurser samordnes bedre for å kunne utnytte kapasitet og kompetanse i kriser.

De forespurte aktørene oppgir to hovedgrunner til at tungregnerressurser bør være plassert i Norge: 1) sikkerhet og beredskap og 2) bygge kompetanse og ekspertise.

Sikkerhet og beredskap

Dersom kriser skulle oppstå er det risiko for at Norge sin tilgang til ressurser i utlandet blir nedprioritert. Sårbarheten vil øke hvis all eller en veldig stor andel lokaliseres i andre land. Samtidig er det flere krisescenarier som tilsier at det er bedre at ressursene eller hvert fall noen av dem ligger utenfor landets grenser da dette også er med på å spre risiko. Men det er viktig ressursene ligger i land vi har gode relasjoner med, spesielt for tungregnerressurser for samfunnskritiske tjenester. Det blir pekt på av flere at sikkerheten vil i mange tilfeller være godt nok ivaretatt så lenge tungregnerressursene er lokalisert i et NATO-land.

Når det gjelder personsensitive data (særlig helsedata) kan det være prinsipper knyttet til personvern. Flere melder imidlertid at det reelt sett ikke er personvernetrusler knyttet til bruk av maskiner utenfor Norge så lenge ressursene er underlagt GDPR. For vurdering av særlig behov for håndtering av helsedata viser vi til Helsedirektoratets arbeid.

FFI vurderer at det ikke nødvendigvis er avgjørende at lokalisering i Norge, men at det viktigste er at det eksisterer en tilstrekkelig grad av kontroll med systemene. I fremtiden kan flernasjonalt samarbeid om tungregnerressurser mellom nære allierte som har et etablert sikkerhetssamarbeid kanskje bli mulig, også på gradert nivå. Vi må uansett forholde oss til de til enhver tid gjeldende lover og forskrifter, og de krav de til enhver tid måtte stille til nasjonal kontroll over og lokalisering av graderte eller samfunnsviktige IT-systemer.

Hvor stor denne andelen av de fremtidige tungregnerressursene som bør lokaliseres i Norge ut fra nasjonale behov for sikkerhet og beredskap forutsetter en grundig vurdering basert på risiko,

²¹ [Sluttrapport - Helhetlig forskningssystem \(forskningssystemet.no\)](https://forskningssystemet.no/sluttrapport)

²² [Vedlegg E i Sluttrapport - Helhetlig forskningssystem](#)

²³ [Oppskalering-av-analysekapasitet-i-kriser--rapport.pdf \(forskningssystemet.no\)](#)

kostnader, mv. Vi har ikke innenfor rammen av dette oppdraget mulighet til å gå i dybden på disse vurderingene.

NSM sitt arbeid med nasjonal skytjeneste

På oppdrag fra Justis- og beredskapsdepartementet (JD) gjennomførte Nasjonal sikkerhetsmyndighet (NSM) en konseptvalgutredning (KVU) for en nasjonal skytjeneste for ugradert, skjermingsverdig informasjon og andre beskyttelsesverdig data. Utredningen var rettet mot statsforvaltningens behov og rapporten²⁴ ble overlevert i januar 2023.

Bakgrunn for oppdraget er at Staten forvalter betydelige mengder ugraderte data og systemer som krever særskilt beskyttelse. Dette er data og systemer som er avgjørende for at staten skal kunne levere nødvendige tjenester av tilstrekkelig kvalitet i hele krisespennet (fred, krise og væpnet konflikt).

NSM mener at Norge trenger en nasjonal skytjeneste fordi det er sårbart å være avhengig av utenlandske skytjenester. Staten forvalter store mengder data og systemer som må beskyttes. Dette kan være personopplysninger, helsedata, finansielle data, ulike registre og informasjon om infrastruktur og transportsystemer i Norge. Manglende nasjonal kontroll over slike data og systemer kan svekke statens evne til å levere sentrale offentlige tjenester, særlig i krise eller væpnet konflikt.

NSM anbefaler at de fleste, om ikke nødvendigvis alle, nodene som inngår i skytjenesten plasseres i Norge. Dette fordi fravær av eller manglende nasjonal kontroll over statens data og IKT-systemer kan medføre betydelige konsekvenser. NSM mener at det er fullt mulig å se for seg at enkelte av nodene i en slik løsning kan være spesielt utformet for å møte krevende og spesialiserte behov som for eksempel tungregning eller KI.

Bygge kompetanse og ekspertise

Et annet viktig argument for å ha tungregnerressurser tilgjengelig i Norge, er for å bygge kompetanse og ekspertise nasjonalt. Alle peker på at det er essensielt at norske miljøer har lokale og nasjonale ressurser med enkel tilgang som treningsarenaer. Nasjonale fasiliteter og støttetjenester er viktig både for forskningsmiljøene, offentlig sektor og næringsliv, samt kobling mellom disse.

For å opprettholde nasjonal spisskompetanse på teknologi og metoder innen tungregning (inkludert KI-relaterte) vil det være nødvendig med nasjonal infrastruktur som muliggjør eksperimentell tungregning/KI forskning. Utviklingen av nasjonal ekspertise og teknologi styrker kunnskapsbasert økonomi og bidrar til teknologisk suverenitet.

Plassering av tungregnerressurser må også ses i lys av klimahensyn og energieffektivitet. Fordi Norge har tilgang på grønn energi og har kompetanse løsninger for gjenbruk av overskuddsvarme fra dataanlegg, kan plassering av anlegg nasjonalt være klimavennlig sammenlignet med plassering andre steder.

Gå til funn og anbefalinger for Tungregnerressurser lokalisert i Norge

Funn og anbefalinger

Tilgjengelige nasjonale ressurser og behov

- Sigma 2 er den nasjonale tilbyderen av tungregnetjenester for forskning. De fire største universitetene er de viktigste brukerne, men svært mange av de forspurte forskningsinstitusjonene benytter seg av Sigma2s tjenester. Også offentlig sektor og private aktører kan få tilgang, men kapasiteten er begrenset.

²⁴ [Konseptvalgutredning for nasjonal skytjeneste - Nasjonal sikkerhetsmyndighet \(nsm.no\)](https://www.nsm.no/konseptvalgutredning-for-nasjonal-skytjeneste)

- Norske aktører har dedikert tilgang til nasjonal kvote på LUMI-maskinen i Finland i tillegg til muligheter for å søke tilgang til EuroHPC-anlegg basert på kvalitet i prosjektet.
- Noen institusjoner har noe egne ressurser, mens andre er avhengige av ressurser gjennom Sigma2, LUMI, EuroHPC eller andre samarbeidspartnere. De institusjonene som har mest egne ressurser, er også de største brukerne av Sigma2, LUMI og EuroHPC (og/eller andre internasjonale ressurser).
- Private norske leverandører av tungregnekraft benyttes ikke/i svært liten grad av forskningsmiljøene, bortsett fra i prosjektsamarbeid med norske industriaktører som har tilgang til egne regneressurser. Kun noen få institusjoner svarer at de utelukkende benytter kommersielle skytjenester, og da de internasjonale.
- Nåværende kapasitet innen tungregnekraft er ikke dekkende for dagens kjente behov. Å holde investeringene på samme nivå som i dag, tilsier at gapet mellom behov og kapasitet vil fortsette å øke.
- Det er økende behov fremover både for tradisjonell tungregning og for KI. Utviklingen dreier mot mer bruk av GPU (ikke bare til KI). 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. Det er knyttet størst usikkerhet knyttet til utviklingen av GPU-behov.

Samfunnsnyttige tjenester

- Flere samfunnsnyttige og samfunnskritiske tjenester utvikles ved bruk av tungregning og samarbeid med forskningsmiljøer.
- Samfunnskritiske tjenester har behov for regneressurser med rask og forutsigbar tilgang og lite nedetid.
- Det må allokeres midler til regnekraft og drift for opprettholdelse og utvikling av samfunnsnyttige/samfunnskritiske tjenester. Dette inkluderer trening og operativ drift av modeller for andre brukere enn forskning, slik som værvarslingstjenester og språkmodeller.
- Nasjonale regneressurser bør samordnes bedre for å kunne utnytte kapasitet og kompetanse i kriser.

Kvanteteknologi

- Vi anbefaler at det utvikles en egen nasjonal strategi for kvanteteknologi.

Næringsliv

- Det er et økende behov for tungregnekraft også i næringslivet, spesielt knyttet til økt bruk av KI.
- De deler av næringslivet som bruker tungregneressurser i dag, bruker i hovedsak egne ressurser eller kommersielle tilbydere. Dette gjelder etter det vi kjenner til også for brukere av regneressurser i deler av offentlig sektor. Forskningsmiljøenes ressurser brukes i noen tilfeller der det er samarbeid.
- De viktigste utfordringene for å ta i bruk av data, KI-verktøy, inkludert tungregning i næringslivet er:
 - manglende kompetanse og innsikt,
 - bekymringer rundt etikk og lovverk (rettigheter til bruk av data og personvern),
 - tilgang til data av god kvalitet.

Vi har stor grunn til å tro at dette også gjelder store deler av offentlig sektor.

- I næringslivet, spesielt SMBer og startup-bedrifter, er det behov for spisskompetanse, opplæring og brukerstøtte for bruk av data, KI og tungregning. Det vil være nødvendig med tettere koblinger til forskningssystemet. Dette vil bidra til å ta forskningen og kompetansen i forskningsinstitusjonene raskere i bruk.

Tungregneressurser på ulike graderingsnivåer

- FFI og NSM har tilgang til tungregneressurser på gradert nivå. Ytterligere informasjon om dette er gradert, det er derfor ikke omtalt i mer detalj i denne rapporten.
- Det vil være et økende behov for tungregnekapasitet i forsvarssektoren, både egne graderte (som blant annet blir brukt til å vurdere ytelse og trusler), men også ugraderte systemer.
- Skjerming av informasjon og FoU er faglig krevende og kostbart. Riktig bruk av graderingsnivå for tungregneressurser kan bidra til økt sikkerhet og er potensielt kostnadsbesparende.
- Det bør være en ambisjon å gjenbruke åpen/felles infrastruktur for tungregning innen forsvarssektoren så langt som mulig, og vurdere behov for å bruke graderte systemer nøye. Det er en forutsetning at det er høyt kunnskapsnivå på å vurdere nødvendigheten av bruk av graderte vs. åpen infrastruktur for tungregning.

Tungregneressurser lokalisert i Norge

- Det er viktig å ha tilgjengelige norske ressurer både lokalt, nasjonalt og internasjonalt for å ivareta ulike behov.
- Vi anbefaler derfor at det fortsatt investeres betydelige beløp i tungregneinfrastruktur lokalisert i Norge. Det er to hovedgrunner til dette:
 - Sikkerhet og beredskap: risiko for at Norges tilgang til internasjonale tungregnekapasitet blir nedprioritert ved usikkerhet og kriser.
 - Bygge kompetanse og ekspertise: norske miljøer må ha enkel tilgang til lokale og nasjonale tungregneressurser som treningsarenaer.
- Det er viktig å ha tilstrekkelig nasjonal kontroll over tungregneressurser, spesielt for oppgaver som er relatert til forsvar, klima, helse, geofarer og matsikkerhet. Ressurser som eventuelt ikke er plassert i Norge bør ligge i land vi har sikkerhetspolitisk samarbeid med
- Nærhet mellom bruker og tungregneressurser er viktig, særlig for mindre erfarne brukere. Dette er viktig for å bygge kompetanse både i forskningsmiljøene og i offentlig sektor og næringsliv og enklere kunne realisere samarbeid mellom ulike fagmiljøer fra ulike sektorer (forskning, offentlig sektor og næringsliv).
- For tilgang til særlig store superdatamaskiner bør Norge samarbeide internasjonalt, fortrinnsvis i Norden og Europa (EuroHPC).
- Plassering av tungregneressurser i Norge kan også være klimavennlig.

Andre funn og anbefalinger

- Det trengs gode støttetjenester som er tilpasset ulike brukergruppers behov (forskningsmiljøene, offentlig sektor og næringsliv) for bedre utnyttelse av regneressursene både nasjonalt, men spesielt ressursene vi får tilgang til gjennom EuroHPC.
- Det er også behov for infrastruktur og verktøy for datahåndtering koplet til tungregneressursene.
- Det er behov for kompetanse og kapasitet for å gjøre data tilgjengelig for bruk.
- Det er behov for kompetanse på juridiske og regulatoriske bestemmelser for bruk av data (inkludert rettigheter, personvern, mm.).
- Det er behov for godt samarbeid mellom lokale, nasjonale og internasjonale tungregneressurser, for bedre utnyttelse av kompetanse og kapasitet.
- Investeringene bør være så klimavennlige som mulig

Ambisjonsnivået i våre naboland

I dette kapittelet vil vi gi en oppsummering av nivået på investering i tungregnekraft i land i Europa.

Vi har innhentet informasjon fra Finland, Sverige, Danmark, Tyskland, Nederland og Frankrike om nivået på investeringer (historiske og planlagte) i tungregnekraft. Informasjon om hvert land er oppsummert i vedlegg 3.

Generelt er det utfordrende å si noe samlet om investeringsnivå mellom landene, da alle land har ulike modeller for finansiering, og ulik grad av sentralisering av satsingen, samt ulikt nivå på investeringer fra private aktører. For eksempel har Finland og Tyskland en langt mer sentralisert modell enn Danmark som har valgt en desentralisert modell. Det er derfor enklere å få oversikt over det totale investeringsnivået i disse to landene.

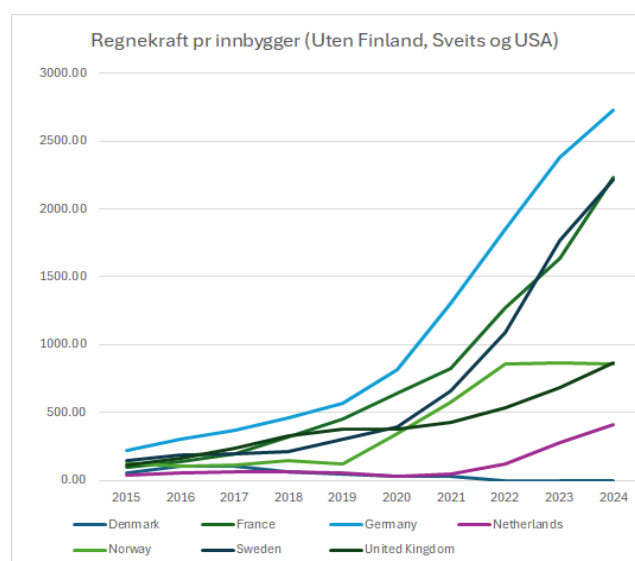
Nivået på fremtidige investeringer samt nåværende tungregneressurser er på et jevnt høyere nivå enn i Norge i de fleste landene. Dette gjelder både investeringer i nasjonale ressurser, samt investeringer gjennom EuroHPC. Vi ser også at organiseringen i de ulike landene påvirker både det strategiske ambisjonsnivået og investeringsnivået. Felles for landene er at de fordeler investeringene mellom EuroHPC, nasjonale og lokale anlegg.

Alle landene oppfatter det europeiske samarbeidet om tungregning, EuroHPC, som strategisk viktig. Finland og Frankrike og etter hvert Tyskland og Sverige vil være vertskap for fire ulike maskiner finansiert delvis av EuroHPC. Landene setter av betydelige summer fremover til investeringer i EuroHPC-maskinene. Nederland har ikke planer om å øke nasjonale HPC investeringene fremover, men skal investere i EuroHPC-anlegg og dekke medfinansieringen i EuroHPC-prosjekter.

Finland og Tyskland har de største investeringene på området, med store investeringer både i nasjonal tungregnekapasitet, men ikke minst store investeringer i EuroHPC-samarbeidet.

Sigma2 har kartlagt regnekraft per innbygger for noen utvalgte land, noe som indirekte sier noe om investeringsnivå på tungregning for de enkelte landene (**Feil! Fant ikke referansebildet.**). Dette er offentlige investeringer som er gjort innen forskning og forvaltning. Norge er blant de tre dårligste landene i dette utvalget. Finland, Sveits og USA er ikke med i denne figuren da de ligger langt over resten av landene. USA investerer mer enn dobbelt så mye pr innbygger som Tyskland, Sveits mer enn tre ganger så mye som Tyskland og Finland mer enn dobbelt så mye som Sveits. I Danmark har Novo Nordisk annonsert at de vil investere i superdatamaskin for FoU. Dette er et offentlig-privat samarbeid for å bygge en superdatamaskin som vil bli blant verdens kraftigste, med en investering på ca. 1 100 millioner kroner. Denne investeringen er ikke med i oversikten, som betyr at den samlede investeringen i Danmark er langt høyere enn det som kommer fram av figuren.

Figur 1: Regnekraft per innbygger



Funn og anbefalinger

- Alle andre sammenlignbare land investerer betydelige offentlige midler i nasjonale tungregneressurser for forskning og forvaltning.
- Alle andre sammenlignbare land i Europa investerer betydelige offentlige midler i EuroHPC.

Deltagelse og muligheter i EuroHPC

Her beskriver vi hvilken tilgang norske forsknings- og innovasjonsmiljøer, inkludert næringslivet, har per nå til forskjellige typer tungregnekraft gjennom internasjonale forsknings- og innovasjonsprogrammer, og norsk deltagelse og investering fremover. Når det gjelder spørsmålet om innretning for å få mest mulig utbytte av investeringen og deltakelsen i det internasjonale partnerskapet EuroHPC, Horisont Europa og DIGITAL, har vi i denne rapporten avgrenset dette til å gjelde tilgang til tungregning og aktiviteter som gjelder utnyttelse av den tilgjengelige tungregnekapasiteten. For generell deltagelse i Horisont Europa og DIGITAL som ikke gjelder tungregning viser vi til regjeringens strategier for deltagelse i de to programmene og de tiltak som er beskrevet der^{25 26}.

Norsk deltagelse og innretning i dag

I dette avsnittet beskriver vi kort hvordan EuroHPC er organisert og hvordan norsk deltagelse er innrettet i dag. Norske aktører har tilgang til ulike tungregneressurser gjennom de internasjonale forsknings- og innovasjonsprogrammene Horisont Europa og programmet for et digitalt Europa (DIGITAL), samt norsk deltagelse i det europeiske partnerskapet for tungregning (EuroHPC). Deltagelse i Horisont Europa og DIGITAL programmene er en forutsetning for utnyttelsen av vår deltagelse i EuroHPC. Medlemskapet i EuroHPC gir norske aktører muligheten til å delta i felles europeiske prosjekter og søke midler gjennom utlysningene som EuroHPC administrerer.

Kort om EuroHPC

EuroHPC Joint Undertaking (EuroHPC) er et europeisk samarbeid om tungregneressurser mellom EU, europeiske land og private aktører, med et budsjett på ca. 7 milliarder euro for perioden 2021-2027. EUs andel utgjør ca. halvparten og kommer fra EUs tre rammeprogrammer: DIGITAL (1,9 milliarder euro), Horisont Europa (HEU, 900 millioner euro) og Connecting Europe Facility (CEF-2, 200 millioner euro). Norge deltar i DIGITAL og HEU, men ikke i CEF-2, noe som begrenser norsk deltagelse i prosjekter finansiert av CEF-2. Forskningsrådet forvalter Norges delegat i EuroHPCs generalforsamling.

EuroHPC har en to-delt rolle;

1. investere i europeisk infrastruktur for tungregning og KI
2. etablere forsknings- og innovasjonsprosjekter som utvikler og bruker tungregneressurser.

EuroHPCs oppdrag er strukturert rundt seks teknologiske søyler: Infrastruktur, Applikasjoner & Data, Internasjonalt Samarbeid, Teknologi, Ferdigheter & Bruk, samt Federasjon & Hyperkonnektivitet.

I mai 2024 vedtok Det Europeiske Råd en endring som inkluderte utvikling og drift av KI-fabrikker (AI factories), med fokus på KI-optimaliserte datamaskiner og systemer som hjelper bedrifter, SMBer og start-ups. Aktivitetene knyttet til KI-fabrikker kan struktureres rundt syv hovedpunkter som også går igjen i budsjettinnspill til DIGITAL-aktiviteter. De to første punktene er investeringer knyttet til infrastruktur. De andre punktene er knyttet til andre KI-aktiviteter.

1. Acquisition & operation of AI dedicated supercomputers (co-located with data centre)
2. Upgrading existing EuroHPC systems with AI capabilities/capacity
3. Provide access for SMEs and start-ups (incl. widening usage)
4. AI supercomputing service centre (algorithms, training- testing models, development of large-scale AI applications, ...)
5. Supercomputer-friendly programming facilities (parallelization, usage optimization, ...)
6. Attracting & pooling talent

²⁵ [Strategi for norsk deltagelse i Programmet for et digitalt Europa – DIGITAL \(regjeringen.no\)](https://www.regjeringen.no)

²⁶ [212540-kd-strategi-horisonteuropa-web.pdf \(regjeringen.no\)](https://www.regjeringen.no)

Tilgjengelig regnekraft gjennom deltakelse i EuroHPC

EuroHPC har investert i ni superdatamaskiner, hvor EuroHPC eier 35-50 prosent, mens konsortier fra deltakerlandene eier resten, eierandelen er proporsjonalt med deres investering.

Gjennom Sigma2 har Norge vært med på etableringen av en av verdens største superdatamaskiner, LUMI, i Kajaani, Finland. Dette er et samarbeid mellom 10 land, og prosjektet har fått i underkant av 1 milliarder kroner fra EuroHPC. Norge har bidratt med 40 millioner kroner via Forskningsrådets finansieringsordning for nasjonal forskningsinfrastruktur (INFRASTRUKTUR). Dette gir Norge tilgang til 2 prosent av LUMIs totale kapasitet. Norge utnytter i dag fullt ut sin kapasitet på LUMI.

EuroHPC utlyser sin andel av kapasiteten. Brukere fra forskningsmiljøer, offentlig sektor og bedrifter i land som er medlem av HEU eller DIGITAL kan søke og kapasiteten har til nå blitt tildelt basert på kvalitet. Dette kan endre seg for de nye AI-utlysningene der det er lagt vekt på at SMBer skal delta.

Norske miljøer søker mest på de største supercomputerne (pre-exascale kapasitet) i EuroHPC fordi dette er et behov som ikke dekkes gjennom nasjonal kapasitet hos Sigma2 (nasjonal mid-range kapasitet). Norske aktører kan også få tilgang til mindre ressurser for å bli kjent med superdatamaskinene, teste programvare og utføre ytelsestester. Næringslivets tilgang til LUMI gis iht. statsstøttereformativet, men EU planlegger større tilgang for utvikling og tidlig trening av KI-modeller gjennom såkalte KI-fabrikker.

Nasjonalt kompetansesenter for HPC

Sigma2, NORCE og SINTEF leder et nasjonalt kompetansesenter for HPC (NCC), rettet mot næringsliv og offentlig sektor³¹. Senteret har som mål å bistå industri og næringsliv, spesielt SMBer, samt offentlig forvaltning med å øke kompetansen innen avansert teknologi, tungregning (HPC), avanserte dataanalyser, kunstig intelligens og maskinlæring. Senteret har EU-finansiering ut 2025 gjennom DIGITAL og nasjonal medfinansiering, gjennom Forskningsrådets ordning IKTPLUSS. Kompetansesenteret er en del av et nasjonalt støtteapparat for norske SMBer, som inkluderer norske innovasjonshuber (EDIHer) som NEMONOOR og Oceanopolis, samt Innovasjon Norge. Det samarbeider også med andre europeiske kompetansesentre og flere Centre of Excellence for HPC. Eksempler på norske aktører fra henholdsvis næringsliv og offentlig sektor som har hatt nytte av det nasjonale kompetansesenteret, er DigiFarm og Tollvesenet.

DigiFarm – digitale verktøy for landbruket

DigiFarm er en oppstartsbedrift som hjelper bønder og landbruksbedrifter med å øke avlinger og redusere kostnader til gjødsel og plantevern.

DigiFarm fikk avgjørende hjelp av det nasjonale kompetansesenter for tungregning (NCC). De hadde behov for større regneressurser enn de kommersielle skyleverandørene kunne tilby. NCC hjalp DigiFarm med å søke Sigma2 om ressurser og ga opplæring i bruk av HPC. Sigma2 bidro også med å optimalisere arbeidsflyten for raskere kjøring, lavere ressursbruk og kostnader.

DigiFarm, med sitt høye forbruk på nasjonale tungregningsressurser, var en naturlig kandidat for kjøretid på LUMI. DigiFarm trengte bare enkel veiledning fra Sigma2s supportteam for å kjøre på LUMI. Dette har resultert i at DigiFarm hittil i år er den største brukeren av Norges kvote på LUMI. DigiFarm har også fått hjelp av NCC til å søke om kapasitet på LUMI fra EuroHPCs kvote, og har blitt tildelt 140.000 GPU-timer til en verdi av 827.000 NOK gjennom «open call for AI and data-intensive applications.»

DigiFarms tilgang til nasjonale ressurser og LUMI har dramatisk effektivisert treningsprosessen for deres KI-modeller. Tilgangen til store mengder sammenkoblede GPUer har gjort det mulig å trene større modeller, detektere nye elementer som avskoging, og spare tid. Ifølge CEO Nils Helseth har dette forkortet veien fra testing til lanseringsklart produkt med cirka 6 måneder. Modellen har forbedret nøyaktigheten med 4,2 prosent på bare noen få måneder, en betydelig prestasjon innen dyplæring, ifølge Helseth.

Kilde: [Sigma2 - nettside](#)

Tollvesenet – sikrere grenser med kunstig intelligens

I 2021 mottok tollvesenet nesten 10 millioner tolldeklarasjoner. Dette skaper enorme mengder data som blir analysert for å oppdage aktiviteter som strider mot regelverket.

Manuelle tolldeklarasjoner inneholder ofte skrivefeil eller variasjoner i hvordan navn eller adresser er skrevet. Når data er unøyaktig, kan det som ser ut som flere ulike adressater, i realiteten være ett og samme firma. En aktør kan gjøre små endringer på skrivemåten til navn og adresser han sender ifra, noe som kan skjule mønsteret i forsendelsene og unngå å bli flagget ved kontroll. Manuell validering av millioner av oppføringer er ikke et alternativ, og det er her kunstig intelligens kommer inn i bildet.

Tollvesenet engasjerte det nasjonale kompetansesenter for HPC for å få bistand til å demonstrere hvordan man kan bruke KI (objekt-identifisering) for å rydde opp i datakvaliteten. Dette er nyttig og tidsbesparende når man slår sammen data fra forskjellige kilder, i deteksjon av svindel og i kundebehandlingsstyring.

Programvaren Splink justerer vurderinger knyttet til om to oppføringer referer til den samme til den finner den beste modellen for dataene. Til å trene større maskinlæringsmodeller trengs kraften fra superdatamaskiner. Nå hjelper kompetansesenteret Tollvesenet med å etablere en maskinlæringsarbeidsflyt for å rydde i dataene og knytte sammen datapunktene som algoritmene ser som samme entitet eller adressat. Slik kan kunstig intelligens hjelpe til med å gjøre grensene våre sikrere.

Kilde: [Sigma2 - nettside](#)

Norsk deltagelse og innretning fremover

Regjeringen ønsker at Norge skal delta aktivt i DIGITAL, og den nye strategien for norsk deltakelse i rammeprogrammet skisserer tydelige mål, prioriteringer og innsatsområder. Den dekker også de ulike satsingsområdene i DIGITAL, inkludert tungregning og superdatamaskiner. Strategien peker på viktigheten av å videreutvikle pågående nasjonale aktiviteter og satsinger. Dette krever koordinering mellom nasjonale og internasjonale virkemidler, og det er nødvendig å bygge opp og vedlikeholde både menneskelige og teknologiske ressurser som støtter deltakelsen i HEU og DIGITAL.

Det ligger store muligheter for norske aktører i DIGITAL og EuroHPC og alle aktører vi har vært i dialog med og mottatt innspill fra påpeker betydningen av det europeiske samarbeidet om tungregning og KI.

Tilgang til infrastruktur for tungregning og KI

Det blir viktig fremover å videreføre LUMI-samarbeidet. Det planlegges en ny LUMI-maskin som i større grad enn tidligere maskiner er KI-optimalisert. For at norske aktører skal få tilgang er det viktig at Norge investerer i denne. Investeringsnivået må være høyt nok til at vi sikrer norske aktører en tilstrekkelig andel av kapasiteten. En endring i EuroHPC-forordningen sikrer at kapasiteten i tungregnemaskiner fremover blir tilgjengelig for et bredere spekter av brukere, både offentlige og private.

Sigma2 estimerer et investeringsnivå for den neste KI-relaterte LUMI-maskinen som utgjør 10-20 prosent av det nasjonale investeringsnivået på HPC/KI. Verter av KI-optimaliserte superdatamaskiner kan motta økonomisk støtte fra EU som dekker opptil 50 prosent av anskaffelses- og driftskostnadene.

Aktiviteter for å nyttiggjøre investeringer i infrastruktur

For å nyttiggjøre seg investeringer i infrastruktur for tungregning er det viktig å delta i andre aktiviteter/utlysninger som skal understøtte bruken av infrastrukturen (nasjonalt kompetansesenter, utvikling av KI-relaterte støttetjenester og programvare m.m.).

I DIGITAL-utlysninger er det krav om egeninnsats. For at Norge skal utnytte de muligheter som ligger i deltagelsen i DIGITAL og EuroHPC er det behov for medfinansiering. Sigma2 har estimert kostnader for aktiviteter de selv og andre aktører ønsker å utnytte i kommende utlysninger. Estimatenes er på mellom 58,8 og 80 millioner kroner for ulike aktiviteter for perioden 2025-2028. Disse kostnadene kommer i tillegg til investeringer i infrastruktur for tungregning. Planlegging og

budsjettering er utfordrende fordi EuroHPC kun har vedtatt arbeidsprogrammet for 2024, og ikke for kommende år. Dette skaper usikkerhet for norske aktører som planlegger aktiviteter gjennom EuroHPC og det er dermed noe usikkerhet knyttet til disse tallene som gjenspeiles i spennet i tallene.

Utviklingen i EuroHPC og beslutninger i styringsorganet skjer raskt. Derfor kan utlysninger komme med veldig kort varsel. For å sikre norske aktører de beste muligheter i partnerskapet er det derfor viktig at det fins nasjonal medfinansiering tilgjengelig som raskt kan settes i arbeid.

Videreføring av Nasjonalt kompetansesenter for HPC

De nasjonale kompetansesentrene for HPC (NCC) vil spille en viktig rolle for KI-fabrikker (AI factories) og de nye norske KI-sentrene som skal etableres i forbindelse med regjeringens KI-satsning. EuroHPC har et økt fokus rettet mot oppstartsbedrifter og SMBer og det nasjonale kompetansesenteret må innrette seg for å møte disse kravene. Det norske kompetansesenteret jobber nå med tre piloter i offentlig sektor og tre i industrien sammen med Kartverket, NGI og NVE.

Det forventes en ny utlysning til nasjonale kompetansesenter gjennom EuroHPC, og norske aktører vil sannsynligvis søke. Videre norsk medfinansiering vil koste cirka 5 millioner NOK årlig (2026-2028).

Samspill med European Open Science Cloud

Bruk av tungregning krever et økosystem av gode datainfrastrukturer og kvalitetsdata. Norge har systematisk investert i datainfrastruktur for forskning gjennom INFRASTRUKTUR-ordningen, noe som har posisjonert norske miljøer godt for utlysninger i HEUs arbeidsprogram for forskningsinfrastruktur, spesielt innen destinasjonen European Open Science Cloud (EOSC). EOSC er et europeisk partnerskap som skal bygge et nettverk for deling og gjenbruk av FAIR forskningsdata. Dette inkluderer også kompetanse om deling av data. Et godt samspill mellom norsk deltakelse i EOSC og EuroHPC kan gi betydelige synergieffekter, slik at forskere får enkel og sikker tilgang til store datasett av høy kvalitet og kan utføre komplekse beregninger gjennom integrerte arbeidsflyter.

Rammevilkår for norske aktører

Retur-EU

Norske fagmiljøer, spesielt i instituttsektoren, som benytter EuroHPC-utlysninger, påpeker at manglende nasjonal medfinansiering, særlig i DIGITAL-utlysninger, er en økonomisk utfordring. Dette fører til at de heller søker Horisont Europa-utlysninger, hvor overheadsatsene er høyere og balansert med nasjonal toppfinansiering gjennom Retur-EU-ordningen. Små og mellomstore organisasjoner, spesielt forskningsinstituttene, er sårbare for lave overhead-rater i DIGITAL-utlysninger, og prosjekterer ofte med tap eller takker nei til invitasjoner til internasjonale prosjektkonsortier. Dette understreker viktigheten av en forutsigbar Retur-EU ordning for Horisont Europa, men også behovet for å adressere økonomiske utfordringer i DIGITAL-utlysninger. Regjeringen vil derfor vurdere etablering av en medfinansieringsordning for norske aktørers deltakelse i DIGITAL, i tråd med DIGITAL-strategien.

Administrativ oppfølging av EuroHPC

Forskningsrådet har delegatansvaret i styringsorganet til EuroHPC. Det avholdes hyppige styremøter og oppfølgingen av norsk deltakelse i EuroHPC er krevende og til tider preget av korte frister og lite tid til forankring og oppfølging med nasjonale eksperter. Vi anbefaler at det settes av minst ett årsverk for oppfølging av EuroHPC. Dette for å sikre god og tilstrekkelig koordinering med relevante aktører og oppfølging av det administrative arbeidet knyttet til aktiviteter i EuroHPC-styret.

Funn og anbefalinger

Investering i tungregneanlegg (store superdatamaskiner)

- For tilgang til særlig store superdatamaskiner bør Norge samarbeide internasjonalt, fortrinnsvis i Norden og Europa (EuroHPC).
- Vi foreslår en samlet investering til den nye KI-relaterte LUMI-maskinen som utgjør 10-20 prosent av den foreslåtte investeringssummen til nasjonal tungregning (Sigma2), i perioden 2026-2027.

Innretning og nyttegjøring av investeringer i EuroHPC

- Utviklingen i EuroHPC og beslutninger i styringsorganet skjer raskt. For å sikre norske aktører de beste muligheter i partnerskapet, anbefaler vi at det settes av en nasjonal EuroHPC-pott, som raskt kan mobiliseres og settes i arbeid.
- Retur-EU ordningen må opprettholdes og være forutsigbar.
- Vi anbefaler at det etableres en medfinansieringsordninger som kompenserer for norske forskningsinstitusjonenes, særlig institutters, manglende fullfinansiering av faktiske kostnader i prosjektene fra DIGITAL. Dette for å sikre forutsigbarhet og konkurransedyktighet hos norske miljøer, i tråd med strategiske prioriteringer i den norske DIGITAL-strategien.
- Vi anbefaler videreføring av nasjonalt kompetansesenter for HPC som tilbyr brukerstøtte og opplæring for offentlig sektor og næringsliv. Senteret må tilpasse aktivitetene for å utnytte nye muligheter i EuroHPC.
- Det kan være muligheter for å være vertskap for et EuroHPC-anlegg på norsk jord, delfinansiert av EuroHPC og et konsortium. Vi anbefaler at det gjøres en kost-nytte-analyse samt en juridisk vurdering av EØS-lands muligheter for å være fysisk vertskap for en forskningsinfrastruktur som delvis er eid av EuroHPC.
- Vi anbefaler at det settes av tilstrekkelige ressurser til administrativ oppfølging av EuroHPC. Dette for å sikre god og tilstrekkelig koordinering med relevante aktører og oppfølging av det administrative arbeidet knyttet til aktiviteter i EuroHPC.

Kostnadsanslag i et femårsperspektiv

Investeringer

Basert på informasjonen vi har innhentet anbefaler vi et totalt investeringsnivå på 2 000 millioner kroner over de neste fem årene for å kunne dekke behovet innenfor FoU som vi kjenner til per i dag. Dette inkluderer:

- Totalt investeringsbehov i tradisjonell tungregningskapasitet for perioden 2025-2029: 1 000 millioner kroner. Dette inkluderer også nytt lagringsanlegg som er en forutsetning for god utnyttelse av tungregneressursene. Anslaget er utarbeidet av Sigma2 basert på historisk utvikling over tid samt brukerundersøkelser.
- KI regnebehov for de nærmeste 3-5 år: 1 000 millioner kroner²⁷. Dette er basert på et forsiktig estimat av Sigma2 som legger til grunn en økning som tilsvarer i overkant av 10 ganger kapasiteten vi har tilgang til per nå. Vi anser dette estimatet som noe mer usikkert da det er begrensede erfaringstall på området, samt stort spenn i forventet behov.
- Av dette foreslår å investere på 200 millioner kroner for 2025.
- Minst 20 prosent av de samlede kostnadene til tungregning er til norsk deltakelse i EuroHPC

Plassering av anleggene og type investeringsmodell vil påvirke kostnadene, for eksempel om det velges en sentralisert eller desentralisert plassering.

Drift

Dagens driftskostnader er totalt ca. 95 millioner kroner årlig, som inkluderer drift- og brukerstøtte, datasenter og utviklingsprosjekter.

Basert på estimater for økning i investeringer og nye områder knyttet til KI og kvanteberegninger legges det opp både en økning direkte i driftsutgifter knyttet til strøm, men også til brukerstøtte og spesielt den mer avanserte brukerstøtten (KI og kvanteberegninger). Estimater på dette er satt til 93 millioner kroner per år, totalt 475 millioner kroner over 5 år. Datasenterkostnader krever at det tas hensyn til innfasing og utfasing av anlegg. Det er i tillegg beregnet et behov på om lag 160 millioner kroner til husleie og strøm. Dette gir totalt sett beregnede driftskostnader på om lag 600 millioner kroner over 5 år.

For 2025 beregnes en årlig kostnad til husleie og strøm på om lag 8 millioner kroner. Dette gir en totalt beregnet behov på om lag 100 millioner kroner i driftskostnader i 2025.

Ut over kostnader til strøm er kostnader til drift uavhengig av størrelsen på investeringer, men mest avhengig av antall ulike systemer som skal driftes. Omfanget og kvaliteten av brukerstøtte vil også ha en betydning. Andelen av kostnader til drift blir mindre desto større selve investeringen er. Det vil også være en stor del kostnadsdekning for drift gjennom bidragsmodell (leiestedsmodell). Avhengig av vurderinger rundt krav til oppetid og redundans, kan det bli behov for flere datasenter lokasjoner. Så langt har oppetidskrav til tungregnesystem for forskning vært moderate og ikke krevd flere lokasjoner.

Totalkostnad i en femårsperiode

På grunn av stor usikkerhet i utviklingen framover vil estimatene endre seg raskt. Vi anbefaler derfor å lage en opptrappingsplan for investeringer og drift på minst 2 600 millioner kroner til tungregneressurser (pluss eventuelle kostnader for språkmodell) for den neste femårsperioden.

²⁷ Dette er basert på en enhetspris per GPU på 380 000 NOK.

Opptappingsplanen bør revideres årlig fordi estimatet trolig vil endre seg mye og raskt både på grunn av usikkerhet i utviklingen i faktisk behov, teknologiutvikling, samt usikre priser på hardware.

Nasjonal infrastruktur som sikrer tilgang til regnressurser for en stor bredde av norsk forskning og forvaltning bør sikres langsiktig og forutsigbar finansiering.

Tabell 1: Forslag til finansieringsplan for en fem-årsperiode (i mill. kroner)

	2025	2026	2027	2028	2029
Investering Sigma2, inkludert GPU-kapasitet til KI	200	300	300	400	400
Drift Sigma2	100	125	125	125	125
Investering gjennom EuroHPC/LUMI		200	200		
Sum	300	625	635	530	540

Felles for våre naboland er at en stor del av deres investeringer i tungregning gjøres gjennom EuroHPC-samarbeidet. Dette samarbeidet muliggjør tilgang til kostbare og toppmoderne tungregnefasiliteter og en kostnadseffektiv utnyttelse av store tungregneressurser. Det ville vært dyrere om landene skulle gjøre tilsvarende investeringer og drifte anleggene selv. Næringslivet har også tilgang til disse fasilitetene. Basert på en forventning om økt ressurseffektivitet og lavere risiko anbefaler at 10 til 20 prosent av Norges investeringer gjøres gjennom EuroHPC-samarbeidet. Størrelsen på andelen bestemmes utfra en kost/nytte-vurdering samt vurderinger rundt nasjonal sikkerhet og beredskap.

Vi anbefaler at det settes av dedikerte midler til utvikling av norsk språkmodell. Midler til dette er ikke inkludert i forslag til finansieringsplanen over. Totalkostnad per år vil avhenge av antall modeller som skal trenes, hyppighet på trening og akseptabel treningstid. Basert på erfaringer fra trening av språkmodeller så langt er kostnader ved bruk av kommersielle regneressurser 5-10 ganger høyere pr regneenhet enn ved bruk av de nasjonale tjenestene og EuroHPC. Vi viser i denne sammenheng til Nasjonalbibliotekets egen behovsanalyse sendt til Kultur- og likestillingsdepartementet (KUD) i mai i år.

Kostandene til investering og drift som beskrevet over inkluderer ikke:

- Kostnader for tungregnekraft på gradert nivå.
- Spesielle tilpasninger for beregninger med bruk av personsensitive data.
- Kostnader for å gjøre data tilgjengelig og klare for bruk.
- Kostnader til utredninger som gjøres for ulike typer data/datasett mht. til rettigheter, lover og reguleringer for bruk.
- Videre investeringer i kvanteteknologi.

Det må også allokeres midler til regnekraft og drift for opprettholdelse og utvikling av andre samfunnsnyttige/samfunnskritiske tjenester. For drift av tjenester som utvikles for bruk til andre formål enn forskning må det allokeres midler fra andre budsjetter enn forskningsbudsjettene, slik at videreutvikling av disse tjenestene sikres tilstrekkelig regnekraft.

Funn og anbefalinger

- Vi anbefaler å lage en opptappingsplan for investeringer og drift på minst 2 600 millioner kroner for den neste femårsperioden for å dekke behovene innen forskning og utvikling.
- Opptappingsplanen bør revideres årlig. Det vurderes at estimatet vil endre seg mye både på grunn av usikkerhet i utviklingen i faktisk behov, teknologiutvikling, samt usikre priser på hardware.
- Vi anbefaler at 10 til 20 prosent av Norges investeringer gjøres gjennom EuroHPC-samarbeidet, til den nye KI-relaterte LUMI-maskinen. Dette er basert på en forventning om økt ressurseffektivitet og lavere risiko. Størrelsen på andelen bestemmes utfra en kost/nytte-vurdering samt vurderinger rundt nasjonal sikkerhet og beredskap.

Vedleggsoversikt

1. Oppdraget fra KD av 05.04.24: Behov for tungregnekraft for forskning og kunstig intelligens.
2. Spørsmål fra spørreundersøkelsene og institusjoner spurt
3. Nivå på investeringer i nordiske land og relevante EU-land

Norges forskningsråd

Besøksadresse: Drammensveien 288

Postboks 564

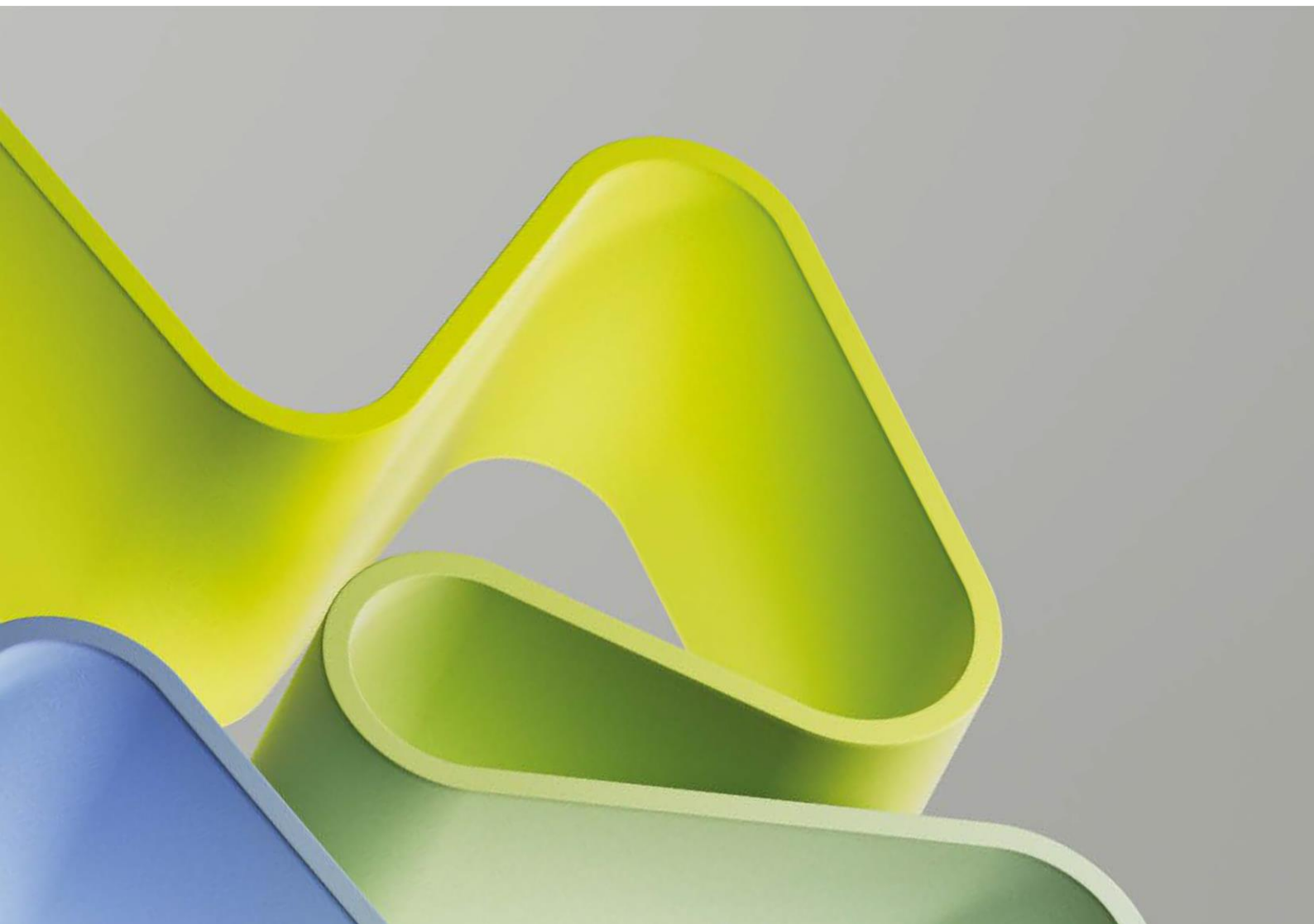
1327 Lysaker

Telefon: 22 03 70 00

Telefaks: 22 03 70 01

post@forskningsradet.no

www.forskningsradet.no





Norges forskningsråd
Postboks 564
1327 LYSAKER

Deres ref

Vår ref
24/1735-

Dato
5.april 2024

Utredningsoppdrag: Behov for tungregnekraft for forskning og kunstig intelligens

Bakgrunn

Behovet for superdatamaskiner og tilhørende infrastruktur for tungregning er stadig økende, særlig i forbindelse med den senere tids utvikling av kunstig intelligens (KI). Verdifulle data frembringes i økende omfang og må ivaretas og gjøres tilgjengelige og gjenbrukes for å utforske og løse fremtidens problemstillinger. Tilstrekkelig tilgang til regnekraft er en absolutt nødvendighet for å drive forskning av høy kvalitet, men også for å drive effektiv og god forvaltning på områder som f.eks. helse, klima, miljø og beredskap. Stortinget har i Innst. 170 S (2022–2023) sluttet seg til regjeringens ambisjon om å sørge for tilstrekkelig nasjonal tungregnekapasitet til å dekke nåværende og fremtidige behov i forskning og forvaltning, jf. Meld. St. 5 (2022–2023) *Langtidsplan for forskning og høyere utdanning*.

Tilgang på regnekraft og grunnmodeller for kunstig intelligens er avgjørende for at Norge skal lykkes med forskning, innovasjon og verdiskaping med KI. Norske grunnmodeller kan videreutvikles av næringsliv eller offentlig sektor for å lage f.eks. språkmodell-baserte KI-verktøy for mer effektive tjenester og produksjon. Som kunngjort da KI-satsingen ble lansert i september i fjor, er regjeringen i gang med et større arbeid på tvers av departementer og sektorer. Målet er å finne tiltak for kompetanseutvikling, datasett og infrastruktur for fremtidens databehandling, inkludert tungregnekapasitet. Dette følges opp i arbeidet med en melding til Stortinget om forskningssystemet.

Økende behov for tungregnekraft er ikke bare knyttet til utviklingen innenfor KI, men er også nødvendig for å sikre en god og nødvendig utvikling på tradisjonelle forskningsområder som fysikk, kjemi, biologi, meteorologi og hav i tillegg til humaniora og samfunnsvitenskapene. Tungregning blir også i økende grad benyttet innenfor forvaltningen og er med på å åpne nye kommersielle muligheter i privat sektor. SMB-er vil kunne ha stor nytte av tilgang til en felles

Postadresse
Postboks 8119 Dep
0032 Oslo
postmottak@kd.dep.no

Kontoradresse
Kirkeg. 18
www.kd.dep.no

Telefon*
22 24 90 90
Org.nr.
872 417 842

Avdeling
Avdeling for høyere
utdanning, forskning
og internasjonalt
arbeid

Saksbehandler
Finn-Hugo
Markussen
22 24 77 28

regnekraftressurs. Kunnskapsdepartementet ber derfor Forskningsrådet om å utrede Norges behov for tungregnekraft til forskning, forvaltning og næring generelt, og til kunstig intelligens spesielt.

Investeringer i de største, nasjonale superdatamaskinene (og datalagringskapasitet) gjøres i dag av det statlige aksjeselskapet Sigma2 AS, finansiert av konkurranseutsatte midler som lyses ut gjennom Forskningsrådets nasjonale ordning for forskningsinfrastruktur. Søknader er basert på dokumenterte behov og prognoser og har i de seneste utlysningene overskredet grensen for hva Forskningsrådet kan tildele. Sigma2 mottar også årlig tilskudd gjennom Forskningsrådet, og superdatamaskinene driftes i samarbeid med UiB, UiO, NTNU og UiT. I tillegg er det andre offentlige virksomheter som disponerer tungregnerressurser. Ettersom behovene forventes å øke både innenfor og utenfor rene forskningsmiljøer, kan det bli nødvendig å se på fremtidig organisering av tungregnerressursene. En mulighet kan være å etablere en nasjonal tungregnerressurs som kan betjene både forskning, forvaltning og bruk av kunstig intelligens i offentlig og privat innovasjon og verdiskaping. Spørsmål om organisering ligger imidlertid utenfor dette utredningsoppdraget og er noe departementet vil komme tilbake til i en egen prosess.

Gjennomføring

Analysen av fremtidige behov for tungregnekraft skal ta utgangspunkt i kjente behov i forskningssektoren, offentlig sektor og næringslivet, i tillegg til vurderinger knyttet til fremtidige ambisjoner. Det vil være naturlig at Norge har et ambisjonsnivå på linje med våre nordiske naboer og relevante EU-land. Analysen skal ta høyde for behov knyttet til skjermingsverdig og gradert FoU på ulike nivåer.

Analysen skal vurdere i hvilken grad behov kan dekkes av Norges deltakelse i internasjonale satsinger, herunder EuroHPC.

I arbeidet med utredningen skal Forskningsrådet ha kontakt med relevante offentlige og private aktører innenfor og utenfor forskningssystemet, inkludert leverandører av skjermingsverdig og gradert FoU.

Utredningen skal følge utredningsinstruksen. Følgende spørsmål skal besvares:

Del 1 – Omfang og sammensetning av tungregnerressursene i dag

- a) Hva slags norskeid offentlig og privat tungregnekraft er i dag tilgjengelig i Norge til
 - tradisjonelle tungregneoppgaver?
 - kunstig intelligens?
- b) Hvilken tilgang har norske forsknings- og innovasjonsmiljøer, inkludert næringslivet, i dag til forskjellige typer tungregnekraft gjennom internasjonale forsknings- og innovasjonsprogrammer?
- c) Hvilken tilgang har forskningssektoren, offentlig sektor og næringslivet til forskjellige typer tungregnekraft på ulike graderingsnivåer?

Del 2 – Forventet utvikling i et femårsperspektiv

- a) Hvor mye og hvilken sammensetning av tungregnekraft vil Norge trenge, gitt ambisjonsnivået om å være på nivå med våre nordiske naboer og relevante EU-land?
- b) Hvordan skal man innrette seg for å få mest mulig utbytte av investeringen og deltakelsen i det internasjonale partnerskapet EuroHPC og internasjonale samarbeidsprogrammer som f.eks. Horisont Europa og DIGITAL?
- c) Hvor stor andel av de fremtidige tungregneressursene bør lokaliseres i Norge ut fra nasjonale behov for sikkerhet og beredskap?
- d) Hvordan kan man dele på tungregneressurser på ulike graderingsnivåer i et nasjonalt perspektiv?
- e) Hva er kostnadsanslagene forbundet med punktene over, fordelt på kostnader til investeringer og drift?

Besvarelse

Kunnskapsdepartementet ber om at det utarbeides en rapport som sendes til postmottak@kd.dep.no innen 2. august 2024. Et foreløpig svar på del 2, spørsmål a og e, bes oversendt som egen forsendelse innen arbeidsdagens slutt mandag 6. mai 2024.

Alle tekniske begreper i tekst, tabeller og figurer skal være på norsk, men veletablerte engelske forkortelser kan også benyttes der det er nødvendig.

Finansiering av virksomhetskostnader

Vi viser til ordinært tildelingsbrev for 2024, hvor kap. 285, post 55 ble økt for å sikre midler til nødvendige utredninger. Det legges til grunn at kostnadene for oppdraget i dette brevet dekkes innenfor denne bevilgningen.

Med hilsen

Anne Line Wold (e.f.)
ekspedisjonssjef

Finn-Hugo Markussen
fagdirektør

Dokumentet er elektronisk signert og har derfor ikke håndskrevne signaturer

Kopi

Digitaliserings- og forvaltningsdepartementet
Forsvarsdepartementet
Kultur- og likestillingsdepartementet
Nærings- og fiskeridepartementet

Vedlegg 2: Spørsmål fra spørreundersøkelsene og institusjoner spurt

UoH mottakere av spørreskjema (10)

Svar (8): UiO, UiB, NTNU, UiT, UiS, NMBU, OsloMet, UiA

Ikke svart: USN, Nord Uni.

Spørsmål:

- A. Hva har dere av egne tungregnerressurser i dag? (Vi er ute etter kapasitet av en viss størrelse/mengde (større og nyere enn (>V100) 16GPUer eller 3000 CPU'er) og som kan brukes sammen slik at de utgjør en betydelig ressurs for et forskermiljø/senter eller andre, enten interne brukere eller eksterne (tilhørende utenfor institusjonen). Et eksempel er Idun på NTNU.)
- B. Hvis aktuelt, kan dere si noe om fordelingen mellom tradisjonell tungregning og KI-relatert kapasitet?
- C. Hvem er dagens brukere av disse tungregnerressursene?
- D. Deltar dere i nasjonal/internasjonal samarbeid om tungregnerressurser eller kjøper tilgang nasjonalt?
- E. Har dere brukere fra industri/næringsliv? Hvor stor andel av totale brukere utgjør brukere fra industri/næringsliv, når det gjelder tradisjonell tungregning? Hvor stor andel av totale brukere utgjør brukere fra industri/næringsliv, når det gjelder KI-relatert tungregning?
- F. Har dere brukere fra offentlig sektor? Hvor stor andel av totale brukere utgjør brukere fra offentlig sektor, når det gjelder tradisjonell tungregning? Hvor stor andel av totale brukere utgjør brukere fra offentlig sektor, når det gjelder KI-relatert tungregning?
- G. Kan dere anslå behovet for tungregnekraft ved din institusjon når det gjelder kapasitet til tradisjonelle/KI-relaterte beregningsoppgaver?
- H. Har dere noen synspunkter på hvordan det kan skapes skikkelig driv og innovasjon rundt KI i Norge, innen 5 år? Gjerne si noe om muligheter som ligger i samarbeid mellom FoU-institusjoner, offentlig sektor og næringsliv.

Universiteter bruker HPC ressurser til forskning fra følgende fag/tema: IKT, data- og elektroteknologi, kvanteberegning, bygg og materialteknologi, konstruksjonsteknikk, samfunnsforskning, livsvitenskap, matematikk og fysikk, astrofysikk, kjemi, biokjemi, geofag, medisin, biologi, bioinformatikk, miljøvitenskap og naturforvaltning, bærekraftig bruk av naturressurser, bioteknologi og matvitenskap, biostatistikk og –informatikk.

Institutt mottakere av spørreskjema (22)

Svar institutt (22): Sintef, Norce, Nofima, Nasjonalbiblioteket, Norsk polarinstitutt (NPI), Meteorologisk institutt(MET), Folkehelseinstituttet (FHI), Norges Geologiske Undersøkelse (NGU), Statistisk Sentralbyrå (SSB), Simula, Norsk Institutt for Luftforskning (NILU), Norsk regnesentral (NR), Transportøkonomisk Institutt (TØI), Norsk Institutt for Vannforskning (NIVA), CICERO Senter for klimaforskning, Norsk institutt for bioøkonomi (NIBIO), Norges Geotekniske Institutt (NGI), Veterinærinstitutt, NORSAR, Forsvarets forskningsinstitutt (FFI), Noregs vassdrags- og energidirektorat (NVE), Norges geologiske undersøkelse (NGU)

Spørsmål:

- A. Hva slags norskeid offentlig og privat tungregnekraft er i dag tilgjengelig i Norge til - tradisjonelle tungregneoppgaver? - kunstig intelligens? (Vi er ute etter kapasitet av en viss størrelse/mengde (større og nyere enn (>V100) 16GPUer eller 3000 CPU'er) og som kan brukes sammen slik at de utgjør en betydelig ressurs for et forskermiljø/senter eller andre, enten interne brukere eller eksterne (tilhørende utenfor institusjonen). Et eksempel er Idun på NTNU.)
- B. Hvilken tilgang har norske forsknings- og innovasjonsmiljøer, inkludert næringslivet, i dag til forskjellige typer tungregnekraft gjennom internasjonale forsknings- og innovasjonsprogrammer?
- C. Kan dere si noe om utviklingen de siste 3-5 årene når det gjelder bruk av tungregnekraft i prosjekter dere har, for/med næringsliv og for/med offentlig sektor?
- D. Kan dere si noe om utviklingen de siste 3-5 årene når det gjelder behov for tradisjonell tungregning versus behov for tungregnekraft til kunstig intelligens?
- E. Leverer dere tjenester av samfunnskritisk betydning som krever bruk av tungregneressurser? I hvilken grad er disse tungregneressursene norskeide og/eller lokalisert i Norge?
- F. Kan dere anslå behovet for tungregnekraft ved deres institusjon når det gjelder kapasitet til tradisjonelle og KI-relaterte beregningsoppgaver?
- G. Fra instituttsektorens perspektiv: har dere noen synspunkter på hvordan Norge skal lykkes med forskning, innovasjon og verdiskapning med KI i Norge, innen 5 år? Gjerne si noe om muligheter som ligger i samarbeid mellom FoU-institusjoner, offentlig sektor og industri/næringsliv.
- H. Fra instituttsektorens perspektiv: har dere noen synspunkter på hvor stor andel av fremtidige, nasjonale tungregneressurser som bør ligge i Norge, med tanke på nasjonal sikkerhet og beredskap?

Instituttene bruker HPC ressurser til forskning fra følgende fag/tema: klimaforskning, atmosfæreforskning, meteorologi, seismologi, hav- og klimamodellering, fjernmåling, teknologi, informasjonsteknologi og anvendt matematikk, forsvar, transport, epidemiologi, bioinformatikk, statistikk, metabolomikk, patologi, genetikk, bildeanalyse, marine områder, karbonkjemi, miljøanalyser, hydrodynamikk, prosessindustri, energi, energisektoren og prosessindustri innen akvatisk miljø.

NHO medlemsundersøkelse

Svar: 200 Abelia medlemmer

Spørsmål:

- A. Benytter din virksomhet tungregning (high-performance computing) i dag?
- B. Planlegger din virksomhet å benytte tungregning de neste fem årene?
- C. Hvordan tror du behovet for tungregningskapasitet i din virksomhet vil endre seg de neste fem årene?

Vedlegg 3: Nivået på investeringer i tungregnekraft hos våre nordiske naboer og relevante EU-land

Alle kostnader er oppgitt i millioner kroner (MNOK) basert på dagens kurser.

Finland

Investeringer gjennom EuroHPC

Finland og CSC¹ er vertskap for den nåværende raskeste superdatamaskinen i Europa, LUMI². I 2019 investerte Finland 500 MNOK i LUMI. Finland har annonsert 2 600 MNOK til investering og drift over fire år (2025-2028) for den neste LUMI-maskinen. LUMI blir også brukt av næringslivet i Finland.

Andre HPC-investeringer

Finland sine nasjonale maskiner er større og kraftigere enn de vi har tilgjengelig i Norge. CSC oppgir at det er budsjettet for 140 MNOK i 2024 i nasjonale investeringer. Dette tallet fordeler seg cirka 50/50 mellom investering/oppgradering og drift. I tillegg mottar CSC ca. 70 MNOK årlig fra Departementet for utdanning og kultur (OKM) til avanserte brukerstøttetjenester for vitenskap og KI-spesifikk metodologi.

Fra 2025 økes det årlige budsjettet til 176 MNOK. Totalt sett vil de investere 470 MNOK i nasjonal kapasitet i den neste femårsperioden. Driftsbudsjettet inkluderer også administrativ drift av superdatamaskiner og skytjenester for forskning. Nasjonale tungregnesystemer må betjene alle vitenskapsområder, men siden LUMI kom i drift, er systemene profilert slik at alle de mest krevende beregningene kjøres på LUMI, og nasjonale systemer betjener mer mellomstore tungregnings arbeidsbelastninger. Det nasjonale systemet har også et mye bredere utvalg av støttet vitenskapelig anvendelse tilgjengelig enn LUMI. De opprettholder også en viss KI-kapasitet i nasjonale systemer, men for tiden kjøres nesten alle KI-arbeidsbelastninger på LUMI.

Økning i HPC-investeringer

Tallene nevnt ovenfor inkluderer en opptapping i Finlands HPC-investeringer. I tillegg har Finland innført en ny finansieringsmodell som innebærer årlige investeringsfond framfor et engangsbetrag

¹ [CSC – IT Center for Science](#) er et finsk kompetansesenter innen informasjonsteknologi eid av den finske staten og høyere utdanningsinstitusjoner.

² [LUMI](#) er den første av en ny klasse av superdatamaskiner satt opp av et konsortium av land i Europa. LUMI er rettet mot KI- og HPC-arbeidsbelastninger som kan dra nytte av GPU-akseleratorer. Den er i tillegg på topp 10 listen over grønne og energi-effektive superdatamaskiner i verden.

hvert 5-6 år. Dette gjør at de kan utvikle nasjonale tungregnetjenester mer strategisk og dynamisk tilpasse seg endrede brukerbehov.

Sverige

Investeringer gjennom EuroHPC

Gjennom EuroHPC skal Sverige investere ca. 850 MNOK i perioden 2025-2031 til den neste tungregnemaskinen, Arrhenius³. Linköping universitet, som er vertskap for NAISS⁴, skal eie maskinen sammen med EuroHPC. Det er en ambisjon at Arrhenius skal bidra til å gi sikker tilgang til HPC-ressurser for næringslivet. Vetenskapsrådet (VR) i Sverige har eget budsjett på 30 MNOK/år for EuroHPC, inkludert 10 MNOK som er satt av til LUMI. Resterende 20 MNOK er for å delta i EuroHPC-prosjekter.

Andre HPC-investeringer

Sveriges største maskin Berzelius⁵ er drøyt 50 ganger kraftigere enn Norges kraftigste maskin Betzy. NAISS (Sigma2 sin søsterorganisasjon) har et årlig budsjett på 120 MNOK/år som skal dekke både drift og investering av nasjonale ressurser. Inkludert i dette budsjettet er 40MNOK som skal brukes på en egen HPC-maskin for sensitive data. Det skal også investeres ca. 25-30 MNOK i egen regnekraft til EISCAT-3D⁶ ved Sunets datasenter, DC Orion i Kalix.

I tillegg har Wallenberg stiftelsen⁷ bidratt til investeringer i KI og HPC siden 2015, og blant annet med et stort bidrag til Berzelius. De bidrar nå med 50 MNOK/år til en maskin som er reservert for KI.

Økning i HPC-investeringer

Vetenskapsrådet har besluttet å gi 100 MNOK ekstra til Arrhenius og NAISS har avsatt 160 MNOK og stilt garanti for resterende 65% medfinansiering i Arrhenius, (35 prosent dekkes av EuroHPC).

I september kommer den nye forskningsproposisjonen (Riksdagens 5-åriga satsning på forskning). VR har bedt om 135 MNOK/år for HPC og 100 MNOK/år for kvantteknologi.

³ [Sweden will host a new EuroHPC supercomputer - Swedish Research Council \(vr.se\)](#)

⁴ [NAISS | The National Academic Infrastructure for Supercomputing in Sweden](#)

⁵ [Sveriges snabbaste superdator för AI är invigd - Linköpings universitet \(liu.se\)](#)

⁶ EISCAT 3D skal bli en internasjonal forskningsinfrastruktur som bruker radarobservasjoner og inkoherente spredningsteknikken for studier av atmosfæren og nærjordmiljøet over den Fenno-Skandinaviske Arktis, samt for støtte til solsystem- og radioastronomivitenskap ([EISCAT 3D - EISCAT Scientific Association](#))

⁷ [Wallenberg AI, Autonomous Systems and Software Program | WASP \(wasp-sweden.org\)](#)

Danmark

DeiC⁸ investerte ca. 63 MNOK i 2023 i tungregning. For 2024 forventes omtrent den samme størrelsen på investeringen, i tillegg er det opprettet en reserve på 12 MNOK. Dette inkluderer investeringer i EuroHPC og LUMI. Det er utfordrende å si noe nøyaktig om Danmarks totale investeringsnivå innenfor tungregning da de ikke har en samlet investeringsstrategi, men en desentralisert modell der universitetene gjør selvstendige investeringer etter egne behov.

I Danmark kom det en KI-strategi i 2019⁹ hvor det foreslås å starte et pilotprosjekt i form av en investeringspott på 31,5 MNOK over fire år, rettet mot virksomheter med KI-baserte forretningsmodeller. Forutsetningen for dette er 50 prosent medfinansiering fra privat sektor, tilsvarende en total investeringspott på 63 MNOK. Det er det danske vekstfondet som skal forvalte fondet. Et offentlig-privat samarbeid bygger superdatamaskinen Gefion, som vil bli blant verdens kraftigste, en investering på ca. 1100 MNOK¹⁰.

Novo Nordisk Fonden og Danmarks Eksport- og Investeringsfond (Eifo) har inngått samarbeid med Nvidia om å utvikle nå supercomputer for forskning og innovasjon. Kostnaden er anslått å være rundt 1 000 MNOK. Novo Nordisk Fonden har opprettet et datterselskap som skal eie og drifte maskinen. Forskere og bedrifter fra offentlig og privat sektor vil kunne få tilgang.

Tyskland

Investeringer gjennom EuroHPC

Tysklands føderale regjering følger en HPC-strategi i tråd med EuroHPC. Den største investeringen framover er JUPITER-maskinen, som vil være den første i EU med exascale-ytelse¹¹. Denne maskinen vil fra i 2025 tilby en enestående GPU-kapasitet for KI-relaterte beregninger. Investeringene i maskinen er på 5 900 MNOK, der halvparten finansieres av Tyskland og resterende av EuroHPC.

Andre HPC-investeringer

I tillegg har Tyskland en nasjonale regnearbeid fordelt på tre nivåer¹²:

Nivå 1: Tyskland med Gauss Center for Supercomputing (GCS)¹³ har en internasjonalt konkurransedyktig, koordinert infrastruktur av høyeste ytelsesklasse, med tre datasentrene. Den totale

⁸ Danish e-infrastructure Consortium er det nasjonale samarbeid med og mellom universitetene om levering av digital forskningsinfrastruktur.

⁹ [The Danish National Strategy for Artificial Intelligence \(digst.dk\)](https://www.digst.dk)

¹⁰ [Denmark to build one of the world's most powerful AI supercomputers, accelerating solutions to societal challenges \(prnewswire.com\)](https://www.prnewswire.com). 150 MNOK finansieres av The Export and Investment Fund of Denmark (EIFO) og 950 MNOK av Novo Nordisk Foundation.

¹¹ [Exascale-datamaskiner kan utføre mer enn 1,000,000,000,000,000,000 \(en eksa\) flyttallsoperasjoner per sekund \(FLOPS\)](#)

¹² [Hoch- und Höchstleistungsrechnen für das digitale Zeitalter – Elektronikforschung](#)

¹³ [Home: Gauss Centre for Supercomputing e.V. \(gauss-centre.eu\)](https://gauss-centre.eu)

ytelsen til de tre HPC-systemene i datasentrene er over 130 PetaFLOPS (per november 2020). KI er en integrert del av superdatamaskinoppdraget, men maskinene er ikke utelukkende dedikert til KI. Infrastrukturen ved de tre GCS-sentrene vil bli utvidet vekselvis i en koordinert smart skaleringsstrategi.

Nivå 2: består i dag av tolv etablerte HPC-sentre på tvers av regioner, med datamaskiner med høy ytelse ved forskningsinstitusjoner og universiteter. Med det føderale statsprogrammet for National High Performance Computing (NHR) vil det for første gang bli etablert et landsdekkende koordinert system av datasentrene ved universitetene, for bruk av universitetsmedlemmer.

Nivå 3: består hovedsakelig av regionale datamaskiner som betjener lokale behov med lavere ytelseskrav. Disse er tilgjengelige og drives på de fleste universiteter og forskningsinstitusjoner. Siden 2008 har Gauss Alliance (GA), en ideell forening for vitenskap og forskning innen HPC, slått sammen det tyske forskningsnettverket med å få samlet de tre sentrene på nivå 1, 16 sentre på nivå 2 og 3. Sammen driver medlemmene av dette nettverket rundt 40 HPC-systemer over hele Tyskland med en nåværende ytelse på rundt 190 PetaFLOPS (per november 2020). GA er også dedikert rådgivning, opplæring og støttetjenester.

Data

I tillegg til tilgang til stadig kraftigere superdatamaskiner, en annen viktig byggestein for å drive med topp-forskning er tilgang til kunnskapslagre, dataarkiver og forskningsresultater. Datasett fra vitenskap og forskning indekseres systematisk og gjøres tilgjengelig gjennom etableringen av National Research Data Infrastructure (NFDI) og European Open Science Cloud (EOSC).

Nederland

Investeringer gjennom EuroHPC

Nederland har en ordning på plass for medfinansiering av EuroHPC-prosjekter som årlig er på ca. 90 MNOK og har i tillegg nylig vedtatt å bruke ca. 120 MNOK på deltakelse i EuroHPC-tungregneanlegg. Nederland ble nylig med i LUMI-konsortiet.

Andre HPC-investeringer

Regjeringen i Nederland investerer årlig strukturerende midler til HPC-anlegg gjennom sitt forskningsråd på ca. 25 MNOK. Dette er dedikerte midler til oppgradering og erstatning av nasjonale HPC-anlegg. I tillegg gjøres noen investeringer til blant annet HPC-anlegg, lagring og utdanning gjennom Sigma2 sin søsterorganisasjon, SURF¹⁴, og deres medlemmer. Medlemmer i SURF består av FoU-institusjoner. SURF skal utarbeide en nederlandsk HPC-strategi for de neste årene.

¹⁴ [SURF is the collaborative organisation for IT in Dutch education and research | SURF.nl](https://www.surf.nl)

Departementene har også mottatt flere henvendelser angående finansiering av andre typer initiativer, som behov knyttet til KI-relaterte tungregneressurser, uten at vi har fått oppgitt noen summer.

Økning i HPC-investeringer

Nederland har ikke planer om å øke investeringene fremover, men viser til nylige vedtatte summer for å investere i EuroHPC-anlegg og dekke medfinansieringen i EuroHPC-prosjekter.

Frankrike

Investeringer gjennom EuroHPC

Den andre exascale-datamaskinen i EuroHPC-samarbeidet i tillegg til Jupiter i Tyskland er under planlegging i Frankrike. Med et totalbudsjett i overkant av 6 000 MNOK vil maskinen bli delfinansiert av EuroHPC, Digital Europe Programme (DIGITAL) og bidrag fra Frankrike og Nederland.

Andre HPC-investeringer

Det estimeres at GENCI¹⁵, søster organisasjon til Sigma2 i Frankrike, har et årlig budsjett (investeringer og drift) for tre nasjonale tungregningssentre på 570 MNOK. I tillegg er investeringen i utvidelsen av ett av sentrene samt og exascale-prosjektet hovedsakelig finansiert av ressurser utenfor det årlige budsjettet til GENCI, som for eksempel France2030¹⁶.

¹⁵ [GENCI | Grand équipement national de calcul intensif](#)

¹⁶ [France 2030 | ANR](#)