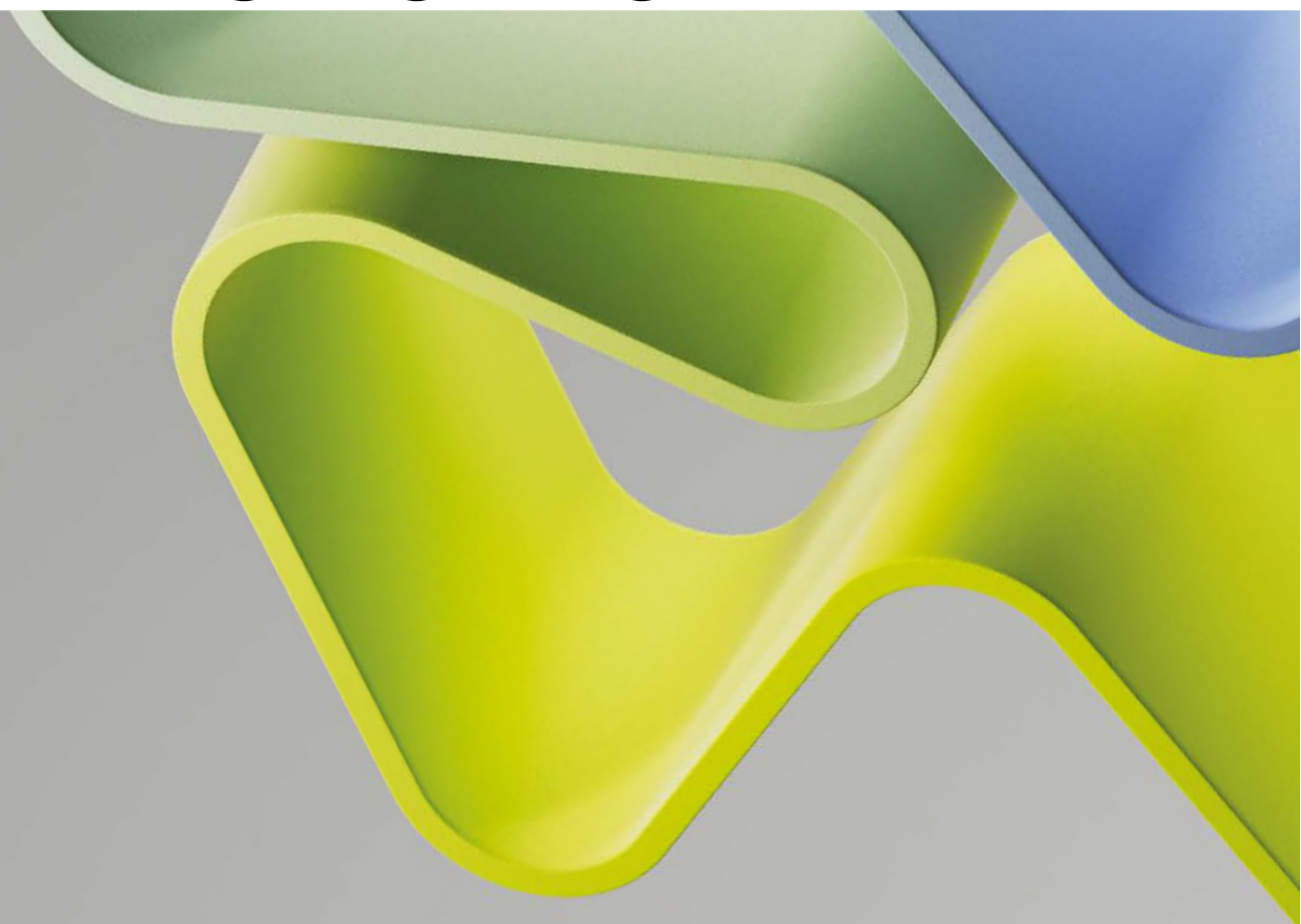


Rapport til Kunnskapsdepartementet

Konseptvalgutredning for tungregning



1	Sammendrag	5
2	Introduksjon	10
2.1	Bakgrunn	10
2.2	Om konseptvalgutredning	10
2.3	Mandat	12
2.4	Sentrale begreper	12
3	Dagens situasjon	15
3.1	Leveransekjeden for tungregneinfrastruktur	15
3.2	Tungregneinfrastruktur for forskning	18
3.3	Forventet utvikling av infrastruktur og behov for tungregning og kunstig intelligens/maskinl�ring	30
4	Problembeskrivelse	36
4.1	Manglende kapasitet for � dekke et �kende behov	36
4.2	Dagens finansieringsmodell gir ikke n�dvendig forutsigbarhet og tilpasningsevne	37
4.3	Dagens organisering m� tilpasses framtidige behov for kapasitet og kompetanse	39
4.4	N�ringsliv	40
4.5	Oppsummering av problembeskrivelsen	40
5	Behovsanalyse	42
5.1	Normative behov	42
5.2	Interessentenes behov	46
5.3	Ettersp�rsebaserte behov	61
5.4	Prosjektutl�sende behov	65
6	Strategiske m�l	67
6.1	Samfunnsm�l	67
6.2	Effektm�l	67
6.3	Resultatm�l	69
7	Rammebetingelser og andre kriterier for konseptvalg	70
8	Mulighetsstudie	75
8.1	Mulighetsrommet	75
8.2	Konsepter	82
8.3	Vurdering av konseptene mot rammebetingelser og utvalgte vurderingskriterier	92
8.4	Oppsummering	95
9	Alternativanalyse	96
9.1	Beskrivelse av konsepter	96
9.2	Rammeverk for vurderinger av konseptene	103
9.3	Samfunns�konomiske virkninger	104
9.4	Forutsetninger til den samfunns�konomiske analysen	106
9.5	Prissatte virkninger	108
9.6	Ikke-prissatte virkninger	112
9.7	Usikkerhetsanalyse	123
9.8	Realopsjoner	127
9.9	Fordelingsvirkninger	127
9.10	Gjennomf�ringsrisiko	128

10	Anbefalt konsept	129
11	Føringer for forprosjekt	132
11.1	Videreføring av utredningsprosjektet	132
11.2	Planlegging og gjennomføring	133
11.3	Suksessfaktorer og fallgruver	137
11.4	Forslag til risikoreduserende tiltak	137
<hr/>		
12	Referanser	140
13	Vedleggsoversikt	142

1 Sammendrag

Introduksjon

Tungregning, også kjent som High Performance Computing (HPC), innebærer bruk av kraftige datamaskiner og systemer for å utføre komplekse beregninger og analyser på store datamengder, noe som er umulig med vanlige datamaskiner.

Det stadig 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 for å videreutvikle og opprettholde forskning i verdensklasse og flytte forskningsfronten.

Tilgang til tungregning er avgjørende for offentlig forvaltning for å kunne levere samfunnskritiske tjenester som helse, sikkerhet og beredskap, kritisk infrastruktur, miljøovervåkning m.m., da det muliggjør avanserte analyser og simuleringer som sikrer effektiv og pålitelig tjenesteyting. Dette bidrar til å opprettholde samfunnets grunnleggende funksjoner og beskytte innbyggerne.

Tilgang på regnekraft og grunnmodeller for kunstig intelligens (KI) 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. Alle som vil utvikle, tilpasse og drifte store språkmodeller eller som vil bygge nye generative KI-tjenester på toppen av disse, f.eks. språkbaserte KI-verktøy for mer effektive tjenester og produksjon, trenger tilgang på regnekraft.

Utfordringer

Nåværende kapasitet innen tungregning er ikke dekkende for dagens kjente fremtidige 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 akterut sammenlignet med våre nordiske naboer. Sverige og Finland har allerede besluttet å gjøre større investeringer i nasjonale tungregneressurser og gjennom EuroHPC for å imøtekomme det sterkt økende behovet.

Det er knyttet stor usikkerhet til fremtidig behov for tungregneressurser, både for tradisjonelle tungregneoppgaver og spesielt for regneressurser optimalisert for KI. Selv om det er stor usikkerhet har alle aktørene vi har hentet informasjon fra pekt på at behovet er stort og raskt voksende. Generativ KI har generelt bidratt til en eksplosiv økning i behov for tungregning i både offentlig og privat sektor.

Det er avgjørende at investeringene i tungregneinfrastruktur 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.

Behov

Med bakgrunn i utfordringene identifisert i konseptvalgutregningen er det behov for å øke investeringene i tungregning. Det innebærer:

- Det er behov for nasjonal infrastruktur som tilrettelegger for og gir tilstrekkelig kapasitet til tungregning for forskning, utvikling, innovasjon, offentlig forvaltning og næringsliv.
- Infrastrukturen gjelder ikke kun tilgjengeliggjøring av kapasitet, men omfatter anskaffelse av ressurser, drift, videreutvikling, kompetanseoppbygging og -deling, samt vanlig og avansert brukerstøtte.
- Behovene blant de ulike brukergruppene er svært ulikt og sammensatt, både i forhold til regnebehov, datamengde, behov for oppetid og grad av sensitivitet. Videre vil behovene endre seg i takt med en svært rask teknologiske utvikling. Investeringene må derfor tilrettelegge for stor fleksibilitet og muligheten til å ta raske beslutninger for videre utvikling. Dette er nødvendig for å ha et tilbud som teknologisk kan oppdateres kontinuerlig for å være relevant for dagens og fremtidens brukere.

- Med flere brukere involvert er det nødvendig med forutsigbare kriterier for tildeling av tungregnerressurser, samt behov for å ha gode mekanismer for deling av regneressursene, for å sikre best mulig ressurseffektivitet.
- For god utnyttelse av kapasitet og kompetanse er det behov for en god kopling både mellom lokale og nasjonale ressurser og mellom de nasjonale og de europeiske ressursene Norge investerer i.

Konseptvalgutredning

Det er utredet ulike konsepter for investeringer i en nasjonal infrastruktur for tungregning, en såkalt konseptvalgutredning, etter oppdrag fra Kunnskapsdepartementet. En konseptvalgutredning er en systematisk prosess for å vurdere ulike konsepter eller løsninger for et prosjekt, med mål om å finne det beste alternativet for videre utvikling. Den gir et beslutningsgrunnlag ved å analysere utfordringer, behov, gevinster, kostnader og risiko ved ulike konsepter. En konseptvalgutredning er en del av statens prosjektmodell for å sikre grundig planlegging og kvalitetssikring av store investeringsprosjekter.

Anbefalt konsept for nasjonal infrastruktur for tungregning

Basert på den samfunnsøkonomiske analysen i utredningen, anbefaler vi at man på sikt beveger seg mot konsept K3 «Tverrsektoriell Sigma3» som et fremtidig mål bilde for nasjonal tungregneinfrastruktur.

I konsept K3 vil både behovene for tungregning innen forskningssektoren og sentrale deler av offentlig sektor dekkes av én leverandør. Leverandøren er en stegvis videreutvikling av dagens Sigma2 slik at leverandøren kan å bli en tverrsektoriell leverandør av tungregning («Sigma3»). Leverandøren vil dekke behovene for tungregning til forskningsdrevet og forvaltningsdrevet FoU, samt tungregning for utvikling, tilpasning og forbedringer av samfunnskritiske tjenester. I tillegg kan leverandøren levere noe tungregnekapasitet, støtte- og kompetansetjenester til næringslivet innenfor statsstøttereguleringen. Samtidig må hver offentlig virksomhet selv skaffe til veie tungregning til å drifte samfunnskritiske tjenester (herunder 24/7-oppetid) og tungregning til bruk og tilpasning av øvrige offentlige tjenester, inkludert kunstig intelligens. Leverandør står fritt til å inngå og endre avtale om kjøp av kapasitet hos EuroHPC og andre samarbeidspartnere, og til å utforske muligheten til å inngå samarbeid med det private leverandørmarkedet for tungregning.

Den nye leverandøren vil motta finansiering til investeringer og grunnleggende driftsstøtte på vegne av departementsfellesskapet som koordineres gjennom Kunnskapsdepartementet, mens bruk av tungregning er helt eller delvis brukerfinansiert. Vi legger til grunn at differensierte modeller for brukerfinansiering må utvikles. Midler til bruken av tungregning i leveranse av samfunnskritiske tjenester kanaliseres til den enkelte forvaltningsområde i årlige budsjettprosesser basert på tverrsektorielle vurderinger av fordeling av midler.

Nyttevirkninger ved implementering av anbefalt konsept

Det anbefalte konseptet K3 kan bidra til en bærekraftig samfunnsutvikling ved å fremme datadrevet innovasjon og effektivisering i både forskning, offentlig forvaltning og privat sektor, og bidra til at Norge blir bedre godt rustet til å møte fremtidens teknologiske og samfunnsmessige utfordringer.

K3 skal gi betydelige samfunnsøkonomiske gevinster ved å tilby en robust og fleksibel infrastruktur for tungregning som dekker både forsknings- og forvaltningsdrevet FoU. Ved å sikre forskningsmiljøer tilgang til nødvendig regnekraft, kan Norge styrke sin posisjon som en ledende kunnskapsnasjon. Dette er avgjørende for å flytte forskningsfronten, fremme innovasjon og tiltrekke seg internasjonale talenter. Forskningsmiljøene vil kunne utføre mer avanserte simuleringer og analyser, noe som vil bidra til å utvikle banebrytende løsninger på komplekse samfunnsutfordringer. Økt tilgang til tungregning kan også muliggjøre mer tverrfaglig forskning og samarbeid, noe som kan føre til nye oppdagelser og teknologiske fremskritt. Dette vil ikke bare styrke Norges konkurransevne, men også bidra til å løse globale utfordringer innen områder som klima, helse og energi.

Offentlige virksomheter vil dra nytte av denne infrastrukturen ved å kunne utvikle og forbedre både samfunnskritiske og øvrige tjenester. For samfunnskritiske tjenester kan det innebære økt

sikkerhet og effektivitet i tjenesteleveransen, inkludert bedre beredskap, optimal ressursallokering og raskere respons i krisesituasjoner. Innenfor andre offentlige tjenester kan tungregning muliggjøre mer presise analyser og bedre beslutningsstøtte, noe som kan føre til mer effektiv forvaltning og høyere kvalitet på tjenestene. Dette inkluderer alt fra automatisering av administrative prosesser til personalisering av tjenester for innbyggerne. Ved å utnytte tungregning kan offentlige virksomheter også forbedre koordineringen mellom ulike sektorer, noe som gir en mer helhetlig og effektiv offentlig tjenesteyting.

Videre vil konseptet legge til rette for økt verdiskaping i næringslivet ved å tilby støtte- og kompetansetjenester innenfor statsstøttereguleringen. Dette kan stimulere til utvikling av nye teknologier og tjenester, og gi små og mellomstore bedrifter muligheten til å konkurrere på et globalt nivå.

K3 kan også fremme samarbeid på tvers av forskning, offentlig forvaltning og næringslivet, noe som gir bedre ressursutnyttelse og styrker Norges teknologiske suverenitet. Ved å integrere data fra ulike kilder kan offentlige og private aktører få et mer helhetlig bilde av utfordringer og muligheter, noe som forbedrer beslutningsgrunnlaget og effektiviteten i tjenesteleveranser og produktutvikling.

Kostnader ved implementering av anbefalt konsept

Konseptet innebærer en total investering på 3,4 mrd. kroner i perioden 2026 til 2030. Dette inkluderer et årlig driftsbudsjett mellom 150 og 185 mill. kroner, beregnet til totalt 850 mill. kroner i analyseperioden. I tillegg til investerings- og driftskostnader vil realiseringen av konseptet ha negative miljøkonsekvenser, særlig i form av økt strømforbruk og utbygging av fysisk infrastruktur.

Vi vil fremheve at det er svært stor usikkerhet knyttet til kostnadsberegningene og forutsetningene som ligger til grunn, herunder behovet for tungregning (både GPUer og CPUer). Usikkerheten er vurdert til å være høy og «venstreskjev», dvs. det kan potensielt bli vesentlig høyere kostnader knyttet til realiseringen av alle konsepter i analysen.

Gjennomføringsrisiko ved implementering av anbefalt konsept

Vi har vurdert gjennomføringsrisikoen i K3 som høy. Årsaken er at leverandøren skal levere tungregnetjenester til flere sektorer, som innebærer at dette er et større tverrsektorielt digitaliseringsprosjekt i offentlig sektor. Erfaringer fra flere andre utviklingsprosjekter tilsier at dette kan være utfordrende. Dette underbygger behovet for en tverrsektoriell tilnærming i de kommende fasene av prosjektet.

Det er avgjørende å sikre at alle brukergrupper over tid får dekket sine behov. En videreutvikling av Sigma2 til «tverrsektoriell Sigma3» hvor tungregnebehovet til flere sektorer dekkes, må ikke gå på bekostning av behovene til forskningsorganisasjonene som får dekket sine behov gjennom Sigma2 i dag. Behovene er ulike mellom brukergruppene og de vil kunne endre seg over tid. Det er derfor en kritisk suksessfaktor at risikoreduerende tiltak identifiseres og adresseres i forprosjektfasen.

Utvikling av leverandøren i anbefalt konsept

Infrastrukturen for tungregning skal så langt som mulig bygge på eksisterende systemer og strukturer. Dette er sentralt for å sikre at nasjonal kompetanse innenfor tungregning ikke forvitrer, men utnyttes på en god måte, og at man kan redusere behovet for nye investeringer og minimere implementeringstiden som bidrar til en kostnadseffektiv utbygging av en nasjonal tungregneinfrastruktur. I konsept K3 innebærer dette en utvidelse og omstrukturering av Sigma2, slik at leverandøren kan bli en tverrsektoriell leverandør av tungregning. Siden en konseptvalgutredning er en strategisk utredning som setter retning for fremtiden, må organiseringen og struktureringen av leverandøren detaljeres i forprosjektet og dokumenteres i det sentrale styringsdokumentet (SSD).

Stegvis utvikling i leveranse av tungregnetjenester i anbefalt konsept

Utviklingen av leverandøren for tungregning må skje gradvis over tid for å redusere gjennomføringsrisiko og ta hensyn til den raske teknologiske utviklingen innenfor tungregning og kunstig intelligens. Den raske utviklingen gjør at behovene for tungregning fort kan endres, og en gradvis utbygging over tid vil sikre at man benytter seg av siste teknologiske nyvinninger

og oppdatert kunnskap om behovene. Det kan bidra til å redusere usikkerhet og hindre feilinvesteringer i infrastruktur. Det er også et behov for å gå opp grenseskiller mot andre utviklingsløp, herunder behandling av gradert informasjon og håndtering av sensitive data.

Overordnet har vi vurdert følgende steg i utviklingen av tjenesteleveranse til leverandøren:

- Steg 1: Leverandøren må først dekke behovet for tungregning til utvikling og testing av språkmodeller og generative nasjonale KI-modeller.
- Steg 2: Gjennomføre kost/nytte-vurderinger for hvert forvaltningsområde og ta stilling til om det enkelte område skal få dekket sine behov for tungregning hos leverandøren eller gjennom andre løsninger/leverandører. Helsesektoren, som er en omfattende og kompleks sektor med særskilte krav til personvern, vil kreve en særskilt utredning.

Disse stegene må detaljeres ytterligere i forprosjektfasen av prosjektet.

Resultatene fra den samfunnsøkonomiske analysen

I den samfunnsøkonomiske analysen i utredningen ble de utvalgte konseptene K2 «Sigma2 pluss» og K3 «Tverrsektoriell Sigma3» vurdert nærmere. Det vil si at det ble vurdert kostnads- og nyttevirkinger ved utbyggingen av en nasjonal infrastruktur for tungregning, samt en kvalitativ vurdering av gjennomføringsrisiko og realopsjoner ved konseptene. Basert på en samlet vurdering av de samfunnsøkonomiske virkningene, gjennomføringsrisiko og realopsjoner er K3 «Tverrsektoriell Sigma3» rangert best av konseptene, jf. Tabell 1.1.

Tabell 1.1 Resultater fra samfunnsøkonomisk analyse av utvalgte konsepter

	Null- alternativet	K2 «Sigma2 pluss»	K3 «Tverrsektoriell Sigma3»
Prissatte virkninger (MNOK, netto nåverdi)			
Investeringskostnad	-170	-2 050	-2 310
Driftskostnad	-610	-1 330	-840
Skattefinansieringskostnad	-160	-680	-640
Sum prissatte virkninger	-940	-4 050	-3 780
Ikke prissatte virkninger			
Mer forskning i internasjonal forskningsfront	-	Middels positiv	Middels positiv
Økt innovasjon og konkurransekraft	-	Meget stor positiv	Meget stor positiv
Forbedret samfunnskritiske tjenester	-	Stor positiv	Meget stor positiv
Forbedret øvrig offentlig tjenesteyting	-	Liten positiv	Liten positiv
Økt verdiskapning	-	Liten positiv	Liten positiv
Negative miljøkonsekvenser	-	Middels negativ	Middels negativ
Risiko og realopsjoner			
Gjennomføringsrisiko	-	Middels	Høy
Realopsjoner	-	Middels	Lav
Rangering basert på samlet vurdering		2	1

Føringer for forprosjekt

Det anbefalte konseptet representerer en utviklingsretning og består av en rekke tiltak som samlet sett skal bringe oss mot et ønsket fremtidig målbilde. Forprosjektet skal etablere et styringsdokument som beskriver hvordan Forskningsrådet og samarbeidspartnere, i og utenfor kunnskapssektoren, skal arbeide sammen for å realisere valgt konsept.

Styringsdokumentet skal beskrive:

- En modell og en prosess for å kartlegge fremtidige behov, samt utarbeide terskelverdier for når det skal investeres i økt kapasitet
- Hvordan sikre at brukergrupper forplikter seg til bruk av kapasitet hos plattformleverandøren
- Finansieringsmodell og hvordan ulik bruk av ulike tjenester skal belastes brukerne.
- Etablere de sentrale prinsippene for hvordan plattformleverandøren skal organiseres.
- Hvordan avansert brukerstøtte skal organiseres fremover og hvordan denne kompetansen skal tilgjengeliggjøres for ulike brukergrupper.
- Hvordan balansen mellom ulike brukergrupper skal ivaretas gjennom styringsmodellen til plattformleverandøren. I tillegg skal man beskrive hvordan kapasitet skal tildeles ulike brukergrupper.

2 Introduksjon

2.1 Bakgrunn

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.

Forskningsmiljøer i Norge har behov for tilstrekkelig tilgang på tungregning og gode støttesystemer for fortsatt å kunne være i forskningsfronten. Flere rapporter¹ og studier peker på banebrytende forskning som en stadig viktigere drivkraft for økt konkurranseevne, og relevant forskningsinfrastruktur er helt avgjørende for at forskningsmiljøer skal hevde seg internasjonalt.

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 samfunnskritiske og -nyttige 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.

2.2 Om konseptvalgutredning

En konseptvalgutredning omfatter et sett av prosesser hvor målet med utredningen er å anbefale et konsept for videre bearbeiding i et forprosjekt. En konseptvalgutredning er med andre ord en strategiprosess der de ulike konseptene som inngår i utredningene representerer ulike strategiske retninger for hvordan de identifiserte behovene skal møtes eller løses. Hvert konsept kan derfor innfri behovene i ulik grad. Noen konsepter vil for eksempel ha et lavt ambisjonsnivå knyttet til endringer fra dagens situasjon. Disse konseptene vil som regel ha lave kostnader, men vil samtidig kanskje ikke møte alle behovene som er identifisert. Motsatt vil et konsept med høyt ambisjonsnivå ha høye kostnader, men samtidig møte de fleste behovene.

En grunnleggende forutsetning for en vellykket konseptvalgutredning er at man har kunnskap om det reelle behovet for tiltaket og hvorfor det er behov for å gjøre tiltaket nå. En grunnleggende feil er at man utreder det konseptet eller løsningen man ønsker seg og i for liten grad hva man egentlig trenger.

Fasene i en konseptvalgutredning er definert i Finansdepartementets rundskriv R-108/23. Rundskrivet beskriver statens prosjektmodell for gjennomføring av store investeringsprosjekter og fastsetter krav til utredning, planlegging og kvalitetssikring av prosjektene. Rundskrivet er en spesifisering av hvordan kravene til utredning, planlegging og styring som følge av utredningsinstruksen og økonomiregelverket skal følges opp i utredning av store statlige

¹ Slik som for eksempel [Draghi-rapporten](#) og [Heitor-rapporten](#).

investeringer. Det angir nærmere blant annet krav til innhold og metode i konseptvalgutredninger. Nedenfor gis en kort beskrivelse av de ulike stegene.

Problembeskrivelsen gir en beskrivelse av dagens situasjon og redegjør for hvilke uløste problemer som foreligger. I tillegg skal det være tydelig hvorfor det offentlige skal iverksette tiltak for å løse problemet. Beskrivelsen av problemet skal blant annet inneholde problemets omfang, hvor alvorlig det er og hvem som er eller blir berørt. Et viktig prinsipp er å ha en nøytral beskrivelse av problemet uten å være løsningsspesifikk – løsningen og ulike tiltak skal først beskrives i mulighetsstudien og alternativanalysen. Problembeskrivelsen inneholder både beskrivelse av dagens situasjon og forventet utvikling. En god problembeskrivelse danner et godt grunnlag for senere beskrivelse av nullalternativet (som benyttes for å vurdere de ulike alternativene).

Behovsanalysen identifiserer aktuelle og konkrete behov vurdert i et samfunnsperspektiv. Gjennom fasen skal det avdekkes om det finnes et prosjektutløsende behov som tilsier at man bør starte planleggingen av et tiltak. Behovet skal være vurdert i et overordnet samfunnsperspektiv. Behovsanalysen inneholder en kartlegging av relevante interessenter og deres behov. Behovsanalysen deles iht. metodikken i (1) normative behov, (2) etterspørselsbaserte behov og (3) interessentbaserte behov. De strategiske målene gir retning for tiltaket og er derfor sentrale for utviklingen av konseptene i senere faser.

De **strategiske målene** defineres på bakgrunn av problembeskrivelsen og behovsanalysen. Formuleringen av de strategiske målene tar utgangspunkt i virkningene av tiltaket og deles inn i samfunns mål og effektmål. Det er også viktig å synliggjøre eventuelle målkonflikter, om tiltaket er i tråd med samfunnets mål og prioriteringer samt relevans opp mot andre tiltak.

Samfunnsmålet beskriver den positive tilstanden tiltaket skal underbygge og er knyttet til tiltakets virkninger for samfunnet. En realisering av samfunnsmålet må delvis kunne tilbakeføres til tiltaket.

Effektmålet er mer direkte knyttet opp til tiltaket og beskriver hvilke virkninger som vil treffe brukerne.

Målene skal være prosjektspesifikke og beskrive relevante egenskaper ved den ønskede tilstanden etter gjennomføring av tiltak. Effektmålene skal være innbyrdes konsistente og ha en tydelig prioritering. Etter avsluttet gjennomføring av tiltaket skal det være mulig å verifisere om effektmålene er oppnådd.

For å danne et godt grunnlag for hvilke konsepter som utarbeides i mulighetsstudien, er det sentralt å utarbeide gode og relevante rammebetingelser for konseptvalget. Rammebetingelsene er et samlet sett av betingelser som skal oppfylles for valg av konseptuell løsning og fremtidig drift av tiltaket.

Rammebetingelsene er sentrale ved at de avgrensner mulighetsrommet for alternative konsepter i mulighetsstudien. Det er derfor viktig at det utarbeides tilstrekkelige rammebetingelser uten at de låser mulighetsrommet og potensielle konsepter for mye.

En bred tilnærming gjør det mulig å identifisere alle vesentlige og aktuelle behov for det aktuelle tiltaket. Analysen må vurdere styrken i de ulike identifiserte behovene og definere hvilket behov som skal legges til grunn for den videre utredningen. I tillegg inneholder analysen identifisering av eventuelle interessekonflikter mellom ulike interessenter og/eller behov. Ved å sammenligne de ulike behovene med dagens situasjon er det mulig å belyse gapene og videre analysere mulighetsrommet for tiltaket.

Mulighetsstudien har som formål å sikre at konseptuelt ulike tiltak videreføres til alternativanalysen. En av de store fallgruvene er å snevre inn mulighetsrommet for tidlig og dermed utarbeide relativt like alternativ. I mulighetsstudien gjøres dette ved å belyse tilstrekkelig mange mulige løsninger (utforsker mulighetsrommet) som kan realisere målene og tilfredsstille rammebetingelsene som er utarbeidet. Mulighetsstudien tar utgangspunkt i allerede identifiserte alternativer. Det er viktig å sikre at det er indre konsistens mellom mulighetsstudien og tidligere arbeid i problembeskrivelsen, behovsanalysen, strategiske mål og rammebetingelser for konseptvalg.

En viktig del av mulighetsstudien inkluderer begrunnelse for hvilke tiltak som ikke videreføres til alternativanalysen. Dette er særdeles viktig for senere forankring og begrunnelse for å unngå omkamper knyttet til anbefalt konsept.

Alternativanalysen tar for seg de ulike konseptene som ble videreført fra mulighetsstudien i en samfunnsøkonomisk analyse. For hvert konsept gjennomføres det en vurdering av kostnader. Analysen skal også inkludere en usikkerhetsanalyse av investeringskostnaden for hvert konsept, som viser forventet verdi for investeringen og sannsynligheten for at kostnaden blir høyere eller lavere enn dette.

I alternativanalysen identifiseres positive og negative nyttevirksomheter av konseptene. Det skilles mellom prissatte og ikke-prissatte nyttevirksomheter. Et viktig prinsipp er at nyttevirksomheter som inngår i analysen er innbyrdes uavhengige og ikke med i analysen av de prissatte virkningene. Dette for å unngå dobbelttelling.

Føringer for forprosjekt beskriver hvordan prosjektet skal videreføres til neste fase i statens prosjektmodell. I forprosjektet skal anbefalt konsept detaljeres for å sikre en vellykket gjennomføring av prosjektet. Føringer for forprosjekt kan eksempelvis være forutsetninger for kostnads- og nytteeffekter, informasjon om usikkerhet og risiko og krav til kompetanse i forprosjektet.

2.3 Mandat

Forskningsrådet mottok utredningsoppdraget fra Kunnskapsdepartementet 4. oktober 2024 der Forskningsrådet ble bedt om å foreta en utredning for investering i tungregning i tråd med statens prosjektmodell, jf. Rundskriv R-108/23. Utredningen skal gjøres med bakgrunn i rapporten «Behov for tungregnekraft for forskning og kunstig intelligens», som Forskningsrådet leverte i august 2024² der det anbefales å investere minst 2,6 mrd. kroner i tungregning i løpet av en 5-årsperiode.

Det overordnede behovet som skal utredes, er å anskaffe IT-utstyr som kan levere tungregning. Det er imidlertid behov for å utdype beslutningsgrunnlaget på noen områder. Utredningen skal vurdere følgende:

- Behovet for tungregning til formålene forskning, offentlig forvaltning og utvikling og bruk av KI sett opp mot hva som finnes i dag.
- Hvordan en eventuell investering i nasjonal infrastruktur bør gjennomføres, herunder gjennom nasjonal anskaffelse, gjennom samarbeidsavtaler med EU eller andre leverandører, eller gjennom en kombinasjon av disse.
- Hvordan en nasjonal infrastruktur for tungregning bør organiseres, herunder hvilke aktører som skal anskaffe og drifte den.
- Hvordan en eventuell investering kan finansieres, dvs. ulike finansieringsmodeller, inkludert muligheten for eventuelt offentlig-privat samarbeid for å tilby regnekraft til aktører i Norge og eventuelt brukerbetaling.

2.4 Sentrale begreper

Tabell 2.1 gir en oversikt over sentrale begreper benyttet i konseptvalgutredningen.

² Forskningsrådet, «Behov for tungregnekraft for forskning og kunstig intelligens».

Tabell 2.1 Sentrale begreper i konseptvalgutredningen

Begrep	Beskrivelse
Tungregning	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.
GPU	En GPU (Graphics Processing Unit) er en spesialisert prosessor designet for å håndtere og akselerere grafikk- og bildebehandling. Innenfor tungregning brukes GPUer til å utføre parallelle beregninger effektivt, noe som gjør dem ideelle for oppgaver som krever høy ytelse, som maskinlæring, kunstig intelligens og vitenskapelige simuleringer.
CPU	En CPU (Central Processing Unit) er en allsidig prosessor som utfører de fleste beregningene i en datamaskin. Den er ansvarlig for å håndtere serielle oppgaver og koordinere komplekse beregningsprosesser ved å utføre aritmetiske, logiske, kontroll- og input/output-operasjoner. CPUer er essensielle for å kjøre programmer og styre systemressurser, og de fungerer ofte sammen med GPUer for å optimalisere ytelsen i beregningsintensive applikasjoner.
EuroHPC	EuroHPC er et offentlig-privat samarbeid som koordinerer ressurser fra EU, medlemsland og private aktører for å utvikle en verdensledende tungregneinfrastruktur i Europa. Målet er å styrke Europas vitenskapelige og industrielle kapasitet, støtte den digitale transformasjonen og sikre teknologisk suverenitet.
Forskningsorganisasjoner	I konseptvalgutredningen er forskningsorganisasjoner en av brukergruppene av tungregning og omfatter: <ul style="list-style-type: none"> • Universiteter, vitenskapelige høyskoler, høyskoler akkreditert på institusjonsnivå av NOKUT • Virksomheter omfattet av regjeringens Strategi for helhetlig instituttpolitikk og virksomheter omfattet av Retningslinjer for statlig grunnfinansiering av forskningsinstitutter og forskningskonsern. • Helseforetak/sykehus med lovpålagte oppgaver innenfor forskning og utvikling og private, ideelle sykehus som inngår i Helse- og omsorgsdepartementets målesystem for forskning
Offentlig forvaltning/ sektor	I konseptvalgutredningen er offentlig forvaltning/ sektor en av brukergruppene av tungregning. Offentlig forvaltning omfatter virksomheter som utfører og ivaretar forvaltningsmessige oppgaver og som er en del av offentlig sektor. Det omfatter både departementer og underliggende virksomheter som direktorater, tilsyn, forvaltningsbedrifter, forvaltningsorganer med særskilte fullmakter, ombud. I tillegg inngår stiftelser, særlovselskap og enkelte statsforetak og statsaksjeselskap.
Forvaltningsområde	Offentlig sektor består av flere sektorer som i stor grad er overlappende med departementenes ansvarsområder, f.eks. energisektoren, utdanningssektoren og utenrikssektoren. I konseptvalgutredningen benytter vi begrepene som betegnelse på disse sektorene.
Næringslivet	I konseptvalgutredningen er næringslivet en av brukergruppene av tungregning. Næringslivet består av bedrifter og næringsdrivende som produserer varer og tjenester i Norge. Det omfatter både private og

	offentlige virksomheter, dvs. hovedvekten av statsforetak og statsaksjeselskap som f.eks. Equinor og Telenor.
Samfunnskritiske tjenester	Samfunnskritiske tjenester er tjenester som er essensiell for å opprettholde samfunnets grunnleggende funksjoner og sikkerhet. I Norge er det definert 14 kritiske samfunnsfunksjoner: Styring og kriseledelse, forsvar, lov og orden, helse og omsorg, redningstjenesten, digital sikkerhet i sivil sektor, natur og miljø, forsyningsikkerhet, vann og avløp, finansielle tjenester, kraftforsyning, elektronisk kommunikasjon, transport og satellittbaserte tjenester. Avbrudd i disse tjenestene kan ha alvorlige konsekvenser for samfunnet.
Leverandører	<p>På leverandørsiden skiller vi mellom plattformleverandører og datasenterleverandører, hvor både private og offentlige leverandører inngår.</p> <p>Plattformleverandører</p> <ul style="list-style-type: none"> • Sky-plattformleverandører (Amazon, Google, Microsoft m.fl.) • Spesialiserte tungregnetjeneste-leverandører (Sigma2, EuroHPC) • Virksomhetsintern leverandør (virksomhetene leverer tungregning til egen virksomhet) <p>Datasenterleverandører</p> <ul style="list-style-type: none"> • Leverandører av datasentertilpasset tungregning (BULK, Lefdal Mining Center m.fl.)
KI inferens	<p>KI-inferens, eller kunstig intelligens-inferens, refererer til prosessen der en trent KI-modell anvender sin lærte kunnskap for å gjøre prediksjoner eller ta beslutninger basert på nye data. Dette er den delen av KI som skjer etter at modellen er trent, og den brukes i praksis for å løse reelle problemer.</p> <p>Her er noen nøkkelpunkter om KI-inferens:</p> <ol style="list-style-type: none"> 1. Inferens innebærer å bruke en modell som allerede har blitt trent på et datasett. Modellen har lært mønstre og sammenhenger fra treningsdataene som den nå anvender på nye data. 2. Under inferens tar modellen inn nye data og gir ut en prediksjon eller beslutning. For eksempel kan en bildegjenkjenningsmodell ta inn et bilde og returnere en etikett som beskriver hva som er i bildet. 3. Inferens må ofte skje raskt og effektivt, spesielt i sanntidsapplikasjoner som autonome kjøretøy eller talegjenkjenning. 4. For å forbedre ytelsen under inferens, kan modeller optimaliseres for å kjøre raskere eller bruke mindre ressurser, for eksempel ved å bruke spesialisert maskinvare som GPUer eller TPUer. <p>KI-inferens er en kritisk komponent i å bringe KI-løsninger fra laboratoriet til praktisk bruk i virkelige applikasjoner. I teksten brukes også «bruk» og «tilpasning» i istedenfor inferens.</p>
Forskningsdrevet FoU	Forskning og utvikling som ikke nødvendigvis har som formål å utvikle en tjeneste for forvaltningen.
Forvaltningsdrevet FoU	Forskning og utvikling initiert på grunnlag av behov i forvaltningen (tilsvarende oppdragsforskning).

3 Dagens situasjon

Beskrivelsen av dagens situasjon og forventet utvikling danner sammen med problembeskrivelsen grunnlaget for nullalternativet. Nullalternativet skal beskrive en forventet utvikling hvis det ikke foretas noen grep knyttet til finansiering, styring og organisering. I dette kapitlet dekker vi følgende temaer:

- **Leveransekjeden for tungregneinfrastruktur.** I dette kapitlet beskriver vi overordnet verdikjeden for tungregneinfrastruktur.
- **Nåsituasjonen for tungregneinfrastruktur for forskning i Norge.** I kapitlet beskriver vi hvilken infrastruktur som blir brukt av ulike forskningsorganisasjoner i Norge. Vi beskriver hvordan leveransene er organisert, styrt og finansiert, samt hvilken tungregnekapasitet som er tilgjengelig.
- **Forventet utvikling for tungregneinfrastruktur og kunstig intelligens/ maskinlæring i Norge og Europa.** Dette kapitlet omfatter både beskrivelse av pågående utviklingstiltak i regi av offentlig sektor, men også forventet utvikling i det private plattformmarkedet. Dette er viktig å ha med seg når vi skal se på alternative konsepter for å løse fremtidige behov og etterspørsel.

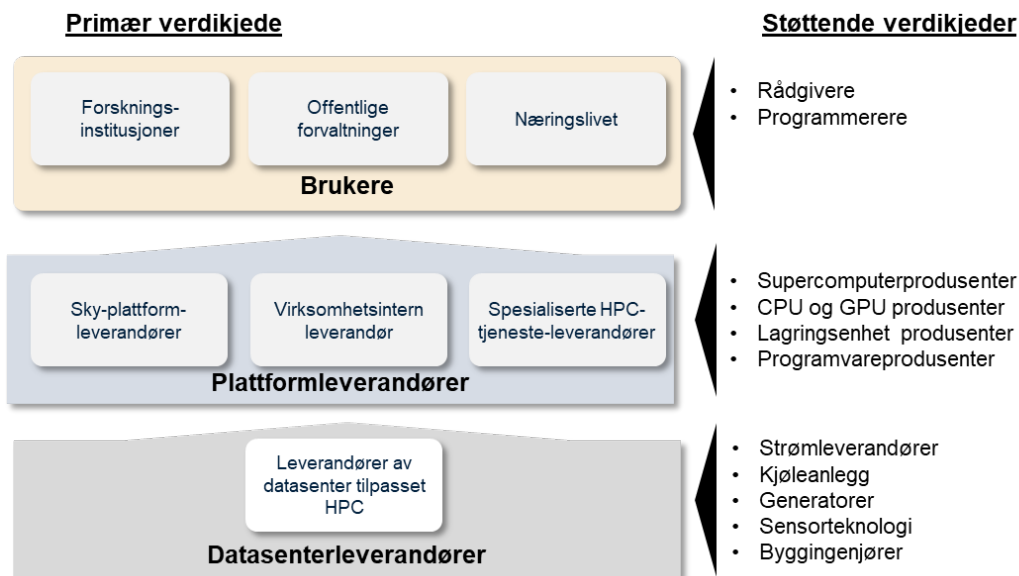
Forventet utvikling av tungregneinfrastruktur for andre brukergrupper i Norge beskrives sammen med fremtidig behov i kapittel 5.2 Interessentenes behov. Her vil vi utdype hvilken tungregneinfrastruktur som blir benyttet av offentlig forvaltning (inkludert offentlige tjenester), næringslivet og det vi benevner som utviklere av KI-modeller.

3.1 Leveransekjeden for tungregneinfrastruktur

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. Disse beregningene krever spesialiserte superdatamaskiner eller beregningsklynger.

Leveransekjeden for tungregneinfrastruktur kan deles inn i et primært verdinettverk og støttende verdikjeder, jf. Figur 3.1.

- Den primære verdikjeden består av de aktørene som leverer datasenter, tilgjengeliggjør tungregneplattformene og til slutt brukerne.
- De sekundære verdikjeder består av ulike leverandører som aktørene i den primære verdikjeden er avhengig av for å etablere sine tjenester.



Figur 3.1 Oversikt over leveransekjeden for tungregning

Datasenterleverandører

Datasenterleverandører spiller en kritisk rolle i verdikjeden for tungregneinfrastruktur (HPC). De leverer og vedlikeholder de fysiske fasilitetene som huser tungregnesystemene, inkludert bygninger, avanserte kjølesystemer for å håndtere store mengder varme som genereres av tungregneutstyr, og pålitelig og avbruddsfri strømforsyning, ofte med redundante systemer for å sikre kontinuerlig drift. Datasenterleverandørene sørger for fysisk sikkerhet med adgangskontroll, overvåking og brannvern. Avslutningsvis tilbyr de høyhastighets internett- og nettverkstilkoblinger som er nødvendig for å støtte overføring av store datamengder og rask databehandling.

Datasenterleverandører er avhengig av flere leveranser (den støttende verdikjeden):

1. **Strømleverandører:** Strømleverandører er essensielle for å sikre kontinuerlig og stabil strømforsyning til datasentre. De leverer elektrisitet som driver alle systemene og utstyret i datasenteret, og sørger for at det ikke oppstår avbrudd som kan føre til datatap eller nedetid.
2. **Kjøleanlegg:** Kjøleanlegg er nødvendige for å opprettholde optimal temperatur og forhindre overoppheting av servere og annet utstyr. Effektive kjølesystemer bidrar til å forlenge levetiden til utstyret og opprettholde ytelsen.
3. **Generatorer og batteribanker:** Generatorer og batteribanker tilknyttet UPSer (Uninterrupted power supply) fungerer som backup-strømkilder i tilfelle strømbrudd. De sikrer at datasenteret kan fortsette å operere uten avbrudd ved å levere nødstrøm til alle kritiske systemer.
4. **Sensorteknologi:** Sensorteknologi brukes til å overvåke ulike parametere i datasenteret, som temperatur, fuktighet, strømforbruk og luftkvalitet. Disse sensorene gir sanntidsdata som hjelper med å oppdage og løse problemer før de påvirker driften.
5. **Byggingeniører:** Byggingeniører spiller en viktig rolle i design og konstruksjon av datasentre. De sørger for at bygningene er strukturelt solide, energieffektive og i stand til å huse alt nødvendig utstyr på en sikker måte.
6. **Internettleverandører:** Internettleverandører er essensielle for å sikre kontinuerlig internett-tilgang. Dette er nødvendig for å støtte overføring av store datamengder og rask databehandling. Også her er redundans viktig med flere uavhengige tilkoblinger til en eller flere leverandører.

Plattformleverandører

Plattformleverandørene er de som installerer, drifter og tilgjengeliggjør tjenester til sluttbrukere. Dette innebærer anskaffelse og installasjon av tungregnemaskiner, leveranse av spesialiserte operativsystemer og mellomvare, programvare for kø- og oppgavestyring, samt lagringssystemer for høyhastighets datahåndtering. De tilbyr også verktøy for dataanalyse og visualisering, utviklingsverktøy for å skrive effektive tungregneapplikasjoner, sikkerhetsløsninger for å beskytte data, og støttetjenester for vedlikehold og teknisk support. Disse leveransene sikrer effektiv drift og maksimal ytelse av superdatamaskiner.

Vi ser tre typer av plattformleverandører som er interessante med hensyn til å sikre tungregning for forskning, offentlig forvaltning og næringsliv:

- **Skyplattformleverandører**, eller «Hyperscalers», som Amazon Web Services (AWS), Microsoft Azure og Google Cloud, er viktige aktører for leveranse av regnekraft. De tilbyr fleksible og skalerbare løsninger som kan tilpasses ulike behov, fra små forskningsprosjekter til store industrielle beregninger. Som for nasjonal og internasjonal offentlig finansiert regnekraft vil kommersiell skybasert tungregning redusere investerings- og driftskostnader for det enkelte prosjekt ved å gi tilgang til avansert maskinvare uten store investeringer som kan skalere opp eller ned etter behov. «Hyperscalers» gir også tilgang til verktøy og plattformer for dataanalyse, maskinlæring og kunstig intelligens, noe som forenkler prosessen med å utnytte superdatakapasitet, men de tilbyr ikke avansert brukerstøtte som en del av tjenestene.
- **Spesialiserte tungregne-tjenesteleverandører** er forskjellige fra skyplattformleverandører ved at de tilbyr skreddersydde løsninger som er optimalisert for komplekse beregningsoppgaver, mens skyplattformleverandører tilbyr fleksible og skalerbare løsninger for en bredere brukerbase. I Norge er Sigma2 en sentral leverandør av tungregnetjenester, med systemer som Betzy, Fram og Saga. I Europa er EuroHPC en sentral aktør, som koordinerer og finansierer utviklingen av superdatamaskiner og kvantemaskiner i ulike land. Noen av de mest kjente systemene inkluderer LUMI i Finland, MareNostrum i Spania, SuperMUC-NG i Tyskland, og Hawk i Storbritannia. Spesialiserte tungregneleverandører har som regel en bedre utnyttelsesgrad på sine systemer og har derfor ofte en lavere pris på sin regnekraft enn «Hyperscalers». Samtidig kan de tilby avansert brukerstøtte for å øke effektiviteten til beregningene som kjøres, ofte med domenekunnskap innenfor forskjellige vitenskapelige domener.
- **Virksomhetsinterne leveranser** er der virksomheter på egen hånd eller i samarbeid med andre aktører oppretter en egen tungregnekapasitet. Et eksempel er Daimler, som hovedsakelig bruker tungregning til kjøretøydesign og automatiserte kjøreteknologier. Dette inkluderer komplekse simuleringer og modellering av sensorer og reiser for autonome kjøretøy. Daimler har etablert sin tungregneplattform sammen med Lefdal Mine Datacenter. Det er også flere små og større private virksomheter som har egne tungregneressurser, som f.eks. EMGS og Equinor som bruker tungregning til Controlled Source Electromagnetic Method (CSEM)³ og seismikkanalyser relatert til oljeleting, eller Nordic Semiconductor som bruker tungregning til simulering av digitale tvillinger i utvikling av mikroprosessorer.

Brukere

Brukere er virksomheter som benytter tungregning til ulike formål. I denne konseptvalgutredningen har vi valgt å dele inn brukerne i tre grupper:

- **Forskningsorganisasjoner** som benytter tungregning til forskning, innovasjon og utvikling. Dette omfatter:
 - Universiteter, vitenskapelige høyskoler, høyskoler akkreditert på institusjonsnivå av NOKUT.

³ Controlled Source Electromagnetic Method er en geofysisk teknikk som brukes til å undersøke undergrunnen, ofte i forbindelse med leting etter olje og gass.

- Virksomheter omfattet av regjeringens Strategi for helhetlig instituttpolitikk⁴ og virksomheter omfattet av Retningslinjer for statlig grunnfinansiering av forskningsinstitutter og forskningskonsern⁵.
- Helseforetak/sykehus med lovpålagte oppgaver innenfor forskning og utvikling og private, ideelle sykehus som inngår i Helse- og omsorgsdepartementets målesystem for forskning.
- **Offentlig forvaltning** er offentlige virksomheter som bruker tungregning til å utvikle sine tjenesteleveranser. Offentlig forvaltning inkluderer virksomheter som utfører og ivaretar forvaltningsmessige oppgaver og som er en del av offentlig sektor. Det omfatter både departementer og underliggende virksomheter som direktorater, tilsyn, forvaltningsbedrifter, forvaltningsorganer med særskilte fullmakter, ombud. I tillegg inngår stiftelser, særlovselskap og enkelte statsforetak og statsaksjeselskap.
- **Næringsliv** er norske virksomheter som bruker tungregning eller KI i utviklingen og leveransen av sine produkter og tjenester. Næringslivet består av bedrifter og næringsdrivende som produserer varer og tjenester i Norge. Det omfatter hovedvekten av statsforetak og statsaksjeselskap som f.eks. Equinor og Telenor, samt oppstartsbedrifter som benytter KI til tjeneste og produktutvikling.

De ulike brukergruppene, deres bruk av tungregning og deres fremtidige behov blir beskrevet nærmere i behovsanalysen i kapittel 5.

3.2 Tungregneinfrastruktur for forskning

I kapittelet beskriver vi dagens infrastruktur og sammensetning av nasjonale tungregneressurser som forskningsorganisasjonene benytter i dag.

3.2.1 Dagens bruk av tungregneinfrastruktur for forskning

I forbindelse med rapporten som Forskningsrådet leverte i august 2024 ble det gjennomført en spørreundersøkelse våren 2024 for å kartlegge dagens bruk av tungregneinfrastruktur for forskningsformål.

Kartleggingen viste at noen forskningsorganisasjoner har egne lokale ressurser, mens andre kun benytter/er avhengige av ressurser gjennom Sigma2, 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).

Åtte universiteter svarte på spørreundersøkelsen, hvor fem av disse har egne lokale tungregneressurser. 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 tungregning (70-90 prosent).

Alle universitetene har tilgang til nasjonale ressurser hos Sigma2. Selv om de største brukerne er fra de fire 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. 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 de 22 instituttene som er spurt svarer ni institutter at de har egne regneressurser (men ikke med veldig stor kapasitet). Flertallet av disse er kun for interne brukere. Over halvparten bruker Sigma2 sine ressurser, seks av instituttene svarer at de bruker EuroHPC-ressurser og noen har

⁴ [Strategi for helhetlig instituttpolitikk - regjeringen.no](https://www.regjeringen.no).

⁵ [Retningslinjer for statlig grunnbevilgning til forskningsinstitutter og forskningskonsern - regjeringen.no](https://www.regjeringen.no).

tilgang til regnekraft gjennom annet nordisk eller europeisk samarbeid. Noen av instituttene benytter også kommersielle skytjenester, men gjerne i tillegg til Sigma2 sine tjenester. Instituttene opplyser at de i hovedsak benytter tungregning 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. På det meste har ca. to prosent av Sigma2 sine ressurser blitt brukt av næringslivet. Alle universitetene 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 tungregning. Blant annet har Nasjonalbiblioteket hatt tilgang til universitetenes ressurser for trening av språkmodeller. 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 tungregning 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.

I prosjektsamarbeid med norske industriaktører benytter forskningsmiljøene i noen tilfeller regneressursene som samarbeidspartnerne har tilgang til. Kun noen få institusjoner svarer at de utelukkende benytter kommersielle skytjenester.

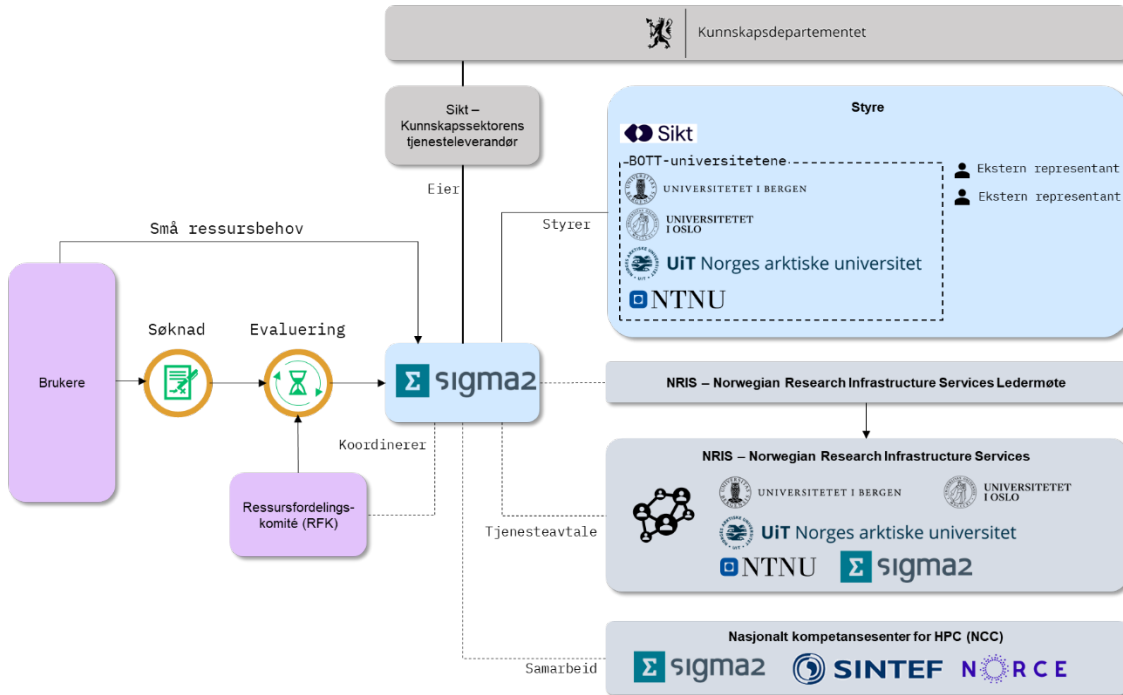
3.2.2 Dagens organisering og styring

Sigma2 AS er et heleid statlig selskap som ble etablert i 2015, da som et datterselskap under Uninett, fra 2022 eid av Sikt – Kunnskapssektorens tjenesteleverandør. Selskapet har det strategiske og operasjonelle ansvaret for anskaffelser og drift av tungregneressurser og datalagring for norske forskningsmiljøer.

Sigma2 er finansiert av Kunnskapsdepartementet gjennom Norges forskningsråd. I tillegg mottar selskapet bidrag fra Universitetet i Oslo (UiO), Universitetet i Bergen (UiB), Universitetet i Tromsø (UiT) og Norges teknisk-naturvitenskapelige universitet (NTNU). Selskapet har også en betalingsmodell⁶, som sørger for inntekt fra brukere. De fire universitetene samarbeider med Sigma2 om å levere og drifte de nasjonale tjenestene. Dette samarbeidet går under navnet Norwegian Research Infrastructure Services (NRIS).

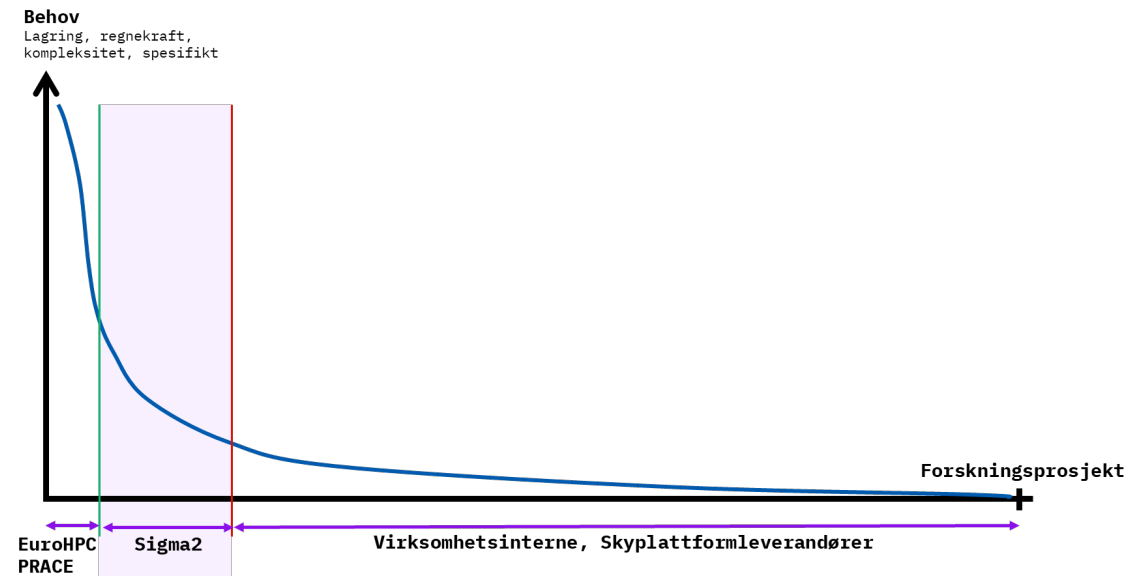
Sigma2 er etablert som en felles plattformleverandør, med ansvar for å anskaffe, etablere og drifte felles tungregneressurser og lagring av vitenskapelige data. I tillegg yter de både grunnleggende og avansert brukerstøtte til de brukerne som skal ta i bruk infrastrukturen. Sigma2 er en liten organisasjon som er avhengig av underleverandører for å levere disse tjenestene. Figur 3.2 gir en oversikt over dagens organisering.

⁶ [Nærmere beskrivelse av brukermodellen.](#)



Figur 3.2 Oversikt over dagens organisering

Sigma2 leverer ikke tungregning og lagringskapasitet til hele spekteret av brukere, men tilbyr først og fremst nasjonal infrastruktur for forskning og utdanning. Slik Figur 3.3 viser, skal Sigma2 dekke behov for tungregning og lagringskapasitet som ikke den enkelte forskningsaktør kan skaffe til veie på egen hånd. Samtidig har man valgt å tilby tjenester til prosjekter som krever mest kapasitet gjennom det internasjonale samarbeidet.



Figur 3.3 Strategisk fokus for Sigma2

Styringsmodell

Sigma2 sitt hovedformål er:

«Økt verdiskapning gjennom å tilby nasjonale e-infrastruktur tjenester innen tungregning og lagring av vitenskapelige data. Primær målgruppe er universiteter, høyskoler og

forskningsinstitusjoner i Norge. Selskapet skal være de samme institusjoners virkemiddel for å utføre oppgaver innen e-infrastruktur i egen regi.»⁷

Generalforsamlingen til Sigma2 består av direktør i Sikt – Kunnskapssektorens tjenesteleverandør og styreleder i Sigma2. I tillegg deltar daglig leder i Sigma2 og Riksrevisjonen. Ordinær generalforsamling avholdes hvert år. Selskapets ordinære generalforsamling behandler:

- Fastsettelse av resultatregnskap og balanse, herunder anvendelse av årsoverskudd eller dekning av årsunderskudd
- Godkjenning av styrets rapport for bruk av selskapets ressurser
- Fastsettelse av godtgjørelse til styrets medlemmer og til revisor
- Valg av styremedlemmer og styrets leder
- Andre saker som i henhold til lov eller vedtekter, hører under generalforsamlingen

Selskapets styre består av styreleder og 6 styremedlemmer etter generalforsamlingens beslutning. 4 av styrets medlemmer skal velges blant ansatte ved NTNU, UiB, UiO og UiT.

De to andre styremedlemmene foreslås av Forskningsrådet og er en representant fra instituttsektor og gjerne fra en utenlandsk institusjon - i dag fra henholdsvis Forsvarets Forskningsinstitutt (FFI) og Uppsala Universitet i Sverige.

Tildelingsmodell

Tildeling av beregnings- og lagringsressurser hos Sigma2 følger en regulert og definert evalueringsprosess. Tilgang til infrastrukturen er underlagt kriterier som er definert i en generell policy⁸ og i detaljerte retningslinjer⁹ for ressurstildeling.

Søknader om beregnings- og lagringsressurser og avansert brukerstøtte for forskningsprosjekter evalueres av en ressursfordelingskomité (RFK). RFK består av ledende norske forskere fra relevante brukergrupper, utnevnt av styret til Sigma2.

RFK evaluerer søknader og tildeler tilgang til Sigma2s ressurser to ganger i året. For å sikre at ressursene blir brukt effektivt, må søknadene vise til vitenskapelig fremdragende arbeid og et tydelig vitenskapelig behov.

Sigma2 er sekretariat for RFK og bistår komiteen i deres arbeid. Tildelingen fra RFK har en pris som dekker driftskostnader. I tillegg leverer Sigma2 også tjenester til andre brukere til full pris. Dette gjøres i tråd med statsstøtteregulativet som har en grense på at 20 prosent av ressursene kan utnyttes i økonomisk aktivitet (se nærmere beskrivelse i avsnitt 3.2.3).

Driftsmodell

Driftsmodell beskriver hvordan felles tungregnings- og lagringsressurser anskaffes, installeres, driftes og vedlikeholdes samt hvordan brukerstøtte gis til brukerne. Sigma2 har inngått egne underleverandøravtaler for leveranse, installasjon og testing av nødvendig utstyr.

Leveransene av tungregnetjenester til primærbrukerne reguleres gjennom en formell tjenesteavtale¹⁰. NRIS består av høyt kvalifisert IT-personell ved universitetene i Bergen, Oslo, Tromsø og NTNU og ansatte i Sigma2. Samlet sett er det omtrent 50 personer som danner et geografisk spredt kompetansenettverk. NRIS har følgende kjerneaktiviteter som utgjør ca. 35 årsverk:

- **Felles drift av de nasjonale superdatamaskinene og lagringssystemene for forskningsdata:** NRIS har ansvaret for felles drift av de nasjonale superdatamaskinene og lagringssystemene for forskningsdata. Dette sikrer at forskere har tilgang til høytytelsesregnerressurser og sikker lagring av sine data.

⁷Vedtekter for Sigma2.

⁸ [Mandate for the Resource Allocation Committee for e-infrastructure, February 2015.](#)

⁹ [Guidelines for Resource Allocation on the National E-infrastructure.](#)

¹⁰ Drifts- og brukerstøtteavtale: Tjenesteavtale om drift av og brukerstøtte for nasjonal eInfrastruktur, Dokument nr: 17/00166-1.

- **Samlet brukerstøtte, inkludert en felles nasjonal helpdesk:** En viktig aktivitet for NRIS er å tilby omfattende brukerstøtte. Dette inkluderer å opprettholde en nasjonal helpdesk som gir assistanse og veiledning til forskere, og hjelper dem med å navigere og bruke de tilgjengelige ressursene effektivt.
- **Avansert brukerstøtte:** I tillegg til grunnleggende brukerstøtte, tilbyr NRIS avansert brukerstøtte for å hjelpe forskere med å håndtere komplekse beregnings- eller datautfordringer og optimalisere bruken av regne- og dataressursene.
- **Delt opplæring og formidling av ressurser og aktiviteter:** NRIS fokuserer også på opplæring og formidling. De tilbyr opplæringsøkter og workshops for å spre kunnskap og ressurser, slik at forskere er godt rustet til å utnytte infrastrukturen til sitt fulle potensial.
- **Deltakelse i internasjonale partnerskap og prosjekter relatert til tungregning:** Sigma2 og partnere deltar aktivt i internasjonale samarbeid og prosjekter relatert til tungregning. Denne deltakelsen bidrar til kunnskapsdeling, oppdatering på globale utviklinger og til fremme av forskningsinfrastruktur på internasjonalt nivå. Spesielt viktig er deltakelsen i EuroHPC. Sigma2 forvalter norsk deltakelse i EuroHPC og har vært med på etableringen av en av verdens største superdatamaskiner, LUMI, i Kajaani, Finland.

Nasjonalt kompetansesenter for tungregning (NCC) er et samarbeid mellom Sigma2, SINTEF og NORCE Norwegian Research Centre. NCC bistår næringsliv, og særlig de små og mellomstore bedrifter, og offentlig forvaltning i å utvikle kompetanse i bruk av avansert teknologi, som tungregning, høykapasitets dataanalyse (HPDA), kunstig Intelligens og maskinlæring. NCC har følgende hovedaktiviteter:

- **Tilbyr ekspertise og teknisk støtte:** NCC tilbyr teknisk ekspertise og støtte til privat næringsliv, offentlig administrasjon og akademia. Dette inkluderer hjelp med å utvikle og optimalisere arbeidsflyter, programmere kode for kjøring av jobber på nasjonale superdatamaskiner, og gi skreddersydd støtte til spesifikke behov.
- **Opplæring og utdanning:** NCC arrangerer opplæringsøkter, workshops og kurs for å øke den teknologiske kompetansen hos brukerne. Dette hjelper forskere og næringsliv med å effektivt benytte seg av ulike avanserte teknologier.
- **Tilgang til tungregnerressurser:** NCC sørger for at brukerne har tilgang til tungregnerressurser. Dette inkluderer tilbudet om Small Scale Explorative Work (SSEW)-tjenester hos Sigma2, som gir brukerne mulighet til å teste ideer eller utføre foreløpig forskning før de investerer i større prosjekter. NCC hjelper også brukere med større behov til å søke om kapasitet hos EuroHPC.
- **Rådgivning:** NCC tilbyr rådgivning for å hjelpe organisasjoner med å forstå potensialet i digitale arbeidsflyter, implementere digitale løsninger og optimalisere bruken av avanserte teknologier.
- **Koordinering og samarbeid:** NCC koordinerer med andre nasjonale og internasjonale kompetansesentre for å dele ekspertise, identifisere mangler og gi skreddersydd støtte til ulike interessenter. Dette inkluderer å øke bevisstheten om fordelene med tungregning og fremme bruken av dette på nasjonalt nivå.
- **Støtte til små og mellomstore bedrifter:** NCC fokuserer på å støtte små og mellomstore bedrifter ved å gi dem den nødvendige ekspertisen for å dra nytte av innovasjonsmuligheter skapt av tungregne-teknologi. Dette inkluderer å hjelpe bedriftene med å skalere opp virksomheten sin og forbedre deres konkurranseevne.

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 korn, 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 tungregning. Sigma2 bidro også med å optimalisere arbeidsflyten for raskere kjøring, lavere ressursbruk og kostnader.

DigiFarm, med sitt høye forbruk på nasjonale tungregneressurser, 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 EuroHPC sin 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.»

DigiFarm sin 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 nesten 8 prosent på bare noen få måneder, en betydelig prestasjon innen dyplæring, ifølge Helseth. DigiFarm ble nominert til 3 priser på den store internasjonale konferansen for tungregning i 2024, hvor de vant prisene for beste innovasjon på tungregning, og for beste samarbeid mellom akademia og industri.

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/person. 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 titalls millioner av oppføringer er ikke et alternativ, og det er her kunstig intelligens kommer inn i bildet.

Tollvesenet engasjerte det nasjonale kompetansesenter for tungregning 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 entiteten 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. Tollvesenet ønsker nå å bruke Sigma2 infrastruktur for å kjøre større transformermodeller som nå kan analysere mønster og finne avvik i tolldeklarasjoner med større nøyaktighet.

Kilde: [Sigma2 - nettside](#).

3.2.3 Dagens finansieringsordning

I dag finansieres Sigma2s investeringer i lagrings- og tungregneressurser gjennom Forskningsrådets finansieringsordning for forskningsinfrastruktur, *Nasjonal satsing på forskningsinfrastruktur* (INFRASTRUKTUR). INFRASTRUKTUR er en nasjonal konkurransearena som bidrar til å sikre god kopling mellom finansiering av forskningsinfrastruktur og annen forskningsfinansiering.

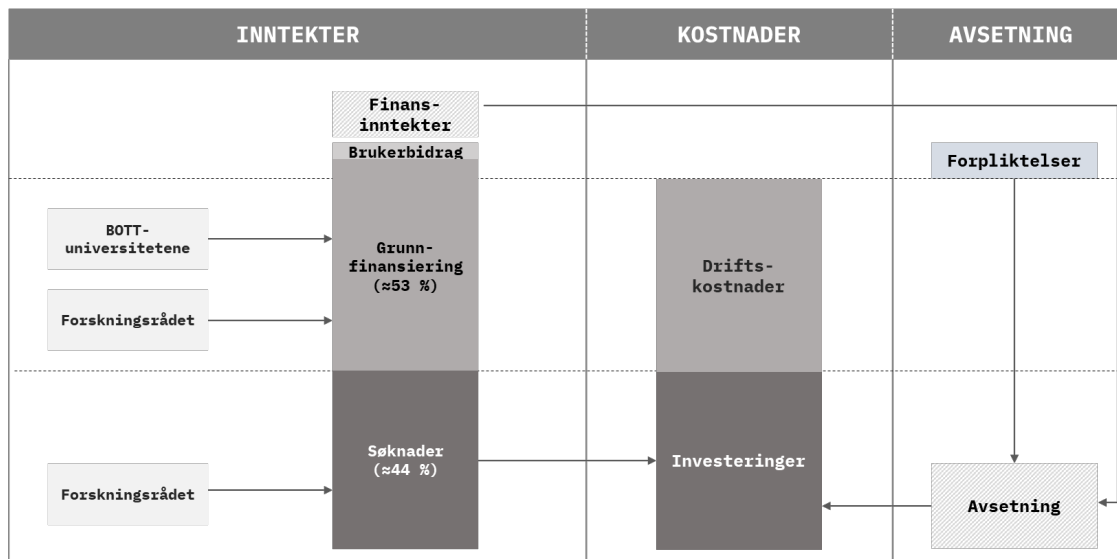
Tildelinger over Forskningsrådets budsjett skal støtte opp under utvikling av nasjonalt prioriterte forskningsområder og nasjonalt viktige næringer med stort behov for forskningsinfrastruktur. Forskningsrådet skal bidra til at institusjonene samordner seg når flere miljøer har behov for forskningsinfrastruktur, men kostnadene er så høye at det er mest hensiktsmessig med samarbeid.

Finansiering gjennom INFRASTRUKTUR støtter primært investeringer i nye, og oppgradering av eksisterende, forskningsinfrastrukturer. Driftskostnadene til infrastrukturene skal så langt det er mulig dekkes av prosjekt som benytter forskningsinfrastrukturene. INFRASTRUKTUR kan bidra til finansiering av drift av infrastrukturer i en oppstartsfasen, men vil være restriktiv med å gi støtte til langsiktig grunnfinansiering av driften til infrastrukturene utover oppstartsfasen. INFRASTRUKTUR vil bare bidra med midler til langsiktig grunnfinansiering av drift utover oppstartsfasen dersom en rekke vilkår er oppfylt.

Finansiering av Sigma2 AS

Finansieringen av Sigma2 er regulert gjennom en samarbeidsavtale som er inngått mellom Sigma2 og BOTT-universitetene. Styret i Sigma2 har ansvar for forvaltning av selskapet jfr. lov om aksjeselskap §6-12. Samarbeidet startet 1.1.2021 og skal avsluttes 1.1.2027. Avtalen kan forlenges med ytterligere 6 år.

Figur 3.4 gir en oversikt over finansieringen av Sigma2.



Figur 3.4 Oversikt over finansieringsmodellen for Sigma2

Sigma2 AS har fire inntektskilder:

- A. **Inntekter fra prosjektsøknader for investering.** Investeringer i anlegg for tungregning og lagring av forskningsdata er finansiert gjennom Forskningsrådets INFRASTRUKTUR-ordning. Sigma2 forplikter seg i henhold til samarbeidsavtalen med BOTT-universitetene og Forskningsrådet til jevnlig å søke om midler til oppgraderinger og investeringer i nytt utstyr gjennom INFRASTRUKTUR-ordningen. I perioden 2016-2024 er det bevilget totalt ca. 683 millioner kroner til Sigma2, hvorav 40 millioner er brukt til å finansiere Norges andel av investeringer i LUMI.

- B. **Grunnfinansiering.** I henhold til samarbeidsavtalen skal Forskningsrådet og BOTT-universitetene bidra med en grunnfinansiering, for å dekke driften av Sigma2. Forskningsrådet bidrar med 50 millioner årlig fra INFRASTRUKTUR og de fire BOTT-universitetene bidrar med totalt ca. 53 millioner årlig (Indeksregulert fra 2023). Grunnfinansiering fra BOTT-universitetene dekker også utgiftene til deres bruk av en del av kapasiteten. BOTT-universitetenes bruk utover det som dekkes av grunnfinansieringen skal dekkes av prosjektene som bruker infrastrukturen. Hver samarbeidsdeltaker må sikre at disse brukerne har dekning for sin bruk og garantere økonomisk for denne bruken. BOTT-prosjekter bruker ca. 70 prosent av tungregnerressursene og 35 prosent av lagringsressursene. (Den institusjonen som bruker mest av lagringsressursene er NORCE (Klima og miljø)).
- C. **Brukerinntekter fra andre brukere enn BOTT-universitetene.** Kapasitet utover det som brukes av BOTT-universitetene dekket av grunnfinansieringen blir tilgjengeliggjort til andre brukere.

Grunnfinansieringen og brukerinntekter følger kost-pluss-metoden. Kost-pluss-metoden defineres på følgende måte:

«...tar utgangspunkt i kostnadene en leverandør pådrar seg for eiendeler overdratt eller tjenester ytet til en nærstående kjøper. Et passende kost-pluss-påslag blir så tillagt for å komme fram til et passende overskudd i lys av de funksjonene som leverandøren utøver og markedsforholdene.»¹¹

Det er inngått en egen avtale mellom Forskningsrådet og Sigma2 AS som regulerer hvordan midlene fra Forskningsrådet skal brukes, samt i hvilken grad kapasiteten kan brukes til økonomisk aktivitet. Støtten blir tildelt fra Forskningsrådet under forutsetning av at den ikke utgjør statsstøtte. Dette innebærer at maksimalt 20 prosent av forskningsinfrastrukturens årlige kapasitet kan bli benyttet i økonomisk aktivitet, slik dette er definert i EFTAs overvåkningsorgan (ESA) sine retningslinjer for statsstøtte til forskning, utvikling og innovasjon, paragraf 20. Resten av forskningsinfrastrukturens kapasitet må brukes i ikke-økonomisk aktivitet. Styret i Sigma2 har satt en grense på at fem prosent av ressursene kan selges ut.

- Med **økonomisk aktivitet** menes blant annet oppdragsforskning, leveranse av forskningstjenester eller utleie av infrastrukturen til foretak. Når infrastrukturen brukes i økonomisk aktivitet, skal det tas markedspris for bruken. Der det ikke er noen markedspris, skal prisen beregnes i henhold til paragraf 25 og 26 i ESAs retningslinjer.
- Med **ikke-økonomisk aktivitet** menes blant annet bruk av forskningsinfrastrukturen i utdanning, uavhengig forskning, inkludert samarbeidsforskning, spredning av forskningsresultater og kunnskapsoverføringsaktiviteter som nærmere og detaljert beskrevet i ESAs retningslinjer paragraf 19.

Finansinntekter. Eventuelt overskudd blir satt av til fremtidige investeringer og oppgraderinger. Regnskapet fra 2023 viser at Sigma2 AS hadde ca. 7.2 mill. kroner i finansinntekter. Selskapet hadde 147,8 mill. kroner i reserverte midler per 31.12.2023 og av disse var 92 mill. kroner reservert anskaffelsen av ny tungregnemaskin A2/Olivia.

Finansiering av Nasjonalt kompetansesenter for tungregning (NCC)

Nasjonalt kompetansesenter for tungregning har EU-finansiering ut 2025 gjennom DIGITAL programmet i EU og nasjonal medfinansiering gjennom Forskningsrådets ordning IKTPLUSS. Kompetansesenteret er en del av et nasjonalt støtteapparat for norske små og mellomstore bedrifter, 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.

¹¹ Ot.prp. nr. 62 (2006-2007).

3.2.4 Deltagelse og muligheter i EuroHPC

I dette kapittelet 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.

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. Forskningsrådet forvalter Norges delegat i EuroHPC sin generalforsamling (Governing Board).

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

EuroHPC sitt oppdrag er strukturert rundt seks teknologiske søyler: Infrastruktur, applikasjoner og data, internasjonalt samarbeid, teknologi, ferdigheter og bruk, samt føderasjon og hyperkonnektivitet.

I mai 2024 vedtok Det Europeiske Råd en endring som inkluderte utvikling og drift av KI-fabrikker («AI factories»). Dette innebærer investering i KI-optimaliserte datamaskiner og systemer som er særlig innrettet mot å bistå små og mellomstore bedrifter og oppstartsbedrifter. 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, mens øvrige punkter 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 etc.)
5. Supercomputer-friendly programming facilities (parallelization, usage optimization etc.)
6. Attracting & pooling talent
7. Interacting with AI-ecosystem at large & other AI initiatives

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.

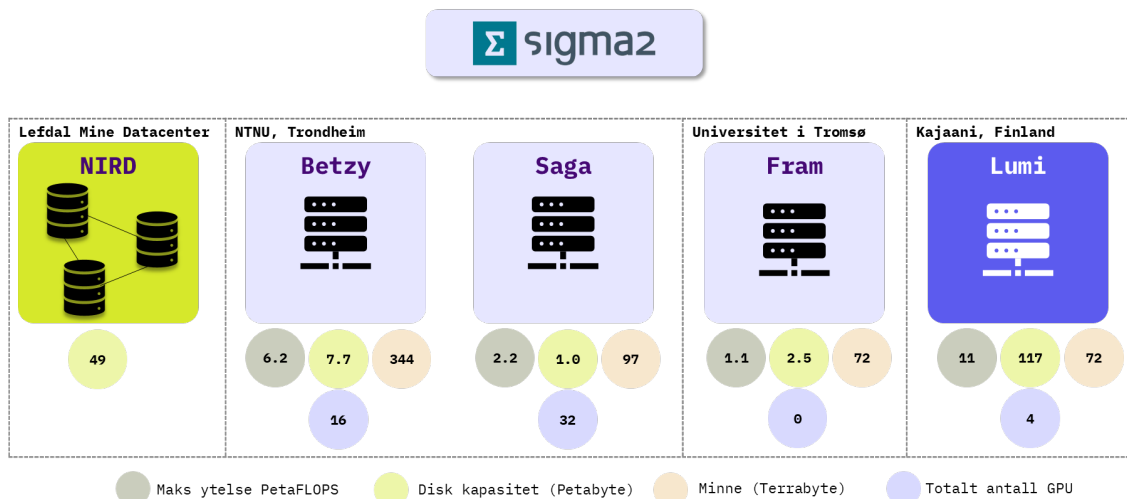
Det planlegges en ny LUMI-maskin som i større grad enn tidligere maskiner er KI-optimalisert. For at norske aktører skal få tilgang er det viktig at Norge investerer i denne. Investeringsnivået må være høyt nok til at vi sikrer norske aktører en tilstrekkelig andel av kapasiteten. En endring i EuroHPC-forordningen 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 innenfor tungregning. Verter av KI-optimaliserte superdatamaskiner kan motta økonomisk støtte fra EU som dekker opptil 50 prosent av anskaffelses- og driftskostnadene. I tillegg gjør størrelsen på investeringene at økonomien blir svært fordelaktig over relativt sett mindre nasjonale investeringer.

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.). EuroHPC har et økt fokus rettet mot oppstartsbedrifter og små og mellomstore bedrifter og det nasjonale kompetansesenteret må innrette seg for å møte disse kravene.

3.2.5 Dagens leveransekapasitet

HPC-systemer gjennom Sigma2

Gjennom Sigma2 har forskere tilgang til tungregnekapasitet fra fire superdatamaskiner, jf. Figur 3.5. Sigma2 disponerer i dag tre superdatasystemer og en moderne lagringsløsning. I tillegg har norske aktører dedikert tilgang til nasjonal kvote på LUMI-maskinen i Finland.



Figur 3.5 Oversikt over tungregnekapasitet og lagring tilgjengelig gjennom Sigma2

Hver av tungregnefasilitetene består av en beregningsressurs (flere beregningsnoder, hver med flere prosessorer og internt delt minne, pluss en sammenkobling som forbinder nodene), en sentral lagringsressurs som er tilgjengelig for alle nodene. Alle fasiliteter bruker varianter av operativsystemet Linux. NIRD (Norwegian Infrastructure for Research Data) er den sentrale infrastrukturen for forskningsdata som ikke aktivt jobbes med, og NIRD utgjør således en felles sekundær lagringsressurs for disse. NIRD har også dedikerte tjenester for enklere analyser og visualisering av forskningsdata, i tillegg til forskningsdata arkivtjeneste. Fremover vil kommende tungregneanlegg samlokaliseres med NIRD i Lefdal Mine Datacenter (LMD). Til sammen tilbyr Sigma2 en teoretisk regnekapasitet på 20.5 petaFLOPS eller 1.383 MCPUh.

Anlegg-Tungregning	Kjerner	2023	2024	2025	2026	2027	2028	2029
Fram	32 000	Extended support						
Saga	16 064	Extended support						
Betzy (Intel equivalenter)	103 219	Ext. Sup.						
A2	64 000							
LUMI-C	6 599							
LUMI-C (2024 ->)	6 679							
Totalt per år (MCPUh)		1 383	1 384	1 664	1 664	760	760	561
Anlegg-Lagring	Kapasitet (Pbytes)	2023	2024	2025	2026	2027	2028	2029
NIRD - Tiered Storage	24							
NIRD - Data Lake	25							
Lagringskapasitet		49	49	49	49	49	-	-

Figur 3.6 gir en oversikt over forventet levetid for de eksisterende anleggene inkludert bevilgede midler til nye anlegg.

Anlegg-Tungregning	Kjerner	2023	2024	2025	2026	2027	2028	2029
Fram	32 000	Extended support						
Saga	16 064			Extended support				
Betzy (Intel equivalenter)	103 219				Ext. Sup.			
A2	64 000							
LUMI-C	6 599							
LUMI-C (2024 ->)	6 679							
Totalt per år (MCPUh)		1 383	1 384	1 664	1 664	760	760	561
Anlegg-Lagring	Kapasitet (Pbytes)	2023	2024	2025	2026	2027	2028	2029
NIRD - Tiered Storage	24							
NIRD - Data Lake	25							
Lagringskapasitet		49	49	49	49	49	-	-

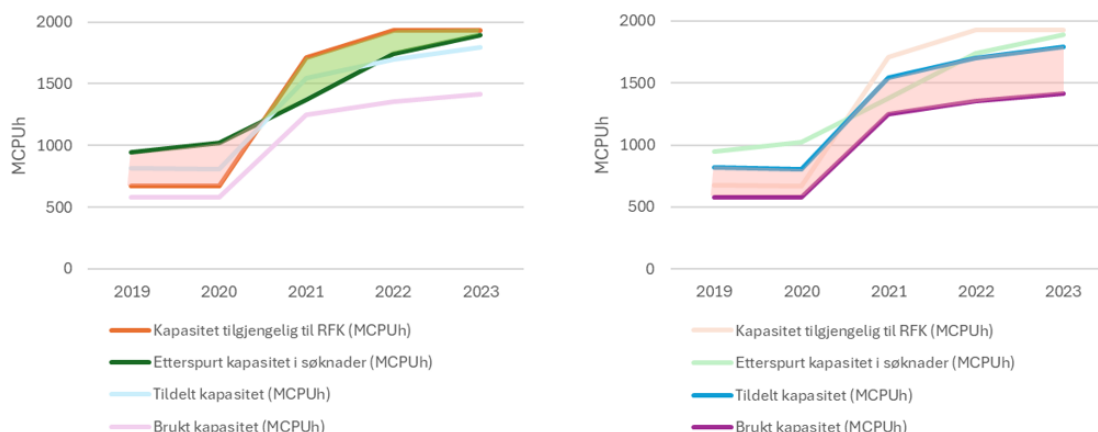
Figur 3.6 Oversikt over estimert levetid for eksisterende anlegg

Kilde: Sigma2.

Figuren viser at det må foretas investeringer i oppgraderinger og ny infrastruktur for tungregning allerede i 2024 for å opprettholde minimum dagens kapasitet. Sigma2 bygger ny kapasitet gjennom A2 anlegget, som vil erstatte og doble kapasiteten som i dag leveres på Fram. Dette vil gi en økning av tungregnekapasiteten i 2025 og 2026 med 20 prosent sammenlignet med dagens situasjon.

Midler tildelt gjennom Forskningsrådets INFRASTRUKTUR-utlysning i 2023 vil blant annet brukes til en delvis finansiering av et nytt anlegg (B2) som vil erstatte deler av kapasiteten til Betzy. Uten denne investeringen vil kapasiteten reduseres til 55 prosent av dagens når Betzy er ute av drift i 2027.

Figur 3.7 gir en oversikt over utviklingen i tilgjengelig tungregnekapasitet gjennom Sigma2 og utvikling i utnyttelsesgrad.



Figur 3.7 Kapasitetsutnyttelse på tungregneressurser tilgjengelig gjennom Sigma2 AS i perioden 2019-2023. Tilgjengelig vs. etterspurt kapasitet (venstre figur) og tildelt vs. brukt kapasitet (høyre figur)

Kilde: Sigma2.

I perioden 2019-2023 økte tilgjengelig kapasitet med 187 prosent, spesielt gjennom investeringer i 2019 og 2020 da den største nasjonale maskinen, Betzy ble anskaffet. Sigma2 fikk omtrent tre ganger så mange kjerner som forventet pga. sterk konkurranse mellom leverandørene. Dette medførte at det i 2021 og 2022 var høyere kapasitet enn etterspørsel. Det vurderes at det er nødvendig å overdimensjonere ved anskaffelse av nye ressurser slik at man har kapasitet for å håndtere vekst i etterspørselen inntil neste anskaffelse kan settes i produksjon. I 2023 har etterspørselen nærmet seg tilgjengelig kapasitet.

Utnyttet kapasitet har i hele perioden ligget lavere enn både tilgjengelig og tildelt kapasitet. De siste årene har utnyttelsesgraden ligget på ca. 80 prosent av den kapasiteten som er tildelt. I 2023 og 2024 lå utnyttelsesgraden på litt under 75 prosent.

Andre faktorer som har bidratt til nivået på utnyttelsesgraden er større planlagte og uplanlagte driftsavbrudd, som har redusert opptiden av utstyret ned mot 95 prosent. Prosenttallene for bruk og etterspurt kapasitet i søknader tar ikke hensyn til nedetid. Modeller for brukerbetaling vil

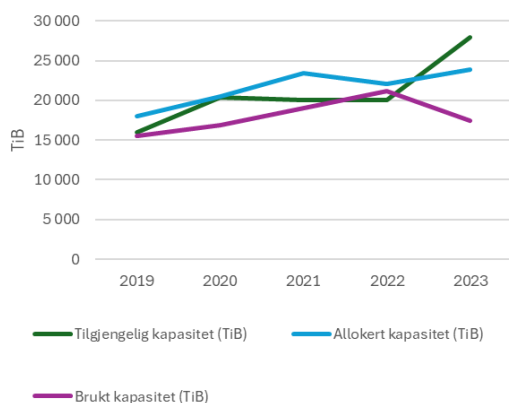
også kunne påvirke etterspørsel av tjenester. For noen forskningsprosjekter vil midler tilgjengelig for finansiering av tungregnetjenester ikke alltid svare til behov/ambisjonsnivået.

Basert på erfaringsutveksling mellom tungregnesentre i Norden er 75 prosent utnyttelse ansett som et rimelig kompromiss mellom utnyttelse av investeringen og ventetid i kø for å kjøre beregningsjobber. Det bør være noe slakk for å ta hensyn til svingninger i lasten. Etterspørselen kan variere sterkt, avhengig av forskernes aktivitet, som publiseringsfrister og tilsvarende.

Lagringssystemer gjennom Sigma2

National Infrastructure for Research Data (NIRD) gir datalagringskapasitet for forskningsprosjekter. Den nåværende NIRD-installasjonen ble anskaffet i 2021, og senere utvidet i 2023. Begge investeringene ble gjort med tildelinger fra Forskningsrådets INFRASTRUKTUR-ordning. Det er indikert fra Sigma2 at dagens NIRD-anlegg vil nå slutten av levetiden i 2028.

Figur 3.8 Figur gir en oversikt over utviklingen i tilgjengelig lagringskapasitet gjennom Sigma2 og hvordan utnyttelsen har utviklet seg.



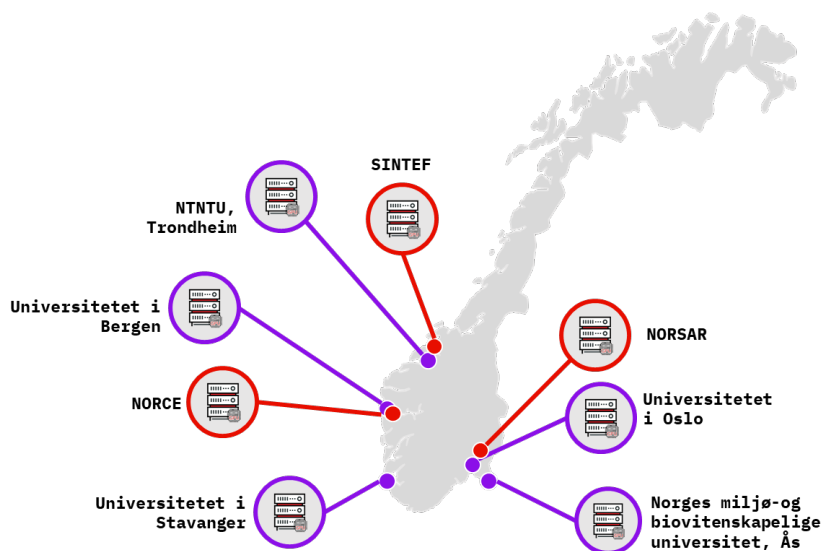
Figur 3.8 Kapasitetsutnyttelse på lagringskapasitet tilgjengelig gjennom Sigma2 AS i perioden 2019-2023. Tilgjengelig vs. etterspurt kapasitet

Kilde: Sigma2.

Mens det virker å være utnyttet kapasitet av tungregning, samsvarer tilgjengelig lagringskapasitet bedre med etterspurt kapasitet.

HPC-systemer gjennom forskningsorganisasjoner

I spørreundersøkelsen som ble gjennomført ifm. rapporten Behov for tungregnekraft for forskning og kunstig intelligens fra Forskningsrådet (2024) angir fem universiteter og åtte forskningsinstitutter at de har tilgang til egne tungregneressurser. Figur 3.9 gir en oversikt over noen av de tungregneressursene som finnes blant forskningsorganisasjonene.



Figur 3.9 Utvalgte forskningsorganisasjoner med egne tungregnerressurser

Disse ressursene er primært tilgjengelige for institusjonenes egne ansatte. Det foreligger ikke en fullstendig oversikt over hvor stor kapasitet hver enkelt ressurs har tilgjengelig.

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 INFRASTRUKTUR-ordningen. Dette gir Norge tilgang til to prosent av LUMI sin 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 KI-utlysningene der det er lagt vekt på at små og mellomstore bedrifter skal delta.

Norske miljøer søker mest på de største supercomputerne i EuroHPC («pre-exascale» kapasitet) fordi dette er et behov som ikke dekkes gjennom nasjonal kapasitet hos Sigma2 (nasjonal «mid-range» kapasitet), jf. Figur 3.1. Norske aktører kan også få tilgang til mindre ressurser for å bli kjent med superdatamaskinene, teste programvare og utføre ytelsestester. For den norske nasjonale andelen av LUMI gjelder statsstøtteregulativet. For EuroHPC-andelen er det ikke fastsatt retningslinjer for andelen som kan brukes i økonomisk aktivitet.

3.3 Forventet utvikling av infrastruktur og behov for tungregning og kunstig intelligens/maskinlæring

I Norge forventes det at tungregning og kunstig intelligens vil få en stadig viktigere rolle innen forskning og offentlig forvaltning. Denne utviklingen drives av flere viktige trender og behov på tvers av ulike sektorer. I dette kapittelet oppsummerer vi noen av de viktigste utviklingstrekkene som vil drive fremtidige behov for infrastruktur.

3.3.1 Forventet utvikling av tjenester innen forskning og offentlig forvaltning

Utviklingen vil bli drevet framover av flere sentrale bruksområder og tjenester. Innen forskning vil tungregning og KI være avgjørende for å håndtere komplekse beregninger og analysere store datamengder. Dette vil gi fremskritt innen områder som klimaendringer, helse, og bioteknologi. For eksempel vil KI kunne brukes til å utvikle presisjonsmedisin ved å analysere genetiske data, mens tungregning vil muliggjøre avanserte klimamodeller for bedre forståelse av miljøendringer. I offentlig sektor vil KI og maskinlæring bidra til å effektivisere tjenester og forbedre beslutningsprosesser. Automatisering av rutineoppgaver, som saksbehandling og dataanalyse, vil frigjøre ressurser og forbedre tjenesteleveransen. Videre vil KI kunne brukes til å forbedre sikkerhet og beredskap ved å analysere trusler og risikoer i sanntid.

I tillegg til de etablerte tungregneressursene på de enkelte institusjoner og gjennom Sigma2, foregår det en rekke utviklingsprosjekter knyttet til tungregning og KI. Noen eksempler er beskrevet nedenfor. Teknologitvillingen vil påvirke både hvilken type tungregning og hvor mye tungregneressurser det vil bli behov for.

Generative KI-modeller - språkmodeller

Teknologirådets rapport «Generativ kunstig intelligens i Norge» 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 til forskere (Google TPU Research Cloud). Men tilgangen er uforutsigbar både i tid, kapasitet og stabilitet/tilgjengelighet.

I statsbudsjettet for 2025¹² bevilget regjeringen til sammen 40 mill. kroner til norske og samiske språkmodeller (grunnmodeller) til bruk i kunstig intelligens. Av disse vil 20 mill. kroner gå over Kultur- og likestillingsdepartementets budsjett til Nasjonalbiblioteket, som får i oppdrag å trene, oppdatere og tilgjengeliggjøre språkmodeller, mens 20 mill. kroner skal gå over Kunnskapsdepartementets budsjett til å finansiere regnekraft til trening av modellene. Regnekraften skal leveres av Sigma2. Satsingen er blant annet en oppfølging av tiltak for språkteknologi i regjeringens handlingsplan for norsk fagspråk.

Norwegian Artificial Intelligence Cloud (NAIC)

NAIC er et femårig prosjekt med treårig finansieringsstøtte fra Forskningsrådets INFRASTRUKTUR-satsing for å etablere infrastruktur for kunstig intelligens og maskinlæring i Norge, som skal fremme KI-utdanning, forskning og innovasjon. NAIC vil gi tilgang til tungregneressurser og støtte til forskere, Ph.d.-studenter, lærere, gründere, oppstartsbedrifter og små og mellomstore bedrifter som er interessert i å utforske KI-metodologier, samt utvikle avanserte KI-modeller og algoritmer. Videre vil prosjektet tilby inkluderende opplæringsprogrammer og ressurser, både for nye og erfarne brukere, enten man søker et introduksjonsnivå av forståelse innen KI, eller om man leter etter mer avanserte kurs for å utnytte infrastrukturen for spesifikke oppgaver.¹³

Prosjektet gjennomføres av et konsortium, som består av en mangfoldig gruppe partnere, inkludert UiO, NORA, NTNU, UiB, SIMULA, UiA, SINTEF, UiT, Sigma2 og NORCE Research. Etter finansieringsperiode vil infrastrukturen bli overført til Sigma2, og inngå i det vanlige nasjonale tjenestetilbudet. NAIC-ressurser kjøpt inn med midler fra Forskningsrådet skal være gratis for norsk statlig finansiert forskning i maskinvarens levetid.

¹² Prop. 1 S (2024-2025), 239.

¹³ [Om prosjektet - NAIC – Norwegian Artificial Intelligence Cloud](#).

3.3.2 Forventet teknologisk utvikling

Teknologiutviklingen vil påvirke både hvilken type tungregning og hvor mye tungregneressurser det vil bli behov for. Exascale databehandling og kvantedatabehandling representerer to viktige fremskritt i utviklingen av høytytelsesdatabehandling, og de vil spille komplementære roller i fremtidens teknologiske landskap.

Eksperimentell infrastruktur for utforskning av exascale-beregninger (eX3)

Simula Research Laboratory og partnere har etablert eX3 med mål om å fylle rollen som en nasjonal ressurs som kan forberede exascale-beregninger i Norge. Den kommende generasjonen av regnekraft vil være basert på et sammensatt samspill mellom tusenvis av sofistikerte prosessorer, hver med et stort antall kjerner, hurtigminner organisert i mange lag og med kraftige akseleratorer. Alt dette vil være koblet sammen i komplekse nettverk for rask kommunikasjon. Slike maskiner vil være helt sentrale for det digitale samfunnet. Den totale kompleksiteten i et system som er konstruert for massive beregninger utført i parallell er en stor og sammensatt teknisk utfordring. Forskningsprosjektet har som mål å bidra til å bygge en sterk kompetanse på den kommende exascale-teknologien, slik at norsk forskning og næringsliv blir godt forberedt.

Prosjektet er finansiert av Forskningsrådets INFRASTRUKTUR-satsing med medfinansiering fra Simula og resten av konsortiet. Foruten Simula deltar Sigma2, forskningsgrupper fra NTNU, Universitetet i Bergen, Universitetet i Tromsø og OsloMet, samt teknologiselskapene Dolphin Interconnect Solutions og Numascale.¹⁴

Kvanteteknologi

En kartlegging Forskningsrådet gjorde på oppdrag fra Kunnskapsdepartementet i 2022, konkluderte med at det er et voksende kvanteteknologisk miljø til stede på internasjonale arenaer, med aktivitet særlig innen kvanteberegning, -simulering og -sensing. 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. Norge mangler også langsiktige industrielle investorer på feltet, som bidrar sterkt til finansiering i våre naboland.

I statsbudsjettet for 2025¹⁵ foreslår regjeringen å styrke forskningen på kvanteteknologi gjennom Forskningsrådet, slik at den årlige løyvinga på sikt blir 70 mill. kroner høyere enn dagens innsats. Satsingen skal innrettes slik at det blir bygd opp solide nasjonale fagmiljø av høy internasjonal kvalitet, som vil kunne konkurrere og lykkes på andre arenaer.

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.

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 Tsjeckia. Sigma2 er norsk partner i konsortiet sammen med SINTEF og Simula. Sigma2 har til nå gått inn med 2 millioner kroner i konsortiet.

¹⁴ [Experimental Infrastructure for Exploration of Exascale Computing - Prosjektbanken](#).

¹⁵ Prop. 1 S (2024-2025), 202.

3.3.3 Forventet utvikling av infrastruktur for tungregning og kunstig intelligens/maskinlæring i det private markedet i Norge

Plattformleverandører

Historisk sett har infrastrukturen for tungregning vært dominert av forskningsorganisasjoner og offentlig finansierte prosjekter. Denne dynamikken endrer seg imidlertid med den økende satsingen av kommersielle aktører. Det er en merkbar økning i kommersielle initiativer drevet av privat sektor. Selskaper på tvers av ulike bransjer, inkludert farmasøytiske produkter, bilindustri og finans, investerer i tungregning for å oppnå konkurransefortrinn. De etablerte store skyleverandørene (Google, Amazon, Microsoft, IBM) utvider sine tilbud med tungregnetjenester, og gjør avanserte dataressurser mer tilgjengelige for bedrifter i alle størrelser.

Tabell 3.1 gir en oversikt over hvilke tjenester som er tilgjengelig gjennom de store skyleverandørene.

Tabell 3.1 Tilgjengelig tjenester fra store skyleverandører

Leverandør	Tjenester	Infrastruktur i Norden
Amazon Web Services (AWS)	AWS tilbyr tungregnetjenester gjennom Amazon EC2, som inkluderer spesialiserte instanser for tungregning, som C5n og P4d. De tilbyr også AWS ParallelCluster for å sette opp og administrere tungregneklynger.	AWS har etablert datasenter i Stockholms-regionen, som en del av deres europeiske nettverk. AWS har kjøpt en tomt i Esbjerg, Danmark, som en del av deres planer om å utvide datasenterkapasiteten i Europa.
Google Cloud Platform (GCP)	GCP tilbyr tungregnetjenester gjennom Compute Engine, med spesialiserte instanser som N2 og C2. De tilbyr også Google Kubernetes Engine (GKE) for å kjøre tungregnearbeidsbelastninger.	GCP har et omfattende nettverk av datasentre globalt, inkludert i Europa. Google Cloud Platform (GCP) har etablert et datasenter i Hamina, Finland, samt kjøpt en tomt i Skien kommune for å utvide sin infrastruktur.
Microsoft	Microsoft tilbyr tungregnetjenester gjennom Azure Virtual Machines, spesielt deres H-serie som er optimalisert for tungregning. Azure Batch og Azure CycleCloud er også tilgjengelige for å administrere tungregnearbeidsbelastninger.	Microsoft har etablert datasentre i Norden, spesielt i Norge og Sverige. I Norge har de datasentre i Oslo-regionen, og i Sverige har de lokasjoner i Gävle, Sandviken og Staffanstorp.
IBM	IBM Cloud tilbyr tungregnetjenester gjennom deres Virtual Servers og Bare Metal Servers, som er optimalisert for tungregnearbeidsbelastninger. De tilbyr også IBM Spectrum LSF for jobbscheduling og ressursstyring.	IBM Cloud har datasentre i flere europeiske land, inkludert Norden. I Norge er datasenter i Fetsund det mest omtalte.

Det forventes at også andre aktører vil lansere KI-fabrikker. Som et eksempel kan vi løfte frem Telenor, som i november 2024 lanserte sin KI-fabrikk¹⁶. KI-fabrikken er designet for å hjelpe bedrifter med å akselerere KI-drevet transformasjon på tvers av ulike bransjer. Som en del av initiativet annonserte Telenor en innledende investering på 100 millioner kroner i februar for å bygge en KI-fabrikk ved å bruke NVIDIAs KI-databehandlingsplattform. Telenors KI-fabrikk har

¹⁶ [Telenor Launches AI Factory with Hive Autonomy as First Customer](#), 14. November 2024.

som mål å forbedre KI-adopsjon både internt og for eksterne kunder, samt tilby lokal KI-databehandlingskapasitet til Norden.

I henhold til en rapport fra Technavio¹⁷ forventes det skybaserte markedet for tungregning vokse med 9.4 prosent hvert år i perioden 2023-2028. I 2028 vil imidlertid fortsatt viksomhetsspesifikke og tungregnespesialiserte installasjoner dominere, med en markedsandel på ca. 14 prosent.

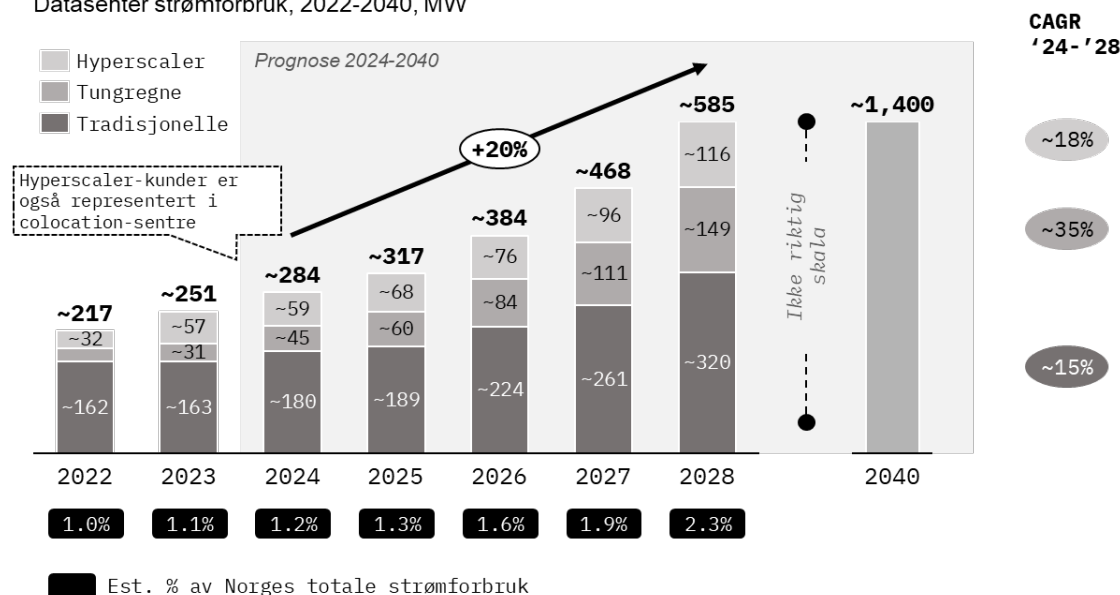
Datasenterleverandører

Norge og de andre nordiske landene er attraktive lokasjoner for å etablere tungregnetjenester på grunn av sitt kjølige klima og tilgang til rimelig, grønn energi. Et stabilt klima med lave temperatur gjør at nordiske datasentre kan oppnå en energieffektivitet som er mer enn 15 prosent lavere enn det man kan oppnå i andre land. Vannkraft som primær kilde til energiforsyning sikrer lave utslipp (36 gCO₂/kWh i Norge vs. 421 gCO₂/kWh i Tyskland) og er godt egnet for energiintensive datasentre.

Kapasiteten i datasenter blir angitt i hvor mye strømkapasitet som er installert. Figur 3.10 gir et estimat på forventet utbygging av datasentre i Norge.

Markedsutvikling av datasentermarkedet i Norge

Datasenter strømforbruk, 2022-2040, MW



Kilder: Norsk datasenter industri, NVE, Danseb Consulting Market Data, PwC Strategy& Analyse

Figur 3.10 Estimat over utviklingen av datasentermarkedet i Norge målt i strømforbruk

Analysen viser at den totale datasenterkapasiteten (målt i strømforbruk) i Norge vil mer enn doble seg i perioden 2024-2028. Den største veksten vil komme fra tungregnekunder der kapasiteten vil øke med 230 prosent, mens kapasiteten hos hyperscalers vil øke med 97 prosent.

¹⁷ Technavio. «Global High Performance Computing (HPC) Market 2024-2028».

3.3.4 Forventet utvikling av infrastruktur for tungregning og kunstig intelligens i Europa

Alle andre sammenlignbare land i Europa investerer betydelige offentlige midler i EuroHPC. Vi har innhentet informasjon fra Finland, Sverige, Danmark, Tyskland, Nederland og Frankrike om nivået på investeringer (historiske og planlagte) i tungregning.

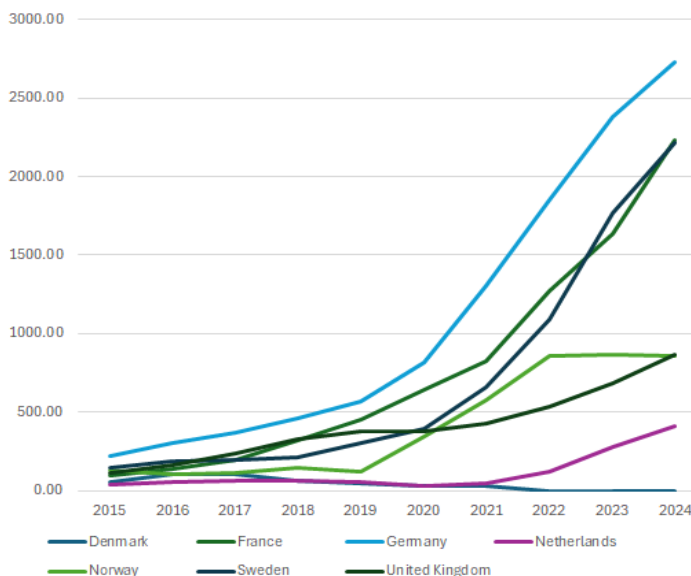
Generelt er det utfordrende å si noe samlet om investeringsnivå mellom landene, da alle land har ulike modeller for finansiering, og ulik grad av sentralisering av satsingen, samt ulikt nivå på investeringer fra private aktører. For eksempel har Finland og Tyskland en langt mer sentralisert modell enn 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, mens Nederland skal investere i EuroHPC-anlegg og dekke medfinansieringen i EuroHPC-prosjekter. Landene setter av betydelige summer fremover til investeringer i EuroHPC-maskinene.

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 regnekapasitet per innbygger for noen utvalgte land, noe som indirekte sier noe om investeringsnivå på tungregning for de enkelte landene. 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 per 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. 1,1 milliarder 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 3.11 Regnekraft per innbygger i utvalgte europeiske land

Kilde: Tallene er hentet fra <https://top500.org/lists/top500/2024/06> og <https://www.imf.org/en/Publications/WE0/weo-database/2024/April>.

4 Problembeskrivelse

I henhold til Finansdepartementets rundskriv R-108/23 Statens prosjektmodell - Krav til utredning, planlegging og kvalitetssikring av store investeringsprosjekter i staten skal problembeskrivelsen omfatte følgende:

Utgangspunktet for ethvert prosjekt er at man står overfor et problem som skal løses. Problembeskrivelsen skal gjøre rede for hvilke uløste problemer man ser på og hva som tilsier at det offentlige bør iverksette tiltak på området. Det skal fremkomme hva som er problemets omfang, hvor alvorlig det er og hvem som er eller blir berørt, uten å gå nærmere inn på hvordan problemet kan løses. Problembeskrivelsen må ta med både dagens problemer og forventet fremtidig utvikling. Det bør gjøres en vurdering av hva som er årsakene til at problemene har oppstått.

Det framtidige behovet er videre utdypet i kapittelet om interessentenes behov og etterspørselsbasert behov.

4.1 Manglende kapasitet for å dekke et økende behov

Som beskrevet over har Sigma2 gjennom investeringene de senere årene økt tilgjengelig kapasitet med 187 prosent. Dette er ikke tilstrekkelig for å dekke behovet for tungregning innen forskning. I undersøkelsen Forskningsrådet gjorde i forbindelse med rapporten som ble levert i august 2024, rapporterer flere universiteter at de har begrensede ressurser for både tradisjonelle beregningsoppgaver og beregninger som krever GPU-ressurser, og at det er et gap mellom hva som er reelt behov og hva forskningsgruppene har finansiering til å bruke.

Med investeringer som er gjort i 2024 vil Sigma2 i 2027-2028 kunne tilby en kapasitet som er 20 prosent høyere i 2025/2026 enn dagens. Uten videre investeringer vil kapasiteten i 2029 være 60 prosent lavere enn dagens.

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 imidlertid estimert et behov for en årlig vekst i kapasitet på 12-15 prosent for CPU, som tilsvarer en økning i perioden 2024-2029 på opp mot 75 prosent.

Behovet for GPU-kapasitet anslås å vokse 40-50 prosent årlig, som tilsvarer en økning i perioden 2024-2029 på opp mot 250 prosent. Det er knyttet størst usikkerhet knyttet til utviklingen av GPU-behov. Det innebærer at å holde investeringene på samme nivå som i dag, tilsier at gapet mellom behov og kapasitet vil fortsette å øke.

Årsaken til økningen i behov er sammensatt, én faktor handler om mer tilgjengelig data. Innen astrofysikk 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.

Økende tilfang av data, inkludert sanntidsdata, kombinert med ny teknologi som KI, gir muligheter for utvikling av flere og bedre tjenester basert på tungregning. Det trengs betydelig tungregnekapasitet både for å utvikle modeller og tjenester, men også til drift og vedlikehold av disse.

Utvikling av kunstig intelligens

Tungregning for KI har fram til nå vært svært marginalt, men flere ser en vridning mot KI de siste årene, og enkelte beskriver en stor økning i 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.

Offentlig forvaltning er i en KI-utprøvningsfase, men mangler tilgang til kostnadseffektiv KI-infrastruktur for utprøving og utvikling av KI-modeller som krever tungregneressurser som kan settes i produksjon.

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.

For KI vil behovet for tungregning og systemløsninger variere med ulike aktivitetstyper. For eksempel, utviklingen av grunnmodeller vil ha behov for stor regnekraft, tilpasninger av modeller med egne data kan dra nytte av lav kostnad, stor fleksibilitet og sikkerhet, mens bruk av ferdige modeller kan dra nytte av høy hastighet og høy tilgjengelighet.

Sikkerhet og beredskap

Det er et økt behov for tungregning for å ivareta sikkerhet og beredskap, som for eksempel å forebygge og håndtere helsekriser, geofarar og klimaendringer. I tillegg gir økt tilgang på data et uutnyttet potensial for mange samfunnsnyttige formål som overvåking av klimaendringer, samfunnsberedskap, og innen forsvar.

Et eksempel på dette er at det under pandemien var krevende å få gjennomført ønskede analyser på grunn av kapasitetsbegrensninger. Det er derfor behov for en langt større kapasitet som kan håndtere personsensitive helsedata med garantert kontinuerlig opptid før neste store helsekrise.

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 tungregning. Det er spesielt krevende å skaffe tilstrekkelig tungregning til klimamodellering, da de eksterne prosjektene ikke har økonomi til å finansiere behovet.

Bruken av skytjenester er økende innenfor offentlig forvaltning, både hos enkeltvirksomheter og på tvers av sektorer. Ønske om fleksibilitet, skalerbarhet og tilgang til moderne teknologi er viktige årsaker til at statlige virksomheter velger skytjenester når de skal utvikle eller anskaffe digitale tjenester. Samtidig eksisterer det ikke noe entydig, klart regelverk for bruk av skytjenester i staten.

Lavere nivå enn sammenlignbare land i Europa

Alle andre sammenlignbare land investerer betydelig mer offentlige midler i nasjonale tungregneressurser og EuroHPC. Nivået på fremtidige investeringer samt nåværende tungregneressurser er på et jevnt høyere nivå i land det er naturlig å sammenligne seg med. Ved å opprettholde et lavere investeringsnivå over tid vil norsk forskning som er avhengig av tungregning sakke akterut.

4.2 Dagens finansieringsmodell gir ikke nødvendig forutsigbarhet og tilpasningsevne

Som beskrevet over finansieres investeringene til Sigma2 i ny og oppgradering av eksisterende nasjonal tungregnekapasitet i dag gjennom Forskningsrådets INFRASTRUKTUR-ordning. I tillegg har Sigma2 anledning til å sette av overskudd fra driften til mindre investeringer. Både

utvalget som evaluerte INFRASTRUKTUR-ordningen¹⁸ og datainfrastrukturutvalget¹⁹ anbefaler at forskningsinfrastrukturer som krever regelmessige, store investeringer og påvirker hele forskningssystemet, bør finansieres sentralt og ikke gjennom konkurranseutsatte INFRASTRUKTUR-midler. Sentral finansiering vil sikre kontinuitet og begrense vanskelige prioriteringer mellom disse infrastrukturene og mer spesialiserte infrastrukturer. En nasjonal tungregneinfrastruktur som sikrer tilgang til regneressurser for en stor bredde av norsk forskning og i forvaltning vil ha behov for en langsiktig og forutsigbar finansiering. En større andel av finansieringen av denne type grunnleggende, generisk infrastruktur for tungregning bør derfor komme gjennom langsiktig finansiering. At en stor andel av finansieringen av slik infrastruktur er konkurranseutsatt skaper usikkerhet og risiko for tap av viktig kompetanse og en reell fare for at viktige forsknings- og forvaltningsoppgaver ikke kan gjennomføres. Nye krav til digital suverenitet, beredskap og sikkerhet taler også for en mer sektorovergripende politikk og finansiering. Med økte investeringer på området er det også nødvendig å innarbeide gode strukturer for kostnadseffektivitet.

Datainfrastrukturutvalget anbefaler at nasjonale infrastrukturer som mottar langsiktig finansiering, må ha klare mål for hvilke tjenester de skal levere, og jevnlig evalueres for å sikre at de leverer i tråd med målene som er satt for dem. Investeringer i nasjonale tungregneinfrastruktur bør sees i sammenheng med norsk deltakelse i internasjonale infrastruktursamarbeid.

Datainfrastrukturutvalgets anbefaling er at finansieringen av drift bør være en *kombinasjon* av

- *Langsiktig grunnfinansiering – stabil og forutsigbar finansiering.* Finansieringen kan for eksempel komme i form av grunnfinansiering fra ett relevant departement eller fra flere departementer etter en fordelingsnøkkel basert på hvilke forvaltningsområder infrastrukturen leverer forskning og forvaltning til. Slik finansiering kan også komme i form av spesifikke forvaltningsoppdrag eller gjennom forpliktende medlemskap eller partnerskap fra institusjoner som bruker infrastrukturen.
- *Brukerfinansiering eller brukerbetaling – finansiering som kommer fra de som bruker tjenestene.* Dette kan enten være i form av abonnement, stykkpris eller forbruksfinansiering, eller en kombinasjon av disse.

Ulike brukergrupper vil ha ulike krav til støttetjenester og spesielle tilpasninger (f.eks. ved bruk av sensitive data). For infrastrukturer som skal betjene brukergrupper innenfor svært ulike fagdisipliner (domener) og på tvers av sektorer (forskning, forvaltning og næringsliv), vil brukernes behov og krav til tjenestene som leveres være svært ulike. Ulike brukergrupper vil f.eks. ha ulike krav til støttetjenester og spesielle tilpasninger (f.eks. ved bruk av sensitive data) og ulik betalingsevne. Det vil derfor kunne være nødvendig med ulike modeller for brukerbetaling for ulike brukere. Lover, regelverk og internasjonale retningslinjer og avtaler kan også legge føringer på hvordan brukerbetalingsmodeller utformes. (Innenfor EU er det for en del infrastrukturtjenester ikke tradisjon for å kreve brukerbetaling. Norske forskningsmiljøer som har behov for slike infrastrukturtjenester for gjennomføring av sine aktiviteter i prosjekter finansiert av EU, vil da ikke ha midler gjennom prosjektfinansieringen til å betale for disse tjenestene.)

I tillegg til ren investering i tungregneressurser er det nødvendig å sikre forutsigbar finansiering for å best mulig utbytte av investeringen, både nasjonalt og internasjonalt. Norske forskningsorganisasjoner, særlig instituttene, mangler finansiering av faktiske kostnader i prosjektene fra DIGITAL. Dette skaper lite forutsigbarhet og konkurransedyktighet hos norske miljøer. Det er videre manglende ressurser til administrativ oppfølging av EuroHPC. Dette hindrer koordinering med relevante aktører, og medfører mangelfull oppfølging av det administrative arbeidet knyttet til aktiviteter i EuroHPC.

¹⁸ Forskningsrådet, «Evaluation of the INFRASTRUKTUR initiative as a funding instrument».

¹⁹ Forskningsrådet, «Investering i infrastrukturer for FAIR forskningsdata og særlig relevante forvaltningsdata for forskning».

4.3 Dagens organisering må tilpasses framtidige behov for kapasitet og kompetanse

Dagens organisering og styringsmodell er hovedsakelig innrettet for å dekke behovene til forskning og ikke offentlig forvaltning. Dette kan skape en unødvendig barriere for prosjekter som er i grenseland til utvikling og utprøving. Det er videre ikke etablert strategiske planer for å sikre en effektiv utbygging av tungregnekapasitet for kunstig intelligens i offentlig forvaltning. Det må vurderes om dagens drift som i stor grad er avhengig av ressurser hos BOTT-universitetene, er tilstrekkelig for å tilby tjenester også til offentlig forvaltning i større skala.

Utvikling og bruk av kunstig intelligens og tungregning er gjennomgripende innen et bredt spekter av fag og disipliner. Behovet er imidlertid ulikt både i forhold til regnebehov, datamengde og grad av sensitivitet. Med et økende antall brukere med ulike behov, samt at behovene vil endre seg i takt med den teknologiske utviklingen, er det nødvendig med større fleksibilitet, og muligheter til å ta raske beslutninger for videre utvikling. Dette er nødvendig for å ha et tilbud som er teknologisk oppdatert for å være relevant for dagens og fremtidens brukere.

Stadig flere fagfelt og sektorer håndterer skjermingsverdige data. Nødvendig sikkerhetsorganisering for tungregning krever spesiell kompetanse og spesiell tilrettelegging. Det må i større grad tas hensyn til variasjonen i datatyper, og at behovet for økte ressurser for behandling av beskyttelsesverdige data hensyntas der dette er nødvendig.

Krav til tilgjengelighet vil variere mellom ulike aktiviteter. Forskning og utvikling vil ofte prioritere lav pris framfor høy oppetid, mens oppetid vil kunne være viktig ved beredskapsaktivitet og kontinuerlige driftsoppgaver. Det er viktig å ivareta muligheten for differensierte tilgjengelighetsnivå og pris basert på krav til oppetid.

Med flere brukere involvert er det nødvendig med forutsigbare kriterier for tildeling av tungregneressurser, samt behov for å ha gode mekanismer for deling av regneressursene, for å sikre best mulig ressurseffektivitet.

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.

Dagens støttetjenester er ikke tilstrekkelig tilpasset/dimensjonert for å svare på alle ulike brukergruppers behov for bedre utnyttelse av regneressursene både nasjonalt, men spesielt ressursene vi får tilgang til gjennom EuroHPC.

Dagens finansieringsmodell gjør det vanskelig å fullt ut utnytte mulighetene i EuroHPC. I dag finansieres alle investeringer Sigma2 gjør, inkludert norsk andel av regneressurser i EuroHPC-anlegg, gjennom Forskningsrådets finansieringsordning for forskningsinfrastruktur. Forskningsrådets utlysninger er ikke tilpasset aktiviteter og utlysninger som skjer i regi av EuroHPC. For å kunne gi de nødvendige garantiene som EuroHPC krever ved nye investeringer eller deltakelse i nye aktiviteter (som KI-fabrikk), må den norske forvalteren av norsk deltakelse i EuroHPC ha større forutsigbarhet når det gjelder finansiering enn den konkurransebaserte tildelingen fra Forskningsrådet gir. Det må også være større fleksibilitet når det gjelder hvilke investeringer som skal gjøres.

Nærhet mellom bruker og tungregneressurser er viktig, særlig for mindre erfarne brukere. Det må derfor i større grad bygges kompetanse både i forskningsmiljøene og i offentlig sektor og næringsliv for å enklere kunne realisere samarbeid mellom ulike fagmiljøer fra ulike sektorer.

For god utnyttelse av kapasitet og kompetanse bør det være en god kopling både mellom lokale og nasjonale ressurser 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»²⁰ bør norskeide regneressurser samordnes bedre for å kunne utnytte kapasitet og kompetanse i kriser.

4.4 Næringsliv

Når det gjelder næringslivet viser til rapporten fra Samfunnsøkonomisk analyse, som i 2023 gjennomførte en verdiskapingsanalyse basert på nasjonalregnskapstall, en spørreundersøkelse rettet mot virksomheter (totalt ca. 5 300 respondenter) og om lag 20 dybdeintervjuer med private virksomheter²¹. 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, 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.

Disse funnene sammenfaller med funnene fra Statistisk sentralbyrås kartlegging av hvilke hindringer næringslivet står ovenfor i bruk av kunstig intelligens²². Nesten syv av ti virksomheter angir mangel på kompetanse som den største hindringen, mens mer enn fire av ti angir utfordringer med tilgangen eller kvaliteten på nødvendige data. Bekymringer om brudd på datavern og personvern og manglende klarhet om de juridiske konsekvensene er også vanlige hindringer.

4.5 Oppsummering av problembeskrivelsen

Tema	Oppsummering
Kapasitet og finansiering	<ul style="list-style-type: none">• Det er i dag begrensede ressurser både innen tradisjonelle beregningsoppgaver og beregninger som krever GPU-ressurser. Det er derfor et gap mellom hva som er reelt behov og hva forskningsgruppene har finansiering til å bruke.• 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, som tilsvarer en økning i perioden 2024-2029 på opp mot 75 prosent.• Behovet for GPU-kapasitet er usikkert, men anslås å vokse 40-50 prosent årlig, som tilsvarer en økning i perioden 2024-2029 på opp mot 250 prosent.• Det innebærer at å holde investeringene på samme nivå som i dag, tilsier at gapet mellom behov og kapasitet vil fortsette å øke.• Årsaken til økningen i behov er større datamengder, og økt bruk av KI i alle sektorer.• Det er behov for økt kapasitet for å ivareta sikkerhet og beredskap.

²⁰ Forskningsrådet, «Forslag til rammeverk for rask oppskalering av analysekapasitet og -infrastruktur i kriser».

²¹ Samfunnsøkonomisk analyse, «Kunstig intelligens i Norge – nytte, muligheter og barrierer».

²² [13272: Hindringer for bruk av kunstig intelligens-teknologi, etter sysselsetting og næring \(SN2007\) \(prosent\) 2021 - 2024. Statistikkbanken.](#)

- Andre sammenlignbare land investerer betydelig mer offentlige midler i nasjonale tungregneressurser og EuroHPC.
-

Organisering

- Dagens organisering og styringsmodell er hovedsakelig innrettet for å dekke behovene til forskning og ikke offentlig forvaltning. Dette kan skape en unødvendig barriere for prosjekter som er i grenseland til utvikling og utprøving.
 - Det må vurderes om dagens drift som i stor grad er avhengig av ressurser hos BOTT-universitetene er tilstrekkelig for å tilby tjenester også til offentlig forvaltning i stor skala.
 - Det er ikke etablert noen formelle strukturer for å sikre en effektiv utbygging av tungregnekapasitet for kunstig intelligens i offentlig forvaltning.
 - Det er behov for bedre støttetjenester som er tilpasset ulike brukergruppers behov for bedre utnyttelse av regneressursene både nasjonalt, men spesielt ressursene vi får tilgang til gjennom EuroHPC.
 - Det er behov for å bygge kompetanse både i forskningsmiljøene og i offentlig sektor og næringsliv for å enklere kunne realisere samarbeid mellom ulike fagmiljøer fra ulike sektorer (forskning, offentlig sektor og næringsliv).
 - Norskeide regneressurser bør samordnes bedre for å kunne utnytte kapasitet og kompetanse i kriser.
-

5 Behovsanalyse

Hensikten med behovsanalysen er å avdekke de reelle behovene som skal gi et godt grunnlag for å anbefale en konseptuell løsning. Iht. Finansdepartementets rundskriv R-108/23 Statens prosjektmodell - Krav til utredning, planlegging og kvalitetssikring av store investeringsprosjekter i staten, skal behovsanalysen omfattende følgende:

Behovsanalysen skal beskrive bredden i aktuelle, konkrete behov relatert til problembeskrivelsen, vurdert i et overordnet samfunnsperspektiv. Den skal inneholde en kartlegging av relevante interessenter/aktører i en interessentanalyse. Analysen skal få frem hvem som berøres av tiltaket og avdekke interessekonflikter. Analysen må inneholde en vurdering av styrken i de ulike identifiserte behovene og det må fremkomme hvilket behov som skal legges til grunn for den videre utredningen.

Kapittelet er iht. rådende konseptvalgmetodikk delt inn i normative, etterspørselsbaserte og interessentenes behov. I dette kapittelet beskrives 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 tungregning og for regnekraft optimalisert for KI.

5.1 Normative behov

De normative behovene er behov som er forankret i politiske mål, lover og forskrifter og deles inn i politiske og juridiske behov. Politiske drevne behov knytter seg til politiske føringer og strategier. Disse føringene og strategiene er ofte formulert i ulike nasjonale og internasjonale politiske dokument som regjeringen stiller seg bak. De juridiske behovene knytter seg i første rekke til juridiske og regulatoriske bestemmelser for bruk av data.

5.1.1 Politiske behov

De politiske behovene belyses gjennom politiske føringer og strategier som tar for seg planlagte og ønskede endringer og tiltak. Gjennom disse politiske strategiene og føringene kommer det frem hvilke behov samfunnet har i dag og i fremtiden som kommer til uttrykk i både stortingsmeldinger, strategier og langtidsplaner, blant annet:

1. Investering i infrastrukturer for FAIR forskningsdata og særlig relevante forvaltningsdata for forskning (Datainfrastrukturutvalget, 2022, innspill til Langtidsplan for forskning og høyere utdanning 2023-2024).
2. Langtidsplan for forskning og høyere utdanning 2023–2032 (Meld. St. 5 (2022–2023))
3. Forsvarsløftet – for Norges trygghet. Langtidsplan for forsvarssektoren 2025–2036 (Prop. 87 S (2023-2024))
4. Nasjonal strategi for kunstig intelligens (2020)
5. Strategi for norsk deltagelse i Programmet for et digitalt Europa – DIGITAL (2024)
6. Fremtidens digitale Norge – Nasjonal digitaliseringsstrategi 2024 – 2030
7. Nasjonal helse- og samhandlingsplan 2024-2027 (Meld. St. 9 (2023–2024))
8. Strategi for å øke næringslivets investeringer i forskning og utvikling

Investering i infrastrukturer for FAIR forskningsdata og særlig relevante forvaltningsdata for forskning²³

Utvalget har følgende anbefaling om ambisjonsnivå for datainfrastruktur i Norge:

- Forskere ved norske universiteter og institutter skal ha tilgang til datainfrastrukturer som muliggjør forskning og utdanning i verdensklasse.
- Innen 2030 skal alle fagområder i Norge tilbys kompetanse, veiledning og kuratering av forskningsdata, enten i form av nasjonale løsninger eller helt eller delvis gjennom deltakelse i europeisk eller internasjonalt infrastruktursamarbeid.
- Norge skal på utvalgte områder ha datainfrastrukturer i verdensklasse som blir foretrukket av internasjonale brukere.
- Norge må ha en opptrappingsplan for organisering og finansiering av datainfrastrukturer som gjør det mulig å hente ut gevinstene av de store datamengdene som vil genereres med offentlig finansiering i årene som kommer.

Langtidsplan for forskning og høyere utdanning 2023–2032 (Meld. St. 5 (2022–2023))²⁴

Regjeringens vurdering er at datainfrastrukturutvalgets foreslåtte ambisjonsnivå er realistisk, samtidig som det gir noe å strekke seg etter. Det er allerede investert mye i datainfrastruktur i Norge som det kan bygges videre på, og det kan samordnes bedre nasjonalt og internasjonalt. Den digitale forskningsgrunnmuren, som består av forskningsnett og superdatamaskiner, må videreutvikles i takt med forskningens og forvaltningens behov.

Regjeringen vil sørge for tilstrekkelig nasjonal tungregnekapasitet (superdatamaskiner) til å dekke nåværende og fremtidige behov i forskning og forvaltning, og legge datainfrastrukturutvalgets anbefalinger til grunn for videre arbeid med datainfrastruktur.

Stortinget har i Innst. 170 S (2022–2023)²⁵ 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.

Ett av målene i Langtidsplanen er «styrket konkurransekraft og innovasjonsevne». Regjeringen uttaler at de vil legge bedre til rette for at både næringslivet og offentlig sektor skal utforske og ta i bruk kunnskap og teknologi. Regjeringen uttaler videre at de vil legge frem en strategi for at forskning og utvikling i næringslivet skal utgjøre 2 prosent av BNP innen 2030 (se nedenfor), og legge mer vekt på å støtte næringslivets egne FoU-prosjekter.

Forsvarsløftet – for Norges trygghet. Langtidsplan for forsvarssektoren 2025–2036²⁶

Regjeringen vil utnytte potensialet i de store datamengdene som produseres til å videreutvikle forsvarssektoren og utvikle forsvarsevnen. 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.

Nasjonal strategi for kunstig intelligens²⁷

²³ Forskningsrådet, «Investering i infrastrukturer for FAIR forskningsdata og særlig relevante forvaltningsdata for forskning», 6.

²⁴ Meld. St. 5 (2022–2023), 89f.

²⁵ Innstilling fra utdannings- og forskningskomiteen om Langtidsplan for forskning og høyere utdanning 2023-2032.

²⁶ Prop. 87 S (2023-2024).

²⁷ Kommunal og moderniseringsdepartementet, «Nasjonal strategi for kunstig intelligens», 30.

Regjeringen vil legge til rette for at Norge skal ha infrastruktur for KI i verdensklasse, i form av digitaliseringsvennlig regelverk, gode språkressurser, raske og robuste kommunikasjonsnett og tilstrekkelig regnekraft.

Utvikling og bruk av KI krever god kommunikasjonsinfrastruktur og tilgang til regnekraft. Tilgang til tilstrekkelig regnekraft skal sikres gjennom bruk av nasjonale og internasjonale ressurser for tungregning.

På mange forsknings- og forvaltningsområder er det en sterk vekst i mengden data. Innenfor forskning er beregninger og datadrevet forskning blitt viktige arbeidsmåter, ved siden av eksperimenter og teoriutvikling. Dette gir et økende behov både for lagring og tilgjengeliggjøring av data, men også for regnekraft til å behandle dataene.

Strategi for norsk deltagelse i Programmet for et digitalt Europa – DIGITAL²⁸

Norge deltar i Programmet for et digitalt Europa (DIGITAL) 2021–2027 som er EUs investerings- og kapasitetsbyggingsprogram for digital omstilling og bruk av innovative digitale teknologier i samfunn og næringsliv. DIGITAL er et viktig verktøy for å realisere ambisjonene knyttet til digitalisering og digital transformasjon som blant annet er formulert i EUs strategier «Det digitale tiåret 2030»²⁹, «EUs hvitebok om kunstig intelligens» (2020), og «EUs datastrategi» (2020).

I april 2024 la regjeringen fram Strategi for norsk deltagelse i Programmet for et digitalt Europa – DIGITAL. Strategien viser hvordan regjeringen vil legge til rette for at Norge skal få mest mulig igjen for denne deltagelsen. Ett av seks innsatsområder i DIGITAL-programmet er «tungregning og superdatamaskiner».

Regjeringen uttaler i strategien at de vil fortsette å legge til rette for at norske aktører skal ha tilgang til tungregnekapasitet gjennom EuroHPC. Regjeringens ambisjon er å stimulere til bærekraftig og etisk forsvarlig datadrevet forskning, forvaltning og verdiskaping i Norge ved å arbeide for oppdatert datainfrastruktur. Regjeringen vil at det bygges videre på tidligere investeringer i datainfrastrukturen, og at det jobbes for å se tungregnekapasiteten nasjonalt i sammenheng med hva norske aktører kan få tilgang til internasjonalt, og legge til rette for videreføring og videreutvikling av Norges oppkobling til europeiske fellesløsninger for grenseoverskridende tjenester. Regjeringen vil legge datainfrastrukturutvalgets anbefalinger til grunn for videre arbeid med datainfrastruktur.

Fremtidens digitale Norge – Nasjonal digitaliseringsstrategi 2024 – 2030³⁰

Regjeringen vil frem mot 2030 få på plass en nasjonal infrastruktur for kunstig intelligens (KI), og Norge skal være i front på etisk og trygg bruk av KI. Næringslivet skal ha gode rammevilkår for å utvikle og bruke KI. Offentlig sektor skal anvende KI for å utvikle bedre tjenester og løse oppgaver mer effektivt.

Utvikling og bruk av KI krever god kommunikasjonsinfrastruktur, kvalitetsdata og tilgang til tilstrekkelig regnekraft. Behovet for superdatamaskiner og tilhørende infrastruktur for tungregning øker sterkt i forbindelse med utvikling av KI, og dekkes gjennom bruk av både nasjonale og internasjonale ressurser for tungregning. Ikke bare store bedrifter, men også oppstartsbedrifter og SMB-er trenger tilgang til tungregneressurser.

Teknologirådet har lagt frem en rapport der de anbefaler flere tiltak for trygg og hensiktsmessig bruk og utvikling av KI i Norge³¹. Det er behov for å etablere en nasjonal infrastruktur for KI som blant annet inkluderer tilgang til regnekraft og språkmodeller som er tilpasset norske og samiske språk og norske samfunnsforhold.

²⁸ Digitaliserings- og forvaltningsdepartementet, «Strategi for norsk deltagelse i Programmet for et digitalt Europa – DIGITAL», 18.

²⁹ EUs betegnelse er «Path to the Digital Decade», omtales som Digital Decade Policy Programme 2030 (COM (2021) 574 final).

³⁰ Digitaliserings- og forvaltningsdepartementet, «Fremtidens digitale Norge – nasjonal digitaliseringsstrategi 2024–2030».

³¹ Teknologirådet, «Generativ kunstig intelligens i Norge».

Superdatamaskiner utgjør en nødvendig infrastruktur for utvikling av norsk KI. Det må svært stor regnekraft til for å trene opp grunnmodeller som næringslivet eller offentlig sektor deretter kan videreutvikle for å lage og bruke KI-verktøy for mer effektive tjenester og mer effektiv produksjon.

Nasjonal helse- og samhandlingsplan 2024-2027 (Meld. St. 9 (2023–2024))³²

Planen peker på at *kunstig intelligens (KI) vil kunne utgjøre et betydelig bidrag til en bærekraftig utvikling av vår felles helsetjeneste*. KI kan bidra til å automatisere arbeidsoppgaver og forbedre arbeidsprosesser, noe som er essensielt for å møte fremtidige utfordringer innen helsepersonellkapasitet og tjenestekvalitet.

Det er utarbeidet en Felles KI-plan for trygg og effektiv bruk av KI i helse og omsorgstjenesten 2024 - 2025³³ for å følge opp stortingsmeldingen. Det overordnede målet med planen er å øke bruken av trygge og effektive KI-systemer i helse- og omsorgstjenesten. Dette skal bidra til tjenester av like god eller bedre kvalitet og frigjøre tid for helsepersonell.

I planen er det et mål om å etablere effektive og fleksible måter å gjøre trening, finjustering og validering av KI-modeller på og som bidrar til trygge løsninger for hele helse- og omsorgstjenesten. Behovet for slik infrastruktur gjelder ulike typer KI-systemer, også språkmodeller. Det kan være ulike behov til hhv. forskning, validering og drift. Arbeidet må gjøres i god dialog med helsetjenestene som har oversikt over mulighetene og avgrensingene i dagens infrastruktur.

Strategi for å øke næringslivets investeringer i forskning og utvikling³⁴

Regjeringen ønsker å legge til rette for FoU i næringslivet i hele landet. Selv om regjeringen har satt et mål for næringslivets forskningsinnsats, vil innsatsen til sist være et resultat av beslutninger tatt av næringslivet selv. Private virksomheter, gjennom egne lønnsomhetsvurderinger, har hovedansvaret for forskning og utvikling i det private næringslivet. Myndighetene har først og fremst mulighet til å påvirke bedriftenes forskningsinvesteringer gjennom rammebetingelser, infrastruktur, regelverk og tilgang til data.

Ett av tiltakene i strategien er at regjeringen vil legge frem et veikart for teknologinæringen som skal bidra til å gi næringslivet gode forutsetninger for å benytte og håndtere muliggjørende og dype teknologier.

5.1.2 Juridiske behov

De juridiske behovene beskriver sentrale lover for forskrifter, og knytter seg i denne sammenheng i første rekke til juridiske og regulatoriske bestemmer for bruk av data som krever tungregning i forskning, og for å levere samfunnsnyttige og samfunnskritiske tjenester. Særlig relevant i denne sammenheng er:

1. Sikkerhetsloven
2. Personopplysningsloven inkl. personvernforordningen (GDPR)

Sikkerhetsloven

Den nye sikkerhetsloven trådte i kraft 1. januar 2019. Den nye loven er tilpasset den teknologiske utviklingen, og vil også sikre informasjon som sendes digitalt. Loven omfatter informasjonssystemer, infrastruktur og objekter av sentral betydning for nasjonal sikkerhet. Loven regulerer både informasjonssystemssikkerhet (kapittel 6) og objekt- og

³² Meld. St. 9 (2023–2024).

³³ Helsedirektoratet, «Felles KI-plan for trygg og effektiv bruk av KI i helse og omsorgstjenesten 2024 – 2025».

³⁴ Nærings- og fiskeridepartementet/Kunnskapsdepartementet, «Strategi for å øke næringslivets investeringer i forskning og utvikling».

infrastruktursikkerhet (kapittel 7) som er relevant for tungregningsinfrastruktur og dataene som behandles i denne infrastrukturen.

Personopplysningsloven inkl. personvernforordningen (GDPR)

Personvern handler om retten til et privatliv og retten til å bestemme over egne personopplysninger. Denne rettigheten balanseres mot at legitime formål ofte krever behandling av personopplysninger. Personopplysningsloven og GDPR åpner for gjenbruk av personopplysninger til arkivformål og forsknings- og statistikkformål som ikke ansees som uforenelig med de opprinnelige behandlingsformålene. Samtidig stiller personopplysningsloven tydelige krav til behandlingen av personopplysninger. For å kunne benytte data må det foreligge behandlingsgrunnlag gjennom for eksempel en lovhjemmel eller at den registrerte samtykker om bruken. Bruk av data som inneholder personopplysninger kan medføre behov for skjermet infrastruktur.

5.2 Interessentenes behov

En interessentanalyse er utført for å avdekke hvilke interessenter som berøres av tiltakene i konseptvalgutredningen og deres opplevde behov, samt for å avdekke mulige interessekonflikter. Interessentene er delt inn i to undergrupper: Primær- og sekundærinteressenter. Primærinteressenter er aktører som har behov eller krav knyttet direkte til tungregning og som vil bli berørt av tiltak i denne konseptvalgutredningen. Sekundære interessenter er aktører som indirekte blir berørt av tiltakene gjennom positive og/eller negative indirekte effekter.

Det foreligger ingen fullstendig kartlegging av nåværende og fremtidig behov for tungregning blant forskningsorganisasjoner, offentlige og private virksomheter i Norge. Det ble utført en kartlegging av behovene for tungregning i forskningsorganisasjoner og offentlige institutter i forbindelse med rapporten som Forskningsrådet leverte i august 2024. I denne konseptvalgutredningen er det gjort en supplerende kartlegging blant offentlige virksomheter og enkelte private virksomheter for å øke innsikten i behovet for tungregning.

5.2.1 Oversikt over interessenter

Primærinteressentene og til dels sekundærinteressentene i denne konseptvalgutredningen er omtalt i beskrivelsen av nåsituasjon i kapittel 3, og problembeskrivelsen i kapittel 4. Tabell 5.1 gir en skjematisk inndeling av de mest relevante interessentgruppene.

Tabell 5.1 Oversikt over interessentgrupper

Interessentgrupper	Interessenter
Primærinteressenter	
Brukere av tungregning	Forskningsorganisasjoner ³⁵ Forskningsorganisasjoner omfatter: <ul style="list-style-type: none">• Universiteter, vitenskapelige høyskoler, høyskoler akkreditert på institusjonsnivå av NOKUT• Virksomheter omfattet av regjeringens Strategi for helhetlig instituttpolitikk og virksomheter omfattet av Retningslinjer for statlig grunnfinansiering av forskningsinstitutter og forskningskonsern.

³⁵ Liste over forskningsorganisasjoner:
<https://www.forskningsradet.no/finansiering/forskningsorganisasjoner/godkjente-forskningsorganisasjoner/>.

- Helseforetak/sykehus med lovpålagte oppgaver innenfor forskning og utvikling og private, ideelle sykehus som inngår i Helse- og omsorgsdepartementets målesystem for forskning

Offentlig forvaltning

Offentlig forvaltning omfatter virksomheter som utfører og ivaretar forvaltningsmessige oppgaver og som er en del av offentlig sektor. Det omfatter både departementer og underliggende virksomheter som direktorater, tilsyn, forvaltningsbedrifter, forvaltningsorganer med særskilte fullmakter, ombud. I tillegg inngår stiftelser, særlovselskap og enkelte statsforetak og statsaksjeselskap.

I omtalen av brukerens behov for tungregning nedenfor har vi valgt å omtale offentlige institutter som har et annet hovedformål enn FoU, samt forskningsinstitutter som får driftsbevilgning direkte fra departementene³⁶, under kategorien offentlig forvaltning.

Næringsliv

Næringslivet består av bedrifter og næringsdrivende som produserer varer og tjenester i Norge. Det omfatter både private og offentlige virksomheter, dvs. hovedvekten av statsforetak og statsaksjeselskap som f.eks. Equinor og Telenor.

Leverandører av tungregning	<p>På leverandørsiden skiller vi mellom plattformleverandører og datasenterleverandører, hvor både private og offentlige leverandører inngår.</p> <p>Plattformleverandører</p> <ul style="list-style-type: none"> • Sky-plattformleverandører (Amazon, Google, Microsoft m.fl.) • Spesialiserte tungregnetjeneste-leverandører (Sigma2, EuroHPC) • Virksomhetsintern leverandør (virksomhetene leverer tungregning til egen virksomhet) <p>Datasenterleverandører</p> <ul style="list-style-type: none"> • Leverandører av datasentertilpasset tungregning (BULK, Lefdal Mining Center m.fl.) <p>Universitetene i Bergen, Oslo, Tromsø og Trondheim (BOTT) involvert på leverandørsiden gjennom samarbeid med Sigma2 AS om å drifte de nasjonale tungregnetjenestene (Norwegian Research Infrastructure Services).</p>
Beslutningstakere knyttet til tungregning	<p>Beslutningstakere er personer eller grupper som har myndigheten til å ta avgjørelser som påvirker offentlig politikk, tjenester og ressursfordeling. Det omfatter både lovverk og regulering, offentlige budsjetter, infrastruktur, planer og strategier m.m.</p> <p>Relevante beslutningstakere knyttet til tungregning i Norge omfatter:</p> <ul style="list-style-type: none"> • Politikere i Norge (Stortinget og regjeringen) • Departementer • EU
Finansierer av tungregning	<p>Investeringer i tungregning – infrastruktur, forskning, utvikling, tjenestetilbud m.m. – finansieres av offentlige og private aktører.</p>
Sekundærinteressenter	
Interesseorganisasjoner	<p>Det eksisterer flere interesseorganisasjoner og interessegrupper i Norge som er opptatt av tungregning. Et utvalg av relevante interesseorganisasjoner er som følger:</p> <ul style="list-style-type: none"> • Forskningsinstituttens fellesarena (FFA) • Universitets- og høyskolerådet (UHR) • IKT-Norge • Abelia

³⁶ Iht. [Strategi for helhetlig instituttpolitikk](#) gjelder dette Norsk polarinstitutt, Meteorologisk institutt, Nasjonalt folkehelseinstitutt, Norges geologiske undersøkelse, Norges vassdrags- og energidirektorat, Arkivverket, Nasjonalbiblioteket, SSBs forskningsvirksomhet, Havforskningsinstituttet, STAMI, Simula, og Forsvarets forskningsinstitutt.

- DigiCat (Norsk Katapultsenter)
- Digital Norway
- Næringslivets hovedorganisasjon (NHO)

Aktører utenfor Norge Det er flere offentlige og private aktører som er involvert i utviklingen av tungregning i europeiske land, utover EuroHPC som omtalt i kapittel 3.2. Et utvalg av relevante aktører er som følger:

- DeIC (Danish e-infrastructure Consortium), Danmark
 - Uddannelses- og Forskningsministeriet, Danmark
 - CSC (IT Center for Science), Finland
 - Finlands Akademi, Finland
 - NAISS (National Academic Infrastructure for Supercomputing), Sverige
 - Vinnova (Sweden's innovation agency), Sverige
 - Vetenskapsrådet, Sverige
 - Federal Ministry of Education and Research, Tyskland
 - Ministry of Economic and Climate Affairs, Nederland
 - Direktoratet for forskning og innovasjon (Dgri), Frankrike
-

5.2.2 Forskningsorganisasjonens behov for tungregning

Forskningsorganisasjoner benytter primært tungregning til å håndtere komplekse beregninger og store datamengder til forskning, innovasjon og utvikling. Dette omfatter både tradisjonelle beregningsoppgaver, beregningsoppgaver relatert til kunstig intelligens (KI) og andre oppgaver.

Tradisjonelle beregningsoppgaver er komplekse simuleringer og analyser som krever betydelig regnekraft for å modellere prosesser på en detaljert og nøyaktig måte. Begrepet "tradisjonell" er knyttet til at man bruker etablerte metoder som er basert på klassiske matematiske og fysikkbaserte tilnærminger, i motsetning til nyere og mer datadrevne metoder som maskinlæring. Fremover vil det fortsatt være behov for tungregning til tradisjonelle beregningsoppgaver, samtidig som det er ventet at regnekraften til slike beregninger i økende grad må tilpasses nye maskinarkitekturer. Det innebærer at det forventes at behovet for tungregning flytter seg fra systemisk enklere prosesser (enkeltprosesser) til parallellprosessering og arbeidsflytprogrammer som er knyttet til å løse enten samfunnsutfordringer eller sammensatte problemstillinger i industrien som kobler ulike prosesser sammen.

Flere fagfelt benytter tungregning til tradisjonelle beregningsoppgaver i dag, som også vil gjelde fremover. Fagfelt som har et høyt behov for tungregning i dag er ulike grener innenfor fysikkfaget slik som astrofysikk³⁷ og partikkelfysikk, bl.a. til å utforske universets strukturer og utvikling, simulering av partikkelinteraksjoner. Innen kjemifaget brukes tungregning bl.a. for analyse av ulike kjemiske prosesser³⁸, numerisk analyse av kjemisk omsetning, optimalisering og drift av kjemiske reaktorer. Innen klimaforskning og atmosfærefysikk brukes tungregning bl.a. til å modellere klimaendringer, værprognoser og studier av atmosfæriske prosesser slik som skydannelse, strålingsbalanse og atmosfærisk kjemi. Mens det innen medisin bl.a. brukes til å forstå genetiske sykdommer og utvikle nye behandlingsmetoder. Innen transportfaget utvikles modeller som simulerer trafikkflyt og kødannelse, og innen materialvitenskap brukes tungregning bl.a. utvikle bærekraftige teknologier og nye materialer. Energi- og miljøforskning bruker tungregning bl.a. til å modellere fornybare energikilder og optimalisering av energisystemer, mens innen marin forskning brukes tungregning bl.a. til å simulere havstrømmer og økosystemer.

Når det gjelder *kunstig intelligens* (KI), benytter forskningsorganisasjoner tungregning til både forskning, utvikling og anvendelse av KI. KI omfatter mange teknologier, fra maskinlæring og

³⁷ Innen astrofysikk er [Rosseland senter for solfysikk \(RoCS\)](#) er stor bruker av tungregning.

³⁸ Innen kjemi er [Hylleraas-senteret](#) en stor bruker av tungregning.

dyp læring til naturlig språkbehandling og datamaskinsyn, og feltet utvikler seg raskt med nye metoder og anvendelser. På tvers av teknologi, er det behov for tungregning i flere faser av utviklingen av kunstig intelligens. Overordnet kan vi dele inn utviklingen av kunstig intelligens i utvikling, trening, tilpasning og bruk. I utviklingsfasen blir ideer og konsepter for KI-modeller utviklet, som inkluderer blant annet forskning, design av algoritmer og utvikling av prototyper. Her er det et stort behov for regnekraft i simuleringer for å teste og validere algoritmer og til å utføre eksperimenter m.m. Når KI-modellene er utviklet, blir modellene trent ved hjelp av store datasett. Det innebærer å justere modellparametere for å optimalisere ytelsen. Her er det et meget stort behov for regnekraft, ofte ved bruk av GPU, til å håndtere store mengder data og komplekse beregninger. Etter at KI-modellene er trent, kan de tilpasses spesifikke bruksområder eller datasett. Det innebærer å finjustere KI-modellen med nye data, gjøre justeringer i algoritmer osv. Her varierer behovet for tungregning alt ettersom omfanget av tilpasningene og størrelsen på de nye datasettene. Etter at KI-modellene er utviklet, testet og evt. tilpasset, tas KI-modellene i bruk i ulike applikasjoner og software. Her vil normalt behovet for tungregning være lavere enn i foregående stegene, hvor behovet vil variere etter modellenes kompleksitet, responstid, bruksområde m.m. I bruksfasen av modellene kan det også være behov for å videreutvikle og ettertrene modellene, som også kan kreve tungregning.

På flere universiteter og institutter i Norge er det forskningsmiljøer som arbeider både med grunnforskning og anvendt forskning innenfor KI. Det eksisterer også spesialiserte forskningssentre som er rettet mot KI. Et fellestrekk ved disse er at det er et stort behov for tungregning til å utvikle, teste og validere KI-algoritmer, naturlig språkbehandling og utvikling av språkmodeller, bildeanalyser, simuleringer og modelleringer av komplekse systemer m.m. Flere av forskningsmiljøene og sentrene samarbeider med både offentlige virksomheter og næringslivet. Noen av de største aktørene i Norge er som følger:

- Norsk Konsortium for Kunstig Intelligens Forskning (NORA) som er et norsk samarbeid mellom åtte universiteter, fem høyskoler og fem forskningsinstitutter innen KI, maskinlæring og robotikk.³⁹ NORA har som mål å styrke norsk forskning, utdanning og innovasjon innen kunstig intelligens, maskinlæring og robotikk, samt annen relevant forskning som støtter utviklingen av applikasjoner for kunstig intelligens.
- Norwegian Open AI Lab er et forskningssamarbeid ledet av NTNU, som fremmer forskning og utvikling innen kunstig intelligens, med fokus på maskinlæring og dyp læring. Laben er et samarbeid mellom NTNU, Telenor og flere andre industripartnere. Laben fokuserer på både grunnleggende forskning og anvendte prosjekter innen områder som helse, kommunikasjon og industri.
- Norsk forskningssenter for KI-innovasjon (NorwAI) er en stor akademiske satsing på KI-innovasjon i Norge. Senteret er finansiert av Forskningsrådet gjennom SFI-ordningen (Sentre for forskningsdrevet innovasjon) og ligger ved NTNU og koordinerer forsknings- og innovasjonsaktiviteter mellom tre universiteter, to forskningsinstitutter og 11 bedrifter. Senteret arbeider med å utvikle avanserte KI-løsninger for ulike sektorer, inkludert helse, energi, og finans, og har som mål å fremme innovasjon og bærekraftig utvikling gjennom KI.
- Simula Research Laboratory er et norsk forskningsinstitutt som fokuserer på forskning innenfor IKT, inkludert KI. De har prosjekter som spenner fra grunnforskning til anvendte løsninger, særlig innen helse, sikkerhet, og transport.
- SINTEF jobber med anvendelser av KI innen blant annet helse, energi, og industri. SINTEF samarbeider ofte med både academia og næringsliv for å utvikle praktiske KI-løsninger.

³⁹ Partnere i NORA-konsortiet: Universitetet i Agder, UiT Norges arktiske universitet, OsloMet – storbyuniversitetet, Universitetet i Bergen, Norges miljø- og biovitenskapelige universitet, Simula Research Laboratory AS, NORCE Norwegian Research Centre AS, Universitetet i Oslo, Universitetet i Stavanger, Universitetet i Sørøst-Norge, Høgskolen Kristiania, Vestlandsforskning, Høgskolen i Østfold, SINTEF, Handelshøgskolen BI, Norsk Regnesentral, Høgskolen i Innlandet og Høgskulen på Vestlandet.

- UiO har flere forskningsgrupper innen KI, inkludert anvendelser innen medisin, språkbehandling og robotikk. Universitetet er involvert i både grunnforskning og utvikling av KI-teknologier.

Forskningsorganisasjonene og -miljøene har behov for både CPU- og GPU-regnekraft. CPU er ideell for numeriske simuleringer og databehandling som ikke kan lett parallelliseres, og brukes normalt til oppgaver som krever sekvensiell prosessering, høy presisjon og kompleks logikk. Tradisjonelle beregningsoppgaver er ofte CPU-intensive. GPU brukes i større grad når oppgavene kan dra nytte av massiv parallellprosessering, som blant annet er nødvendige for å trene store nevralt nettverk. Dette inkluderer oppgaver som krever store mengder data som kan behandles samtidig, som i visse typer simuleringer og dataanalyse. KI-relaterte beregningsoppgaver er som oftest GPU-intensive på grunn av behovet for å prosessere store datamengder parallelt, mens tradisjonelle beregningsoppgaver kan også dra nytte av GPU for spesifikke parallele beregninger. For øvrig kombinerer mange systemer både CPU og GPU for å dra nytte av styrkene til begge.

Når det gjelder det fremtidige behovet for tungregning, oppgir forskningsinstitusjonene at omfanget av behovet for fremtidig tungregning er høyt og svært usikkert. Institusjonene påpeker at det ikke bare er behov for økt regnekraft (CPU og GPU), men også behovet for kunnskap og kompetanse innenfor feltet. Dette gjelder særlig innenfor KI, hvor det er viktig å ha oppsett og arbeidsflyter, samt programvare og brukerstøtte for KI-beregninger, og ikke bare maskinvare. Behovet for tungregning er nærmere omtalt under etterspørselsbaserte behov.

5.2.3 Offentlig forvaltning sitt behov for tungregning

Offentlig forvaltning har behov for tungregning til å håndtere komplekse beregninger og store datamengder til ulike formål innenfor flere sektorer. Ett sentralt formål er utvikling og leveranse av tjenester av samfunnskritisk betydning. I Norge er det definert 14 kritiske samfunnsfunksjoner: Styring og kriseledelse, forsvar, lov og orden, helse og omsorg, redningstjenesten, digital sikkerhet, natur og miljø, forsyningssikkerhet, vann og avløp, finansielle tjenester, kraftforsyning, elektronisk kommunikasjon, transport og satellittbaserte tjenester.⁴⁰ En relativt stor andel offentlige virksomheter leverer samfunnskritiske tjenester innenfor disse funksjonene, hvor flere benytter tungregning. Et viktig fellestrekk er at disse virksomhetene har behov for regneressurser med rask og forutsigbar tilgang og lite nedetid. Dette er videre utdypet under kapittel 5.3.3 Etterspørselen etter tungregneressurser til andre brukere.

For å få oversikt over bruk og behov for tungregning i offentlig forvaltning er det gjennomført en spørreundersøkelse som ble sendt ut til totalt 57 offentlige virksomheter, inkludert alle direktorater, de fleste tilsyn, helseforetakene, statsforetak og utvalgte statlige aksjeselskap. Av disse svarte totalt 38 virksomheter (se vedlegg for fullstendig oversikt over respondenter og spørsmål). Av disse svarer 11 at de bruker tungregning i dag. Av de 27 som ikke bruker tungregning i dag svarer 15 at de ser behov for å benytte tungregning i framtiden. Av de 11 som bruker tungregning i dag har 6 egne, interne ressurser, mens de øvrige kjøper tungregneressurser i markedet eller av norske forskningsmiljøer.

Vår kartlegging viser at det er nåværende eller potensielle brukere av tungregning i alle sektorer, men forsvarssektoren, helse- og omsorgssektoren, miljø- og klima og samferdselssektoren skiller seg ut med særlige behov innen tungregning.

Utover samfunnskritiske tjenester, benytter flere offentlige virksomheter også tungregning til utvikling og forbedring av ikke-kritiske tjenester og interne prosesser. En økende andel offentlige virksomheter har også begynt å utforske og anvende kunstig intelligens til å forbedre offentlige tjenester og effektivisere forvaltningen, som potensielt også vil kreve tungregneressurser.

Overordnet svarer respondentene i spørreundersøkelsen at behovet for tradisjonell tungregning vil være på samme nivå som i dag, eller øke som følge av at mer data blir tilgjengelig. Når det

⁴⁰ Regjeringen, «Liste over virksomheter med kritisk samfunnsfunksjon og nøkkelpersonell».

gjelder tungregning til KI-relaterte beregninger anslår flertallet at denne vil øke betydelig, men det er vanskelig å anslå nøyaktig behov.

I det videre følger en omtale av behovet for tungregning i utvalgte sektorer, inkludert samfunnskritiske tjenester, etterfulgt av en sektorovergripende omtale av kunstig intelligens. Informasjonen er hentet fra spørreundersøkelsene samt intervjuer av utvalgte virksomheter.

Forsvarssektoren

Innenfor forsvarssektoren benyttes ugradert tungregning av flere virksomheter, herunder Forsvarets forskningsinstitutt (FFI) og Nasjonal Sikkerhetsmyndighet (NSM). Ugradert tungregning benyttes til å gjøre avanserte beregninger og analyser, hvor beregningene blant annet blir brukt til å vurdere ytelse og trusler.⁴¹ Analysene inngår som en del av satsningen mot et sterkere datadrevet forsvar, jf. omtale under. I de tilfellene hvor fenomenet på det som skal vurderes er graderte, eller simuleringsresultatene er graderte, er det helt nødvendig med graderte tungregneressurser. Disse behovene kan ikke detaljeres i denne konseptvalgutregningen.

Fremover kan behovet for tungregning potensielt øke betydelig i forsvarssektoren. I Langtidsplan for forsvarssektoren 2025-2036⁴² stadfester regjeringen at de ønsker å utnytte potensialet i de store datamengdene som produseres til å videreutvikle forsvarssektoren og den norske forsvarsevnen gjennom et datadrevet forsvar. Forsvaret skal utnytte potensialet i data for operative formål, og regjeringen ønsker å 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 medføre et økt behov for tungregnekapasitet. Det gjelder både i form av ugraderte og graderte systemer.

NSM opplyser at staten forvalter betydelige mengder graderte 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).

Det er viktig at disse dataene og systemenes konfidensialitet, tilgjengelighet og integritet er tilstrekkelig ivaretatt og under nasjonal kontroll. Ved bruk av utenlandske skyleverandører av skytjenester kan dette være særskilt utfordrende.

NSM vurderer videre at en del av behovet for tungregning i sektoren kan dekkes gjennom ugradert tungregning, slik at fellesløsninger 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 tungregneressurser.

Stabil og sikker tilgang til åpne tungregneressurser er imidlertid en forutsetning for å bruke fellesløsninger for å dekke behovene i forsvarssektoren. 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 gradert vs. åpen infrastruktur for tungregning. Vi viser i denne sammenheng til anbefalingen om at det etableres en systematikk for å identifisere flerbruksmuligheter (dual-use) og skjermet forskning⁴³.

⁴¹ 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.

⁴² Prop. 87 S (2023–2024), kapittel 10.

⁴³ Forskningsrådet, «Et helhetlig forsyningssystem for åpen, skjermet og gradert forskning».

Det pågår for tiden et tilgrensende oppdrag som gjennomføres av Forskningsrådet, FFI og NSM om å utarbeide et kunnskapsgrunnlag for vurdering av sensitive teknologier (KVA⁴⁴). Oppdraget skal bidra til å etablere et mer helhetlig kunnskapsgrunnlag om hvilke teknologiområder og konkrete teknologier som til enhver tid vurderes som særlig sensitive for nasjonal sikkerhet, og hvordan forskningsfronten utvikler seg over tid. Teknologiområder som inkluderes i vurderingen er blant annet kunstig intelligens, kvanteteknologier, avanserte sensorteknologier, mv, der tungregning er en sentral infrastruktur. Dette kunnskapsgrunnlaget vil kunne bidra til ytterligere kunnskap og anbefalinger om bruk av gradert og skjermingsverdig informasjon innen tungregning.

Justis- og beredskapssektoren

I justis- og beredskapssektoren har vi fått informasjon fra Politiets IT-enhet (PIT), Kriminalomsorgsdirektoratet, Kripos og Økokrim. Samtlige av disse svarer at de ikke bruker tungregning i dag, men at de ser behov for å bruke tungregning for å løse forvaltningsoppdraget de nærmeste årene.

Kripos uttaler at utvikling og bruk av kunstig intelligens blir bare viktigere og viktigere i politiets arbeid med å sørge for rettssikkerhet, trygghet og velferd i samfunnet gjennom forebygging, avdekke og stanse kriminell virksomhet. I 2024 startet Kripos etablering av en organisatorisk enhet for kunstig intelligens og anvendt teknologi som skal samkjøre utviklingen og bruken i norsk politi. I denne sammenheng er tilgang på tungregning i dataplattformene en viktig forutsetning for utvikling og bruk av kunstig intelligens. Kripos er behandlingsansvarlig for sentrale registre i politiet, i tillegg til selv å ha, og tilgang på, store mengder politidata, og er videre engasjert i regelverksutvikling, slik at politiet kan utvikle og bruke kunstig intelligens til det fulle fremover.

PIT opplyser at svært mye av politiets informasjon er «fri tekst», tale (avhør, telefonsamtaler, og så videre), og video. Det er naturlig nok ikke mange tilgjengelige løsninger tilpasset norsk politi, og mye må lages. I praksis er all informasjon Politiet sitter på personsensitiv. Bruk av kommersielle aktører er derfor krevende, men ikke umulig. Kriminalomsorgsdirektoratet opplyser at tungregning bli viktig for utvikling og fremtidig bruk av bl.a. KI i kriminalomsorgen, mens Økokrim vil etter all sannsynlighet komme til å trenge kraftig økt kapasitet for behandling og analyse av store strukturerte - og ustrukturerte datasett for bekjempelse av økonomisk kriminalitet.

Helse- og omsorgssektoren

I helse- og omsorgssektoren jobber Helsedirektoratet med ulike tiltak for trygg og effektiv bruk av KI i helse- og omsorgstjenesten. På oppdrag fra Helse- og omsorgsdepartementet leder Helsedirektoratet nå arbeidet med en felles plan for trygg og sikker innføring av KI i helse- og omsorgssektoren.⁴⁵ Som en del av dette arbeidet, inngår vurderinger behovet for regnekraft og hva som bør gjøres nasjonalt og gjennom europeisk samarbeid.

Helsedirektoratet samarbeider med Nasjonalbiblioteket om hvordan det bør tilrettelegges for språkmodell av god kvalitet. Det er p.t. under avklaring om det skal lages en stor nasjonal språkmodell eller sektorspesifikke språkmodeller. De utforsker også ulike helsemodeller som er tilpasset norsk klinisk praksis, inkludert prinsipper for bruk av forskningsdata. Små modeller skreddersydd for en oppgave kan være like effektive som de store modellene. Hvilken modell som velges vil være avgjørende for hvor mye regnekraft som trengs til dette arbeidet, samt hvilken organisering av regnekraften som er hensiktsmessig for dette formålet. Helsedirektoratet beskriver også tilstrekkelig regnekraft som en viktig risiko for bruk av språkmodeller i helsevesenet. Helsedirektoratet trekker i tillegg fram KI-sentrene som vil komme ut av KI-milliarden⁴⁶ der et potensielt KI-senter innen helse- og omsorgssektoren vil generere behov for tungregning.

⁴⁴ [Oppdragsbeskrivelse - Kunnskapsgrunnlag for vurdering av sensitive teknologier](#).

⁴⁵ Helsedirektoratet, «Felles KI-plan for trygg og effektiv bruk av KI i helse og omsorgstjenesten 2024-2025».

⁴⁶ Mer informasjon om satsingen: [Satsing på kunstig intelligens](#).

De regionale helseforetakene (RHF) har tilgang til tungregning på egne servere, hvor de har sensitive data på sikre plattformer. Det er ingen sentral koordinering av disse tungregneressursene, men helseforetakene skal legge til rette for samarbeid på tvers av foretak. Mange av sykehusene arbeider tett med forskningsmiljøer og deltar i forskningsprosjekter med tilgang på tungregneressurser hos Sigma2.

Helse Midt-Norge RHF, ved St. Olavs hospital HF, opplyser at de anskaffer tungregning hos sin skyleverandør. IT tjenester ved St. Olavs hospital og de andre helseforetakene i Midt-Norge leveres av Hemit HF. Framover har de behov for validering og bruk av KI verktøy innen helsehjelp og utvikling og validering av KI verktøy innen forskning.

De beskriver at for at Norge skal lykkes med helseinnovasjon og verdiskaping med KI trengs det bedre verktøy for en sikker flyt av helsedata mellom helseforetakene og universitetene, lett tilgjengelige ressurser for tungregning i henhold til helseforetakene sine krav for informasjonssikkerhet (sensitive data), samt sikre testmiljø ("sandboxes") for forskning på KI for helsehjelp som gjenspeiler de kliniske miljøene. De har videre behov for 100 prosent oppetid av ressursene.

Helse Nord RHF beskriver at forskningsmiljøene i helseforetakene og Helse Nord IKT bruker ulike tungregneressurser etter behov. Framtidig behov vil øke i takt med at vi har mye og stadig mer data vi ønsker å bruke til styringsdata for å utvikle en enda mer effektiv og kvalitativ god helsetjeneste. De trekker fram at etablering av Nasjonalt genomsenter (oppdrag fra Helse- og omsorgsdepartementet i 2024 til de regionale helseforetakene) kreve betydelig mer behov for tungregning.

Helse Nord RHF er midt i en prosjektperiode for SIKTH-prosjektet (som er lagt til Helse Nord IKT) som jobber for å etablere infrastruktur for KI, samt lagring og analyse av helseforskningsdata⁴⁷. Dette innebærer at behov for tungregnekapasitet til KI er økende framover.

Helse Nord RHF peker på at penger, prioritering, kultur for dette i egne organisasjoner er viktige stikkord for å lykkes med innovasjon og verdiskaping med KI. I sin innovasjonsstrategi har de også tydeliggjort at: «For å finne frem til, og realisere, de beste løsningene kreves samarbeid både internt i helseforetak, mellom helseforetak og på tvers av sektorer. Næringslivssamarbeid er særlig viktig for å øke innovasjonsgraden i helsetjenesten og for å møte utfordringene vi står overfor. Samarbeid med brukerne er viktig også i innovasjonsaktiviteter, enten gjennom brukerdrevet innovasjon eller brukermedvirkning i ulike faser av innovasjonsprosjekter».

Når det gjelder krav til tungregneressurser peker Helse Nord på håndtering av sensitive data, helsedata, personidentifiserbare data i sykehus, og i adm. prosesser er det bl.a. persondata, innkjøpsdata, kontrakter, prissensitiv informasjon m.m. Tilgang hele døgnet er viktig i sykehusene.

Folkehelseinstituttet (FHI) opplyser at biobankene og genetiske data sammen med ordinære helseregistre og helseundersøkelser, i økende grad blir avgjørende for fremtidens kunnskapsproduksjon. De ser dette i helseforskningen, men ikke minst i kunnskapssystemene knyttet til helseberedskapen generelt og smittevernberedskapen spesielt. FHI vil ha en sterk vekst av etterspørsel til tungregneressurser til tradisjonelle beregningsoppgaver. De ser videre for seg at økende del av rådgivningen knyttet til utvikling av helsetjenesten, tiltak ved helsekriser, osv., vil benytte KI-verktøy for effektivisering av arbeidet. De ser en eksponentiell etterspørsel og har grunn til å tro at behovet kommer til å være umettelig. Norge har helsedata i verdensklasse som kan gi verdifull forskning, innovasjon og verdiskaping hvis man samler kreftene på tvers av sektorer. Dette krever imidlertid en tydelig organisatorisk, juridisk og semantisk utvikling utover den tekniske infrastrukturen.

I beredskapssammenheng er det behov for tungregning innenfor helsesektoren. Under koronapandemien benyttet FHI tungregning til å estimere og predikere reproduksjonstallet for koronasmitten i befolkningen. Det var imidlertid krevende å få gjennomført ønskede analyser på grunn av kapasitetsbegrensninger. I fremtiden er det behov for en langt større kapasitet dersom

⁴⁷ Mer informasjon om en tjeneste som nylig er lansert for forskere i Helse Nord. <https://www.helsenordikt.no/nyheter/nytt-fagomrade-i-hn-ikt/> og <https://www.helse-nord.no/nyheter/ny-tjeneste-for-forskning-og-kunstig-intelligens-i-helse-nord/>.

pandemier eller større helsekriser inntreffer. I den sammenheng opplyser FHI at det er behov for miljøer som kan håndtere personsensitive helsedata med garantert kontinuerlig oppetid under helsekriser.

Direktoratet for medisinske produkter (DMP) bruker ikke tungregning i dag, men forventer å utvikle og/eller ta bruk av KI på flere områder i løpet av de nærmeste årene, men har ingen konkrete planer om å ta i bruk tungregning. DMP antar i første rekke at vårt behov for tungregning i dette langt på vei vil dekkes av eksterne leverandører som Microsoft, Google, etc.

DMP vil likevel trekke fram noen områder som det potensielt er aktuelt å benytte avansert KI eller tungregning til i løpet av de nærmeste årene. Områdene hvor vi ser at det kan være spesielt behov for dette, er signaldeteksjon innen medisinsk utstyr, samt forsyningsikkerhet og mangelsituasjoner (både hva gjelder legemidler og medisinsk utstyr).

Videre ser de at det kan være behov for tungregning knyttet registerdata, men siden registerdata ligger under Folkehelseinstituttet (FHI) sitt ansvarsområde, er det også det FHI som vil være behovseier for eventuell utvikling eller tilrettelegging for dette.

Miljø- og klimasektoren

Innen miljø- og klimasektoren er Meteorologisk institutt (MET) den største brukeren av tungregning og en av de største brukerne i hele offentlig sektor.

MET bruker tungregning både til forskning og forvaltning, og tungregning er nødvendig for METs samfunnskritiske tjenester. Dette omfatter drift og forskning og videreutvikling av operasjonell vær- og havvarsling, samt forutsi værmønstre og klimaendringer. 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.

MET har historisk benyttet tungregning til tradisjonelle beregningsmodeller, og behovet har økt med rundt 30 prosent per år de siste årene. MET vil framover fortsatt ha behov for tungregning til tradisjonelle beregningsmodeller, men også i økende grad til datadrevne modeller i tilknytning til KI. Det innebærer at MET har behov for både CPU- og GPU-ressurser. MET vil også ha stor nytte av en norsk språkmodell, som kan erstatte mye manuelt arbeid og gi effektiviseringsgevinster.

MET legger vekt på at det er svært nyttig med samarbeid om 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. Effektiv bruk av dataene krever dedikert innsats og investeringer fra forskerne. Til gjengjeld er dataene gode, og det er mulig å få svært gode synergieffekter gjennom et felles rammeverk. MET beskriver at behovet for tungregning 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 tungregning. Det er spesielt krevende å skaffe tilstrekkelig tungregning.

Den norske jordsystemmodellen NorESM har vært utviklet siden 2007 og vært et viktig verktøy for norske klimaforskere i studiet av fortidens, dagens og fremtidens klima. NorESM har også bidratt med klimasimuleringer som har vært benyttet i FNs klimapanelens femte og sjette hovedrapporter. Klimasimuleringer med NorESM krever tilgang til kraftige tungregneressurser. Videreutvikling av NorESM skjer gjennom prosjektet Infrastructure for Norwegian Earth System modeling (INES) og vil bidra til at norske forskere også i årene fremover har tilgang til en cutting-edge jordsystemmodell med evne til klimavarsling opptil 10 år frem i tid, inkludert modellering av nye prosesser på høye breddegrader og evne til langsiktig framskrivning av havnivå. INES finansieres gjennom Forskningsrådets infrastruktursatsing og ledes av forskningsinstituttet NORCE i samarbeid med flere institutter og universiteter (MET, NILU, UiB, UiO og NERSC).

Norges Geotekniske Institutt (NGI) benytter ikke tungregning i dag, men beskriver at det er et utnyttet potensial for tungregning som kan gi bedre tjenester for eksempel knyttet til håndtering av geofarar 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. geofarar, eller allmenn tilgjengelige tungregnesystemer

internasjonalt som kan operere opp mot akutt situasjoner. 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.

Riksantikvaren benytter også tungregning, og kjøper kapasitet i markedet. Fremover vil de ha behov for tungregning for å kunne kjøre mer komplekse algoritmer på geografiske data med hensikt å skape bedre statistikk og mer innsikt, og de vil trenge tungregning for å kunne utvikle domenemodell for sitt område. De peker videre på sektorsamarbeid, bygge opp felles nasjonal infrastruktur, og kunne kopiere suksesshistorier til andre innen offentlig sektor som suksesskriterier for å skal lykkes med innovasjon og verdiskapning med KI i Norge.

I tillegg opplyser Miljødirektoratet at de ikke bruker tungregning i dag, men ser framtidige behov særlig innenfor kart/stedfestet informasjon, samt til prediksjoner.

Næring og fiskerisektoren

I nærings- og fiskerisektoren har Havforskningsinstitutt (HI) ansvar for «Trafikklyssystemet for bærekraftig akvakultur». Mer enn 90 prosent av beregningsoppgavene med tungregnerressurser som HI utfører, er knyttet til prosjekter under dette oppdraget. De siste 3-5 årene har omfanget omtrent doblet seg.

Norsk romsenter opplyser at de finansierer nasjonalt bakkeselement for satellittdata har lagringskapasitet lokalisert på MET. Dette sikrer god tilgang til satellittdata fra Copernicus over norske interesseområder for norske brukere.

I Norge er det flere prosjekter i gang for å bruke maskinlæring, kunstig intelligens og stordataanalyse sammen med data fra Copernicus. Det skal brukes til blant annet å overvåke faren for, og gi varslere om, snøskred for fastlands-Norge, Svalbard og Nord-Sverige. Kunstig intelligens skal også benyttes i prosjektet Foundation Models for Earth Observation, som Norsk Regnesentral har sammen med Universitetet i Tromsø og Norges vassdrags- og energidirektorat (NVE).

Satellittdatamengden fra Copernicus vil økes med 700 prosent fram mot 2034, og behovet for tungregnekapasiteten antas å øke tilsvarende eller mer.

For Norsk romsenter er det særlig viktig med integrasjon av KI, stordata analyse og bruk av satellittdata innen anvendelser der Norge er sterke. Nyttverdien for Norge vil være spesielt stor når det gjelder kartlegging og overvåking av nordområdene, norske havområder, miljø- og ressursforvaltning, infrastruktur, klima, forskning, maritim sikkerhet, geofare, forsvar og nasjonal beredskap. De peker på at eksisterende nasjonal infrastruktur bør bindes sammen og forsterkes betydelig, og kontinuerlig tilgang hele døgnet er viktig. Håndtering av sensitive data og gradert informasjon er spesielt viktig for anvendelser innen forsvar og sikkerhet, og for nasjonalt kritiske funksjoner.

Norges geologiske undersøkelse (NGU) finansierer sammen med Norsk romsenter og NVE lagringskapasitet for InSAR Norge. InSAR Norge publiserer landsomfattende nedsynkingsdata i en gratis, nettbasert 3D-karttjeneste. Ved bruk av InSAR Norge kan man kartlegge deformasjoner i landskapet, som for eksempel nedsynking i byer og bevegelser i ustabile fjellpartier. Nedsynking kan føre til setningsskader i bygninger, problemer med rørsystemer for vann- og avløp og alvorlige skader på veier og annen infrastruktur. InSAR Norge-konsortiet er også sentrale i Copernicus European Ground Motion Service der de produserer data for 10 land i Europa, og har data for hele Europa lagret lokalt.

Utviklingen innen kunstig intelligens for jordobservasjon vil være avgjørende for neste generasjons maritime overvåkingstjenester basert på satellittdata. NTNU og NGU samarbeider om prosjektet Data Analytics for Ground Motion Monitoring (DARIO) for neste generasjons tjenester innen overvåking av nedsynking.

KI-relaterte beregninger vil komme på toppen av eller som en integrert del av prosesseringsbehovet for nasjonalt bakkeselement og InSAR Norge, i tillegg til integrasjon i nedstrøms bruk av satellittdata og satellittbaserte tjenester.

Kongsberg Satellite Services (KSAT) har betydelige tungregnerressurser opp mot sine oppgaver innen nedlesing og sanntidsprosessering av satellittdata. 55 prosent av dataene i Copernicus

blir nedlastet på Svalbard, samt alle europeiske og amerikanske værstatellitter i polarbane. Satellittdata fra Sentinel-1 blir prosessert i sanntid og tilgjengeliggjort for norske brukere over norske interesseområder. Andre store offentlige brukere av satellittdata er for eksempel NVE og NIBIO.

Samferdselssektoren

Jernbanedirektoratet kjøper tungregnerressurser av Microsoft/Databricks. De er i gang med å etablere en ny avdeling for digitalisering, innovasjon og teknologi, der er formålet blant annet at organisasjonen skal kunne ta en større rolle som pådriver for digitalisering i jernbane- og transportsektoren. De ser for seg at det vil være et økende behov for KI-relaterte beregninger de kommende årene. Behovet vil trolig ikke bare være internt, men på tvers av jernbane-, og transportsektoren.

Avhengig av formål vil det kunne være behov for håndtering av sensitive data. Både i form av konkurransesensitive data, som inntektstall, og sensitive personopplysninger knyttet til reiseadferd.

For å lykkes i jernbanesektoren med innovasjon og KI beskriver de at jernbanesektoren består av mange aktører, samtidig er det store systemavhengigheter. Dette gir behov for effektiv datadeling mellom aktører, etablering av effektivt regelverk, harmonisert tolkning av regelverk, samt «spilleregler» for deling av data, standardisering og harmonisering av dataformater for effektiv bruk, nasjonale virkemidler/finansiering som kan benyttes for utvikling av nasjonale/sektorielle løsninger. Disse må kunne aksesserer på kort tid og være orientert mot funksjonalitet (legge til rette for smidig utvikling av løsninger). De viser til at Norge må utvikle/sikre at kapitalmiljøer (offentlige/private) investerer i selskaper (med KI-teknologi) for skalering og vekst.

Statens vegvesen bruker tungregning i dag, og har egne interne ressurser. De gjør en god del tunge tradisjonelle utregninger som de håper å kunne kjøre på beregningsklynger i fremtiden. Eksempler inkluderer støyberegning, beregning innenfor planlegging av vei og simuleringer i forbindelse med konstruksjoner. De arbeider med å utvikle maskinsyn og vil følgelig ha behov for å prosessere for eksempel bilder fra og ved vei. På lengre sikt skal både bilder og video benyttes, disse initiativene vil trolig kreve tyngre regnekraft. Behovet for tungregning er derfor økende, med det er et kost-nytte spørsmål der de må prioritere hvor mye som brukes på beregning opp mot hvor mye som brukes på å løse problemet som de finner ved tungregning. Statens vegvesen opplyser at det vil være behov for å jobbe med data som de vurderer som skjermingsverdige og da er det nyttig at de har kapasiteten i Norge.

Trafikkøkonomisk institutt (TØI) har behov for tungregnerressurser til å foreta trafikksimuleringer og avanserte analyser. Bruken omfatter både tradisjonelle regneoppgaver og kunstig intelligens. TØI ser et økende behov for tungregning særlig innenfor maskinlæring, og at behovet for GPU-ressurser vil øke fremover. De benytter i dag ressurser hos Statens vegvesen.

Ruter AS/TET Digital AS benytter også tungregnerressurser, de anskaffer tjenester/kapasitet hos sin skyleverandør. De har primært behov knyttet til trening av maskinlæringsmodeller, og ser at behovet vil være sterkt stigende framover. De beskriver at for å lykkes med innovasjon og verdiskaping med KI i Norge er vi avhengig av tett samarbeid mellom akademia, næringsliv og offentlig sektor, der aktørene får frem hvilke områder det er nødvendig med forskning, for å bygge evner som kan realisere forretningsbehov. Det blir viktig å fokusere smalt, men sterkt, for å tilstrekkelig kapasitet og kompetanse til å kunne konkurrere med andre land og aktører.

Avinor opplyser at de ikke bruker tungregning i dag, men at det kan dukke opp brukstilfeller der det kan være behov for tungregning de nærmeste årene, men de har ikke identifisert slike konkrete brukstilfeller per nå.

Utdanningssektoren

I utdanningssektoren opplyser Lånekassen at de i forbindelse med tidligere utvikling på KI-området har benyttet større regnekapasitet i sky. Fremtidig bruk er, slik de ser det, tilknyttet behov for å skalere opp regnekraft for å støtte aktiviteter rundt KI. Nærmere bestemt modellutvikling og modellbaserte beregninger. De anser at den kapasiteten de har behov for er skalerbar innenfor den Azure-infrastrukturen vi allerede er tilknyttet.

Direktoratet for høyere utdanning og kompetanse (HK-dir) opplyser at de utforsker hvordan de skal utnytte potensialet i KI.

Digitaliserings- og forvaltningssektoren

Direktoratet for forvaltning og økonomistyring (DFØ) opplyser at de i fremtiden vil utvikle og drifte egne KI-modeller, samt kunne ha behov for tungregning ved prosessering og behandling av egne datasett.

Olje- og energisektoren

Sokkeldirektoratet beskriver at petroleumsnæringen, som Sokkeldirektoratet skal forvalte, jobber med store datafiler på tvers av hele verdikjeden (f.eks. undergrunnsdata som seismikk). Industrien bruker mer og mer KI i deres daglige arbeid blant annet ved å bygge opp digitale tvillinger av installasjoner på norsk sokkel eller ved å lage store mega-cuber av seismikk. For at Sokkeldirektoratet skal kunne utarbeide beslutningsgrunnlag, må de kunne replikere arbeidsflyter eller bruke samme datasettet brukt i bransjen. Det vil kreve tilgang til tungregnerressurser. Videre vil Sokkeldirektoratet selv bruke det omfattende datasettet som er tilgjengelig per i dag (17 TBt) for avanserte analyser som understøtter forsvarlig ressursforvaltning på norsk sokkel.

Statkraft opplyser at de har egne ressurser i form av 300 CPUer som i hovedsak brukes til predikative algoritmer og modeller. Bortsett fra moderate oppgraderinger er dette dekkende for behovet framover.

Utenrikssektoren

Norsk utenrikspolitisk institutt (NUPI) opplyser at de har et økende behov for regnekraft for å behandle stordata i forskningen.

Kommunal- og distriktssektoren

Kartverket leverer kart som kan anses som samfunnskritiske. De opplyser at de benytter egne ressurser for tungregning som de bruker mest til arbeid med tradisjonell tungregning, og noe KI-relatert forskning. De bruker i tillegg tungregnerressurser fra Universitetet i Agder. De anslår at behovet for tradisjonell tungregning vil være på samme nivå som i dag, mens behovet for tungregning når det gjelder kapasitet til KI-relaterte beregninger vil øke. De legger vekt på at det er viktig at det bygges en arkitektur som kan håndtere store mengder data (dataflyt), og at det bør finnes noe i Norge på dette for å håndtere sensitive data.

Direktoratet for byggkvalitet opplyser at de ikke bruker tungregning per i dag, men ser framtidig behov innen automatisk saksbehandling, rapporter, analysevirksomhet, og publikumstjenester.

Kultur- og likestillingssektoren

Arkivverket har investert i egne tungregnerressurser (3 GPUer) for å dekke sitt behov for tungregnekapasitet, i tillegg til at de bruker Hunt Cloud⁴⁸ til noen eksterne prosjekt, og indirekte bruker Sigma2 i et enkelt forskningsprosjekt. De benytter GPUer og tungregning hovedsakelig til KI-relaterte oppgaver.

Arkivverket forvalter en stor mengde data som inneholder personopplysninger og er underlagt GDPR. Det er sentralt at tungregnerressurser kan håndtere og brukes for denne typen data. Dette er også en av grunnene til at de har valgt å investere i egne ressurser. Investeringen er foreløpig moderat, men de ser at den typisk vil vokse.

De anslår at det vil være behov framover for anslagsvis 10 GPUer for å dekke interne behov, men ser også mulighet for at behovet kan bli større siden Digitalarkivet er en fellesløsning og

⁴⁸ NTNU har fra 2013 etablert og kontinuerlig videreutviklet egen løsning for sensitive data, HUNT Cloud. HUNT Cloud har siden 2018 også levert sikre analyserom med GPU. Aktiviteten er nå utvidet med NORTRE-samarbeidet, hvor TSD (UiO), SAFE (UiB) og HUNT Cloud samarbeider om å levere løsninger for helsedata ("Gjenbrukssporet").

skal utvikles som en forskningsinfrastruktur. Det kan være aktuelt å tilby tjenester som krever tungregning som en tjeneste til disse brukerne.

Arkivverket beskriver at offentlig sektor sammen må løse utfordringer som kan gi delt nytte for bruk av KI, samtidig som man tydelig må beskrive og se på løsninger for konkrete utfordringer. Kompetanse på bruk (både der man kan og der man ikke bør bruke) av KI er viktig, men også hvordan man skal løse utfordringer rundt forklarbarhet, transparens, tillit m.m. Det bør være en kombinasjon av forskning og uttesting, men også å se ulike fagområder i sammenheng og gi ansvar og kobling mot premissgjivere og fagmyndigheter på området. Arkivverket er i denne sammenheng en viktig deltager i arbeidet for å tydeliggjøre behov, krav og å utforske mulighetsrom.

Arbeids- og sosialsektoren

Statens arbeidsmiljøinstitutt (STAMI) opplyser at de ikke bruker tungregning per i dag, men at det vil kunne dukke opp behov for å bruke tungregning til forskningsformål de nærmeste årene. Behovene er ikke klare enda.

5.2.4 Kunstig intelligens i offentlig sektor

Når det gjelder *kunstig intelligens*, er det flere offentlige virksomheter på tvers av sektorer som har begynt å utforske og eksperimentere med KI-modeller. Anvendelse av KI-verktøy som f.eks. Microsoft Copilot, tilpassede GPT-applikasjoner m.m., for å effektivisere saksbehandling og interne virksomhetsprosesser krever imidlertid ikke tungregneressurser. Virksomheter som leverer samfunnskritiske tjenester eller andre tjenester som er særs data- og analyseintensive, har derimot behov for tungregning til å utvikle og teste datadrevne modeller. Eksempel som er nevnt ovenfor, er MET innenfor miljø- og klimasektoren og FFI innenfor forsvarssektoren. Flere offentlige virksomheter kan potensielt få økt behov for KI-relatert tungregning til å forbedre og utvikle nye offentlige tjenester, om enn i en mindre skala, i fremtiden. Flere av instituttene i Forskningsrådets kartlegging fremhever at utvikling og anvendelse av KI fortsatt er på eksperimentstadiet og at behovet for KI-relatert tungregning er uklart i dag, men at det forventes et økende behov i tiden framover.

En viktig anvendelse av tungregning relatert til KI, er utvikling, tilpassing og drift av store *språkmodeller* og generative KI-tjenester som er bygget på toppen av disse.⁴⁹ I offentlig sektor er Nasjonalbiblioteket en sentral aktør i utviklingen og testing av store språkmodeller. Nasjonalbiblioteket har fått en bevilgning på 40 mill. kroner i statsbudsjettet for 2025⁵⁰ som skal dekke lokal infrastruktur, regnekraft via Sigma2 og til å etablere en organisasjon for trening, oppdatering og tilgjengeliggjøring av norske og samiske språkmodeller. Nasjonalbiblioteket anslår at nivået på bevilgningen vil være tilstrekkelig for å dekke behovet i 2025 siden de er i en oppstartsfasen, men anslår at budsjettet for 2026 bør økes til 50-60 mill. kroner for å dekke behovet, siden det vil være behov for flere ulike modelltyper i ulike størrelser og konfigurasjoner.

Et relevant pågående prosjekt hos Nasjonalbiblioteket er «Språkbanken», som tilbyr store datasett med norsk tekst og tale for utvikling av språkteknologi. Dette inkluderer verktøy som NB-Whisper, som er basert på OpenAI's Whisper-teknologi og brukes til å produsere de beste transkripsjonene på begge målformer. 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 som offentlig sektor i så fall må tilgjengeliggjøre. Helsedirektoratet beskriver også tilstrekkelig regnekraft som en viktig risiko for bruk av språkmodeller i helsevesenet.

Praksis viser at det er vanskelig å beregne fremtidig behov og krav til regnekraft for å trene språkmodeller, og enda vanskeligere for bruk av språkmodellene (inferens). 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

⁴⁹ Teknologirådet, «Generativ kunstig intelligens i Norge».

⁵⁰ Prop. 1 S (2024–2025), 193.

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.

5.2.5 Næringslivets behov for tungregning

Private bedrifter og organisasjoner som produserer varer og selger tjenester, har behov for tungregning til flere formål. Næringslivet har, som forskningsorganisasjoner og offentlige virksomheter, behov for å håndtere komplekse beregninger og store datamengder til innovasjon og utvikling innenfor flere bransjer og områder. Til forskjell fra offentlige virksomheter, søker næringslivet i større grad å produsere nye produkter og tjenester som kan selges i et marked, samtidig som en del private bedrifter også leverer samfunnskritiske tjenester.

I forbindelse med Forskningsrådets kartlegging våren 2024 spurte Abelia deres medlemmer om bruk av tungregning i dag og i fremtiden. Det var relativt få som rapporterte at de benyttet tungregning i dag, hvor hovedvekten av dagens brukere var IKT-virksomheter og større virksomheter. Fremover var mange av de spurte bedriftene usikre på hvor omfattende det fremtidige behovet for tungregning er. Samtidig svarte over ni av ti at behovet vil øke i virksomheten de neste fem årene, og over halvparten trodde at behovet vil øke kraftig, særlig blant brukere av tungregning i dag.

Et særskilt teknologisk satsingsområde for private virksomheter er *kunstig intelligens*. 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.

En spørreundersøkelse gjennomført av samfunnsøkonomisk analyse viste at én av fire norske virksomheter benytter KI i dag og at bruken er økende. Mange av virksomhetene oppga at de var i test- og piloteringsfasen og omtrent halvparten av virksomhetene sier at de har eller vil utvikle en KI-strategi.

Økt bruk av KI-modeller i næringslivet, vil kreve tungregneressurser. KI-modeller må utvikles og trenes, som er særlig regnekraftintensive faser. De fleste private virksomheter vil derimot høyst sannsynlig ikke utvikle egne KI-modeller, men ta i bruk generiske KI-modeller eller tilpasse og justere utviklede KI-modeller til sitt formål. I disse prosessene vil behovet for tungregning være vesentlig lavere enn i foregående stegene, hvor behovet vil variere etter modellenes kompleksitet, responstid, bruksområde m.m. Rapporten til Samfunnsøkonomisk analyse gir innblikk i de vanligste bruksområdene for KI i dag. Oversikten viser at det er et stort spenn av bruksområder, men hvor behovet for avansert tungregning vil være begrenset:

- Effektivisere arbeidsprosesser (66 prosent) og forbedre innsikt/beslutningsstøtte (43 prosent). Dette sammenfaller med funnene fra Statistisk sentralbyrås undersøkelse av bruk av IKT i næringslivet⁵¹.
- Effektivisere produksjon eller drift, herunder logistikk (27 prosent). Effektivisering av produksjon er særlig fremtredende i primærnæringene som jordbruk, skogbruk, fiske og havbruk, samt i kultur, underholdning og teknologiindustrien. Equinor er et eksempel på en virksomhet som bruker KI for å optimalisere drift og utføre prediktivt vedlikehold. OptoScale er et annet eksempel. De hjelper oppdrettsnæringen med å effektivisere produksjonen ved hjelp av KI. Kunder som SalMar og Lerøy forbedrer tidsbruk og gir ansatte bedre beslutningsgrunnlag, noe som øker innsikten i produksjonen og forbedrer beslutninger om føring, slakting og avlusing.
- Forbedre salgsprosesser (21 prosent) og forbedret og effektivisert kundedialog (23 prosent). Optimalisering av salgsprosesser er spesielt utbredt i detalj- og engroshandel. Bedre kundedialog er mest vanlig i reiselivsbransjen, men også i detalj- og

⁵¹ 113265: Bruk av kunstig intelligens-teknologi, etter sysselsetting og næring (SN2007) (prosent) 2021 - 2024. Statistikkbanken.

engroshandel. For eksempel bruker Strawberry KI for å forbedre prisstrategiene sine, noe som gir dem mer tid til å fokusere på kundedialog. Sparebank 1 SMN oppnår også besparelser i kundekommunikasjon ved hjelp av chatbots.

- Redusere risiko og feil (23 prosent). Industrien, spesielt innen vareproduksjon, samt primærnæringene og finanssektoren, fokuserer sterkt på å minimere risiko og feil. Et eksempel er Sparebank 1 SMN som bruker maskinlæring for å oppdage svindel og hvitvasking, som reduserer risikoen for brudd på lover og regler.

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 oppstartsbedrifter 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 tungregning 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.

5.2.6 Behovet til leverandører av tungregning

Leveransekjeden for tungregneinfrastruktur består av datasenterleverandører og plattformleverandører.

Datasenterleverandørene har behov for:

- Kontinuerlig og stabil strømforsyning til datasentrene for å unngå avbrudd som kan føre til datatap eller nedetid.
- Effektive kjølesystemer for å forlenge levetiden til utstyret og opprettholde ytelsen.
- Tilgang til generatorer som fungerer som backup-strømkilder i tilfelle strømbrytning. De sikrer at datasenteret kan fortsette å operere uten avbrudd ved å levere nødstrøm til alle kritiske systemer.
- Tilgang til sensorteknologi for å kunne overvåke ulike parametere i datasenteret, som temperatur, fuktighet, strømforbruk og luftkvalitet. Disse sensorene gir sanntidsdata som hjelper med å oppdage og løse problemer før de påvirker driften.
- Tilgang på kompetente byggingeniører som arbeider med design og konstruksjon av datasentre. De sørger for at bygningene er strukturelt solide, energieffektive og i stand til å huse alt nødvendig utstyr på en sikker måte.
- Forutsigbar lovgivning på området.

Plattformleverandørene kan deles inn i kategoriene skyplattformleverandører, spesialiserte tungregnetjeneste-leverandører og virksomhetsinterne leveransere. De har behov for:

- Tilgang til kompetanse og ressurser. Evnen til å tilpasse løsninger for spesifikke bruksområder, sektorer eller applikasjoner er avgjørende. Dette kan innebære spesialiserte programvareløsninger eller tilpassede maskinvarekonfigurasjoner.
- Tilgang til sterke partnerskap for å sikre konkurransedyktige priser på hardware (supercomputere, CPUer, GPUer, lagringsenheter). Samarbeid med

programvareleverandører og andre teknologipartnere er nødvendig for å utvide tilbudet og forbedre tjenestene.

- Tilgang til sikre, pålitelige og utvidbare datasenterløsninger for å sikre muligheten for å utvide plattformen og tjenestene ettersom etterspørselen øker.
- Tilgang til finansiering for å kunne investere i kapasitet i forkant av etterspørselen.

5.2.7 Behovet til beslutningstakere knyttet til tungregning

Behovene til beslutningstakerne er i stor grad overlappende med normative behov som er beskrevet over. Det er blant annet behov for:

- Tilstrekkelig nasjonal tungregnekapasitet (superdatamaskiner) til å dekke nåværende og fremtidige behov i forskning og forvaltning.
- Utnytte potensialet i de store datamengdene som produseres til å videreutvikle sektorene.
- Legge til rette for at Norge skal ha infrastruktur for KI i verdensklasse, i form av digitaliseringsvennlig regelverk, gode språkressurser, raske og robuste kommunikasjonsnett og tilstrekkelig regnekraft,
- Legge til rette for at norske aktører skal ha tilgang til tungregnekapasitet gjennom EuroHPC.

5.2.8 Behovet til sekundærinteressentene

Sekundærinteressentene er delt inn i interesseorganisasjoner og aktører utenfor Norge.

Interesseorganisasjoner har behov for at det legges til rette for god involvering, og muligheter til å støtte sine medlemmer i arbeidet med å få tilstrekkelig kapasitet og gode tjenester innen tungregning og utvikling av KI.

Aktørene utenfor Norge har behov for gode samarbeidsmuligheter og erfaringsutveksling med norske aktører.

5.3 Etterspørselsbaserte behov

Etterspørselsbaserte behov kan forstås som behov knyttet til den fremtidige etterspørselen etter tungregning i Norge. Kvalitativ beskrivelse av behovet for ulike interessenter er beskrevet i kapittel 5.2, mens her utdyper vi behovet for behovet for tungregning kvantitativt.

5.3.1 Hoveddrivere for fremtidig etterspørsel

Etterspørselsbaserte behov kan forstås som behov knyttet til den fremtidige etterspørselen etter tilbudet til tungregning i Norge. Figur 5.1 presenterer en oversikt over de viktigste driverne for etterspørsel til tungregneinfrastruktur fremover til forskning og offentlig sektor.

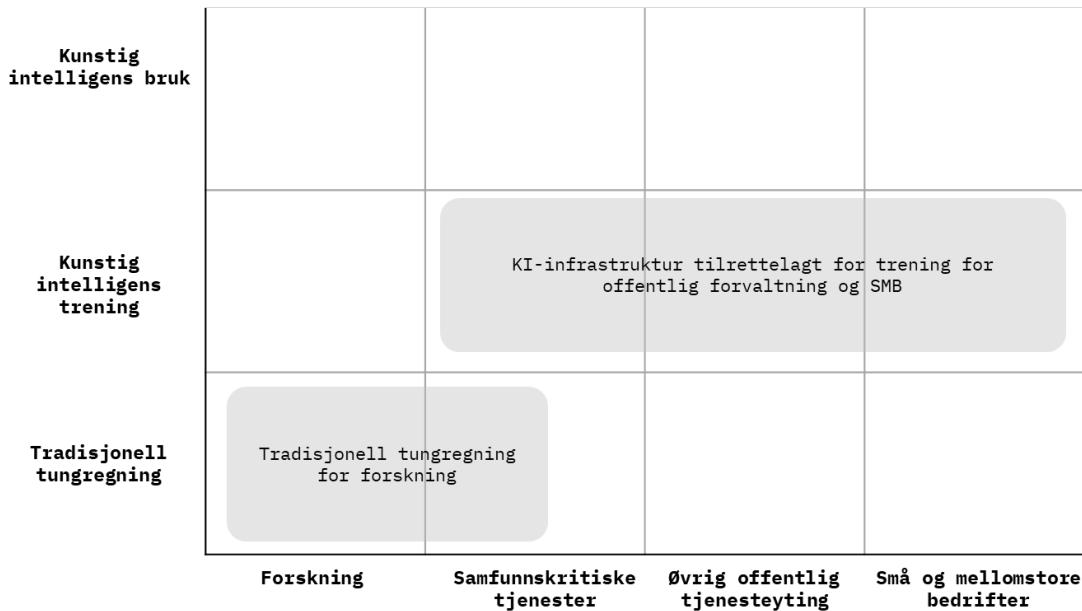


Figur 5.1. Drivere for etterspørsel av tungregneinfrastruktur

Disse driverne representerer en stigende grad av kompleksitet og setter nye krav til tungregneinfrastrukturen. Fra det grunnleggende behovet for å håndtere store datamengder til avanserte anvendelser innen forskning og offentlig forvaltning.

Der etterspørselen av tradisjonell tungregning kan kvantifiseres basert på historisk utvikling og innspill fra forskningsmiljøer, vil behovet for tungregneressurser til komplekse simuleringer og kunstig intelligens/maskinlæring være mer usikker i en periode på tre til fem år.

Basert på interessentenes behov, har vi systematisert behovet for tungregning i brukersegmenter som vi vil kvantifisere etterspørselen etter, jf. Figur 5.2.

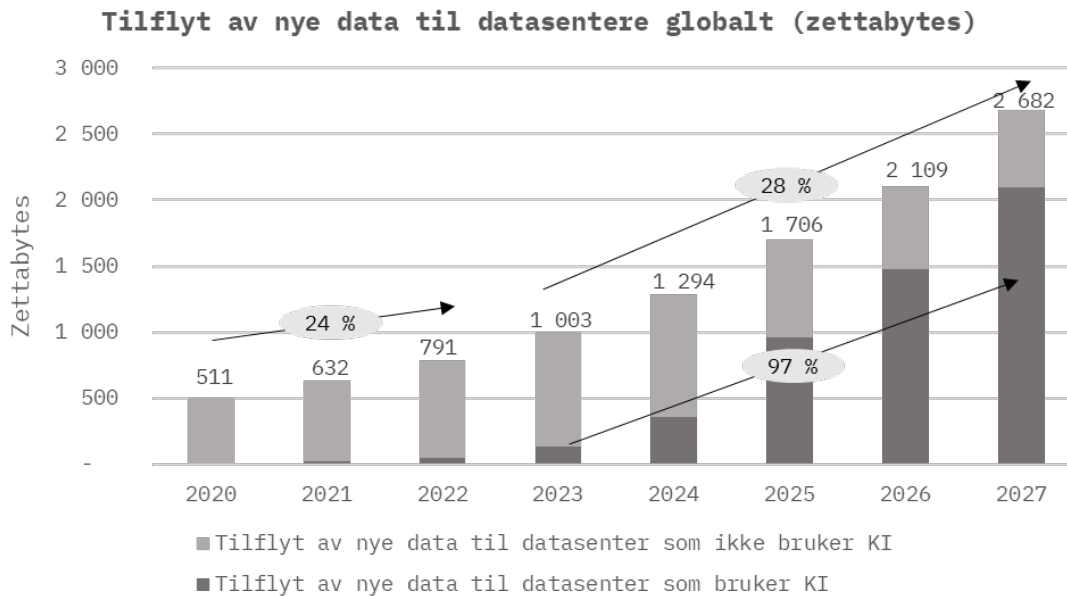


Figur 5.2 Oversikt over brukersegmenter med behov for tungregning

5.3.2 Etterspørselen av tungregning for forskning

Det blir et stadig mindre skille mellom tradisjonell tungregning, der det brukes ren CPU-basert infrastruktur, og tungregning til KI, der det brukes GPU-basert infrastruktur i tillegg. Når det gjelder behov for tradisjonell tungregning versus behov for tungregning til KI beskriver flere at behovet for tradisjonell tungregning har opprettholdt et stabilt nivå, mens andre beskriver en økning på dette området. Tungregning for KI har vært svært marginalt, men flere ser en vridning mot KI de siste årene, og enkelte beskriver en stor økning ila. 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. Nye sensorer, satellitter og observatorier produserer 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.

Figur 5.3 illustrerer hvordan mengden nye data vil flyte til datasentre, samt hvor stor andel som vil kreve tungregning for KI for å behandles.



Figur 5.3. Prognose på mengden nye data som vil tilflytte datasentre. Merknader: Andelen indikerer funksjonell bruk av KI-algoritmer for inferens eller trening minst én gang

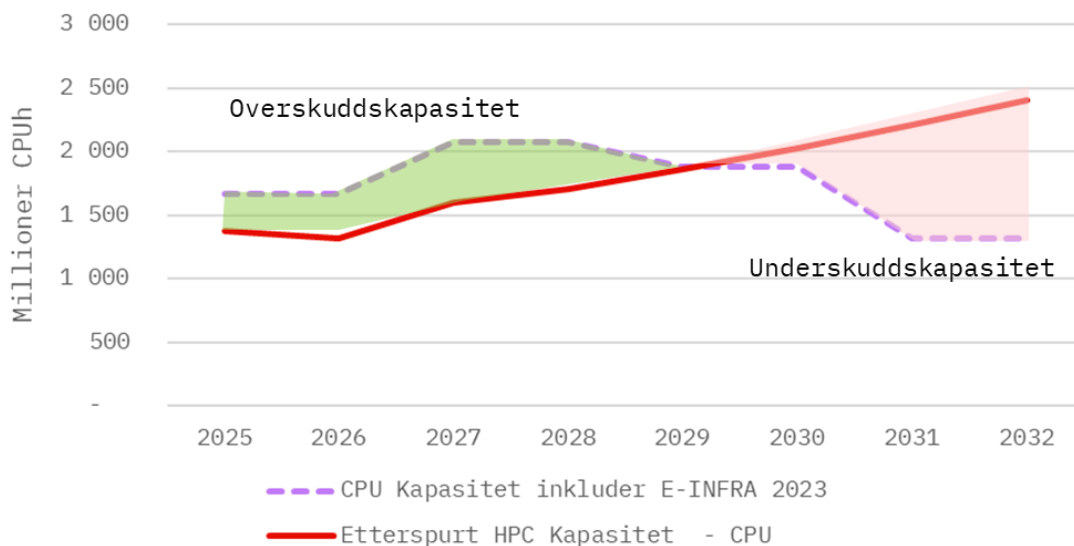
Kilde: PwC Strategy & prognose.

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.

Etterspørselen for kapasitet til tradisjonell tungregning for forskning innhentes fra Sigma2 AS⁵². Etterspørselen blir vurdert på to måter. Den første metoden (A) estimerer fremtidige behov gjennom ekstrapolering av historiske bruksdata. Den andre metoden (B) er gjennom brukerundersøkelser.

Figur 5.4 oppsummerer resultatene fra estimeringen av etterspørselen fra forskningsinstitusjonene, sammenlignet med tilgjengelig kapasitet fra Sigma2. Tilgjengelig kapasitet omfatter også investeringer i en B2 maskin som er inkludert i prosjektsøknaden til INFRASTRUKTUR-utlysningen i 2023.

⁵² Sigma2, «Project description E-INFRA 2023 – A National e-infrastructure for science».



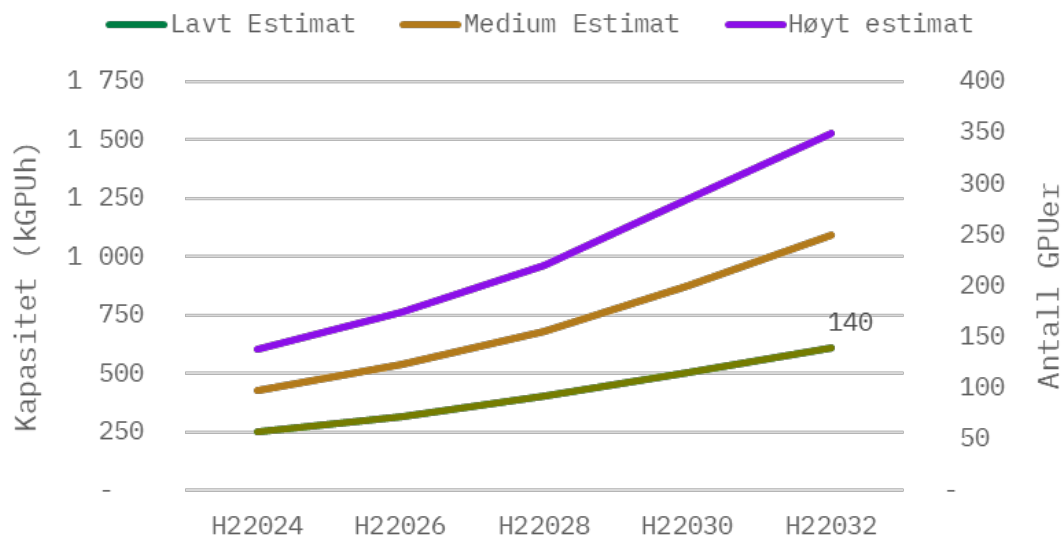
Figur 5.4. Etterspørsel tungregnekapasitet for forskning (CPU-basert)

Kilde: Project description E-INFRA 2023 – A National e-infrastructure for science.

En bekymring er at den økende etterspørselen etter GPU-ressurser ikke fanges opp i Figur 5.4. 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 framtidige 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.

I forbindelse med Sigma2 sin anskaffelse av A2-maskinen ble det foretatt en vurdering av etterspørselen etter GPU-ressurser til forskning.



Figur 5.5. Estimert GPU-behov for å dekke etterspørsel til forskning

Kilde: Project description E-INFRA 2023 – A National e-infrastructure for science.

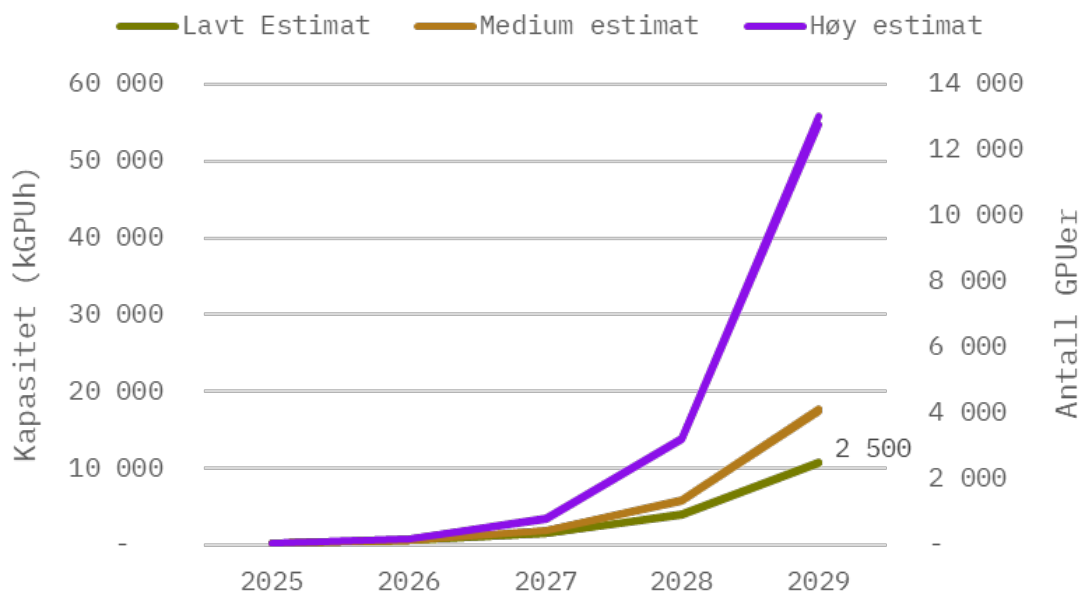
Figur 5.5 er resultatet av GPU-behovsanalysen utført i forbindelse med Sigma2 sin anskaffelse av A2, som illustrerer tre estimater for antall GPU-timer som trengs i den nasjonale infrastrukturen i de kommende årene. GPU-analysen er kun basert på regnskap fra de

nasjonale systemene. Vi vet at det gjennomføres en betydelig opplæring av språkmodeller ved universitetene. Denne belastningen, og mulig økt etterspørsel etter GPU for opplæring av språkmodeller, er i liten grad inkludert i GPU-analysen ovenfor. I sin søknad til INFRASTRUKTUR-utlysningen *Project description E-INFRA 2023 – A National e-infrastructure for science* la Sigma2 til grunn et lavt estimat.

5.3.3 Etterspørselen etter tungregnerressurser til KI-trening/maskinlæring

Det er stor usikkerhet knyttet til å kvantifisere etterspørselen etter tungregnekapasitet KI-trening og maskinlæring. Foreløpige estimater er fra de ledende KI-miljøer i Norge. Estimatenes fra disse uavhengige kilder er sammenfallende, men har ganske stor spredning knyttet til usikkerhet og hvor avanserte modeller som skal trenes, dvs. i størrelsesorden 10-100 ganger kapasiteten som er tilgjengelig i dag.

Vi estimerer etterspørselen med utgangspunkt i dagens GPU-kapasitet gjennom LUMI (Sigma2 har i dag kapasitet tilsvarende 250 GPUer). I vårt lave estimat vil vi nå en kapasitet tilsvarende 2 500 GPUer i 2029, med en initiell kapasitet på 50 GPUer i 2025. Dette gir en gjennomsnittlig årlig vekst på 170 prosent. I vårt høye estimat legger vi til grunn en årlig vekst på 300 prosent, mens i medium estimatet 200 prosent.



Figur 5.6 Prognose på etterspurt GPU kapasitet

Kilde: Project description E-INFRA 2023 – A National e-infrastructure for science.

I det videre arbeidet legger vi til grunn lavt estimat for å beregne investeringsbehovet.

5.4 Prosjektutløsende behov

Med prosjektutløsende behov menes det samfunnsbehovet som utløser planlegging av tiltak som offentlig sektor skal bidra med, til et bestemt tidspunkt.

Nåværende kapasitet innen tungregning er ikke dekkende for dagens kjente behov innenfor forskning og utvikling (FoU) og utvikling av offentlig sektor. 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 tungregnerressurser 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.

Det økende volumet og mangfoldet av data kombinert med nye og mer avanserte algoritmer gir store muligheter for forsknings- og innovasjonsmiljøer, offentlig sektor og næringsliv 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 tungregning.

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 tungregnekapasitet, 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 regneressurser optimalisert for 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.

Med bakgrunn i disse utfordringene og mulighetene er det behov for å øke investeringene i nasjonal infrastruktur for tungregning. Dette utgjør til sammen følgende prosjektutløsende behov:

- Det er behov for en nasjonal infrastruktur som tilrettelegger for, og gir tilstrekkelig kapasitet til tungregning for forskning, utvikling, innovasjon, offentlig forvaltning og næringsliv.
- Det er et økt behov for tungregning innen samfunnskritiske tjenester for å ivareta sikkerhet og beredskap, som for eksempel å forebygge og håndtere helsekriser, geofarar og klimaendringer.
- Infrastrukturen gjelder ikke kun tilgjengeliggjøring av kapasitet, men omfatter anskaffelse av ressurser, drift, videreutvikling, kompetanseoppbygging og -deling, avansert brukerstøtte og vanlig brukerstøtte m.m.
- Behovene blant de ulike brukergruppene er svært ulikt og sammensatt, både i forhold til regnebehov, datamengde, behov for oppetid og grad av sensitivitet. Videre vil behovene endre seg i takt med en svært rask teknologiske utvikling. Det er derfor behov for at investeringen tilrettelegger for stor fleksibilitet, og muligheter til å ta raske beslutninger for videre utvikling. Dette er nødvendig for å ha et tilbud som er teknologisk oppdatert for å være relevant for dagens og fremtidens brukere.
- Med flere brukere involvert er det nødvendig med forutsigbare kriterier for tildeling av tungregneressurser, samt behov for å ha gode mekanismer for deling av regneressursene, for å sikre best mulig ressurseffektivitet.
- For god utnyttelse av kapasitet og kompetanse er det behov for en god kopling både mellom lokale og nasjonale ressurser og mellom de nasjonale og de europeiske ressursene vi investerer i.
- Det er behov for økt kompetanse på bruk (både der man kan og der man ikke bør bruke) av KI i offentlig forvaltning, men også hvordan man skal løse utfordringer rundt forklarbarhet, transparens, tillit m.m.

6 Strategiske mål

Dette kapittelet gir en beskrivelse av samfunnsmålet og effektmålene for prosjektet. Samfunnsmålet gir uttrykk for nytten og verdiskapningen tiltaket skal gi for samfunnet. Målet skal vise intensjonen og ambisjonen med tiltaket, og samtidig reflektere det prosjektutløsende behov. Effektmålene bygger opp under samfunnsmålet og er en konkretisering av hva som ønskes oppnådd for tiltakets målgrupper. De bør i størst mulig grad uttrykke ønskede virkninger av eventuelle tiltak på området, og utformet slik at de beskriver relevante egenskaper ved den ønskede tilstanden etter gjennomføring av tiltaket. Hvor mange effektmål det er hensiktsmessig å sette opp, vil variere fra analyse til analyse. Dersom det utformes flere effektmål, bør de være innbyrdes konsistente og prioritering mellom målene skal fremgå.

6.1 Samfunns mål

I en verden preget av rask teknologisk utvikling og økende digitalisering og tilgang til data, er tungregning en viktig komponent for å sikre nasjonal konkurranseevne og høy innovasjonskraft. Norge står overfor betydelige samfunnsutfordringer, inkludert klimaendringer, aldrende befolkning og behovet for bærekraftig ressursforvaltning. For å møte disse utfordringene er det nødvendig med en infrastruktur som kan støtte avansert forskning og utvikling av kunstig intelligens. En forutsigbar og fleksibel tungregneinfrastruktur vil ikke bare styrke Norges posisjon som en ledende kunnskapsnasjon, men også fremme datadrevet innovasjon og effektivisering i offentlig forvaltning. Dette kan bidra til å skape et mer bærekraftig samfunn, hvor teknologiske fremskritt kan anvendes til å forbedre livskvaliteten og sikre en trygg fremtid for kommende generasjoner. På denne bakgrunn er samfunnsmålet for denne konseptvalgutredningen definert som følger:

En forutsigbar, fleksibel og bærekraftig infrastruktur for tungregning, som muliggjør forskning i verdensklasse, en offentlig sektor som leverer gode og effektive tjenester og utvikling av kunstig intelligens, som bidrar til å løse samfunnsutfordringer og fremme innovasjon i Norge.

6.2 Effektmål

Vi har definert totalt fire effektmål for denne konseptvalgutredningen. Effektmålene bygger opp under samfunnsmålet. Flere av målene konkretiserer av hva som ønskes oppnådd for tiltakets primære målgrupper, det vil si forskningsorganisasjoner og offentlig forvaltning. I tillegg har vi formulert et effektmål som omfatter kunstig intelligens og et effektmål knyttet til miljø og bærekraft.

Effektmål 1

Forskningsorganisasjoner er hjørnesteiner i utviklingen av ny kunnskap og teknologi. For å kunne utføre forskning av høy kvalitet, er det sentralt at disse institusjonene har tilgang til tungregneinfrastruktur. Det gir forskerne muligheten til å utføre komplekse simuleringer og analyser som er nødvendige for å forstå og løse utfordringer innenfor ulike fagområder som medisin, klima, teknologi m.m. Videre vil tilstrekkelig tungregnekapasitet støtte utviklingen av nye metoder og verktøy som kan drive innovasjon fremover. Ved å sikre at forskningsmiljøene har tilstrekkelige tungregnekapasitet, kan vi også tiltrekke og beholde talentfulle forskere, noe som er avgjørende for å opprettholde Norges konkurranseevne på den globale forskningsarenaen. På denne bakgrunn er effektmål 1 for denne konseptvalgutredningen definert som følger:

Forskningsorganisasjoner har tilgang til tilstrekkelig og skalerbar infrastruktur for tungregning som styrker forskningskapasiteten, fremmer innovasjon og bidrar å utvikle ny kunnskap.

Effektmål 1 har høy prioritet blant effektmålene.

Mulige indikatorer for måloppnåelse:

- Antall forskningsprosjekter som har tilgang til offentlig finansierte tungregneressurser.
- Antall publiserte forskningsresultater hvor tungregning har vært benyttet.
- Brukertilfredshet blant forskere med tilgang til regneressurser.

Effektmål 2

Offentlig forvaltning står overfor stadig mer komplekse oppgaver som krever effektive og pålitelige løsninger. Tilgang til sikker og forutsigbar tungregneinfrastruktur er avgjørende for å kunne levere samfunnskritiske tjenester som helse, sikkerhet og miljøovervåkning. I tillegg gir kunstig intelligens muligheter for å forbedre tjenestekvaliteten og effektivisere prosesser, noe som kan føre til bedre ressursutnyttelse og bedre tjenester til innbyggerne. Ved å investere i nødvendig tungregneinfrastruktur, kan vi sikre at offentlige virksomheter er rustet til å møte fremtidige utfordringer og tilpasse seg raskt skiftende behov. Dette vil også bidra til å styrke tilliten til offentlige tjenester. På denne bakgrunn er effektmål 2 for denne konseptvalgutredningen definert som følger:

Offentlig forvaltning har tilgang til sikker og forutsigbar infrastruktur for tungregning slik at offentlige virksomheter kan levere tjenester av samfunnskritisk betydning, samt utnytte kunstig intelligens til å forbedre offentlige tjenester og effektivisere offentlig forvaltning.

Effektmål 2 har høy prioritet blant effektmålene.

Mulige indikatorer for måloppnåelse:

- Antall offentlige tjenester som forbedres ved bruk av tungregning.
- Reduksjon i behandlingstid for offentlige prosesser ved hjelp av KI.

Effektmål 3

Næringslivet er en sentral drivkraft for økonomisk vekst og innovasjon i et samfunn. For å kunne utvikle nye teknologier, produkter og tjenester, er det avgjørende at bedrifter har tilgang til pålitelig tungregneinfrastruktur. Det gjelder både store, mellomstore og små bedrifter som ønsker å forbedre sine interne prosesser og øke sin konkurransevne. Ved å legge til rette for at næringslivet kan utnytte avansert tungregneteknologi, kan vi stimulere til økt verdiskapning og innovasjon. Dette vil også bidra til å skape nye arbeidsplasser og styrke Norges posisjon i det globale markedet. På denne bakgrunn er effektmål 3 for denne konseptvalgutredningen definert som følger:

Norsk næringsliv har tilgang til infrastruktur for tungregning der det private leverandørmarkedet ikke strekker til slik at næringslivet kan øke verdiskapingen gjennom utvikling av nye teknologier, forretningsmodeller, produkter og tjenester.

Effektmål 3 har lavest prioritet blant effektmålene. Årsaken til dette er at det ikke er en offentlig oppgave å dekke næringslivet sitt samlede behov for tungregning.

Mulige indikatorer for måloppnåelse:

- Antall bedrifter som benytter tilrettelagt tungregneinfrastruktur for innovasjon og utvikling av tjenester og produkter.
- Antall nye produkter og tjenester utviklet hvor det har blitt benyttet tilrettelagt tungregneinfrastruktur
- Tilfredshet blant næringslivsaktører med tilgang til regnekraft.

Effektmål 4

I en tid hvor bærekraft og miljøhensyn er i fokus, er det viktig at tungregneinfrastrukturen benytter fornybar energi og energieffektive løsninger. Dette vil bidra til å redusere karbonavtrykket og støtte Norges forpliktelser til nasjonale og internasjonale klimamål. Ved å

fokusere på miljøvennlige løsninger, kan vi sikre at utviklingen av tungregning skjer på en ansvarlig måte som ivaretar både nåværende og fremtidige generasjoners behov. Dette vil også fremme en bærekraftig forvaltning av ressurser og styrke Norges omdømme som en leder innen grønn teknologi. På denne bakgrunn er effektmål 4 for denne konseptvalgutredningen definert som følger:

Infrastrukturen for tungregneknkraft benytter fornybar energi og energieffektive løsninger, som reduserer karbonavtrykket og fremmer bærekraftig forvaltning.

Effektmål 4 har middels prioritet blant effektmålene.

Mulige indikatorer for måloppnåelse:

- Andel av energiforbruk fra fornybare kilder i tungregneinfrastrukturen.
- Reduksjon i karbonavtrykk fra driften av regneressurser.
- Implementering av energieffektive løsninger i nye og eksisterende anlegg.

6.3 Resultatmål

Resultatmål viser alle tiltak som skal bidra til at effektmålene innfris. Hvis det ikke er en klar sammenheng mellom målene og problembeskrivelsen, er det en fare for at man ender opp med tiltak som i liten grad bidrar til å løse problemet. I tillegg er det viktig at målene ikke utformes så snevert at de virker begrensende på muligheten til å finne relevante tiltak. Derfor er det viktig å fokusere på hva man ønsker å oppnå med tiltaket, og ikke på selve løsningen. Samtidig er det viktig å ikke formulere målene på en så vag måte at det man ønsker å oppnå blir utydelig. Da blir det utfordrende å vurdere relevansen av de ulike tiltakene.

Tid, økonomi og kvalitet utgjør de sentrale styringsmessige parameterne i prosjektet. Parameterne henger sammen og påvirker hverandre gjennom prosjektets livssyklus. For eksempel vil et større ambisjonsnivå i et prosjekt ofte resultere i økte kostnader og/eller tid, mens motsatt vil et prosjekt som ønskes levert raskere normalt sett innebære lavere kvalitet i form av mindre omfang eller høyere ytelse. Samtidig vil et press på tid skape press på økte kostnader eller redusert kvalitet. Denne styringsmodellen og disse avhengighetene skaper styrbarhet i prosjekter, og vil gi utslag i bedre prosjektstyring underveis med nødvendig avveining i alle fasene av prosjektene.

7 Rammebetingelser og andre kriterier for konseptvalg

Rammebetingelsene omfatter et samlet sett betingelser som skal oppfylles for valg av konsept og fremtidig drift. Rammebetingelser deles i to typer; rammebetingelser som kan utledes av samfunns- og effektmålene og rammebetingelser som relateres til andre ikke-prosjektspesifikke mål og prinsipielle spørsmål. De identifiserte rammebetingelsene i denne konseptvalgutredningen er presentert i Tabell 7.1. Etter tabellen beskrives rammebetingelsene nærmere.

Tabell 7.1 Oversikt over rammebetingelser

#	Beskrivelse
R1	Infrastrukturen for tungregning må være pålitelig og robust
R2	Infrastrukturen for tungregning må oppfylle sikkerhetskrav , herunder tungregneressurser som ikke er plassert i Norge må ligge i land vi har sikkerhetspolitisk samarbeid med
R3	En tilstrekkelig andel av tungregneressursene må lokaliseres i Norge pga. sikkerhetshensyn og kompetanse
R4	Infrastrukturen for tungregning må være skalerbar og fleksibel til fremtidige endringer i behov for tungregnekapasitet
R5	Infrastrukturen for tungregning må oppfylle statens krav til energieffektivitet
R6	Det må legges til rette for gode støttetjenester som er tilpasset <i>ulike</i> brukergruppers behov for å kunne utnytte tungregneressursene både nasjonalt og internasjonalt
R7	Infrastrukturen for tungregning skal så langt som mulig bygge på eksisterende systemer og strukturer
#	Andre vurderingskriterier
A1	Infrastrukturen for tungregning bør ha potensial for høy måloppnåelse for samtlige effektmål
A2	Infrastrukturen for tungregning bør være så ressurseffektiv som mulig
A3	Gjennomføringsrisikoen for investeringene og organiseringen av infrastrukturen for tungregning bør være lav

R1 - Infrastrukturen for tungregning må være pålitelig og robust

Pålitelighet og robusthet er fundamentale egenskaper for tungregneinfrastruktur, uavhengig av bruksområde. Disse egenskapene sikrer at systemene kan håndtere store datamengder og komplekse beregninger uten avbrudd. Robusthet kjennetegnes ved systemets evne til å tåle feil og fortsatt fungere effektivt, mens pålitelighet sikrer kontinuerlig tilgjengelighet og stabil ytelse. Effektiv overvåking, vedlikehold og evnen til å identifisere og løse problemer raskt (responstid) er essensielt for å sikre oppetid på systemet. Krav til oppetid og responstid vil imidlertid variere avhengig av bruksområde:

- **Forskningsdrevet FoU:** Moderat til høy oppetid er viktig, da avbrudd kan forsinke forskningsprosjekter. Noe fleksibilitet kan være akseptabelt, avhengig av prosjektets tidslinje.
- **Forvaltningsdrevet FoU:** Høy oppetid er kritisk for å sikre kontinuerlig tilgang til data og analyser som støtter beslutningsprosesser og politikktutforming.
- **KI utvikling og trening:** Høy oppetid er nødvendig for å unngå avbrudd i treningsprosesser som kan ta lang tid og kreve betydelige ressurser.
- **KI bruk og tilpasning (inferens):** Ekstremt høy oppetid er essensielt, spesielt for sanntidsapplikasjoner, hvor selv korte avbrudd kan påvirke brukererfaringen og funksjonaliteten negativt.

Disse variasjonene reflekterer de spesifikke behovene og toleransene for nedetid i hvert bruksområde.

R2 - Infrastrukturen for tungregning må oppfylle sikkerhetskrav, herunder tungregnerressurser som ikke er plassert i Norge må ligge i land vi har sikkerhetspolitisk samarbeid med

Infrastrukturen for tungregning må oppfylle strenge sikkerhetskrav, og dette reguleres av flere lover og forskrifter i Norge og EU. En sentral lov er Nasjonal sikkerhetslov, som krever at virksomheter beskytter kritisk infrastruktur mot trusler som kan påvirke nasjonal sikkerhet. Denne loven pålegger virksomheter å gjennomføre risikoanalyser og implementere nødvendige sikkerhetstiltak for å beskytte informasjonssystemer.

Ekomloven er også viktig, da den regulerer elektronisk kommunikasjon og stiller krav til sikkerhet og beredskap i nettverk og tjenester. Dette innebærer at tilbydere av elektroniske kommunikasjonstjenester må sikre sine systemer mot cybertrusler og sørge for at data er beskyttet mot uautorisert tilgang.

På EU-nivå er Personvernforordningen (GDPR) avgjørende for å sikre at personopplysninger behandles på en sikker måte. GDPR krever at data som behandles av tungregneinfrastruktur er beskyttet med passende tekniske og organisatoriske tiltak, som kryptering og tilgangskontroller, for å forhindre databrudd.

NIS-direktivet (Network and Information Security Directive) stiller krav til medlemslandene om å beskytte kritisk infrastruktur mot cybertrusler. Dette direktivet krever at operatører av essensielle tjenester, som tungregning, implementerer sikkerhetstiltak og rapporterer sikkerhetshendelser til relevante myndigheter.

I tillegg til disse lovene, er internasjonale standarder som ISO/IEC 27001 ofte brukt som rammeverk for å etablere og vedlikeholde et effektivt styringssystem for informasjonssikkerhet. Selv om dette ikke er en lov, gir det en strukturert tilnærming til å håndtere sikkerhetsrisikoer.

For å sikre samsvar med disse kravene, må organisasjoner regelmessig evaluere og oppdatere sine sikkerhetstiltak. Dette inkluderer å utføre sikkerhetsrevisjoner, trene ansatte i sikkerhetsprosedyrer, og implementere teknologiske løsninger som brannmurer og inntrengingsdeteksjonssystemer.

Samlet sett sikrer disse lovene og reguleringene at tungregneinfrastruktur i Norge er robust, pålitelig og i stand til å beskytte mot et bredt spekter av sikkerhetstrusler. Dette er avgjørende for å opprettholde tillit og sikre at kritiske data og tjenester er beskyttet.

Tungregneressurser utenfor Norge må ligge i land med sikkerhetspolitisk samarbeid. Dette er primært styrt av nasjonal sikkerhetslovgivning og internasjonale avtaler. Nasjonal sikkerhetslov krever beskyttelse av kritisk infrastruktur, og dette inkluderer vurdering av hvor data lagres og behandles. Loven understreker behovet for å minimere risikoen for utenlandsk tilgang til sensitive data. Videre, Ekomloven stiller krav til sikkerhet og beredskap i elektronisk kommunikasjon, noe som påvirker valg av lokasjon for datalagring.

R3 - En tilstrekkelig andel av tungregneressursene må lokaliseres i Norge pga. sikkerheshensyn og kompetanse

Å lokalisere en tilstrekkelig andel av tungregneressursene i Norge er viktig av flere grunner. **Sikkerheshensyn** er en primær driver, da lokal lagring og behandling av data reduserer risikoen for utenlandsk tilgang og potensielle sikkerhetsbrudd. Dette er spesielt kritisk for sensitive data knyttet til nasjonal sikkerhet og personvern. Videre bidrar lokal infrastruktur til å **oppretholde og utvikle nasjonal kompetanse innen IT og teknologi**, noe som er essensielt for innovasjon og konkurranseevne. Det gir også bedre kontroll over databehandlingsprosesser og sikrer samsvar med norske lover og reguleringer. I tillegg kan lokal tilstedeværelse redusere latens og forbedre ytelsen for brukere i Norge. Samlet sett styrker dette landets digitale suverenitet og evne til å beskytte kritisk infrastruktur.

R4 - Infrastrukturen for tungregning må være skalerbar og fleksibel til fremtidige endringer i behov for tungregnekapasitet.

Som beskrevet i behovsanalysen er det knyttet stor usikkerhet til hvordan behovet for tungregnekapasitet vil utvikle seg de neste årene. Spesielt er etterspørselen for GPU-basert kapasitet usikker. Prognosene varierer mellom 10 og 100 ganger dagens kapasitet. Det er derfor essensielt at valgt konsept kan håndtere denne usikkerheten. Det betyr følgende:

- For å sikre fleksibilitet i investeringer i ny kapasitet, bør en nasjonal plattformleverandør utforme en **finansieringsmodell** som er både dynamisk og tilpasningsdyktig. Dette er utfordrende i dagens modell, der Sigma2 søker om midler til nye eller erstatning av eksisterende anlegg gjennom INFRASTRUKTUR-ordningen. Sigma2 får tildelt midler oppad begrenset til 200 millioner kroner. I tildelingen gis det føringer om midlene skal brukes til anleggsinvesteringer i Norge eller gjennom EuroHPC.

En fremtidig finansieringsmodell må legge til rette for at ny leverandør kan tilføres midler til fremtidige investeringer, slik at de kan justere kapasiteten i takt med endrede behov og teknologiske fremskritt. Dette gir mulighet for rask omfordeling av midler når nye krav oppstår.

Det bør innføres krav om langsiktig planlegging. Ved å utvikle en strategisk, scenariobasert plan (for eksempel 4-årig), kan leverandøren forutse og tilpasse seg fremtidige krav mer effektivt. Denne helhetlige tilnærmingen til finansiering vil gi en statlig plattformleverandør den nødvendige fleksibiliteten til å møte skiftende behov og teknologiske utfordringer på en kostnadseffektiv måte.

- **Samarbeid og partnerskap** kan betydelig forbedre skalerbarhet og fleksibilitet ved å gi tilgang til delte ressurser og ekspertise. Gjennom samarbeid kan man dra nytte av felles infrastruktur og teknologi, noe som reduserer kostnader og øker kapasiteten. Partnerskap med teknologileverandører gir tilgang til de nyeste innovasjonene og muligheten til å raskt tilpasse seg endringer. Dette samarbeidet fremmer også kunnskapsdeling, som styrker evnen til å håndtere komplekse utfordringer. Samlet sett gir dette en mer dynamisk og responsiv infrastruktur.

Norge deltar i Programmet for et digitalt Europa (DIGITAL) 2021–2027 – EUs investerings- og kapasitetsbyggingsprogram for digital omstilling og bruk av innovative digitale teknologier i samfunn og næringsliv. EuroHPC er et av satsingsområdene som har som mål å styrke EUs kapasitet på tungregning, og utvikle et økosystem for tungregning som lar Europa konkurrere globalt. Gjennom sin deltakelse i EuroHPC forplikter Norge seg til en nasjonal koordinering av innsatsen innen tungregneinfrastruktur. Sigma2, i tett samarbeid med SINTEF og NORCE, er i dag nasjonalt kompetansesenter for tungregning i EuroHPC-samarbeidet (NCC).

- Hvordan etablering og tilgjengeliggjøring av tungregnekapasitet **organiseres og styres** påvirker evnen til å skalere og fleksibiliteten til å omdisponere kapasiteten. Sentralisert

styring kan gi bedre kontroll og koordinering, noe som forenkler beslutningsprosesser for skalering. Det gir også mulighet for standardisering av teknologier og prosesser. Desentralisert styring kan gi større fleksibilitet lokalt, slik at ulike enheter kan tilpasse seg spesifikke behov raskere.

R5 - Infrastrukturen for tungregning må oppfylle statens krav til energieffektivitet

Infrastrukturen for tungregning må oppfylle statens krav til energieffektivitet, noe som er avgjørende for å redusere miljøpåvirkningen og driftskostnadene. Dette innebærer at datasentre og andre fasiliteter må designes og drives med fokus på å minimere energiforbruket. Bruk av energieffektive maskinvarekomponenter, som prosessorer og kjølesystemer, er essensielt for å oppnå dette. Videre bør det implementeres avanserte energistyringssystemer som overvåker og optimaliserer energibruken i sanntid. Spillvarme fra tungregneinfrastruktur bør utnyttes som en ressurs for å forbedre energieffektiviteten. Datasentre genererer betydelige mengder varme under drift, og denne varmen kan gjenvinnes og brukes til oppvarming av bygninger eller i industrielle prosesser. Dette fører til at man ikke bare reduserer karbonutslipp, men også bidrar til lokal energiforsyning.

Statlige krav kan også inkludere spesifikke mål for reduksjon av karbonutslipp, noe som kan oppnås ved å integrere fornybare energikilder som sol- og vindkraft i energiforsyningen. I tillegg kan bruk av teknologi som virtualisering bidra til å øke utnyttelsesgraden av eksisterende ressurser, noe som reduserer behovet for ekstra maskinvare og dermed energiforbruk.

Energieffektivitet er også en del av bærekraftstrategier, hvor det er viktig å rapportere og dokumentere energibruken for å sikre samsvar med nasjonale og internasjonale standarder. Dette kan inkludere sertifiseringer som ISO 50001, som gir en strukturert tilnærming til energiledelse. Ved å oppfylle disse kravene, bidrar tungregneinfrastrukturen til statens overordnede mål om bærekraft og miljøvern, samtidig som det sikrer kostnadseffektiv drift.

R6 - Det må legges til rette for gode støttetjenester som er tilpasset ulike brukergruppers behov for å kunne utnytte tungregneressursene både nasjonalt og internasjonalt

For å maksimere utnyttelsen av tungregneressurser både nasjonalt og internasjonalt, er det avgjørende å legge til rette for gode støttetjenester som er tilpasset ulike brukergruppers behov. Dette innebærer å tilby omfattende kompetanseutvikling og avansert brukerstøtte som kan hjelpe brukere med å navigere komplekse systemer og teknologier.

En viktig del av denne rammebetingelsen er å sikre at brukerne har tilgang til opplæring og ressurser som kan forbedre deres ferdigheter i å bruke tungregneinfrastruktur effektivt. Dette kan inkludere kurs, workshops og online læringsplattformer som dekker alt fra grunnleggende bruk til avanserte teknikker innen dataanalyse og modellering.

Avansert brukerstøtte er også kritisk. Dette innebærer å ha et dedikert team av eksperter som kan tilby teknisk støtte og veiledning, samt hjelpe til med feilsøking og optimalisering av beregningsprosesser. Støttetjenestene bør være tilgjengelige på flere nivåer, fra grunnleggende hjelp til spesialisert rådgivning for komplekse prosjekter.

For å møte behovene til ulike brukergrupper, bør støttetjenestene være fleksible og tilpasset spesifikke sektorer, som forskning, industri og offentlig sektor. Dette kan innebære å utvikle spesialiserte verktøy og ressurser som er skreddersydd for de unike kravene i hvert forvaltningsområde.

Internasjonalt samarbeid er også en viktig faktor. Ved å delta i globale nettverk og partnerskap kan norske institusjoner dra nytte av internasjonal ekspertise og beste praksis, noe som kan forbedre kvaliteten på støttetjenestene. Dette samarbeidet kan også bidra til å standardisere prosesser og verktøy, noe som gjør det enklere for brukere å operere på tvers av landegrenser.

Til slutt, kontinuerlig evaluering og tilbakemelding fra brukerne er essensielt for å forbedre og tilpasse støttetjenestene over tid. Ved å forstå brukernes behov og utfordringer kan tjenestene utvikles for å sikre at alle brukere, uansett nivå, kan utnytte tungregneressursene fullt ut. Dette bidrar til å styrke Norges posisjon som en ledende aktør innenfor tungregning og avansert databehandling.

R7 - Infrastrukturen for tungregning skal så langt som mulig bygge på eksisterende systemer og strukturer

Rammebetingelsen om at infrastrukturen for tungregning skal bygge på eksisterende systemer og strukturer, handler om å utnytte og integrere nåværende ressurser for å oppnå effektivitet og kostnadsbesparelser. Ved å bygge videre på eksisterende infrastruktur, kan man redusere behovet for nye investeringer og minimere implementeringstiden.

Dette innebærer å evaluere og optimalisere nåværende teknologier og prosesser for å sikre at de kan støtte økte krav til beregningskapasitet og behovene til nye brukergrupper. Det kan inkludere oppgradering av maskinvare, forbedring av nettverkskapasitet, og implementering av programvareløsninger som kan integreres sømløst med eksisterende systemer.

A1 - Infrastrukturen for tungregning bør ha potensial for høy måloppnåelse for samtlige effektmål

Konseptene som har høyest måloppnåelse på effektmålene vil ha prioritet.

A2 - Infrastrukturen for tungregning bør være så ressurseffektiv som mulig

Med ressurseffektiv menes at vi gjør en kvalitativ evaluering av investeringskostnadene og samlede driftskostnader.

A3 - Gjennomføringsrisikoen for investeringene og organiseringen av infrastrukturen for tungregning bør være lav

Gjennomføringsrisikoen er knyttet til i hvilken grad det må gjennomføres endringer i eierstruktur og styringsmodell, endring i organisering og eventuell utvidelse/etablering av avansert brukerstøtte.

8 Mulighetsstudie

Hensikten med mulighetsstudien er å identifisere mulige tiltak, og grunnlaget for dette arbeidet er problembeskrivelsen, behovsanalysen og de strategiske målene som er beskrevet i sine respektive kapitler.

Denne mulighetsstudien følger vanlig metodikk innenfor konseptvalgutredninger. Først går vi bredt ut, dvs. at vi utforsker mulighetsrommet. Vi starter med å identifisere ulike mulighetsdimensjoner og definere ulike valgmuligheter innenfor hver dimensjon. På bakgrunn av dimensjonene konkretiseres ulike løsninger for tungregning, som i henhold til konseptvalgutrednings-metodikken er kjent som konsepter. I dette steget er det vanlig å gå bredt ut og utforme mange ulike konsepter for å vise spennet av ulike løsningsalternativer. Hele spekteret av mulige løsninger inngår i mulighetsrommet. Et viktig premiss i mulighetsrommet er å vise bredden av ulike konsepter som kan løse problemet, uavhengig av kompleksitet og sannsynlighet for realisering. I tillegg er det en sentral forutsetning for analysen at konseptene som skal vurderes er konseptuelt ulike fra hverandre, og ikke en variasjon av samme konsept. I neste steg blir alle de utarbeidede konseptene vurdert opp mot rammebetingelsene som er satt for konseptene. Konseptene som innfrir rammebetingelsene vurderes videre opp mot et sett av andre vurderingskriterier. Denne prosessen er kjent som nedvalg eller siling av konsepter. På bakgrunn av helhetsvurderingen blir de mest hensiktsmessige konseptene tatt med videre til analyse av virkningene av konseptene. Analysen er kjent som alternativanalysen, og presenteres i neste kapittel.

8.1 Mulighetsrommet

Med utgangspunkt i problembeskrivelsen, behovsanalysen og mål og krav identifiseres de mulighetsdimensjoner som i størst mulig grad «spenner ut lerretet». I beskrivelsen av hver enkelt mulighetsdimensjon gjennomføres en faglig vurdering av hva som definerer mulighetsrommet for hver enkelt mulighetsdimensjon isolert sett uavhengig av andre dimensjoner.

Følgende mulighetsdimensjoner er identifisert:

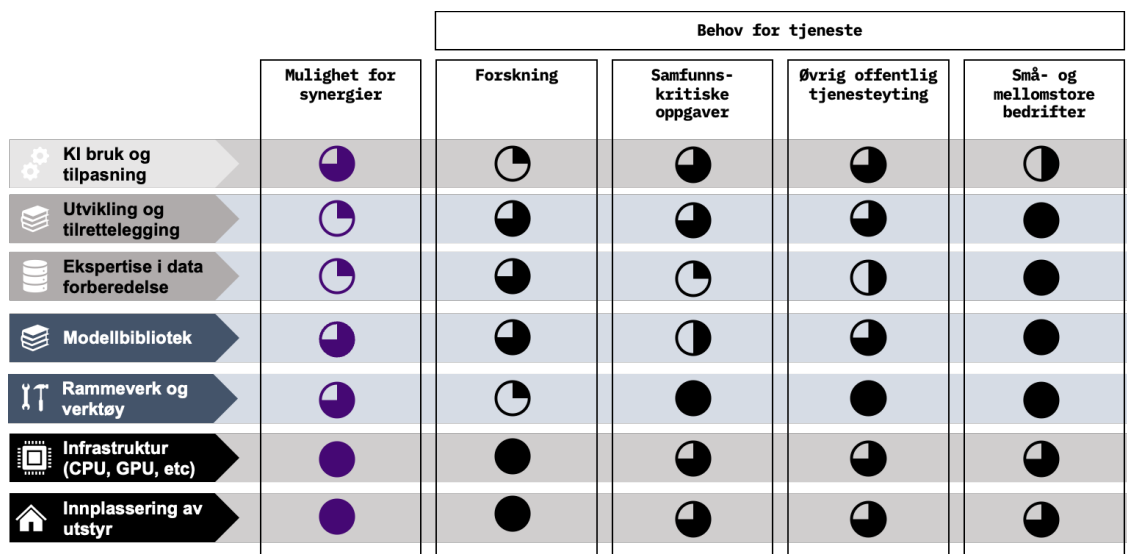
- A. **Formål og bruk.** Denne mulighetsdimensjonen beskriver omfanget av brukere og formål for tungregning og lagring som skal omfattes av konseptet. Implisitt i dimensjonen ligger også behovet for å håndtere og beskytte sensitive data. For eksempel vil utvikling av generiske språkmodeller ha lavt behov for beskyttelse, mens utvikling av KI-modeller for bruk innen helsetjenesten ha stort behov for beskyttelse.
- B. **Tildelingsmodell.** Denne mulighetsdimensjonen beskriver hvordan tungregnekapasitet skal tildeles til de ulike brukergruppene som skal ha tilgang til tungregnekapasitet og lagring.
- C. **Leveransemodell inkl. grunnleggende brukerstøtte.** Denne mulighetsdimensjonen beskriver hvordan anskaffelse, innstallering og drift av infrastrukturen og grunnleggende brukerstøtte organiseres. Leveranse av infrastrukturen omfatter å bygge ut nødvendig kapasitet, enten denne er i Norge eller gjennom EuroHPC, samt å sørge for at infrastrukturen har høy oppetid og sikkerhet iht. brukernes krav.
- D. **Organisering av avansert brukerstøtte (AUS).** Denne mulighetsdimensjonen beskriver hvordan avansert brukerstøtte organiseres. Avansert brukerstøtte tilbyr hjelp som går utover grunnleggende brukerstøtte, for å gi spesialisert og mer dyptgående kompetanse til en virksomhet eller brukergruppe. Den avanserte brukerstøtten bidrar til effektiv og optimal bruk av de nasjonale data- og datalagringsfasilitetene.
- E. **Lokalisering av ressurser.** Denne mulighetsdimensjonen beskriver hvor tungregnings- og lagringskapasiteten lokaliseres og reflekterer behovet for selvberging og håndtering av sensitive data.

F. **Finansiering og eierskap.** Denne mulighetsdimensjonen beskriver hvordan investering i ny kapasitet skal finansieres og eies, samt hvordan de fortløpende driftskostnadene skal dekkes.

I de følgende avsnittene beskrives kortfattet de ulike mulighetsdimensjonene.

8.1.1 Mulighetsdimensjon A: Formål og bruk

Denne mulighetsdimensjonen beskriver omfanget av brukere og formål for tungregning og lagring som skal omfattes av konseptet. Mulighetsrommet for denne dimensjonen er i stor grad styrt av hvilke synergier som er mulig å hente ut ved å inkludere flere brukergrupper til å dele på den samme infrastrukturen og tjenesten. Figur 8.1 gir en oversikt over kartlegging av hvilke tjenester de ulike brukergruppene har behov for og muligheten for synergier på tvers av brukergruppene. Kartleggingen er basert på behovsanalysen.



Figur 8.1 Oversikt over kartlegging av hvilke tjenester de ulike brukergruppene har behov for og muligheten for synergier på tvers av brukergruppene

De største synergiene på tvers av brukergruppene er relatert til anskaffelse, drift og innplassering av utstyr. Et samarbeid her vil kunne gi storskalafordeler i forbindelse med anskaffelser og drift. Høyere opp i stakken vil vi ha noen synergier ved bruk av felles plattformer (programvare, bibliotek og verktøy) men selve utnyttelsen av plattformene vil både trenge domenekompetanse og kanskje forskjellige verktøy.

Vi ser fra behovsanalysen at store bedrifter i mindre grad har behov for felles infrastruktur og datasenterleverandører for innplassering av utstyr. De fleste av disse har allerede etablerte avtaler med egne leverandører.

Mulighetsrommet for formål og bruk er beskrevet nærmere i

Tabell 8.1.

Tabell 8.1 Beskrivelse av Formål og bruk

Dekker forsknings- og forvaltningsdrevet FoU	Dekker forsknings- og forvaltningsdrevet FoU, samt kapasitet til å bruke og tilpasse samfunnskritiske tjenester	Dekker forsknings- og forvaltningsdrevet FoU, samt kapasitet til å bruke og tilpasse alle offentlige tjenester
<ul style="list-style-type: none"> Dekker behovet for tungregning der problemstillingen enten er initiert fra forskningsordinasjoner, eller fra offentlig forvaltningen Legges til rette for samarbeid med næringslivet 	<ul style="list-style-type: none"> Dekker behovet for tungregning der problemstillingen enten er initiert fra forskningsorganisasjoner, eller fra offentlig forvaltningen Dekker behovet for at samfunnskritiske tjenester kan bruke kapasitet til å levere tjenester Legges til rette for samarbeid med næringslivet 	<ul style="list-style-type: none"> Dekker behovet for tungregning der problemstillingen enten er initiert fra forskningsorganisasjoner, eller fra offentlig forvaltningen Dekker behovet for at alle i offentlig forvaltning kan bruke kapasitet til å levere tjenester Legges til rette for samarbeid med næringslivet

8.1.2 Mulighetsdimensjon B: Tildelingsmodell

Denne mulighetsdimensjonen beskriver hvordan tungregnekapasitet skal tildeles til de ulike brukergruppene som skal ha tilgang til tungregnekapasitet og lagring.

Innenfor forskningsmiljøene er det allment akseptert at ressurser blir tildelt basert på kvaliteten på prosjektet som søker om å få tildelt kapasitet. Dette er en ordning som ikke vil la seg gjennomføre i de andre brukergruppene. Mulighetsrommet ligger i å etablere et kvotesystem, slik at en brukergruppe får reservert en gitt kapasitet. Den enkelte brukergruppen kan selv etablere en tildelingsmodell som er tilpasset denne gruppens preferanser. I tillegg kan det i noen alternativer åpnes opp for at virksomheter eller samling av virksomheter selv kan anskaffe utstyr å få innplassert i felles datasenter, eller anskaffer egen kapasitet i en privat sky eller av en kommersiell skyleverandør.

Mulighetsrommet for tildelingsmodell er beskrevet nærmere i Tabell 8.2.

Tabell 8.2 Beskrivelse av Tildelingsmodell

Søknadsbasert allokering av felles ressurser	Kvotearlokering av felles ressurser	Kvotearlokering av felles ressurser og innplassering av eget utstyr	Kvotearlokering av felles ressurser og anskaffelse av egen kapasitet
<ul style="list-style-type: none"> I dagens situasjon tildeles ressurser basert på en evaluering av kvaliteten på det prosjektet som har behov for kapasitet. 	<ul style="list-style-type: none"> I dette alternativet blir hvert forvaltnings-område og virksomhet tildelt kapasitet på felles ressurser etter det behov. Hvert forvaltnings-område eller virksomhet kan deretter selv organisere hvordan de velger å utnytte den tildelte kapasiteten. 	<ul style="list-style-type: none"> I dette alternativet kan et forvaltningsområde eller virksomhet i tillegg til å få tildelt kapasitet på felles ressurser, anskaffe eget utstyr som de kan få innplassert hos en felles driftsorganisasjon. 	<ul style="list-style-type: none"> I dette alternativet kan et forvaltningsområde eller virksomhet i tillegg til å få tildelt kapasitet på felles ressurser, anskaffe egen kapasitet i en privat sky eller av en kommersiell skyleverandør.

8.1.3 Mulighetsdimensjon C. Leveransemodell inkl. grunnleggende brukerstøtte

Denne mulighetsdimensjonen beskriver hvordan anskaffelse, innstallering og drift av infrastrukturen og grunnleggende brukerstøtte organiseres. Leveranse av infrastrukturen omfatter å bygge ut nødvendig kapasitet, enten denne er i Norge eller gjennom EuroHPC, samt å sørge for at infrastrukturen har høy oppetid og sikkerhet iht. brukernes krav.

Innen forskningsdomenet er det etablert tjenesteavtaler mellom Sigma2 og BOTT-universitetene for å sikre kompetanse og kapasitet til sentrale driftsoppgaver, felles generell brukerstøtte og avansert brukerstøtte. Dette er en samarbeidsmodell, som fungerer i dagens situasjon, men som ikke vil være tilstrekkelig i en modell hvor også andre brukergrupper skal utnytte felles infrastruktur. Mulighetsrommet ligger i å enten transformere Sigma2 til en felles plattformleverandør med egen kapasitet til å håndtere driftsoppgaver, eller å etablere en ny plattformleverandør. Kapasiteten til å håndtere driftsoppgaver er ikke tilknyttet noen av brukerne.

Universitetene har en stor grad av autonomi innenfor sine budsjetter. Det er derfor viktig å opprettholde muligheten for disse å selv å kunne anskaffe og drifte egen infrastruktur. Mulighetsrommet ligger i om det skal pålegges universiteter (og andre virksomheter) i å inngå avtale om å dele lokale ressurser ved behov. Dette kan være tilfelle ved eventuelle kriser, lik den som inntraff under pandemien.

Mulighetsrommet for tildelingsmodell er beskrevet nærmere i Tabell 8.3.

Tabell 8.3 Beskrivelse av Leveransemodell

Videreføring/utvidelse av dagens konsept	En felles nasjonal plattformleverandør for drift av avanserte ressurser	Avanserte ressurser sikres i samarbeid
<ul style="list-style-type: none">I dette alternativet vil man opprettholde dagens modell der tungregnekapasitet etableres av hver virksomhet/forvaltnings-område for å sikre bruk og tilpasning av tjenesterEn leverandør (Sigma2) videreføres for å sikre avansert kapasitet til forsknings- og forvaltningsdrevet FoU.Mindre avanserte behov løses av den enkelte virksomhet.	<ul style="list-style-type: none">I dette alternativet etableres en felles nasjonal leverandør som sikrer kapasitet til FoU og eventuelt bruk og tilpasning av tjenester.	<ul style="list-style-type: none">I dette alternativet sikres avanserte ressurser gjennom samarbeid med Europa og/eller med private aktører.

8.1.4 Mulighetsdimensjon D. Organisering av avansert brukerstøtte (AUS)

Denne mulighetsdimensjonen beskriver hvordan avansert brukerstøtte organiseres. Avansert brukerstøtte tilbyr hjelp som går utover grunnleggende brukerstøtte, for å gi spesialisert og mer dyptgående kompetanse til en virksomhet eller brukergruppe. I dagens situasjon er avansert brukerstøtte en del av tjenesteavtalen som Sigma2 har etablert med BOTT-ene og er derfor tett koblet til selve driften. I tillegg er Norsk kompetansesenter for tungregning (NCC) etablert til å støtte andre virksomheter. Som vist i Figur 8.1 vil det være store forskjeller mellom brukergruppene når det gjelder behov for programvare, KI bibliotek og verktøy. Mulighetsrommet ligger i å bygge videre på denne modellen, men å frikoble denne støtten fra selve driftsoppgavene.

Mulighetsrommet for tildelingsmodell er beskrevet nærmere i

Tabell 8.4.

Tabell 8.4 Beskrivelse av Organisering av avansert brukerstøtte

Virtuelle kompetansesentre tett koblet til drift	Distribuerte og løst sammenkoblede kompetansesentre	Distribuerte og tett sammenkoblede kompetansesentre
<ul style="list-style-type: none"> I dagens situasjon er det etablert to kompetansesentre. NRIS som understøtter forskningsinstitusjonene og NCC som bistår næringslivet, med særlig vekt på små og mellomstore bedrifter, og offentlig forvaltning i å utvikle kompetanse i bruk av avansert teknologi. 	<ul style="list-style-type: none"> Hvert forvaltningsområde eller virksomhet etablerer dedikerte kompetansesentre for å understøtte sine virksomheter. Det inngås en avtale mellom kompetansesentre for å sikre samarbeid, spesielt knyttet til kompetanseoverføring og kompetanseutvikling. 	<ul style="list-style-type: none"> Hvert forvaltningsområde eller virksomhet etablerer dedikerte kompetansesentre for å understøtte sine virksomheter. Kompetansesentre styres gjennom en felles styringsmodell, med felles rekrutterings- og kompetansestrategier, utvikling av felles prosesser og opplæringsprogram

8.1.5 Mulighetsdimensjon E. Lokalisering av ressurser

Denne mulighetsdimensjonen beskriver hvor tungregnings- og lagringskapasiteten lokaliseres og reflekterer behovet for selvberging og håndtering av sensitive data.

Denne mulighetsdimensjonen er knyttet til i hvor stor grad Norge skal legge vekt på autonomi og selvbergingsbehov. Mulighetsrommet ligger teoretisk sett i at vi enten etablerer all infrastruktur på norsk jord, eller at vi utvider samarbeidet med EU og kjøper all fremtidig kapasitet gjennom dette samarbeidet. Et styrket samarbeid med EU vil kunne gi oss en større fleksibilitet i når vi utvider infrastrukturen i Norge.

8.1.6 Mulighetsdimensjon F. Finansiering og eierskap.

Denne mulighetsdimensjonen beskriver hvordan investering i ny kapasitet skal finansieres og eies, samt hvordan de fortløpende driftskostnadene skal dekkes.

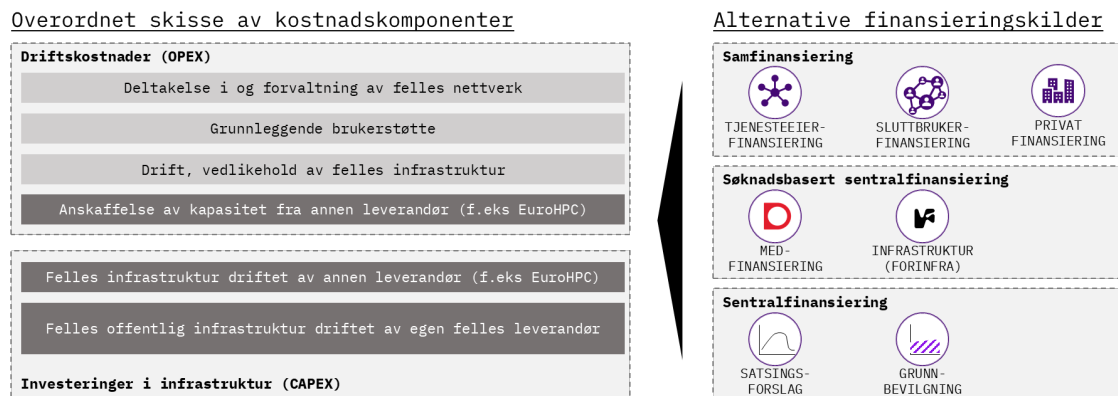
Mulighetsrommet i denne dimensjonen defineres av to faktorer:

- Hvordan sikre forutsigbar finansiering med hensyn til å etablere og opprettholde felles infrastruktur som dekker behovet? I dag skjer dette innen forskning ved at Sigma2 søker på midler fra INFRASTRUKTUR-ordningen. Dette kan Sigma2 gjøre annethvert år. Tildelingen er begrenset oppad til 200 millioner kroner, som inntil nylig har vært tilstrekkelig. Som beskrevet i problembeskrivelsen er dette imidlertid ikke tilstrekkelig for å møte behovet for forskning alene, og enda mindre tilstrekkelig å ivareta behovene til andre brukergrupper.
- Hvordan plassere eierskapet slik at infrastrukturen og leveransene blir innrettet slik at de understøtter behovene også utenfor forskningssektoren? I dag er Sigma2 eid av Sikt – Kunnskapssektorens tjenesteleverandør, som også har styrets leder. I tillegg består styret av representanter av BOTT-universitetene, samt to eksterne. Denne modellen har fungert godt med hensyn til å sikre at dagens infrastruktur innrettes til de behov forskningen har, samt å sikre en leveransemodell som utnytter kompetansen og ressursene som BOTT-universitetene har. I en situasjon der felles infrastruktur skal brukes av flere brukergrupper til formål utenfor forskning, vil det være nødvendig å se på en annen eierskapsmodell.

En arbeidsgruppe under Skates⁵³ arbeidsutvalg utarbeidet i 2023 et rammeverk som skal være til hjelp for forvaltere og tjenesteeiere i offentlig sektor for å lage gode finansieringsmodeller for fellesløsninger, fra konsept til utfasing⁵⁴. Hovedformålet med finansieringsmodellen er:

- Å sikre en forutsigbar og langsiktig investeringsramme for utvikling/anskaffelse av tungregneinfrastruktur.
- Å gi gode insentiver for samfunnsøkonomisk riktig bruk av fellesløsningen.
- Å understøtte kostnadseffektiv utvikling og drift.
- Å være enkel å forstå og å administrere, med forutsigbarhet for alle parter.

Figur 8.2 gir en oversikt over alternative finansieringskilder.



Figur 8.2. Oversikt over alternative finansieringskilder

Mulighetsrommet ligger i hvilken grad man balanserer sentralfinansiering og samfinansiering. Tabell 8.5 oppsummerer de konseptuelle alternativene for denne dimensjonen.

Tabell 8.5 Beskrivelse av Finansiering og eierskap

Investeringer gjennom Forskningsrådet, brukerfinansiert drift basert på kost-pluss modell	Sentrale og felles ressurser eies som nasjonal kritisk infrastruktur. Sentral-finansierte investeringer, samfinansiert drift	Sentrale og felles ressurser eies som nasjonal kritisk infrastruktur. Samfinansierte investeringer og drift
<ul style="list-style-type: none"> • Dagens eierskapsstruktur opprettholdes. • Investeringsmidler til ny kapasitet blir søkt om gjennom Forskningsrådet. • Driften dekkes gjennom brukerbetaling. 	<ul style="list-style-type: none"> • Nasjonal leverandør eies av et departement som kan ivareta tverrsektorielle behov. • Det etableres en langsiktig investeringsplan (satsingsforslag eller grunnbevilgning) som sikrer forutsigbarhet. • Driften finansieres gjennom avtaler med brukerne (tjenesteeier- eller sluttbrukerfinansiering) 	<ul style="list-style-type: none"> • Nasjonal leverandør eies av et departement som kan ivareta tverrsektorielle behov. • Det inngås avtaler med de virksomheter som skal bruke infrastrukturen. Tjenesteeiere bidrar med finansiering av investeringer og drift (tjenesteeierfinansiering)

⁵³ Skate (Styring og koordinering av tjenester i e-forvaltning) er et strategisk samarbeidsråd og rådgivende organ til Digitaliseringsdirektoratet og kommunal- og distriktsministeren. Organet er sammensatt av topledere fra offentlige virksomheter.

⁵⁴ Skate, «Rammeverk for utforming av finansieringsmodeller for nasjonale og sektorielle fellesløsninger».










8.2 Konsepter

Med bakgrunn i mulighetsdimensjonene, er det utarbeidet unike konsepter for tungregning i et norsk perspektiv. Vi velger å først bruke de dimensjonene som er mest dimensjonerende for konseptene til å definere ulike konseptkategorier. Innenfor hver konseptkategori bruker vi de øvrige dimensjonene til å utarbeide konseptuelt forskjellige konsepter.

De mulighetsdimensjonene som er de som er mest dimensjonerende for tiltaket er:

- **Formål og bruk.** Hvilket formål og hvilken bruk som skal legges til grunn for felles infrastruktur vil være drivende for størrelsen på investeringen og ha en stor innvirkning på valg av tildelingsmodell, organisering av avansert brukerstøtte (AUS) og eierskap og finansieringsmodell. Behovet for tungregning til bruk og tilpasning for offentlige virksomheter generelt blir vurdert som usikkert. Det vil være naturlig at offentlige virksomheter vil utvide sine nåværende driftsavtaler for å dekke dette behovet. Derfor er det ikke laget konsepter som dekker dette behovet.
- **Leveransemodell inkl. grunnleggende brukerstøtte.** Valget på leveransemodell vil være drivende for hvor stor endring i eksisterende strukturer som må foretas og ha en stor innvirkning på valg av eierskap og finansieringsmodell.

Figur 8.3 gir en oversikt over hvilke konseptkategorier som er aktuelle når vi vurderer de dimensjonerende mulighetsdimensjonene.

FORMÅL OG BRUK	FORSKNINGSDREVET FoU (1) FORVALTNINGSDREVET FoU (2) BRUK OG TILPASNING SAMFUNNSKR. TJENESTER (3)	 Ikke mulig uten større endringer fra dagens konsept	 «En nasjonal leverandør» Leverandør dekker (1+2+3)	 «Europeisk leverandør» Dekker (1+2+3) Sigma2 forvalter og kompetansesenter
	FORSKNINGSDREVET FoU (1) FORVALTNINGSDREVET FoU (2)	 «Nullpluss-alternativet» Sigma2 dekker (1) og noe (2)	 «To leverandører: Slank Sigma2 og ny off. leverandør» Sigma2 dekker (1) Ny leverandør dekker (2)+(3)	 «To leverandører: Slank Sigma2 og ny leverandør basert på OPS» Sigma2 dekker (1) Ny leverandør dekker (2)+(3)
		 «Sigma3» Sigma2 dekker (1) og(2) Sentralfinansiering Dekke eget behov (3)	 «Sigma2 plus» Sigma2 dekker (1) og(2) Sektorfinansiering Dekke eget behov (3)	 Ikke mulig uten større endringer fra dagens konsept
	VIDEREFØRING AV DAGENS KONSEPT	EN FELLE NASJONAL PLATTFORMLEVERANDØR FOR DRIFT AV INFRASTRUKTUR	INFRASTRUKTUR SIKRES I SAMARBEID	
	LEVERANSEMODELL INKL. GRUNNLEGGENDE BRUKERSTØTTE «HVORDAN LEVERE KAPASITET?»			

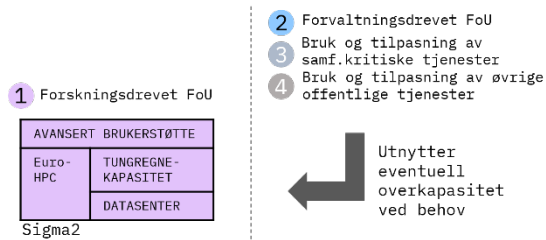
Figur 8.3 Oversikt over konseptkategorier og konsepter

Det er identifisert tre konseptkategorier. Disse representerer strategisk forskjellige valg:

- **Videreføring/utvidelse av dagens konsept.** Hovedinnretningen for disse konseptene er at man utsetter beslutningen om å etablere en felles løsning for alle brukergrupper. En eventuell utvidelse av dagens konsept skjer innenfor rammen av hvert forvaltningsområde.
- **Nasjonale plattformkonsepter.** Hovedinnretningen er at vi etablerer en felles plattformleverandør som skal dekke et bredt behov for infrastruktur på tvers av forskning, samfunnskritiske tjenester, og offentlig forvaltning.
- **Samarbeidskonsepter.** Hovedinnretningen på konseptene er at vi bygger ny infrastruktur som også dekker noe av næringslivets behov basert på ulike former for samarbeid. I disse konseptene er det ikke nødvendigvis én plattformleverandør som leverer felles løsning.

8.2.1 Nullalternativet

OVERORDNET BESKRIVELSE OG INNRETNING AV KONSEPTET



- Dette er en videreføring av dagens konsept.
- Sigma2 er største offentlige tilbyder av tungregning i Norge.
- Sigma2 er primært rettet mot forskningsdrevet FoU.
- Offentlig virksomheter får tilgang til eventuell overkapasitet til forvaltningsdrevet FoU i samarbeid med forskningsorganisasjoner.

FORMÅL OG BRUK

- Sigma2 dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) i forskningsdrevet FoU.
- Forskningsorganisasjoner og offentlige virksomheter (til bruk og tilpasning av offentlige tjenester) etablerer/holder egne ressurser.
- Offentlige virksomheter og næringsliv (særlig SMBer) kan samarbeide med NCC og Sigma2 for utprøving.

ORGANISERING: LEVERANSEMODELL

- Sigma2 disponerer investeringsmidler for bygging av ny tungregnekapasitet i eget datasenter eller gjennom EuroHPC.
- Sigma2 har inngått tjenesteavtaler med BOTT-universitetene for å få levert driftstjenester og grunnleggende brukerstøtte.
- Sigma2 kan gjennom sin leveransemodell ikke garantere 24/7 drift.

ORGANISERING: AVANSERT BRUKERSTØTTE

- Dagens ordning med NRIS (tjenesteavtale med BOTT-universitetene) for avansert brukerstøtte til forskningsorganisasjoner videreføres.
- Norwegian Competence Center (NCC) bistår næringsliv, med særlig søkelys på små og mellomstore bedrifter, og offentlig forvaltning i å utvikle kompetanse i bruk av avansert teknologi.

ORGANISERING: MODELL FOR TILDELING AV KAPASITET

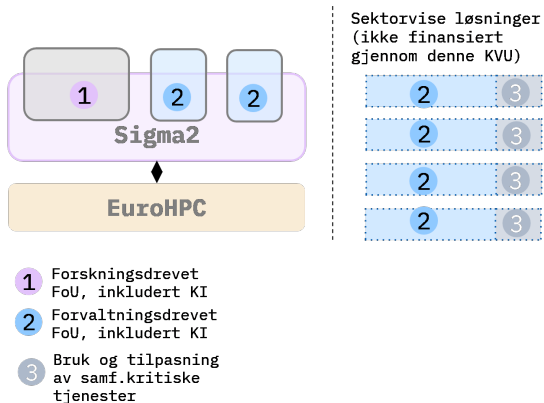
- For forskningsorganisasjoner skjer tildeling av kapasitet gjennom en evalueringsprosess.
- Sigma2 disponerer eventuell overkapasitet til å tildele til prosjekter utenfor forskningsorganisasjoner.
- Andel av kapasiteten som kan brukes til andre formål enn forskning blir bestemt av styret i Sigma2.

EIERSKAP OG FINANSIERING

- Dagens eiermodell, med SIKT som eier av Sigma2 videreføres.
- Styret i Sigma2 beslutter andel av kapasiteten som skal gjøres tilgjengelig til andre brukere og formål enn forskning.
- Dagens finansieringsmodell av Sigma2 videreføres, dvs.:
 - Søknadsbasert sentralfinansiering: Investeringsmidler for ny kapasitet eller erstatning av eksisterende kapasitet søkes om fra INFRASTRUKTUR-ordningen.
 - Sluttbrukerfinansiering: Løpende drift dekkes basert på brukerbetaling (kostplussmodellen) i tillegg til noe grunnleggende driftsstøtte

8.2.2 K1 «Nullplussalternativet»

OVERORDNET BESKRIVELSE OG INNRETNING AV KONSEPTET



- Sigma2 er største offentlige tilbyder av tungregning i Norge, og bevilges midler til å utvide sin kapasitet til å møte det økende behovet for tungregning for både forsknings- og forvaltningsdrevet FoU.
- Det enkelte forvaltningsområde må bestemme hvordan de skal dekke behovet for tungregning til utvikling av modeller, bruk og tilpasning av alle offentlige tjenester, inkl. KI (internasjonale leverandører, bygge ut egen kapasitet, samarbeid med andre virksomheter eller sektorer).

FORMÅL OG BRUK

- Sigma2 dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) til både forsknings- og forvaltningsdrevet FoU.
- Det enkelte forvaltningsområde må etablere egne løsninger for bruk og tilpasning av samfunnskritiske og øvrige tjenester (alene eller samarbeid med andre sektorer). Foreløpig vil ikke dette dekkes gjennom tiltaket som omfattes av denne KVUen.

ORGANISERING: LEVERANSEMODELL

- Sigma2 disponerer investeringsmidler for bygging av ny tungregnekapasitet i eget datasenter eller gjennom EuroHPC.
- Sigma2 har inngått tjenesteavtaler med BOTT-universitetene for å få levert driftstjenester og grunnleggende brukerstøtte.
- Sigma2 kan gjennom sin leveransemodell ikke garantere 24/7 drift.

ORGANISERING: AVANSERT BRUKERSTØTTE

- Dagens ordning med NRIS (tjenesteavtale med BOTT-universitetene) for avansert brukerstøtte til forskningsorganisasjoner videreføres.
- Norwegian Competence Center (NCC) bistår næringsliv, med særlig vekt på små og mellomstore bedrifter, og offentlig forvaltning i å utvikle kompetanse i bruk av avansert teknologi.
- Norwegian Competence Center (NCC) får ansvaret til å støtte nye brukere, eventuelt bistår i å etablere egne kompetansenettverk for andre sektorer.

ORGANISERING: MODELL FOR TILDELING AV KAPASITET

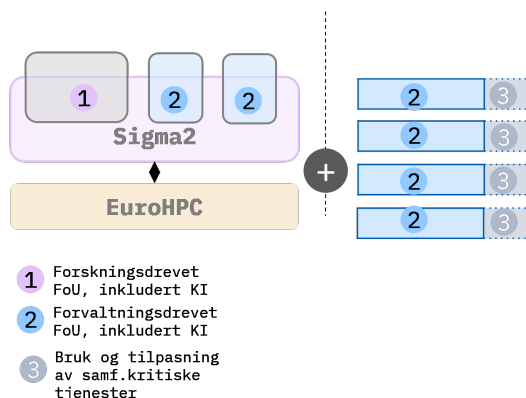
- For forskningsorganisasjoner skjer tildeling av kapasitet gjennom en evalueringsprosess.
- Sigma2 disponerer overkapasitet til å tildele til prosjekter utenfor forskningsorganisasjoner.
- Andel kapasitet som kan brukes til andre formål enn forskning vedtas av styret.
- Andre sektorer har dedikert kapasitet som de selv disponerer.

EIERSKAP OG FINANSIERING

- Dagens eiermodell med SIKT som eier av Sigma2 videreføres.
- Finansieringsmodellen for Sigma2 endres slik at investeringer i infrastruktur finansieres gjennom grunnbevilgning. Det etableres en langsiktig investeringsplan (fireårig langtidsplan for infrastruktur).
- Løpende drift dekkes basert på brukerbetaling (kost-plussmodellen) i tillegg til noe grunnleggende driftsstøtte.
- Hvert forvaltningsområde finansierer investering og drift av kapasitet for å dekke den enkelte sektors behov.

8.2.3 K2 «Sigma2-pluss»

OVERORDNET BESKRIVELSE OG INNRETNING AV KONSEPTET



- Sigma2 er største offentlige tilbyder av tungregning i Norge, det bevilges midler til å utvide sin kapasitet til å møte det økende behovet for tungregning for både forsknings- og forvaltningsdrevet FoU.
- Det bevilges midler til det enkelte departement som selvstendig bestemmer hvordan de skal dekke behovet for tungregning til utvikling av modeller, bruk og tilpasning av alle offentlige tjenester innen sitt forvaltningsområde, inkl. KI (internasjonale leverandører, bygge ut egen kapasitet, samarbeid med andre virksomheter eller forvaltningsområder).

FORMÅL OG BRUK

- Sigma2 dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) til både forsknings- og forvaltningsdrevet FoU. I tillegg dekker Sigma2 visse behov innenfor forvaltningsdrevet FoU som har nasjonal betydning, herunder utvikling og testing av norske språkmodeller og KI-modeller.
- Det enkelte departement får bevilget midler til å etablere løsninger for utvikling av modeller, og bruk og tilpasning av samfunnskritiske og øvrige tjenester innen sitt forvaltningsområde (alene eller samarbeid med andre sektorer). Disse midlene kan brukes til å bygge kapasitet hos Sigma2.

ORGANISERING: LEVERANSEMODELL

- Sigma2 disponerer investeringsmidler for bygging av ny tungregnekapasitet i eget datasenter eller gjennom EuroHPC.
- Sigma2 har inngått tjenesteavtaler med BOTT-universitetene for å få levert driftstjenester og grunnleggende brukerstøtte.
- Sigma2 kan gjennom sin leveranse-modell ikke garantere 24/7 drift.
- Sigma2 utvider kapasiteten til å kunne håndtere at andre sektorer ønsker å utnytte deres leverandøravtaler (utstyr, innplassering).

ORGANISERING: AVANSERT BRUKERSTØTTE

- Dagens ordning med NRIS (tjenesteavtale med BOTT-universitetene) for avansert brukerstøtte til forskningsorganisasjoner videreføres.
- Norwegian Competence Center (NCC) bistår næringsliv, med særlig vekt på små og mellomstore bedrifter, og offentlig forvaltning i å utvikle kompetanse i bruk av avansert teknologi.
- Norwegian Competence Center (NCC) får ansvaret til å støtte nye brukere, eventuelt bistår i å etablere egne kompetansenettverk for andre forvaltningsområder.

ORGANISERING: MODELL FOR TILDELING AV KAPASITET

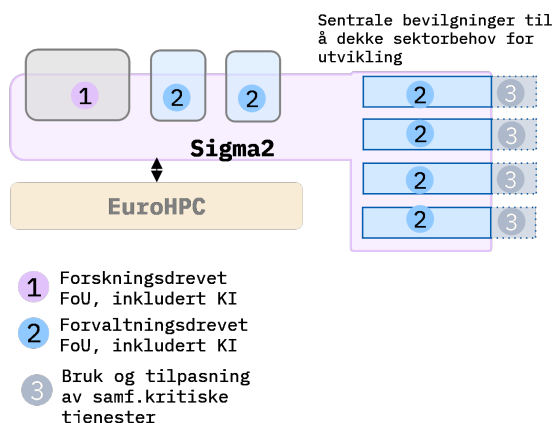
- For forskningsorganisasjoner skjer tildeling av kapasitet gjennom en evalueringsprosess.
- Sigma2 disponerer eventuell overkapasitet til å tildele til prosjekter utenfor forskningsorganisasjoner.
- Andel kapasitet som kan brukes til andre formål enn forskning vedtas av styret.
- Andre sektorer har dedikert kapasitet som de selv disponerer, tilsvarende den investeringen de bidrar med.

EIERSKAP OG FINANSIERING

- Dagens eiermodell med SIKT som eier av Sigma2 videreføres.
- Finansieringsmodellen for Sigma2 endres slik at investeringer i infrastruktur finansieres gjennom grunnbevilgning. Det etableres en langsiktig investeringsplan (fireårig langtidspan for infrastruktur).
- Løpende drift dekkes basert på brukerbetaling (kost-plussmodellen) i tillegg til noe grunnleggende driftsstøtte.
- Hvert departement bevilges midler og finansierer investering og drift av kapasitet til behov i sitt forvaltningsområde.

8.2.4 K3 «Tverrsektoriell Sigma3»

OVERORDNET BESKRIVELSE OG INNRETNING AV KONSEPTET



- Sigma2 er største offentlige tilbyder av tungregning i Norge, og bevilges midler til å utvide sin kapasitet til å møte det økende behovet for tungregning for både forsknings- og forvaltningsdrevet FoU. Bevilgningen til Sigma2 skal også dekke investeringer for tungregneinfrastruktur til andre sektorer.
- Sigma2 videreutvikles for å kunne tilby tungregnekapasitet som en tjeneste til alle sektorer.
- Den enkelte off. virksomhet er selv ansvarlig for å dekke behovet for tungregning til bruk og tilpasning av øvrige offentlige tjenester, inkl. KI.

FORMÅL OG BRUK

- Sigma2 dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) i både forsknings- og forvaltningsdrevet FoU.
- Sigma2 får bevilget midler til å dekke behovet for tungregning til utvikling og tilpasning av samfunnskritiske tjenester.

ORGANISERING: LEVERANSEMODELL

- Sigma2 disponerer investeringsmidler for bygging av ny tungregnekapasitet i eget datasenter eller gjennom EuroHPC.
- Sigma2 har inngått tjenesteavtaler med BOTT-universitetene for å få levert driftstjenester og grunnleggende brukerstøtte.
- Sigma2 kan gjennom sin leveransemodell ikke garantere 24/7 drift.
- Sigma2 utvides til å kunne håndtere at andre sektorer bruker infrastruktur hos dem.

ORGANISERING: AVANSERT BRUKERSTØTTE

- Dagens ordning med NRIS (tjenesteavtale med BOTT-universitetene) for avansert brukerstøtte til forskningsorganisasjoner videreføres.
- Norwegian Competence Center (NCC) bistår næringslivet, med særlig søkelys på små og mellomstore bedrifter, og offentlig forvaltning i å utvikle kompetanse i bruk av avansert teknologi
- Sigma2 samarbeider med deltakende virksomheter for å bygge opp nødvendige tjenester for avansert brukerstøtte.

ORGANISERING: MODELL FOR TILDELING AV KAPASITET

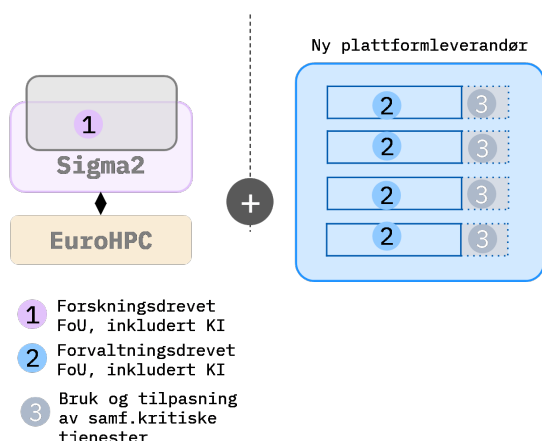
- Det etableres et system der ulike virksomheter/sektorer «bestiller» kapasitet hos Sigma2, som de kan disponere gjennom året.
- For forskningsorganisasjoner skjer tildeling av kapasitet gjennom en evalueringssprosess.

EIERSKAP OG FINANSIERING

- Kunnskapsdepartementet eier Sigma2 på vegne av departementsfellesskapet.
- Finansieringsmodellen for Sigma2 endres slik at investeringer i infrastruktur finansieres gjennom grunnbevilgning. Det etableres en langsiktig investeringsplan (fireårig langtidsplan for infrastruktur).
- Løpende drift dekkes gjennom en kombinasjon av grunnleggende driftsstøtte og brukerbetaling.

8.2.5 K4 «Sigma2 og ny offentlig leverandør»

OVERORDNET BESKRIVELSE OG INNRETNING AV KONSEPTET



- Utvider det offentlige leverandørmarkedet til to leverandører for å ivareta flere brukere av tungregning.
- Sigma2 bevilges midler til å utvide sin kapasitet til å møte det økende behovet for tungregning for forskningsdrevet FoU.
- Det etableres en ny felles plattformleverandør som skal sikre felles kapasitet til offentlige virksomheter til forvaltningsdrevet FoU (inkludert KI utvikling og trening), bruk og tilpasning av samfunnskritiske tjenester.

FORMÅL OG BRUK

- Sigma2 dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) i forskningsdrevet FoU.
- Det etableres ny felles plattformleverandør som fokuserer på å sikre felles kapasitet til offentlige virksomheter til forvaltningsdrevet FoU (inkludert KI utvikling og trening), bruk og tilpasning av samfunnskritiske tjenester.
- Den enkelte off. virksomhet er selv ansvarlig for å dekke behovet for tungregning til bruk og tilpasning av øvrige offentlige tjenester, inkl. KI.

ORGANISERING: LEVERANSEMODELL

- Sigma2 fortsetter med ansvar for å anskaffe og drifte tungregneressursene, samt å forvalte investeringene i EuroHPC knyttet til forskning.
- Ny felles leverandør disponerer investeringsmidler for å anskaffe og drifte kapasitet. I tillegg etablerer de en driftsorganisasjon som sikrer 24/7 tilgjengelighet på infrastrukturen.
- Det inngås samarbeid mellom Sigma2 og ny felles leverandør for å sikre volumfordeler i innkjøp av utstyr og tjenester.

ORGANISERING: AVNSERT BRUKERSTØTTE

- Videreføring av agens ordning med NRIS for avansert brukerstøtte til forskningsmiljøer.
- Norwegian Competence Center (NCC) får ansvaret til å støtte nye brukere, eventuelt bistår i å etablere egne kompetansenettverk for andre sektorer.
- Ny felles leverandør samarbeider med deltakende virksomheter for å bygge opp nødvendige tjenester for avansert brukerstøtte.

ORGANISERING: MODELL FOR TILDELING AV KAPASITET

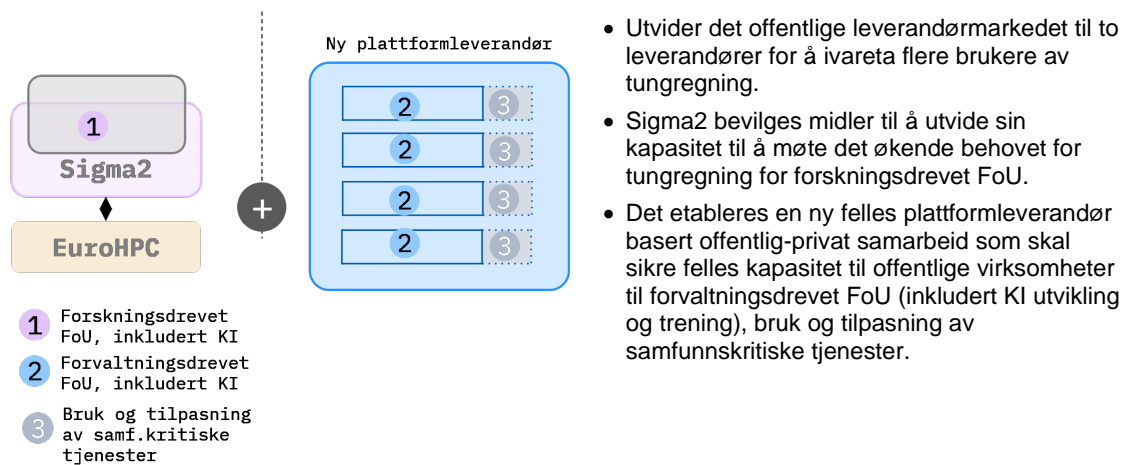
- For forskningsorganisasjoner skjer tildeling av kapasitet gjennom en evalueringsprosess.
- Sigma2 disponerer eventuell overkapasitet til å tildele til prosjekter utenfor forskningsorganisasjoner.
- Virksomheter som skal bruke ny plattformleverandør «bestiller» og får tildelt dedikert kapasitet, som de disponerer 24/7.

EIERSKAP OG FINANSIERING

- Dagens eiermodell, med SIKT som eier av Sigma2 videreføres
- Finansieringsmodellen for Sigma2 endres slik at investeringer i infrastruktur finansieres gjennom grunnbevilgning.
- Eierskap til ny felles offentlig leverandør plasseres til et departement som kan ivareta tverrsektorielle interesser.
- Ny plattformleverandør finansieres basert på en tjenesteeierfinansiering (samfinansiering) der deltakende departementer betaler i henhold til bestilt kapasitet.

8.2.6 K5 «Sigma2 og ny leverandør basert på offentlig-privat samarbeid»

OVERORDNET BESKRIVELSE OG INNRETNING AV KONSEPTET



- Utvider det offentlige leverandørmarkedet til to leverandører for å ivareta flere brukere av tungregning.
- Sigma2 bevilges midler til å utvide sin kapasitet til å møte det økende behovet for tungregning for forskningsdrevet FoU.
- Det etableres en ny felles plattformleverandør basert offentlig-privat samarbeid som skal sikre felles kapasitet til offentlige virksomheter til forvaltningsdrevet FoU (inkludert KI utvikling og trening), bruk og tilpasning av samfunnskritiske tjenester.

FORMÅL OG BRUK

- Sigma2 dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) i forskningsdrevet FoU.
- Det etableres ny felles plattformleverandør basert på offentlig-privat samarbeid som fokuserer på å sikre felles kapasitet til offentlige virksomheter til forvaltningsdrevet FoU (inkludert KI utvikling og trening), bruk og tilpasning av samfunnskritiske tjenester.
- Den enkelte off. virksomhet er selv ansvarlig for å dekke behovet for tungregning til bruk og tilpasning av øvrige offentlige tjenester, inkl. KI.

ORGANISERING: LEVERANSEMODELL

- Sigma2 fortsetter med ansvar for å anskaffe og drifte tungregnerressursene, samt å forvalte investeringene i EuroHPC knyttet til forskning.
- Ny felles leverandør disponerer investeringsmidler for å anskaffe og drifte kapasitet. I tillegg etablerer de en driftsorganisasjon som sikrer 24/7 tilgjengelighet på infrastrukturen.
- Det inngås samarbeid mellom Sigma2 og ny felles leverandør for å sikre volumfordeler i innkjøp av utstyr og tjenester.

ORGANISERING: AVANSERT BRUKERSTØTTE

- Videreføring av dagens ordning med NRIS for avansert bruker støtte til forskningsmiljøer.
- Norwegian Competence Center (NCC) får ansvaret til å støtte nye brukere, eventuelt bistår i å etablere egne kompetansenettverk for andre sektorer.
- Ny felles leverandør samarbeider med deltakende virksomheter for å bygge opp nødvendige tjenester for avansert bruker støtte.

ORGANISERING: MODELL FOR TILDELING AV KAPASITET

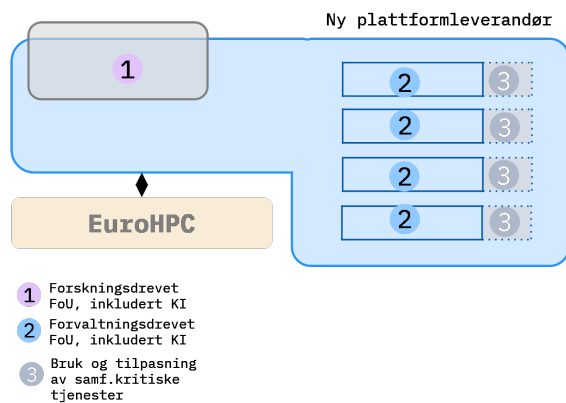
- For forskningsorganisasjoner skjer tildeling av kapasitet gjennom en evalueringsspross
- Sigma2 disponerer eventuell overkapasitet til å tildele til prosjekter utenfor forskningsorganisasjoner.
- Virksomheter som skal bruke ny plattformleverandør «bestiller» og får tildelt dedikert kapasitet, som de disponerer 24/7

EIERSKAP OG FINANSIERING

- Dagens eiermodell, med SIKT som eier av Sigma2 videreføres
- Finansieringsmodellen for Sigma2 endres slik at investeringer i infrastruktur finansieres gjennom grunnbevilgning.
- Eierskapsmodellen av ny leverandør er delt mellom offentlig og privat. Eierskapet til det offentlige deles mellom ett eller flere departement som sikrer tverrsektorielle interesser.
- Deltakende departementer bidrar med finansiering for sine kvoter gjennom en tjenesteeierfinansiering (samfinansiering).

8.2.7 K6 «En nasjonal leverandør»

OVERORDNET BESKRIVELSE OG INNRETNING AV KONSEPTET



- Det etableres en ny plattformleverandør som dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) til både forsknings- og forvaltningsdrevet FoU og behovet til offentlige virksomheter som har behov for tungregnekapasitet til bruk og tilpasning av samfunnskritiske tjenester.

FORMÅL OG BRUK

- Det etableres ny plattformleverandør som dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) i både forsknings- og forvaltningsdrevet FoU, samt kan tilby kapasitet til virksomheter som har behov for tungregnekapasitet til bruk og tilpasning av tjenester.
- Den enkelte off. virksomhet er selv ansvarlig for å dekke behovet for tungregning til bruk og tilpasning av øvrige offentlige tjenester, inkl. KI
- Forskningsorganisasjoner og offentlige virksomheter kan beholde og etablere egne ressurser om ønskelig.

ORGANISERING: LEVERANSEMODELL

- Plattformleverandøren har ansvar for å anskaffe og drifte tungregnerressursene, samt å forvalte investeringene i EuroHPC.
- Ny kapasitet til andre formål etableres gjennom en utvidelse av plattformleverandøren sin infrastruktur.
- Plattformleverandøren etablerer en driftsorganisasjon som sikrer 24/7-tjenester.

ORGANISERING: AVANSERT BRUKERSTØTTE

- Dagens ordning med NRIS (tjenesteavtale med BOTT-universitetene) for avansert brukerstøtte til forskningsorganisasjoner videreføres.
- Norwegian Competence Center (NCC) bistår privat industri, med særlig søkelys på små og mellomstore bedrifter, og offentlig forvaltning i å utvikle kompetanse i bruk av avansert teknologi.
- Ny felles leverandør samarbeider med deltakende virksomheter for å bygge opp nødvendige tjenester for avansert brukerstøtte.

ORGANISERING: MODELL FOR TILDELING AV KAPASITET

- Det etableres et system der ulike virksomheter/ sektorer «bestiller» kapasitet hos Sigma2, som de kan disponere gjennom året.
- For forskningsorganisasjoner skjer tildeling av kapasitet gjennom en evalueringsprosess.

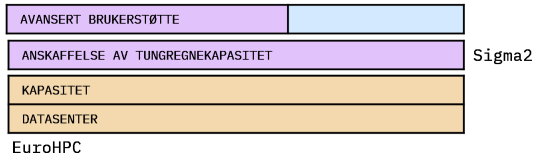
EIERSKAP OG FINANSIERING

- Plattformleverandør eies av ett eller flere departement som sikrer tverrsektorielle interesser.
- Deltakende departementer bidrar med finansiering for sine kvoter gjennom en tjenesteeierfinansiering (samfinansiering).
- Det etableres en langsiktig investeringsplan (fireårig langtidsplan for infrastruktur).

8.2.8 K7 «Europeisk leverandør»

OVERORDNET BESKRIVELSE OG INNRETNING AV KONSEPTET

- 1 Forskningsdrevet FoU
- 2 Forvaltningsdrevet FoU
- 3 Bruk og tilpasning av samf. kritiske tjenester



- All nasjonal kapasitet legges til EuroHPC, som innebærer at Sigma2 ikke investerer og drifter egne ressurser, men er ansvarlig for innkjøp og forvaltning av kontrakter.
- EuroHPC dekker behovene for tungregning til forsknings- og forvaltningsdrevet FoU og samfunnskritiske tjenester (inkl. KI). Øvrig behov er den enkelte off. virksomhet ansvarlig for å dekke.

FORMÅL OG BRUK

- Kapasitet anskaffes gjennom EuroHPC-samarbeidet. Denne dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) i både forsknings- og forvaltningsdrevet FoU, samt off. virksomheter som har behov for tungregnekapasitet til bruk og tilpasning av samfunnskritiske tjenester.
- Off. virksomheter med behov for tungregnekapasitet til øvrige tjenester må finne egne løsninger.
- Forskningsorganisasjoner og offentlige virksomheter kan beholde og etablere egne ressurser om ønskelig.

ORGANISERING: LEVERANSEMODELL

- Sigma2 AS vil ikke lenger ha ansvar for å drifte tungregnerressurser, men vil bli en innkjøpsorganisasjon.
- Sigma2 AS får ansvar til å anskaffe ny kapasitet gjennom samarbeidet i EuroHPC og det kommersielle leverandørmarkedet for å sikre innkjøpsvolum.

ORGANISERING: AVANSERT BRUKERSTØTTE

- Dagens ordning med NRIS (tjenesteavtale med BOTT-universitetene) for avansert brukerstøtte til forskningsorganisasjoner videreføres.
- Norwegian Competence Center (NCC) bistår privat industri, med særlig søkelys på små og mellomstore bedrifter, og offentlig forvaltning i å utvikle kompetanse i bruk av avansert teknologi.
- Norwegian Competence Center (NCC) får ansvaret til å støtte nye brukere, eventuelt bistår i å etablere egne kompetansenettverk for andre sektorer.

ORGANISERING: MODELL FOR TILDELING AV KAPASITET

- Det etableres et system der ulike virksomheter/ sektorer «bestiller» kapasitet hos Sigma2, som de kan disponere gjennom året.
- For forskningsorganisasjoner skjer tildeling av kapasitet gjennom en evalueringsprosess.

EIERSKAP OG FINANSIERING

- Sigma2 eies av SIKT og et departement som sikrer tverrsektorielle interesser.
- Deltakelsen i EuroHPC og kompetansesenter for brukerne finansieres gjennom grunnbevilgninger.
- Bruk av kapasitet i EuroHPC gjennom brukerbetaling i tråd med retningslinjer fra EuroHPC.

8.3 Vurdering av konseptene mot rammebetingelser og utvalgte vurderingskriterier

De ulike konseptene for en nasjonal infrastruktur for tungregneinfrastruktur har blitt vurdert for å avdekke konseptenes egnethet, realisme og måloppnåelse. Det innebærer at konseptene er vurdert opp mot rammebetingelsene for konseptvalg og utvalgte vurderingskriterier. Konseptene som oppfyller alle rammebetingelsene og vurderingskriteriene «går videre» og inngår i den samfunnsøkonomiske analysen i kapittel 8.

Tabell 8.6 viser resultatene fra vurderingene av konseptene. Grønn hake illustrerer at rammebetingelsen/vurderingskriteriet er oppfylt, rødt kryss illustrer at rammebetingelsen/vurderingskriteriet ikke er oppfylt og gul sirkel illustrer at rammebetingelsen/vurderingskriteriet er oppfylt, men med en begrensning.

Tabell 8.6 Oversikt over vurdering av konseptene opp mot rammebetingelser og utvalgte vurderingskriterier

#	K1	K2	K3	K4	K5	K6	K7
R1	✓	✓	✓	✓	✓	✓	✓
R2	✓	✓	●	✓	✓	✓	✓
R3	✓	✓	✓	✓	✓	✓	✗
R4	✓	✓	✓	●	●	●	✗
R5	✓	✓	✓	✓	✓	✓	✓
R6	✓	✓	●	●	●	●	✗
R7	✓	✓	✓	✗	✗	●	✗
A1	✗	●	✓	✓	✓	✓	✓
A2	✓	✓	✓	●	●	✗	●
A3	✓	✓	✓	✗	✗	✗	✗

Under følger en begrunnelse for vurdering av rammebetingelsene og vurderingskriterier.

R1 Infrastrukturen for tungregning må være pålitelig og robust

I alle konseptene inngår det driftstjenester som sikrer oppetid tilpasset det formål og bruk som konseptene er designet for. Konseptene K2-K3 innebærer en videreutvikling av Sigma2 for å sikre 24/7 oppetid, som er nødvendige for at tungregneressursene er pålitelige også for KI utvikling og trening av samfunnskritiske tjenester. I konseptene K4-K7 vil de nye leverandørene etableres med en leveransemodell som sikrer pålitelig og robust drift 24/7.

Alle konseptene innebærer at brukere får kapasitet hos EuroHPC. LUMI, som er en av verdens kraftigste superdatamaskiner, tilbyr tungregning som kan brukes til KI bruk og tilpasning. Den er designet for å håndtere komplekse beregningsoppgaver, inkludert både utvikling og trening av KI-modeller. Når det gjelder oppetid, er LUMI konstruert for å tilby høy tilgjengelighet, inkludert

24/7-oppetid, for å sikre kontinuerlig tilgang til beregningsressurser. Dette er viktig for brukere som trenger pålitelig tilgang til tungregning for kritiske applikasjoner.

R2 Infrastrukturen for tungregning må oppfylle sikkerhetskrav, herunder tungregneressurser som ikke er plassert i Norge må ligge i land vi har sikkerhetspolitisk samarbeid med

I alle konseptene der det velges å plassere tungregneressurser utenfor Norge vil gjøres primært gjennom EuroHPC-samarbeidet. Vi har sikkerhetspolitisk samarbeid med alle land der tungregneressursene til EuroHPC er plassert.

Konseptene K3-K7 innebærer at infrastrukturen skal ivareta sikkerhetskrav for å bruke sensitive data, f.eks. for helsesektoren. Det er blitt gjort flere forsøk å etablere løsninger for helseanalyse (f.eks. Helseanalyseplattformen). Sistnevnte ble stoppet hovedsakelig på grunn av juridiske utfordringer knyttet til Schems II-dommen. Denne dommen påvirket overføringen av personopplysninger til land utenfor EU/EØS. Dette vil ikke være tilfellet for konseptene K3-K7, ettersom disse konseptene innebærer å etablere og bruke en «privat sky», dvs. at databehandlingsansvarlig er en norsk leverandør, eller en offentlig leverandør som er underlagt den samme lovgivningen.

LUMI beskytter sensitive data ved å bruke beste praksis for krypteringsteknikker og gjennomføre regelmessige og omfattende penetrasjonstester med uavhengige eksperter. De overholder global og regional databeskyttelseslovgivning og har et robust, reviderbart og eksternt verifisert rammeverk av kontroller for å opprettholde konfidensialitet, integritet og tilgjengelighet av kundeinformasjon og personopplysninger. LUMI er også sertifisert i henhold til ISO/IEC 27001:2013 og SOC2 internasjonale standarder.

Uavhengig av dette vil det fortsatt være krevende å etablere en modell for tilgangsstyring til disse dataene. Dette vil slå ut primære på det økonomiske kriteriet (A2), men vi velger å vurdere at det kan være en begrensning for konseptene K3-K7 opp mot denne rammebetingelsen.

R3 En tilstrekkelig andel av tungregneressursene må lokaliseres i Norge pga. sikkerhetskonsens og kompetanse

I alle konseptene med unntak av K7 vil det etableres og tilgjengeliggjøres tilstrekkelig med tungregnekapasitet i Norge, under kontroll av det offentlige.

R4 Infrastrukturen for tungregning må være skalerbar og fleksibel til fremtidige endringer i behov for tungregnekapasitet

Det er kun K7 som ikke oppfyller rammebetingelsen fullt ut. I dette konseptet vil Norge konkurrere med øvrige europeiske land om å få tilgang til kapasitet og vil være låst til de planprosessene som EuroHPC har med hensyn til å «bestille» kapasitet. Det vil ikke være like enkelt å utvide kapasiteten fortløpende ettersom man ikke har kontroll på investeringsplanene.

Konseptene K1, K2 og K3 oppfyller fullt og helt de grunnleggende kriteriene for å være skalerbar og fleksibel til fremtidige endringer i behov. I disse tre konseptene ligger et samarbeid med EuroHPC til grunn for å kunne anskaffe mer kapasitet eller redusere kapasiteten etter at behovene endrer seg. Konseptene tillater at virksomheter kan beholde eller anskaffe egen kapasitet utenfor felles leverandør.

Det er en viss usikkerhet knyttet til hvorvidt konseptene K4, K5 og K6 vil kunne tilby den samme fleksibiliteten, rett og slett for det store omfanget av brukere som vil dekkes.

For alle konseptene vil finansieringsmodellen endres slik hele eller deler av investeringene baserer seg på grunnbevilgninger basert på en strategisk langtidsplan.

R5 Infrastrukturen for tungregning må oppfylle statens krav til energieffektivitet

Vi vurderer at denne rammebetingelsen vil kunne oppfylles i alle prosjektene.

R6 Det må legges til rette for gode støttetjenester som er tilpasset ulike brukergruppers behov for å kunne utnytte tungregneressursene både nasjonalt og internasjonalt

I alle konseptene må det etableres infrastrukturtenester og plattformtenester (software, verktøy) inkludert kompetanse og avansert brukerstøtte som dekker et bredt spekter av brukere.

I konseptene K1 og K2 mener vi imidlertid at det er mulig å bygge videre på etablerte kompetansenettverk og kompetansemiljøer for å få dette til. I de øvrige konseptene (K3-K7) må det bygges ny kompetanse for å dekke behovene for å understøtte bruk og tilpasning i offentlig forvaltning og næringsliv.

R7 Infrastrukturen for tungregning skal så langt som mulig bygge på eksisterende systemer og strukturer

I konseptene K1, K2, K3 og K6 vil man bygge videre på de systemer og strukturer som er etablert gjennom Sigma2, NRIS og NCC. Det er imidlertid knyttet en usikkerhet til i hvor stor grad det vil være behov for å gjøre endringer og tilpasninger for å kunne realisere K6. I de øvrige konseptene (K4, K5 og K7) vil det innebære etableringen av helt ny leverandør utover Sigma2. Disse konseptene oppfyller derfor ikke rammebetingelsen.

A1 Infrastrukturen for tungregning bør ha potensial for høy måloppnåelse for samtlige effektmål

Tabell 8.7 gir en oversikt over den kvalitative vurderingen av oppnåelsen av effektmålene for de ulike konseptene. Effektmålene er gjennomgått i kapittel 6.2.

Tabell 8.7 Vurdering av måloppnåelse (A2) i konseptene

#	K1	K2	K3	K4	K5	K6	K7
Effektmål 1	H	H	H	H	H	H	H
Effektmål 2	L	H	H	H	H	H	H
Effektmål 3	M	M	M	M	M	M	M
Effektmål 4	H	H	H	H	H	H	H

Alle konseptene har potensiale for høy måloppnåelse på alle effektmålene med unntak av K1. I dette konseptet vil hovedinnretningen av samarbeidet med næringslivet være særlig mot små og mellom store bedrifter, samt man i mindre grad legger til rette for forvaltningsdrevet FoU.

A2 Infrastrukturen for tungregning bør være så ressurseffektiv som mulig

Det er vurdert at omstillingskostnaden for å realisere alle konseptene er håndterbar for alle konseptene med unntak av K6. I dette konseptet vil det være store kostnader i forbindelse med etableringen av virksomhetene, omstillingen av eksisterende organisasjon, samt etableringen av et bredt tjenestespekter som dekker alle brukergruppene.

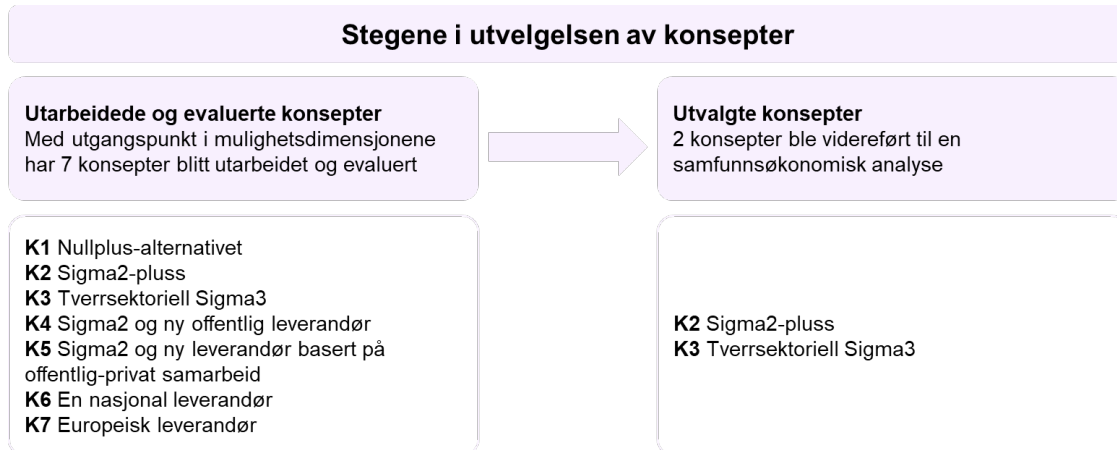
A3 Gjennomføringsrisikoen for investeringene og organiseringen av infrastrukturen for tungregning bør være lav

Gjennomføringsrisikoen er primært knyttet til etableringen av den plattformleverandør som skal realisere konseptet. Det er vurdert at gjennomføringsrisikoen er håndterbar for alle konseptene med unntak av K4, K5, K6 og K7. I disse fire konseptene vil det være behov for mange aktører å bli enig om finansieringsmodell og styringsmodell. I tillegg vil det være stor risiko ved å etablere et bredt tjenestespekter som dekker alle brukergruppene. I K7 vil det også være stor usikkerhet knyttet til å endre oppgavene til Sigma2, fra å være en plattformleverandør til å kun være en anskaffelsesorganisasjon.

8.4 Oppsummering

Basert på den samlede analysen videreføres de to konseptene som oppfyller alle rammebetingelsene og utvalgte vurderingskriterier til alternativanalysen:

- K2 «Sigma2-pluss»
- K3 «Tverrsektoriell Sigma3»



Figur 8.4 Oppsummering av utvelgelsen av konsepter

9 Alternativanalyse

I alternativanalysen utføres en samfunnsøkonomisk analyse av de utvalgte konseptene for tungregning, dvs. K2 «Sigma2-pluss» og K3 «Tverrsektoriell Sigma3». En samfunnsøkonomisk analyse lar oss se helhetlig på virkningene av hvert konsept. Det innebærer en vurdering av både negative virkninger (kostnader) og positive virkninger (nytte) som følge av konseptene sammenlignet med nullalternativet, og en samlet vurdering av konseptenes samfunnsøkonomiske lønnsomhet.

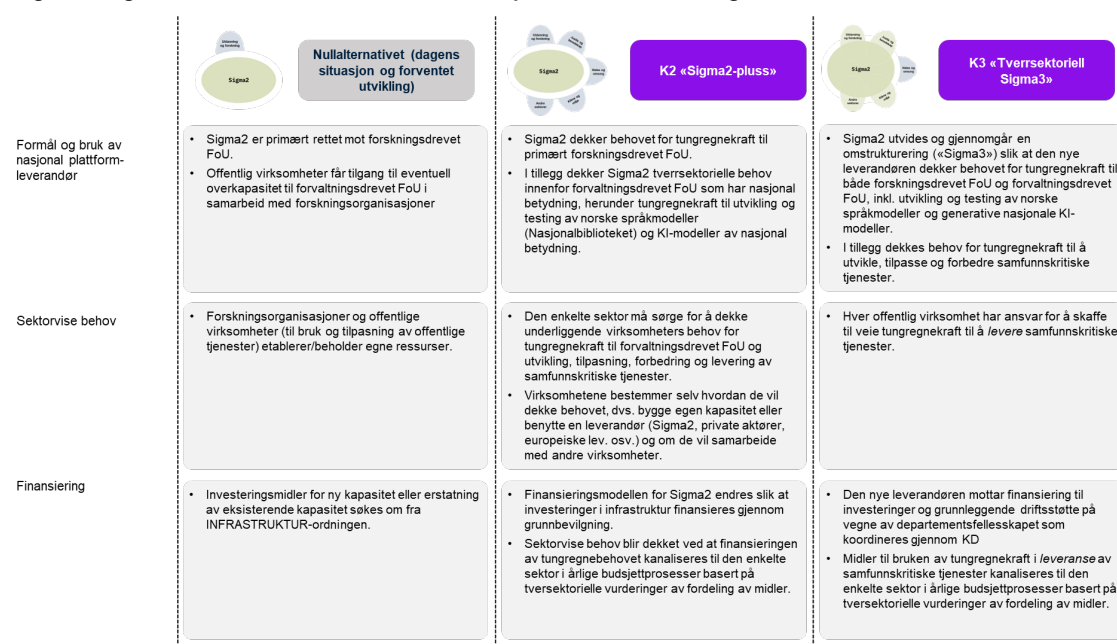
9.1 Beskrivelse av konsepter

I dette delkapittelet gis en nærmere omtale av de utvalgte konseptene K2 og K3 i tillegg til nullalternativet.

9.1.1 Sentrale forskjeller mellom de utvalgte konseptene

Hovedforskjellen mellom de to konseptene K2 og K3 ligger i omfanget av brukere som skal betjenes av en felles plattformleverandør og hvordan denne plattformleverandøren eies, styres, finansieres og organiseres.

Figur 9.1 gir en oversikt over sentrale forskjeller mellom K2 og K3.



Figur 9.1. Oversikt over sentrale forskjeller mellom K2 og K3 sammenlignet mellom nullalternativet

9.1.2 Nullalternativet

I nullalternativet opprettholdes Sigma2 som tilbyder av tungregning til primært forskningssektoren, mens øvrige sektorer har etablert egne løsninger for å dekke sektorenes behov for tungregning.

Brukere og behov for tungregning

Sigma2 dekker behovet for tungregnekapasitet (inkludert GPU-basert tungregning) til forskningsdrevet FoU og til en viss grad forvaltningsdrevet FoU.

Det enkelte forvaltningsområde i offentlig sektor etablerer egne løsninger for bruk og tilpasning av samfunnskritiske og øvrige tjenester (alene eller samarbeid med andre sektorer). Foreløpig vil ikke dette dekkes gjennom tiltaket som omfattes av denne konseptvalgutredningen.

Offentlige virksomheter beholder/etablerer egne ressurser om ønskelig, men det er krav om samordning innenfor det enkelte forvaltningsområde.

Investeringer i tungregnekapasitet

I nullalternativet investeres det ca. 80 000 CPU-kjerner i B2. Det investeres også i en utvidelse av CPU-kapasitet i A2. Figur 9.2 viser at med denne investeringen sikrer at dagens kapasitet opprettholdes.

Sigma2-Tungregning (CPU)	Kjerner	2025	2026	2027	2028	2029	2030
Saga	16 064	Utdet support					
Betzy (Intel equivalenter)	103 219	Utv. Sup.					
A2	64 000						
B2	80 000						
LUMI-C (2024 ->)	6 679						
Totalt per år (MCPUh)		1 664	1 664	1 664	1 664	1 261	701

Figur 9.2 Investering i tungregnekapasitet i nullalternativet

Investeringer i GPU-kapasitet

Det investeres i en utvidelse i GPU-kapasitet i A2, men dette gir en begrenset økning sammenlignet med dagens situasjon.

Sigma2-Tungregning (GPU)	Antall	2025	2026	2027	2028	2029	2030
LUMI-G (Eksisterende kapasitet)		250	250	250	250	250	250
Utvidelse A2 (GPU and CPU)		-	37	74	74	74	74
GPU-kapasitet gjennom Sigma2		250	287	324	324	324	324

Figur 9.3 Investering i GPU-kapasitet i nullalternativet

Investering i nettverk og lagringskapasitet

Det investeres noe i utvidelse og oppgradering av infrastruktur. Det investeres ikke i ytterligere lagringskapasitet.

Finansieringsmodell

Dagens finansieringsmodell av Sigma2 videreføres, dvs. investeringsmidler for ny kapasitet eller erstatning av eksisterende kapasitet søkes om fra INFRASTRUKTUR-ordningen. Løpende drift dekkes av en begrenset grunnbevilgning kombinert med brukerbetaling (kost-plussmodellen).

Styringsmodell inkludert modell for tildeling av kapasitet

I dag eies Sigma2 av Sikt – Kunnskapssektorens tjenesteleverandør. Sikt har også styreleder. I tillegg er de fire BOTT-universitetene representert i styret sammen med to eksterne representanter. Styret i Sigma2 fastsetter i dag årlig hvordan kapasiteten skal fordeles til ulike bruksområder. Ettersom det allerede i dette konseptet vil bygges infrastruktur som ikke kun har sin kilde i forskningsdrevet FoU, vil det være nødvendig å se på hvordan man kan ivareta at den overordnede allokeringen av kapasitet speiler denne endringen.

Tildelingen av kapasitet til forskningsdrevet FoU skjer gjennom søknader til Ressursfordelingskomiteen (RFK) og tildeles basert på prosjektenes kvalitet. Dette er en ordning som er bredt akseptert blant forskningsmiljøer.

Leveransemodell inkl. grunnleggende brukerstøtte

Sigma2 har i dag inngått tjenesteavtaler med BOTT-universitetene for å få levert driftstjenester og grunnleggende brukerstøtte.

Organisering av avansert brukerstøtte

Dagens ordning med NRIS (tjenesteavtale med BOTT-universitetene) for avansert brukerstøtte til forskningsorganisasjoner gjør at forskere, særlig ved BOTT-universitetene raskt får tilgang til høyt kvalifisert kompetanse.

Norwegian Competence Center (NCC) er en forpliktelse Norge har som følge av EuroHPC-samarbeidet. NCC bistår privat industri, med særlig søkelys på små og mellomstore bedrifter, og offentlig forvaltning i å utvikle kompetanse i bruk av avansert teknologi.

9.1.3 K2 «Sigma2-pluss»

I dette konseptet er det en leverandør av tungregning (Sigma2) til forskningssektoren og utvikling av nasjonale generative KI-modeller, mens øvrige sektorer i offentlig sektor har etablert egne løsninger for å dekke sektorenes behov for tungregning.

Brukere og behov for tungregning

Sigma2 dekker behovet for tungregning til primært forskningsdrevet FoU. I tillegg dekker Sigma2 tverrsektorielle behov innenfor forvaltningsdrevet FoU som har nasjonal betydning, herunder tungregning til utvikling og testing av norske språkmodeller (Nasjonalbiblioteket) og KI-modeller av nasjonal betydning.

Det enkelte forvaltningsområde i offentlig sektor må sørge for å dekke underliggende virksomheters behov for tungregning til forvaltningsdrevet FoU og utvikling, tilpasning, forbedring og levering av samfunnskritiske tjenester. Virksomhetene bestemmer selv hvordan de vil dekke behovet, dvs. bygge egen kapasitet eller benytte en leverandør (Sigma2, private aktører, europeiske leverandører, osv.) og om de vil samarbeide med andre virksomheter. Finansieringen av tungregnebehovet kanaliseres til det enkelte forvaltningsområde i årlige budsjettprosesser basert på tverrsektorielle vurderinger av fordeling av midler.

Investeringer i tungregnekapasitet

I konseptet investeres det i tradisjonell tungregning tilsvarende 300 000 CPU-kjerner i tillegg til det som eksisterer i nullalternativet. Figur 9.4 viser at dette til sammen vil gi en mer enn dobling i kapasiteten fra og med 2027 sammenlignet med dagens situasjon.

Sigma2-Tungregning (CPU)	Kjerner	2025	2026	2027	2028	2029	2030
Saga	16 064	Utvidet support					
Betzy (Intel equivalenter)	103 219		Utv. Sup.				
A2	64 000						
B2	150 000						
A3	130 000						
LUMI-C (2024 ->)	6 679						
LUMI-C (2026 ->)	75 000						
LUMI-C (2027 ->)	75 000						
Totalt per år (MCPUh)		1 664	1 606	4 527	4 527	3 189	3 767

Figur 9.4 Investering i tungregnekapasitet i K2

Sigma2 får tildelt midler de har søkt om i INFRASTRUKTUR-utlysningen fra 2023 for å investere i et nytt anlegg, B2. I tillegg bevilges det midler for å utvide EuroHPC-samarbeidet gjennom investeringer i et nytt LUMI-C anlegg. Denne kapasiteten forventes å bli tilgjengelig i 2026 og 2027.

Det er også lagt inn investeringer i et nytt anlegg, HPC System A3, som vil måtte settes i drift om lag 2030. Det er planlagt å søke om midler til dette anlegget under INFRASTRUKTUR-utlysningen som er forventet i 2025. Det vil ligge i samme kostnadsområde som B2, men med en noe annen fordeling av GPU og CPU.

Investeringer i GPU-kapasitet

Gjennom EuroHPC-samarbeidet har Norge tilgang til en kapasitet på 250 GPUer på LUMI i dag. Gjennom investeringen i B2 økes kapasiteten med 140 GPUer fra og med 2027. Det utvidede samarbeidet med EuroHPC gir tilgang til 150 GPUer i 2026 og ytterligere 150 GPUer i 2027. Det ligger også inne prosjektsøknader fra henholdsvis NORA (Norwegian Artificial Intelligence Research Consortium) og Openpower foundation som til sammen vil gi en kapasitet på 138 GPUer i 2026 og en dobling av denne kapasiteten i 2027.

Når det gjelder å dekke behovet for annen KI-utvikling og trening har vi tatt utgangspunkt i det anslåtte estimatet fra rapporten som Forskningsrådet leverte i august 2024⁵⁵. I denne tar man utgangspunkt i et forsiktig anslag fra de ledende KI-miljøene i Norge om at det minimum vil trenge en økning av kapasiteten med ca. 2 500 GPUer (10 ganger dagens kapasitet fra LUMI). Denne investeringen bør komme de neste 2-3 årene. Gitt at det allerede ligger inne en vesentlig økning gjennom B2, LUMI-G og andre prosjektsøknader velger vi å fordele økningen gjennom årene 2026-2029. I dette konseptet legger vi til grunn at halvparten av midlene bør fordeles til Sigma2, mens den andre halvparten bør disponeres av de ulike sektorene som har behov. Det vil være opp til hvert enkelt forvaltningsområde selv å finne gode løsninger, enten det er å bygge egen kapasitet, anskaffe kapasitet i leverandørmarkedet, eller gjennom å samarbeide med Sigma2. Sigma2 har gjennom årene opparbeidet seg god kompetanse i å anskaffe og etablere nye kapasitet og har etablert en rekke leverandøravtaler. Vi anslår derfor at ca. 40 prosent av volumet som er dedikert til sektorvise investeringer vil bli kanalisert til Sigma2. Figur 9.5 gir en oversikt over hvordan utbyggingen av de 2 500 GPUene vil fordeles over 2026-2029 og mellom ulike aktører.

Antall nye GPUer	Andel	2025	2026	2027	2028	2029	2030
Antall nye GPUer		-	500	500	500	500	500
Sigma2	50 %	-	250	250	250	250	250
Sigma2 via sektor	20 %	-	100	100	100	100	100
Ute i markedet	30 %	-	150	150	150	150	150

Figur 9.5 Utbygging av GPU-ressurser i K2

Figur 9.6 viser en anslått utviklingen av GPU-kapasitet i Norge gjennom midler fra dette tiltaket. I praksis bør det legges opp til en finansiering som sikrer at det ikke bygges ut ny kapasitet før den kapasiteten som er der utnyttes.

Sigma2-Tungregning (GPU)	Antall	2025	2026	2027	2028	2029	2030
LUMI-G (Eksisterende kapasitet)		250	250	250	250	250	250
B2		-	28	140	140	140	140
LUMI-G (Ny kapasitet)		-	150	300	300	300	300
Prosjektsøknader		-	138	277	277	277	277
KI-kapasitet (direkte bevilgning)		-	250	500	750	1 000	1 250
KI-kapasitet (via sektor)		-	100	200	300	400	500
GPU-kapasitet gjennom Sigma2		250	916	1 667	2 017	2 367	2 717
Sektorvis GPU-kapasitet	Antall	2025	2026	2027	2028	2029	2030
KI-kapasitet (sektorvise investeringer)		-	150	300	450	600	750
GPU-kapasitet gjennom sektorvise løsninger		-	150	300	450	600	750
Samlet GPU-kapasitet		250	1 066	1 967	2 467	2 967	3 467

Figur 9.6 Utvikling i GPU-kapasitet i K2

Investering i nettverk og lagringskapasitet

Det vil også være behov for å investere i erstatning av dagens lagringsanlegg og videre forbedre kapasiteten i nettverket. Ved slutten av sin levetid (2028) er det estimert at nåværende

⁵⁵ Forskningsrådet, «Behov for tungregnekraft for forskning og kunstig intelligens».

lagringsanlegg vil inneholde om lag 80 PB med data. For å erstatte dette, og legge til rette for de neste to års vekst (15 prosent gjennomsnittlig årlig vekstrate), må man ta høyde for om lag 106 PB ved anskaffelsen.

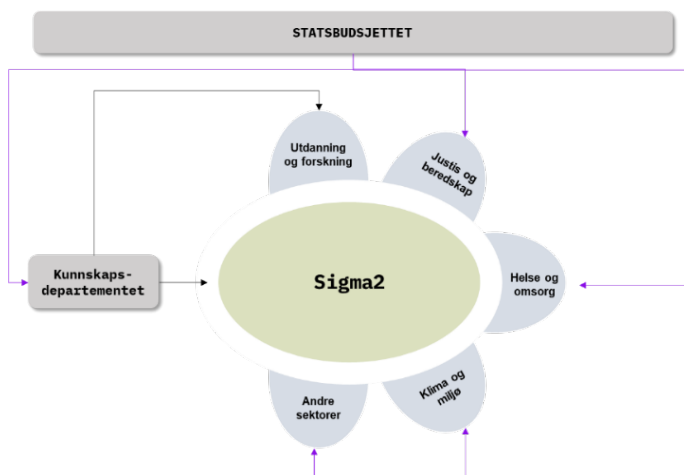
Finansieringsmodell

I dette konseptet vil det være nødvendig å gi Sigma2 en bedre forutsigbarhet når det gjelder tilgjengelige investeringsmidler. Finansieringsmodellen for Sigma2 endres slik at investeringer i infrastruktur finansieres gjennom grunnbevilgning. Det etableres en langsiktig investeringsplan (fireårig rullerende langtidsplan for infrastruktur) som ligger til grunn for bevilgningene gjennom den ordinære budsjettprosessen.

Løpende drift dekkes av en begrenset grunnbevilgning kombinert med brukerbetaling (kost-plussmodellen).

Det enkelte forvaltningsområde i offentlig sektor finansierer anskaffelse og drift av tungregning til områdets behov. Det bevilges midler gjennom den ordinære budsjettprosessen. Sektorvise behov tas med i utarbeidelsen av den langsiktige investeringsplanen. Virksomhetene bestemmer selv hvordan de vil dekke behovet, dvs. bygge egen kapasitet eller benytte en leverandør (Sigma2, private aktører, europeiske leverandører osv.) og om de vil samarbeide med andre virksomheter.

Figur 9.7 gir en forenklet oversikt over hvordan tungregnekapasitet finansieres i dette konseptet.



Figur 9.7 Illustrasjon av finansieringsmodell i K2

Styringsmodell inkludert modell for tildeling av kapasitet

Ettersom det allerede i dette konseptet også vil bygges infrastruktur som ikke kun har sin kilde i forskningsdrevet FoU, vil det være nødvendig å se på hvordan man kan ivareta at den overordnede allokeringen av kapasitet speiler denne endringen.

- Tildelingen av kapasitet til forskningsdrevet FoU skjer gjennom søknader til Ressursfordelingskomiteen (RFK) og tildeles basert på prosjektenes kvalitet. Dette er en ordning som er bredt akseptert blant forskningsmiljøer og videreføres derfor for den kapasiteten som allokeres til dette formål.
- Når det gjelder forvaltningsdrevet FoU etableres det en ordning der Sigma2 i større grad tildeler kapasitet basert på andre kriterier. Dette vil være en oppgave for forprosjektet å designe denne modellen.

Leveransemodell inkl. grunnleggende brukerstøtte

For å sikre at investeringsmidlene utnyttes mest mulig effektivt, får Sigma2 fullmakter til å disponere tilgjengelige investeringsmidler for bygging av ny tungregnekapasitet mer fleksibelt enn i dag. Sigma2 kan da velge om ny kapasitet skal etableres i eget datasenter eller gjennom EuroHPC, basert på kost-nytte vurderinger og behovet for å opprettholde norsk suverenitet.

Sigma2 har i dag inngått tjenesteavtaler med BOTT-universitetene for å få levert driftstjenester og grunnleggende brukerstøtte, som på en god måte dekker dagens behov. Ved at det over tid

vil være behov for å yte disse tjenestene til flere brukere utenfor BOTT-universitetene, vil det være nødvendig å ytterligere profesjonalisere leveranseprosessene.

Sigma2 anskaffer utstyr og tjenester primært for eget bruk. I dette konseptet er det avgjørende at Sigma2 utvides til å kunne håndtere at andre sektorer ønsker å utnytte deres leverandøravtaler (utstyr, innplassering).

Organisering av avansert brukerstøtte

Tilgang til kvalifisert kompetanse er løftet frem av flere aktører som et av de største hindringene når det gjelder å utnytte KI. Det er derfor nødvendig å se på hvordan avansert brukerstøtte kan organiseres, slik at den dekker fremtidige behov samtidig som man opprettholder tjenestekvaliteten overfor de aller største brukerne.

- Dagens ordning med NRIS (tjenesteavtale med BOTT-universitetene) for avansert brukerstøtte til forskningsorganisasjoner videreføres.
- Norwegian Competence Center (NCC) er en forpliktelse Norge har som følge av EuroHPC-samarbeidet og videreutvikles. Norwegian Competence Center (NCC) får ansvaret til å støtte nye brukere, eventuelt bistår i å etablere egne kompetansenettverk for andre sektorer.

9.1.4 K3 «Tverrsektoriell Sigma3»

I dette konseptet er det en større leverandør av tungregning som dekker et tverrsektorielt behov for tungregning i offentlig sektor. Leverandøren har oppstått gjennom en utvidelse og omstrukturering av Sigma2 («Sigma3»).

Brukere og behov for tungregning

Leverandøren av tungregning («Sigma3») dekker behovet for tungregning til både forskningsdrevet FoU og forvaltningsdrevet FoU, inkl. utvikling og testing av norske språkmodeller og generative nasjonale KI-modeller. I tillegg dekkes behov for tungregning til utvikling, tilpasse og forbedre samfunnskritiske tjenester.

Den enkelte offentlig virksomhet har imidlertid ansvar for å selv anskaffe og finansiere tungregning til å *levere* samfunnskritiske tjenester.

Investeringer i tungregnekapasitet

I dette konseptet vil det investeres i tradisjonell tungregning tilsvarende 300 000 CPU-kjerner i tillegg til det som eksisterer i nullalternativet. Figur 9.8 viser at dette til sammen vil gi en mer enn dobling i kapasiteten fra og med 2027 sammenlignet med dagens situasjon.

Sigma2-Tungregning (CPU)	Kjerner	2025	2026	2027	2028	2029	2030
Saga	16 064	Utvidet support					
Betzy (Intel equivalenter)	103 219	Utv. Sup.					
A2	64 000						
B2	150 000						
A3	130 000						
LUMI-C (2024 ->)	6 679						
LUMI-C (2026 ->)	75 000						
LUMI-C (2027 ->)	75 000						
Totalt per år (MCPUh)		1 664	1 606	4 527	4 527	3 189	3 767

Figur 9.8 Investering i tungregnekapasitet i K3

Det er søkt om midler for å investere i et nytt anlegg, B2, i INFRASTRUKTUR-utlysningen fra 2023, som er under behandling. I tillegg er det behov for å utvide EuroHPC-samarbeidet gjennom investeringer i et nytt LUMI-C anlegg. Denne kapasiteten forventes å bli tilgjengelig i 2026 og 2027.

Det er også lagt inn investeringer i et nytt anlegg, HPC System A3, som vil måtte settes i drift om lag 2030. Det er planlagt å søke om midler til dette anlegget under INFRASTRUKTUR-

utlysningen som er forventet i 2025. Det vil ligge i samme kostnadsområde som B2, men med en noe annen fordeling av GPU og CPU.

Investeringer i GPU-kapasitet

Gjennom EuroHPC-samarbeidet har Norge tilgang til en kapasitet på 250 GPUer på LUMI i dag. Gjennom investeringen i B2 vil man øke kapasiteten med 140 GPUer fra og med 2027. Det utvidede samarbeidet med EuroHPC vil gi tilgang til 150 GPUer i 2026 og ytterligere 150 GPUer i 2027. Det ligger også inne prosjektsøknader fra henholdsvis NORA (Norwegian Artificial Intelligence Research Consortium) og Openpower foundation som til sammen vil gi en kapasitet på 138 GPUer i 2026, og en dobling av denne kapasiteten i 2027.

Når det gjelder å dekke behovet for annen KI-utvikling og trening har vi tatt utgangspunkt i det anslåtte estimatet i rapporten som Forskningsrådet leverte i august 2024⁵⁶. I denne tar man utgangspunkt i et forsiktig anslag fra de ledende KI-miljøene i Norge om at det minimum vil trenge en økning av kapasiteten med ca. 2 500 GPUer (10 ganger dagens kapasitet fra LUMI). Denne investeringen bør komme de neste 2-3 årene. Gitt at det allerede ligger inne en vesentlig økning gjennom B2, LUMI-G og andre prosjektsøknader velger vi å fordele økningen gjennom årene 2026-2029. I dette konseptet legger vi til grunn at disse midlene går til Sigma2. Figur 9.9 gir en oversikt over hvordan utbyggingen av de 2 500 GPUene vil fordeles over 2026-2029.

Antall nye GPUer	Andel	2025	2026	2027	2028	2029	2030
Antall nye GPUer		-	500	500	500	500	500
Sigma2	100 %	-	500	500	500	500	500
Sigma2 via sektor	0 %	-	-	-	-	-	-
Ute i markedet	0 %	-	-	-	-	-	-

Figur 9.9 Utbygging av GPU-ressurser i K3

Figur 9.10 viser en anslått utviklingen av GPU-kapasitet i Norge gjennom midler fra dette tiltaket. I praksis bør det legges opp til en finansiering som sikrer at det ikke bygges ut ny kapasitet før den kapasiteten som er der utnyttes.

Sigma2-Tungregning (GPU)	Antall	2025	2026	2027	2028	2029	2030
LUMI-G (Eksisterende kapasitet)		250	250	250	250	250	250
B2		-	28	140	140	140	140
LUMI-G (Ny kapasitet)		-	150	300	300	300	300
Prosjektsøknader		-	138	277	277	277	277
KI-kapasitet (direkte bevilgning)		-	500	1 000	1 500	2 000	2 500
GPU-kapasitet gjennom Sigma2		250	1 066	1 967	2 467	2 967	3 467
Samlet GPU-kapasitet		250	1 066	1 967	2 467	2 967	3 467

Figur 9.10 Utvikling i GPU-kapasitet i K3

Investering i nettverk og lagringskapasitet

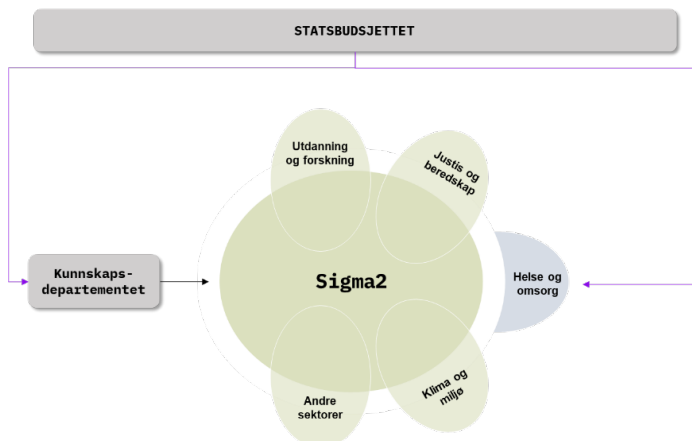
Det vil også være behov for å investere i erstatning av dagens lagringsanlegg og videre forbedre kapasiteten i nettverket. Ved slutten av sin levetid (2028) er det estimert at nåværende lagringsanlegg vil inneholde om lag 80 PB med data. For å erstatte dette, og legge til rette for de neste to års vekst (15 prosent gjennomsnittlig årlig vekstrate), må man ta høyde for om lag 106 PB ved anskaffelsen.

Finansieringsmodell

Leverandøren mottar finansiering til investeringer og grunnleggende driftsstøtte på vegne av departementsfellesskapet som koordineres gjennom Kunnskapsdepartementet, mens bruken av tungregning er brukerfinansiert (differensierte modeller for brukerfinansiering må utvikles). Midler til bruken av tungregning i leveransen av samfunnskritiske tjenester kanaliseres til det enkelte forvaltningsområde i årlige budsjettprosesser basert på tversektorielle vurderinger av fordeling av midler.

⁵⁶ Forskningsrådet, «Behov for tungregnekraft for forskning og kunstig intelligens».

Figur 9.11 gir en forenklet oversikt over hvordan tungregnekapasitet vil finansieres i dette konseptet.



Figur 9.11 Illustrasjon av finansieringsmodell i K3

Styringsmodell inkludert modell for tildeling av kapasitet.

Tildelingen av kapasitet til forskningsdrevet FoU skjer gjennom søknader til Ressursfordelingskomiteen (RFK) og tildeles basert på prosjektenes kvalitet. Dette er en ordning som er bredt akseptert blant forskningsmiljøer og derfor bør videreføres. Når det gjelder forvaltningsdrevet FoU etableres det en ordning der Sigma2 i større grad tildeler kapasitet basert på andre kriterier. Dette vil være en oppgave for forprosjektet å designe denne modellen.

Leveransemodell inkl. grunnleggende brukerstøtte

For å sikre at investeringsmidlene utnyttes mest mulig effektivt, får Sigma2 fullmakter til å disponere tilgjengelige investeringsmidler for bygging av ny tungregnekapasitet mer fleksibelt enn i dag. «Sigma3» kan da velge om ny kapasitet skal etableres i eget datasenter eller gjennom EuroHPC, basert på kost-nytte vurderinger og behovet for å opprettholde norsk suverenitet.

Konseptet innebærer at en større andel av kapasiteten vil brukes av andre brukere enn det som er forskningsdrevet FoU. I dette konseptet vil en ytterligere profesjonalisering av tjenesteleveransene være nødvendig. Det vil etableres en presis og differensiert tjenestekatalog som ivaretar behovene til ulike brukere. Tjenesteavtalene med BOTT-universitetene gjennomgås for å ytterligere profesjonalisere driftstjenester og grunnleggende brukerstøtte som leveres. «Sigma3» utvides med egen kapasitet for å dekke behovene som ikke kan leveres gjennom avtalene med BOTT-universitetene.

Sigma2 anskaffer utstyr og tjenester primært for eget bruk. I dette konseptet er det avgjørende at Sigma2 utvides til å kunne håndtere at andre sektorer ønsker å utnytte deres leverandøravtaler (utstyr, innplassering).

Organisering av avansert brukerstøtte

Avansert brukerstøtte utvides i Sigma2 for å ivareta behov fra nye sektorer som har tungregnekapasitet via Sigma2. Dagens ordning med NRIS (tjenesteavtale med BOTT-universitetene) for avansert brukerstøtte til forskningsorganisasjoner videreføres og utvides. I tillegg videreutvikles Norwegian Competence Center (NCC) for å ivareta det økte behovet for avansert brukerstøtte.

9.2 Rammeverk for vurderinger av konseptene

Samfunnsøkonomiske virkninger av de utvalgte konseptene for tungregning er endringer fra dagens situasjon som øker eller reduserer velferden for én eller flere grupper i samfunnet. I analysene skilles det mellom virkninger som er verdsatt i kroner («prissatte virkninger») og virkninger som er verdsatt kvalitativt («ikke-prissatte virkninger»). Ifølge veilederen for

samfunnsøkonomiske analyser utarbeidet av DFØ skal samfunnsøkonomiske virkninger verdsettes i kroner så langt det er mulig og hensiktsmessig.

For å vurdere de prissatte virkningene er det lagt til grunn enkelte forutsetninger som er beskrevet nærmere i kapittel 9.4.

For å vurdere de ikke-prissatte virkningene har vi benyttet verdimatrixemetoden beskrevet i veilederen for samfunnsøkonomisk analyse. Metoden er en systematisk og dokumenterbar vurdering av de ikke-prissatte virkningene. For å finne den samfunnsøkonomiske verdien vurderes først *kvantum* ved å vurdere hvor mange som berøres av virkningen og hvor stor påvirkningen er på hver av de berørte, og deretter vurderes *enhetsverdi* som tilsvarer betalingsvilligheten i befolkningen for virkningen. Prosessen er illustrert i Figur 9.12.



Figur 9.12 Illustrasjon av verdimatrixemetoden

Kvantumet vurderes ut ifra en syv-punkt skala fra stort negativt til stort positivt, og enhetsverdi vurderes ut ifra en tre-punkt skala lav, middels og høy. Ved å sammenstille vurderingen av kvantum og enhetsverdi, kan man vurdere den samfunnsøkonomiske verdien i DFØs verdimatrix, jf. Figur 9.13. Eksempelvis vil en ikke-prissatt virkning med lite positivt kvantum og høy enhetsverdi gi en middels positiv samfunnsøkonomisk verdi.

Kvantum \ Enhetsverdi	Liten	Middels	Høy
Stort positivt	Middels positiv	Stor positiv	Meget stor positiv
Middels positivt	Lite positiv	Middels positiv	Stor positiv
Lite positivt	Ubetydelig / ingen virkning	Lite positiv	Middels positiv
Hverken positivt eller negativt	Ubetydelig / ingen virkning	Ubetydelig / ingen virkning	Ubetydelig / ingen virkning
Lite negativt	Ubetydelig / ingen virkning	Lite negativ	Middels negativ
Middels negativt	Lite negativ	Middels negativ	Stor negativ
Stort negativt	Middels negativ	Stor negativ	Meget stor negativ

Figur 9.13 Verdimatrise til verdsetting av ikke-prissatte virkninger

9.3 Samfunnsøkonomiske virkninger

For å identifisere samfunnsøkonomiske virkninger av tiltaket har vi fulgt de anbefalte stegene i veilederen for samfunnsøkonomisk analyse fra DFØ:

1. Identifisert de berørte gruppene
2. Spesifisert egenskapene som driver virkningene for de berørte gruppene
3. Brukt årsaksforhold til å finne de samfunnsøkonomiske virkningene
4. Tatt en ekstra sjekk – er det noe som er fordelingsvirkninger eller telt dobbelt?

Figur 9.14 gir oversikt over de identifiserte berørte gruppene, hva tiltaket innebærer, de virkningsdrivende egenskapene ved tiltaket og overordnede årsak-virkningsforhold som fører frem til de utvalgte samfunnsøkonomiske virkningene og øvrige sekundærvirkninger. Vi har valgt å gi en overordnet fremstilling av årsak-virkningsforholdene slik at det skal være mulig å forstå hvordan vi har identifisert de samfunnsøkonomiske virkningene i analysen.

Berørte grupper	Hva innebærer tiltaket?	Virkningsdrivende egenskaper ved tiltaket	Hva fører dette til?			Samfunnsøkonomiske virkninger
<ul style="list-style-type: none"> Brukere Forskningsorganisasjoner Offentlige virksomheter Næringslivet 	Bygger offentlig finansiert infrastruktur for tungregning i Norge	<ul style="list-style-type: none"> Investeringer i tungregnekraft Investeringer i lagringskapasitet Investeringer i kompetanse Investeringer i profesjonell bruker støtte og tilgjengeliggjøring Investeringer i innsatsfaktorer (strøm, fysisk og digital sikring, arbeidskraft m.m.) Investeringer i software 	Forskere får tilgang til tungregnekraft og avansert bruker støtte	Utføre avanserte simuleringer og dataanalyser ved bruk av GPU og CPU	Økt kvalitet i forskningen og flytting av forskningsfronten i Norge	Mer forskning i internasjonal forskningsfront
Off. virksomheter får tilgang til tungregnekraft og avansert bruker støtte			Økt kunnskap og frambringning av ny kunnskap		Økt innovasjon og konkurransekraft	
Næringslivet får tilgang til tungregnekraft og avansert bruker støtte			Økt datainnsikt, beslutningsstøtte og kunnskapsgrunnlag samf.kritiske tj.		Forbedret samfunns-kritiske tjenester	
			Optimalisert beslutningsstøtte og ressursutnyttelse for off. styring og tjenester		Forbedret øvrige offentlig tjenesteyting	
<ul style="list-style-type: none"> Plattform-leverandører Avansert bruker støtte Datasentre 			Utvikle og trene KI-modeller	Økt kompetanse om bruk og nytte av tungregning		
			Bruke og tilpasse KI-modell		Bedre kunnskapsgrunnlag og forbedrede/nye prod. og tjenester	Økt verdiskapning
			Økte kostnader til investeringer og drift			Økte offentlige utgifter
			Økt kompetanse knyttet til tungregnekraft	Økt kvalitet på tjenester til brukere	Økt konkurransefortrinn	Økt verdiskapning
			Økt energiforbruk			Negative miljøkonsekvenser
			Økt behov for fysisk plass			

Figur 9.14 Identifisering av samfunnsøkonomiske virkninger i analysen

Nærmere forklaring av figuren:

Berørte grupper er primært brukere av tungregning og leverandører av tungregning, hvor sistnevnte består av plattformleverandører, tilbydere av avansert bruker støtte og datasentre (lagring). Samfunnet er også en relevant gruppe, som blir berørt av de samfunnsøkonomiske virkningene.

Tiltaket innebærer å bygge en offentlig finansiert infrastruktur for tungregning i Norge. Norge er også medfinansør av tungregneinfrastruktur i Europa, som inngår i tiltaket, men den samfunnsøkonomiske analysen omfatter virkninger i Norge (som anbefalt i veilederen for samfunnsøkonomisk analyse).

De virkningsdrivende egenskapene ved tiltaket er primært investeringene som gjøres ved utbygging av en tungregneinfrastruktur. Investeringene omfatter både den primære verdikjeden som består av de aktørene som leverer datasentre og tilgjengeliggjør tungregneplattformene og de sekundære verdikjedene som består av ulike leverandører som aktørene i den primære verdikjeden er avhengig av for å etablere sine tjenester, jf. kapittel 3.1.

Tiltaket innebærer investeringer i en tungregneinfrastruktur som *fører til* at brukerne får tilgang til tungregning (både kapasitet og kompetanse gjennom avansert bruker støtte). Denne tilgangen på tungregning, alt annet likt, gjør at brukerne kan utføre avanserte simuleringer og utføre dataanalyser (tradisjonelle beregninger og datadrevne beregninger ved bruk av kunstig intelligens), samt utvikle, trene, tilpasse og bruke KI-modeller, jf. omtale i kapittel 5.2. Forskernes tilgang og bruk av tungregning som tiltaket innebærer, fører til økt kunnskap og frambringning av ny kunnskap innenfor flere forskningsområder. I tillegg bidrar det til å øke kvaliteten i forskningen og flytte forskningsfronten, noe som gir økt konkurransekraft. Offentlige virksomheters tilgang og bruk av tungregning som tiltaket innebærer fører til et bedre kunnskapsgrunnlag, og således beslutningsunderlag, for utvikling av samfunnskritiske tjenester, øvrige offentlige tjenester og generelt styringsprosesser, herunder arbeidsprosesser m.m. Næringslivets tilgang og bruk av tungregning som tiltaket innebærer fører til at private bedrifter får bedre kunnskapsgrunnlag og kan forbedre og produsere nye produkter og tjenester som leveres i markedet.

De samfunnsøkonomiske virkningene som er identifisert i analysen er mer forskning i internasjonal forskningsfront og økt innovasjon og konkurransekraft som følge av forskere får tilgang til tungregning. Andre samfunnsøkonomiske virkninger er forbedret samfunnskritiske tjenester og øvrig offentlig tjenesteyting som følge av at offentlige virksomheter får tilgang til tungregning og økt verdiskapning som følge av at næringslivet får tilgang til tungregning. Årsak-virkningsforholdene er ytterligere beskrevet under den enkelte virkning i kapittel 9.6.

I den samfunnsøkonomiske analysen har vi prissatt en av virkningene og kvalitativt vurdert de øvrige virkningene (såkalte «ikke-prissatte» virkninger). Figur 9.15 gir en oversikt over virkningene.

Mer forskning i internasjonal forskningsfront	<ul style="list-style-type: none"> • Virkning som følge av at forskere får tilgang til tungregning, inkl. avansert brukerstøtte • Vurdert som <i>ikke-prissatt</i> virkning
Økt innovasjon og konkurransekraft	<ul style="list-style-type: none"> • Virkning som følge av at forskere får tilgang til tungregning, inkl. avansert brukerstøtte • Vurdert som <i>ikke-prissatt</i> virkning
Forbedret samfunnskritiske tjenester	<ul style="list-style-type: none"> • Virkning som følge av at offentlige virksomheter får tilgang til tungregning, inkl. avansert brukerstøtte • Vurdert som <i>ikke-prissatt</i> virkning
Forbedret øvrig offentlig tjenesteyting	<ul style="list-style-type: none"> • Virkning som følge av at offentlige virksomheter får tilgang til tungregning, inkl. avansert brukerstøtte • Vurdert som <i>ikke-prissatt</i> virkning
Økt verdiskapning	<ul style="list-style-type: none"> • Virkning som følge av at private virksomheter får tilgang til tungregning, inkl. avansert brukerstøtte • Vurdert som <i>ikke-prissatt</i> virkning
Økte offentlige utgifter	<ul style="list-style-type: none"> • Virkning som følge av investeringer i offentlig finansiert infrastruktur for tungregning • Vurdert som <i>prissatt</i> virkning
Miljøkonsekvenser	<ul style="list-style-type: none"> • Virkning som følge av investeringer i infrastruktur for tungregning og bruk av tungregning (strøm) • Vurdert som <i>ikke-prissatt</i> virkning

Figur 9.15 Oversikt over samfunnsøkonomiske virkninger

9.4 Forutsetninger til den samfunnsøkonomiske analysen

Forutsetninger for prissatte virkninger

Tabell 9.1 Forutsetninger for prissatte virkninger

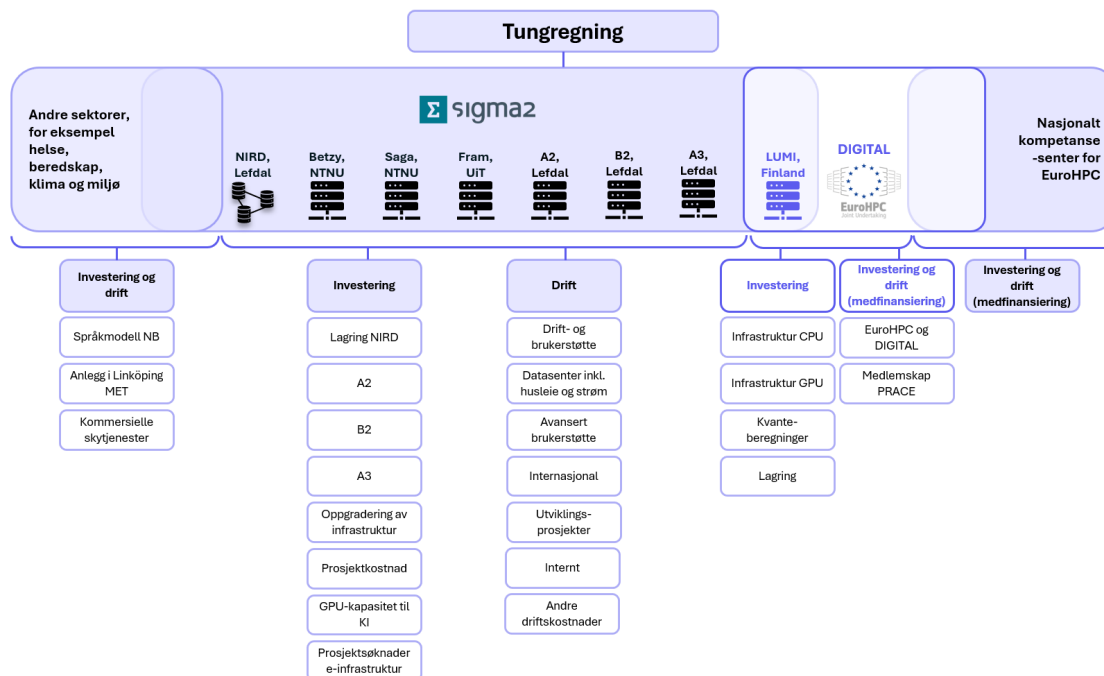
Felles forutsetning	Verdi	Begrunnelse	Kilde
Kalkulasjonsrente	4 prosent	Diskontering benyttes for å kunne sammenligne alternativer med prissatte nytte- og kostnadsvirkninger som oppstår ved ulike tidspunkt. Dette gjøres ved å omregne virkningene i en periode, gjerne over mange år, til verdien i et gitt år ved å benytte en diskonteringsrente eller kalkulasjonsrente. Kalkulasjonsrenten settes til 4 prosent for en periode fra 0 til 40 år, i henhold til Finansdepartementets rundskriv R-109/21 for	DFØ (2023). Veileder i samfunnsøkonomiske analyser, tabell 3.7

samfunnsøkonomiske analyser for investeringer.

Priser og prisutvikling		Estimatene er basert på prognoser fram i tid, og kildene som er brukt har tatt høyde for prisutvikling. Unntaket er priser for GPUer fra kommersielle skytjenester. Her er det inflasjonsjustert tall fra 2025 til 2026 kr ved bruk av KPI til SSB på 2,7prosent.	SSB (2024). Konjunkturtendensene.
Analyseperiode	2026-2030	KVUen har som mandat å se på forventet utvikling i et femårsperspektiv. Ettersom tidligste mulig finansiering er fra 2026, er valgt analyseperiode fra 2026 til 2030.	
Skattekostnader	20 øre per krone	Det antas at all finansiering skjer gjennom bevilgninger over statsbudsjettet eller andre offentlige budsjetter, og at det således påløper skattekostnader. Skattekostnader representerer den marginale kostnaden knyttet til å hente inn en ekstra skattekrone. Skattekostnader settes til 20 øre per krone i henhold til Finansdepartementets veileder. Beregningsgrunnlaget for skattekostnaden er finansieringsbehovet eksklusive merverdiavgift. Skattefinansieringskostnader inngår i den samfunnsøkonomiske analysen, men påvirker ikke finansieringsbehovet.	DFØ (2023). Veileder i samfunnsøkonomiske analyser, kapittel 3.4.8.
Merverdiavgift		Merverdiavgift for prosjektets finansieringsbehov beregnes som 25 prosent av forventningsverdien til alle momsbelagte elementer i analysen. Det tillegges ikke en egen usikkerhet på merverdiavgiften, men beløpet vil avhenge av usikkerhet i elementer som det beregnes merverdiavgift på. Usikkerhetsfaktorer virker, med mindre annet er spesifisert, på beløp inkludert merverdiavgift. Innsatsfaktorer oppgis inklusive avgifter, herunder merverdiavgift. Den samfunnsøkonomiske analysen gjøres uten merverdiavgift da dette ikke direkte representerer en kostnad for samfunnet, men en overføring. Finansieringsbehovet oppgis med merverdiavgift.	

9.5 Prissatte virkninger

Prissatte virkninger for utbygging av en infrastruktur for tungregning består overordnet av investerings- og driftskostnader. Kostnadene er fordelt på aktørene i dagens Sigma2, andre nasjonale sektorer (som for eksempel helse, beredskap, klima og miljø), og internasjonale samarbeidsplattformer inkludert EuroHPC. Figur 9.16 Prosjektnedbrytningsstrukturen for kostnader som er lagt til grunn for kalkylene av de ulike konseptene. Figur 9.16 viser prosjektnedbrytningsstrukturen for prissatte virkninger i konseptene.



Figur 9.16 Prosjektnedbrytningsstrukturen for kostnader som er lagt til grunn for kalkylene av de ulike konseptene

Aktører i første rad har blå bakgrunnsfarge for å markere en nasjonal satsing, og hvit farge for å markere internasjonal satsing. Andre rad i består av overordnede kostnader som investering, drift, finansiering og medfinansiering. Kostnadselementene i hver kolonne som følger er det den overordnede kostnaden består av.

Sigma2 sine kostnader består av investerings- og driftskostnader. Investeringskostnader inkluderer utvidet lagringskapasitet hos NIRD, etablering av superdatamaskinene A2, B2, og A3, oppgradering av eksisterende infrastruktur, prosjektkostnader (personalkostnader for gjennomføring av anskaffelse og utviklingsprosjekter), etablering av GPU-kapasitet til KI og kapasitet for å dekke kjente behov for tungregning som følge av prosjekter finansiert gjennom INFRASTRUKTUR-ordningen. Driftskostnader for Sigma2 inkluderer, drift- og brukerstøtte (tilknyttet superdatamaskiner, lagring og tjenestepattform), datasentre inkludert husleie og strøm, avansert brukerstøtte, internasjonale kostnader i forbindelse med personalkostnader og eventuelle kontingenter for deltakelse i internasjonale prosjekter, utviklingsprosjekter, interne kostnader (som for eksempel lønnskostnader), og andre driftskostnader.

Andre sektorer sine kostnader består av prosjekter knyttet til tungregning. De prissatte kostnadene som er inkludert er utvikling av norsk språkmodell hos Nasjonalbiblioteket, investering og drift av tungregneanlegg for klimamodellering (et samarbeid mellom MET og andre internasjonale meteorologiske institusjoner) og kjøp av kommersielle skytjenester til tungregning (som for eksempel Google TPU Research Cloud, IBM og D-Wave). Avhengig av konsept kan kostnadene overlappe mellom andre sektorer og Sigma2, som er illustrert i prosjektnedbrytningsstrukturen.

For at Norge skal kunne ha nytte av europeisk samarbeid i satsingen for tungregning, kreves det norsk investering og medfinansiering i Europa. Dette inkluderer investeringskostnader i superdatamaskinen LUMI i Finland via EuroHPC, som består av etablering av infrastruktur til

CPU, GPU, kvanteberegninger og lagring. Medfinansiering inkluderer kostnader tilknyttet deltagelse i EuroHPC, DIGITAL og medlemskap i PRACE. For EuroHPC gjelder dette kostnader for aktiviteter som Sigma2 og andre aktører ønsker å utnytte i kommende utlysninger. Det anbefales at det etableres fleksible medfinansieringsordninger som gir bedre muligheter for norske aktørers deltakelse i DIGITAL/EuroHPC. Sigma2 har per nå medlemskap i PRACE, og denne kostnaden er inkludert i prosjektnedbrytningsstrukturen. Det forventes også en ny utlysning til nasjonale kompetansesenter gjennom EuroHPC, og norske aktører vil sannsynligvis søke. Dette er illustrert i prosjektnedbrytningsstrukturen som en overlapp av nasjonal og internasjonal satsing, som medfinansiering i et nasjonalt kompetansesenter for EuroHPC.

Beregningen av prissatte virkninger er basert på data og forutsetninger fra flere kilder:

- Cloud GPU configuration (Google Cloud, 2024)
- Azure Machine Learning pricing (Microsoft, 2024)
- Behov for tungregnekraft for forskning og kunstig intelligens (Forskningsrådet 2024)
- 40 millioner til styrking av norsk og samisk språk i KI (Regjeringen, 2024)
- 2024-04-29 Sigma2 Investeringsbehov 2025 Oppdatert (Sigma2, 2024)
- Langtidsbudsjett 2024-2033 (Sigma2, 2024)
- Revised Downscaled investment budget for project 350171 (Sigma2, 2023)
- Revised Minimum proposal for NeIC – 350178 (Sigma2, 2023)
- Sak-51-24 Strategidiskusjon internasjonale investeringer (Sigma2, 2024)
- NVIDIA RTX A1000 (TechPowerUp, 2024)

9.5.1 Årlige kostnader for nullalternativet, K2 og K3

Tabell 9.2, Tabell 9.3 og Tabell 9.4 viser årlige kostnader for nullalternativet, K2 og K3 i analyseperioden 2026 til 2030. Kostnadselementene er i hovedsak like, med litt variasjon knyttet «inkludert GPU-kapasitet til KI» og «annen sektorfinansiering». Dette er grunnet forskjeller i finansiering og ansvarsfordeling for disse to elementene i K0, K2 og K3.

Videre i kapittelet er en overordnet beskrivelse av kostnadselementer og trender i kostnadene. For en detaljert beskrivelse av konseptene, se kapittel 9.1.

Årlige kostnader for null-alternativet

Tabell 9.2 Årlige kostnader i MNOK, K0 Nullalternativet

Kostnadselement	2026	2027	2028	2029	2030	2026-30
Investering Sigma2, inkludert GPU-kapasitet til KI	58	121	0	0	0	179
Drift Sigma2	138	138	126	128	128	658
Investering gjennom EuroHPC/LUMI	0	0	0	0	0	0
Sum	196	259	126	128	128	837

I K0 legges til grunn det reviderte investeringsbudsjettet til Sigma2, som inkluderer etablering av superdatamaskinen B2, utvidelse av superdatamaskinen A2, oppgradering av infrastruktur og andre prosjektkostnader. Disse tar over for superdatamaskinene Fram, Saga og Betzy som vil gå ut av drift i løpet av analyseperioden. For drift av Sigma2 legges til grunn Sigma2 sitt langtidsbudsjett per i dag. I K0 vil det ikke være en investering i EuroHPC/LUMI.

Grunnet investeringskostnader Sigma2 er det en økning i totale kostnader i 2026 og 2027, og ellers stabilt ca. 130 MNOK per år. Total kostnad i perioden 2026-30 er ca. 840 MNOK.

Årlige kostnader for K2 «Sigma2-pluss»

Tabell 9.3 Årlige kostnader i MNOK, K2 «Sigma2-pluss»

Kostnadselement K2	2026	2027	2028	2029	2030	2026-30
Investering Sigma2	102	253	182	255	0	791
Drift Sigma2	138	155	161	167	170	792
Investering gjennom EuroHPC/LUMI	193	193	85	20	20	510
Annen sektorfinansiering, inkludert GPU-kapasitet til KI	298	289	276	317	358	1 537
Sum	731	889	705	759	547	3 630

I K2 legges til grunn Sigma2 sitt fremtidige investeringsbehov, som inkluderer anskaffelse av lagringskapasitet ved NIRD, økte midler sammenliknet med K0 til etablering av B2 og utvidelse av A2, etablering av superdatamaskinen A3, oppgradering av infrastruktur og andre prosjektkostnader. Drift av Sigma2 er basert på langtidsbudsjettet til Sigma2 per i dag, samt en økning i datasenterkostnader inkl. husleie og strøm for ny tungregne- og lagringskapasitet. Det investeres gjennom internasjonalt gjennom EuroHPC/LUMI. Under annen sektorfinansiering, inkludert GPU-kapasitet til KI, består av følgende kostnader: GPU-kapasitet til KI, prosjektsøknader tungregning, norsk språkmodell, MET investering og drift av anlegg i Linköping og midler til bruk av kommersielle tjenester.

I K2 er kostnadene høyere i perioden 2026-2029 (mellom 700 og 900 MNOK) grunnet økte kostnader for investering Sigma2 og investering gjennom EuroHPC/LUMI. I 2030 er estimert total kostnad ca. 550 MNOK. Total kostnad i perioden 2026-30 er ca. 3 630 MNOK.

Årlige kostnader for K3 «Tverrsektoriell Sigma3»

Tabell 9.4 Årlige kostnader i MNOK, K3 «Tverrsektoriell Sigma3»

Kostnadselement K3	2026	2027	2028	2029	2030	2026-30
Investering Sigma2, inkludert GPU-kapasitet til KI	395	496	372	445	190	1 898
Drift Sigma2	153	170	176	182	170	852
Investering gjennom EuroHPC/LUMI	193	193	85	20	20	510
Annen sektorfinansiering	21	21	21	21	21	105
Sum	762	879	654	668	401	3 365

I K3 består følgende av investering Sigma2, inkludert GPU-kapasitet til KI: Sigma2 sitt fremtidige investeringsbehov (som inkluderer anskaffelse av lagringskapasitet ved NIRD, økte midler sammenliknet med K0 til etablering av B2 og utvidelse av A2, etablering av superdatamaskinen A3, oppgradering av infrastruktur og andre prosjektkostnader), GPU-kapasitet til KI, prosjektsøknader tungregning og norsk språkmodell. Drift Sigma2 er basert på langtidsbudsjettet til Sigma2 per i dag, samt en økning i datasenterkostnader inkl. husleie og strøm for ny tungregne- og lagringskapasitet, og en økning i kostnader for avansert brukerstøtte. Det investeres internasjonalt gjennom EuroHPC/LUMI. Annen sektorfinansiering består av MET sin investering og drift av anlegg i Linköping.

I K3 er kostnadene høyere i perioden 2026-2029 (mellom 650 og 900 MNOK) grunnet økte kostnader for investering Sigma2 og investering gjennom EuroHPC/LUMI. I 2030 er estimert total kostnad ca. 400 MNOK. Total kostnad i perioden 2026-30 er ca. 3 370 MNOK.

Oversikt prissatte virkninger 2026-2030

Tabell 9.5 Oversikt prissatte virkninger 2026-2030

Prissatte virkninger (nåverdi MNOK, avrundet til nærmeste 10 MNOK)	K0 Absolutte tall (brutto)	K2 Endring fra K0 (netto)	K3 Endring fra K0 (netto)
Investeringskostnad	-170	-1 880	-2 140
Driftskostnad	-610	-720	-230
Skattefinansieringskostnad	-160	-520	-470
Sum prissatte virkninger	-940	-3 110	-2 840

Tabell 9.5 viser nåverdien av prissatte virkninger for konseptene. Prissatte virkninger i tabellen er delt opp i investeringskostnad, driftskostnad og skattefinansieringskostnad. For K2 og K3 vises tallene i netto, altså endring fra K0 i perioden 2026-2030.

Forskjellen i investeringskostnad og driftskostnad mellom K2 og K3 er primært tilknyttet fordelingen av kostnader knyttet til GPU-kapasitet til KI. I K2 er utvidet GPU-kapasitet til KI både en investering via Sigma2 og finansiering av andre sektorer. Andre sektorer kan velge å bruke kapasitet hos Sigma2, eller bruke bevilgede midler til å kjøpe andre tungregnetjenester som kommersielle skytjenester, som detaljert i kapittel 9.1.3. Investering i GPU-kapasitet i Sigma2

regnes som en investeringskostnad, og kjøp av andre tungregnetjenester som kommersielle skytjenester regnes som en driftskostnad. I K3 er utvidet GPU-kapasitet til KI en investering hos Sigma2, som beskrevet i kapittel 9.1.4. Dermed er det en høyere investeringskostnad i K3 enn i K2, og en lavere driftskostnad i K3 enn i K2.

Skattefinansieringskostnad er satt til 20 øre per krone per prissatt virkning, altså ytterligere 20 prosent av investerings- og driftskostnad, som begrunnet i kapittel 9.4.

Dette gir en sum prissatte virkninger i endring fra K0 på omtrent -3 110 MNOK for K2 og -2 840 MNOK for K3, dermed litt høyere for K2. Dette skyldes primært at det er estimert dyrere for andre sektorer å kjøpe kommersielle skytjenester i K2 sammenliknet med å investere i økt GPU-kapasitet i Sigma2 i K3.

9.6 Ikke-prissatte virkninger

For å vurdere de ikke-prissatte virkningene har vi benyttet verdimatrisemetoden beskrevet i veilederen for samfunnsøkonomisk analyse, jf. omtale i kapittel 9.2. Under følger en omtale av verdsettingen av den enkelte ikke-prissatte virkning.

9.6.1 Vurdering av «Mer forskning i internasjonal forskningsfront»

Å *flytte forskningsfronten* refererer til det å frembringe ny kunnskap eller innsikt som betydelig endrer forståelsen innenfor ulike fagfelt. Flere rapporter og studier peker på banebrytende forskning som en stadig viktigere drivkraft for økt konkurranseevne⁵⁷, og relevant forskningsinfrastruktur er helt avgjørende for at forskningsmiljøer skal hevde seg internasjonalt.

Dette handler om å utvikle banebrytende teorier, metoder, eller empiriske funn som transformerer eller redefinerer eksisterende viten. Denne typen forskning er mer drevet av å ligge i kunnskapsfronten enn av samfunnsmessig nytte, men slik forskning har vist seg å ofte ha det største potensiale for å frembringe store gjennombrudd i utviklingen av samfunnet og bidrar til radikale innovasjoner.

Med den raske utviklingen i potensialet for å bruke KI også i vitenskapelig utvikling, er tilstrekkelig og forutsigbar tilgang til tungregning en *forutsetning* for at forskere i Norge skal lykkes med å være forskningsfronten og ledende internasjonalt innen sitt fagfelt. Dersom Norge skal klare å tiltrekke seg de best kvalifiserte forskertalentene er det å ha tilgang på ressurser knyttet til tungregnekapasitet og kompetanse avgjørende.

Forskning i internasjonal forskningsfront er videre en forutsetning for å oppnå øvrige samfunnsøkonomiske virkninger, herunder økt innovasjon og konkurransekraft i samfunnet.

Basert på de virkningsdrivende egenskaper ved tiltaket og årsak-virkningsforholdene som vist i Figur 9.14 og beskrevet over, har vi verdsett den samfunnsøkonomiske verdien av «mer forskning i internasjonal forskningsfront» i de utvalgte konseptene. Analysen er vist i Tabell 9.6.

⁵⁷ Slik som for eksempel [Draghi-rapporten](#) og [Heitor-rapporten](#).

Tabell 9.6 Verdsetting av virkningen «Mer forskning i internasjonal forskningsfront»

Konsept	Kvantum	Enhetsverdi	Samfunns-økonomisk verdi
K2, K3	<p><i>Hvor mange blir berørt av virkningen?</i></p> <p>Mer forskning i internasjonal forskningsfront kan påvirke alle forskere og ansatte i forskningsorganisasjoner. Antall personer som blir berørt i et samfunnsperspektiv er derfor vurdert til lavt.</p> <p><i>Hvor stor er påvirkningen på hver enkelt berørt?</i></p> <p>I både K2 og K3 får forskere tilgang til infrastruktur for tungregning som er avgjørende for at forskningsorganisasjoner får brukt tungregning til forskning og utvikling (FoU), både grunnforskning og anvendt forskning. Dette er en forutsetning for å flytte forskningsfronten i Norge, og dermed oppnå mer forskning i internasjonal forskningsfront. Den direkte positive effekten av dette vil være stort positivt for det forskningsfeltet det gjelder, og vurderes til middels positivt for øvrige forskningsfelt.</p> <p>Samlet vurderer vi kvantum til lite positivt.</p>	<p><i>Hva er verdien av virkningen?</i></p> <p>Mer forskning i internasjonal forskningsfront har størst potensiale for å bli grensesprengende og frembringe radikale innovasjoner, og er avgjørende for å generere ny kunnskap og flytte forskningsfronten innenfor flere forskningsfelt.</p> <p>Samlet vurderer vi enhetsverdien av mer forskning i internasjonal forskningsfront som høy.</p>	<p><i>Samlet vurdering av virkningen</i></p> <p>Middels positiv virkning i K2 og K3 basert på verdimatrisen.</p>

9.6.2 Vurdering av «Økt innovasjon og konkurransekraft»

Basert på antakelsene om virkningsdrivende egenskaper og årsak-virkningsforholdene vist i Tabell 9.7, har vi verdsett den samfunnsøkonomiske verdien av «økt innovasjon og konkurransekraft» i samfunnet. Analysen er vist Tabell 9.7.

Tilstrekkelig tungregneressurser er en sentral komponent i en moderne forskningsinfrastruktur, og er en sentral drivkraft i et velfungerende forskningssystem. Investering i tungregning gir forskere tilstrekkelig tilgang til infrastruktur som er avgjørende for at forskningsorganisasjoner får brukt tungregning til forskning og utvikling (FoU). Dette bidrar til å øke kvaliteten på forskningen og gjør det mulig å utforske nye ideer og teknologier. Dette vil bidra til nye og forbedrede produkter og tjenester.

Investering i tungregning fremmer samarbeid mellom forskningsinstitusjoner, næringsliv og offentlig forvaltning, og muliggjør at forskere fra forskjellige fagfelt samarbeider i tverrfaglige prosjekter. Dette er videre en forutsetning for internasjonalt samarbeid og kompetansebygging. Dette samarbeidet fører til kunnskapsoverføring og synergier som kan resultere i nye innovasjoner og løsninger på samfunnsutfordringer.

Innovasjon og konkurransekraft er avgjørende for et samfunnsøkonomiske vekst, omstilling, konkurransevne og evne til å løse komplekse utfordringer, som legger grunnlaget for et kunnskapsbasert og bærekraftig velferdssamfunn.

Basert på de virkningsdrivende egenskaper ved tiltaket og årsak-virkningsforholdene som vist i Figur 9.14 og beskrevet over, har vi verdsett den samfunnsøkonomiske verdien av «økt innovasjon og konkurransekraft» i de utvalgte konseptene. Analysen er vist i Tabell 9.7.

Tabell 9.7 Verdsetting av virkningen «Økt innovasjon og konkurransekraft»

Konsept	Kvantum	Enhetsverdi	Samfunns-økonomisk verdi
K2, K3	<p><i>Hvor mange blir berørt av virkningen?</i></p> <p>Økt innovasjon og konkurransekraft kan påvirke hele samfunnet, dvs. alle innbyggere i Norge. Antall personer som blir berørt er derfor vurdert til høy.</p> <p><i>Hvor stor er påvirkningen på hver enkelt berørt?</i></p> <p>Påvirkningen på de berørte er avhengig av forskeres tilgang til tungregning. I både K2 og K3 får forskere tilgang til infrastruktur for tungregning som er avgjørende for at forsknings- organisasjoner får brukt tungregning til forskning og utvikling (FoU). Økt forskning som innebærer økt kunnskap og frambringning av ny kunnskap, bidrar til økt innovasjon og konkurransekraft i samfunnet. Påvirkningen av økt innovasjon og konkurransekraft kan variere mellom innbyggere og befolknings-grupper alt ettersom hvor den økte eller nye kunnskapen generes. Samtidig vil store deler av forskningen gi sektorovergripende gevinster som vil gange alle samfunnslag positivt.</p> <p>Samlet vurderer vi kvantum i K2 og K3 til stor positivt.</p>	<p><i>Hva er verdien av virkningen?</i></p> <p>Innovasjon og konkurransekraft er avgjørende for et samfunns økonomiske vekst, omstilling, konkurranseevne og evne til å løse komplekse utfordringer, som legger grunnlaget for et kunnskapsbasert og bærekraftig velferdssamfunn.</p> <p>Samlet vurderer vi enhetsverdien av økt innovasjon og konkurransekraft i K2 og K3 som høy.</p>	<p><i>Samlet vurdering av virkningen</i></p> <p>Meget stor positiv virkning i K2 og K3 basert på verdimatrisen.</p>

9.6.3 Vurdering av «Forbedret samfunnskritiske tjenester»

Tiltaket i stort innebærer investeringer i en tungregneinfrastruktur som innebærer at offentlige virksomheter får tilgang til tungregning slik at de kan utføre avanserte simuleringer og utføre dataanalyser (tradisjonelle beregninger og datadrevne beregninger ved bruk av kunstig intelligens), samt utvikle, trene, tilpasse og bruke KI-modeller.

Ett sentralt behovsområde for tungregning er utvikling og leveranse av tjenester av samfunnskritisk betydning. I Norge er det definert 14 kritiske samfunnsfunksjoner: Styring og kriseledelse, forsvar, lov og orden, helse og omsorg, redningstjenesten, digital sikkerhet i sivil sektor, natur og miljø, forsyningssikkerhet, vann og avløp, finansielle tjenester, kraftforsyning, elektronisk kommunikasjon, transport og satellittbaserte tjenester.⁵⁸ En relativt stor andel offentlige virksomheter leverer samfunnskritiske tjenester innenfor disse funksjonene, hvor flere benytter tungregning. Både kapasitet og kompetanse innenfor tungregning er nødvendige innsatsfaktorer.

⁵⁸ Regjeringen, «Liste over virksomheter med kritisk samfunnsfunksjon og nøkkelpersonell».

Tilgang på tungregning kan bidra til å forbedre samfunnskritiske tjenester gjennom flere prosesser. Vi har identifisert flere potensielle prosesser avhengig av fagområde og forvaltningsområde:

- **Forbedret datainnsikt:** Tungregning gir mulighet til å analysere store og komplekse datasett raskt, noe som gir dypere innsikt i trender og mønstre som kan påvirke tjenesteleveransen. For eksempel kan helsemyndigheter bruke dette til å analysere pasientdata for tidlig identifisering av sykdomsutbrudd, noe som forbedrer responsen på helsekriser.
- **Presis modellering og simulering:** Evnen til å utføre avanserte simuleringer gir bedre forståelse av potensielle scenarier, noe som kan forbedre planlegging og forebygging. Innen energiforsyning kan dette brukes til å modellere strømmnettets respons på ulike belastninger, noe som sikrer stabil og effektiv energiforsyning.
- **Raskere beslutningsprosesser:** Med mer nøyaktige og omfattende dataanalyser kan beslutninger tas raskere og med større sikkerhet. Dette er kritisk iblant annet beredskapssituasjoner, som ved naturkatastrofer, hvor rask respons kan redde liv og minimere skader.
- **Økt automatisering:** Integrering av kunstig intelligens og maskinlæring kan automatisere rutineoppgaver og forbedre tjenestekvaliteten. For eksempel kan IKT-sikkerhetssystemer bruke KI til å oppdage og respondere på trusler automatisk, noe som øker sikkerheten.
- **Optimalisering av ressurser:** Tungregning kan bidra til mer effektiv ressursallokering ved å identifisere de mest effektive måtene å bruke tilgjengelige ressurser på. I helsevesenet kan dette bety bedre fordeling av medisinsk utstyr og personell under pandemier.
- **Bedre risikostyring:** Evnen til å forutsi og modellere risikoer mer nøyaktig kan forbedre beredskapsplanlegging og redusere potensielle negative konsekvenser. Meteorologiske tjenester kan bruke tungregning til å forutsi ekstremvær mer presist, noe som gir bedre forberedelse og respons.
- **Forbedret beslutningsgrunnlag:** Ved å kombinere data fra ulike kilder kan offentlige virksomheter få et mer helhetlig bilde av situasjoner, noe som gir et bedre grunnlag for strategiske beslutninger. Dette kan være avgjørende i komplekse situasjoner som krever koordinert innsats fra flere sektorer, som i nasjonal sikkerhet og beredskap.

Gjennom disse prosessene kan tilgang på tungregning bidra til å forbedre samfunnskritiske tjenester. Å forbedre samfunnskritiske tjenester omfatter flere forhold, blant annet forbedret nøyaktighet, effektivitet og kvalitet i tjenesteleveransene, bedre koordinering mellom sektorer, bedre ressursallokering og effektiv bruk av offentlige ressurser, økt pålitelighet og sikkerhet i kritiske systemer og økt pålitelighet og sikkerhet i tjenesteleveranser.

Basert på de virkningsdrivende egenskaper ved tiltaket og årsak-virkningsforholdene som vist i Figur 9.14 og beskrevet over, har vi verdsett den samfunnsøkonomiske verdien av «forbedret samfunnskritiske tjenester» i de utvalgte konseptene. Analysen er vist i Tabell 9.8.

Tabell 9.8 Verdssetting av virkningen «Forbedret samfunnskritiske tjenester»

Konsept	Kvantum	Enhetsverdi	Samfunnsøkonomisk verdi
K2	<p><i>Hvor mange blir berørt av virkningen?</i></p> <p>Samfunnskritiske tjenester kan påvirke alle innbyggere i samfunnet. Antall personer som blir berørt er derfor vurdert til høy.</p> <p><i>Hvor stor er påvirkningen på hver enkelt berørt?</i></p>	<p><i>Hva er verdien av virkningen?</i></p> <p>Samfunnskritiske tjenester er funksjoner som er nødvendig for å opprettholde samfunnets grunnleggende behov, som mat, vann, helse, sikkerhet</p>	<p><i>Samlet vurdering av virkningen</i></p> <p>Stor positiv virkning i K2 basert på verdimatrisen.</p>

Påvirkningen på de berørte er avhengig av offentlige virksomheter sin tilgang på tungregning for å utvikle, tilpasse og levere å levere samfunnskritiske tjenester.

I K2 må det enkelte forvaltningsområde selv må sørge for å dekke underliggende virksomheters behov for tungregning til å utvikle, tilpasse, forbedre og levere samfunnskritiske tjenester. Det kan være positivt i form av at løsningene kan tilpasses forvaltningsområdene sine behov. Det vil typisk være en fordel for større sektorer med særskilte behov, f.eks. helsesektoren. På den andre siden kan K2 resultere i en mer fragmentert tungregneinfrastruktur med varierende kvalitet og hvor den enkelte virksomhet kan risikere å bli nedprioritert hos private leverandører. Det gjelder særlig mindre virksomheter. I tillegg får ikke offentlig sektor effektivisert bruken av tungregning på tvers av offentlige virksomheter mtp. fleksibilitet, skalerbarhet og ressursbruk.

Samlet sett vurderer vi at den totale påvirkningen på hver enkelt berørt vil være positiv siden offentlige virksomheter som leverer samfunnskritiske tjenester får finansiering til å dekke disse behovene. Samtidig vil effekten bli dempet av en fragmentert infrastruktur med større risiko for at virksomhetene får tilgang til tungregneressurser.

Samlet vurderer vi kvantum i K2 til **middels positivt**.

og styringsevne, og for at samfunnet skal fungere. I Norge er det definert 14 kritiske samfunnsfunksjoner, inkludert helseberedskap, energiforsyning, IKT-sikkerhet og styring og ledelse.

Samlet vurderer vi enhetsverdien av forbedret samfunnskritiske tjenester i K2 som **høy**.

K3	<p><i>Hvor mange blir berørt av virkningen?</i></p> <p>I K3 kan de samme bli berørt av virkningen som i K2, dvs. alle innbyggerne i samfunnet. Antall personer som blir berørt er derfor vurdert til høy.</p> <p><i>Hvor stor er påvirkningen på hver enkelt berørt?</i></p> <p>I K3 vil offentlige virksomheter som har ansvar for samfunnskritiske tjenester bli tilbudt kapasitet gjennom «Sigma3». Det kan være positivt i form av virksomhetene er garantert tungregnekapasitet og kompetanse. En stor leverandør vil også ha større markedsmakt og midler til å investere på vegne av flere. I tillegg vil løsningen økt fleksibilitet og koordinering på tvers av sektorer og bedre utnyttelse av tungregning på tvers av virksomheter og over tid. På den andre siden vil K3 innebære mindre valgfrihet for den enkelte virksomhet og potensielt mindre skreddersydde løsninger.</p> <p>Samlet sett vurderer vi at den totale påvirkningen på hver enkelt berørt vil være positiv og at en tverrsektoriell tilnærming kan føre til mer helhetlige robuste og effektive løsninger som gagnar offentlige virksomheter som leverer samfunnskritiske tjenester.</p> <p>Samlet vurderer vi kvantum i K3 til stor positivt.</p>	<p><i>Hva er verdien av virkningen?</i></p> <p>Enhetsverdien av forbedret samfunnskritiske tjenester er lik i K2 og K3, og vurderes derfor til høy.</p>	<p><i>Samlet vurdering av virkningen</i></p> <p>Meget stor positiv virkning i K3 basert på verdimatrisen.</p>
----	--	--	--

9.6.4 Vurdering av «Forbedret øvrig offentlig tjenesteyting»

Utover samfunnskritiske tjenester omtalt i kapittel 9.6.4, benytter flere offentlige virksomheter tungregning til utvikling og forbedring av ikke-kritiske offentlige tjenester og i interne prosesser. En økende andel offentlige virksomheter har også begynt å utforske og anvende kunstig intelligens til å forbedre offentlige tjenester og effektivisere forvaltningen, som potensielt også vil kreve tungregneressurser.

Tilgang på tungregning kan bidra til å forbedre den offentlige tjenesteytingen utover samfunnskritiske tjenester. Vi har identifisert flere potensielle prosesser:

- Økt automatisering: Tungregning muliggjør bruk av kunstig intelligens for å automatisere rutineoppgaver, noe som kan frigjøre tid og ressurser. For eksempel kan automatisert saksbehandling i statlige, fylkeskommunale og kommunale tjenester redusere behandlingstiden og forbedre brukeropplevelsen hos innbyggerne.
- Optimalisering av ressurser: Tungregning gjennom analysing av store datamengder kan hjelpe offentlige virksomheter å identifisere ineffektive prosesser og optimalisere ressursbruken. Dette kan gi kostnadsbesparelser og mer effektiv drift.
- Forbedret beslutningsstøtte: Avanserte dataanalyser kan gi bedre grunnlag for beslutninger, noe som kan forbedre planlegging og gjennomføre offentlige prosjekter og forvaltningsoppgaver. For eksempel kan byplanlegging dra nytte av simuleringer for å utvikle mer bærekraftige og effektive løsninger, og offentlige etater kan få bedre digital bistand i behandling av søknader til støtteordninger.
- Personalisering av tjenester: Ved å bruke dataanalyser kan offentlige tjenester i større grad tilpasses individuelle behov, noe som forbedrer tjenestekvaliteten. For eksempel kan utdanningssektoren bruke analyser for å tilpasse læringsopplegg til elevens spesifikke behov og læringsstiler.
- Effektivisering av arbeidsflyt: Tungregning kan bidra til å strømlinjeforme arbeidsprosesser ved å identifisere flaskehalsar og foreslå prosessforbedringer. Det kan føre til raskere saksbehandlingstid og gjennomføring av utviklingstiltak.
- Forbedret dataintegrasjon: Ved å samle og analysere data fra ulike kilder kan offentlige virksomheter få bedre tverrsektorielle styringsinformasjon, noe som forbedrer koordinering og samarbeid på tvers av forvaltningsorganer.
- Kvalitetsforbedring: Gjennom avanserte analyser kan offentlige tjenester kontinuerlig overvåkes og evalueres, noe som gir mulighet for løpende forbedringer og høyere kvalitet på tjenestene som tilbys.
- Redusert feilmargen: Ved å bruke presise dataanalyser kan offentlige virksomheter redusere feil i beslutningsprosesser, noe som øker påliteligheten og tilliten til offentlige tjenester.

Gjennom de overnevnte, og potensielt flere, prosesser kan tilgang på tungregning bidra til å forbedre den offentlige tjenesteytingen. Det det mener vi økt effektivitet og reduserte kostnader i tjenesteproduksjonen, forbedret kvalitet og tilpasning av tjenester til brukernes behov, raskere og mer presise beslutningsprosesser, økt brukertilfredshet gjennom mer responsive og tilgjengelige tjenester og forbedret ressursallokering basert på datadrevne innsikter.

Basert på de virkningsdrivende egenskaper ved tiltaket og årsak-virkningsforholdene som vist i Figur 9.14 og beskrevet over, har vi verdsett den samfunnsøkonomiske verdien av «forbedret øvrig offentlig tjenesteyting» i de utvalgte konseptene. Analysen er vist i Tabell 9.9.

Tabell 9.9 Verdsetting av virkningen «Forbedret øvrig offentlig tjenesteyting»

Konsept	Kvantum	Enhetsverdi	Samfunns- økonomisk verdi
K2, K3	<p><i>Hvor mange blir berørt av virkningen?</i></p> <p>Øvrig offentlig tjenesteyting, dvs. offentlig produserte tjenester utover samfunnskritiske tjenester, kan påvirke alle innbyggere i samfunnet. Antall personer som blir berørt er derfor vurdert til høy.</p> <p><i>Hvor stor er påvirkningen på hver enkelt berørt?</i></p> <p>Påvirkningen på de berørte er avhengig av offentlige virksomheter sin tilgang på tungregning for å tilpasse og levere øvrige offentlige tjenester.</p> <p>I både K2 og K3 får ikke offentlige virksomheter særskilte midler til å dekke sine behov for tungregneressurser (både kapasitet og kompetanse) til å tilpasse og levere øvrige offentlige tjenester. Dette omfatter både tilpasning og bruk av KI-modeller til å forbedre kvaliteten på tjenester og virksomhetsinterne prosesser, arbeidsmetoder osv. For å dekke sine behov for tungregning må den enkelte virksomhet prioritere å benytte tildelte midler til dette formålet. Ut fra kartleggingen i utredningen legger vi til grunn at mange virksomheter vil til en viss grad benytte midler på dette. Det vil gi noe økt kvalitet i tjenestene og forbedrede interne prosesser, som kan bidra til å forbedre øvrig offentlig tjenester. Den direkte påvirkning på den enkelte innbygger vil avhengig av hvilke tjenester de tar i bruk.</p> <p>Samlet vurderer vi kvantum i K2 til middels positivt.</p>	<p><i>Hva er verdien av virkningen?</i></p> <p>Innbyggerne sin villighet til å betale for øvrige offentlige tjenester er avhengig av hva det enkelte individ vurderer som viktig. I stort er grunn til å anta at befolkningen i stort er mest opptatt av de essensielle tjenestene for å opprettholde samfunnets grunnleggende funksjoner, dvs. samfunnskritiske tjenester. Samtidig tilbyr offentlige virksomheter tjenester som trolig har verdi for innbyggerne. F.eks. kulturtilbud, fritidsaktiviteter for barn og unge, omsorgstilbud for eldre, bibliotekstjenester og enkelte utdanningstilbud.</p> <p>Samlet vurderer vi enhetsverdien av øvrig offentlig tjenesteyting i K2 og K3 som middels.</p>	<p><i>Samlet vurdering av virkningen</i></p> <p>Liten positiv virkning i K2 og K3 basert på verdimatrisen.</p>

9.6.5 Vurdering av «Økt verdiskapning»

Som beskrevet i kapittel 5.2.5, har næringslivet, som forskningsorganisasjoner og offentlige virksomheter, behov for å håndtere komplekse beregninger og store datamengder til innovasjon og utvikling innenfor flere bransjer og områder. Et særskilt teknologisk satsingsområde for private virksomheter er *kunstig intelligens*. Næringslivet er opptatt av mulighetene for å ta KI i bruk og å bygge KI inn som funksjonalitet i sine produkter og tjenester. Økt bruk av KI-modeller i næringslivet, vil kreve tungregneressurser. KI-modeller må utvikles og trenes, som er særlig regnekraftintensive faser. De fleste private virksomheter vil derimot høyst sannsynlig ikke utvikle egne KI-modeller, men ta i bruk generiske KI-modeller eller tilpasse og justere utviklede KI-modeller til sitt formål. I disse prosessene vil behovet for tungregning være vesentlig lavere enn

i foregående stegene, hvor behovet vil variere etter modellenes kompleksitet, responstid, bruksområde m.m. De bedriftene som benytter seg av tungregnerressurser fra Sigma2 og EuroHPC i dag springer ofte ut av forskningsmiljøer eller er forskningsintensive.

Ved at flere bedrifter får tilgang til tungregnerressurser, kan det føre til at bedrifter som ikke har egne midler til å utvikle, trene, bruke og tilpasse egne KI-modeller, får utført dette. Dette kan føre til bedre kunnskapsgrunnlag og forbedrede/nye privatproduserte varer og tjenester. Noen eksempler inkluderer forbedret innsikt/beslutningsstøtte og identifisering/utvikling av nye varer og tjenester gjennom maskinlæring for dataanalyse, og redusert risiko og feil i vareproduksjon ved bruk av kunstig intelligens for å automatisere prosesser.

Forbedrede og/eller nye privatproduserte varer og tjenester kan videre bidra til økt verdiskapning gjennom nye arbeidsplasser, økte inntekter, økt skatteinngang og forbedret levestandard. For eksempel vil bedrifter som kan tilby forbedrede/nye varer og tjenester oppleve økt markedsandel og salg, som kan føre til økte inntekter. Økte inntekter for en bedrift betyr økte skatteinntekter for staten. I tillegg kan økte inntekter bety at en bedrift tar et strategisk valg hvor de ønsker å vokse, og ansetter dermed flere. Bedriften kan også øke lønn for ansatte for å tiltrekke kompetanse. Forbedrede/nye varer og tjenester kan øke levestandarden på flere måter, for eksempel innen helse (tilgang til nye medisiner/vaksiner, raskere og mer presise diagnoser), transport/logistikk (optimaliserte ruter, autonome fremkomstmidler) og mat (bedre metoder for produksjon av mat, forbedret matsikkerhet).

Basert på de virkningsdrivende egenskaper ved tiltaket og årsak-virkningsforholdene som vist i Figur 9.14 og beskrevet over, har vi verdsett den samfunnsøkonomiske verdien av «forbedret øvrig offentlig tjenesteyting» i de utvalgte konseptene. Analysen er vist i Tabell 9.10.

Tabell 9.10 Verdsetting av virkningen «Økt verdiskapning»

Konsept	Kvantum	Enhetsverdi	Samfunnsøkonomisk verdi
K2, K3	<p><i>Hvor mange blir berørt av virkningen?</i></p> <p>Økt verdiskapning som følge av privatproduserte varer og tjenester kan påvirke et helt samfunn, dvs. både enkeltindivider, private bedrifter og offentlige virksomheter ved at det kan bli skapt flere arbeidsplasser, økte inntekter, økte skatteinngang, forbedret levestandard m.m. Antall personer som blir berørt er derfor vurdert til høy.</p> <p><i>Hvor stor er påvirkningen på hver enkelt berørt?</i></p> <p>Påvirkningen på de berørte er avhengig av private virksomheter sin tilgang på tungregning til å utvikle, tilpasse, forbedre og levere produkter og tjenester.</p> <p>Virksomheter i privat sektor er i utgangspunktet ansvarlig for å dekke sine behov for tungregning i det private leverandørmarkedet. Dersom de ikke klarer dette, f.eks. ved at de blir nedprioritert, kan Sigma2 (K2) eller «Sigma3» (K3) tilby noe tungregnekapasitet, støtte- og kompetansetjenester innenfor statsstøtteregulverket. Vi legger vi til grunn at de fleste private virksomheter vil klare å dekke sine behov på egenhånd, men at særlig</p>	<p><i>Hva er verdien av virkningen?</i></p> <p>Innbyggerne sin villighet til å betale for privatproduserte varer og tjenester er avhengig av hva det enkelte individ vurderer som viktig. Fra et samfunnsperspektiv vil varer og tjenester produsert av næringslivet ha stor betydning ettersom privat sektor produserer hovedvekten av varer og tjenester i et samfunn. Et velfungerende næringsliv er derfor viktig for et samfunn.</p> <p>Samlet vurderer vi enhetsverdien av økt verdiskapning i K2 og K3 som middels.</p>	<p><i>Samlet vurdering av virkningen</i></p> <p>Liten positiv virkning i K2 og K3 basert på verdimatrisen.</p>

SMB-segmentet og svært sektorspesifikke virksomheter med høy grad av FoU vil etterspørre disse tjenestene. Ved å få tilgang til tungregneinfrastruktur vil private virksomheter kunne økt kvaliteten og utvikle nye produkter og tjenester, som vil gi økt verdiskapning for samfunnet. Den direkte påvirkning på den enkelte innbygger vil avhengig av hvilke varer og tjenester de går til innkjøp av.

Samlet vurderer vi kvantum i K2 og K3 til **middels positivt**.

9.6.6 Vurdering av «Negative miljøkonsekvenser»

Miljø er en sammensetning av alle mulige forhold som eksisterer på et sted. Dette kan inkludere klima, topografi, naturmangfoldet, hvordan det er å bo et sted og alt annet som inngår i omgivelsene.

Utbygging og bruk av en infrastruktur for tungregning vil få negative miljøkonsekvenser. Den største påvirkningen skjer gjennom økt energiforbruk og økt behov for fysisk plass. Særlig er strøm en viktig faktor for bruk av tungregning. I Norge er strømproduksjonen hovedsakelig basert på fornybar energi med vannkraft som den dominerende kilden, supplert av vindkraft og solenergi. Strømproduksjonen i Norge kan likevel ha negative miljøkonsekvenser som påvirkning på økosystemer og biologisk mangfold fra vannkraftutbygging, samt støyrelaterte konsekvenser fra vindkraftanlegg.

I tillegg krever strømproduksjon og utbygging av datasentre m.m. en fysisk infrastruktur. Dette kan gå på bekostning av naturareal med verdifulle økosystemer og biomangfold. Fysisk infrastruktur medfører også miljøutslipp ved utbygging og vedlikehold.

Basert på de virkningsdrivende egenskaper ved tiltaket og årsak-virkningsforholdene som vist i Figur 9.14 og beskrevet over, har vi verdsett den samfunnsøkonomiske verdien av «forbedret samfunnskritiske tjenester» i de utvalgte konseptene. Analysen er vist i

Tabell 9.11.

Tabell 9.11 Verdsetting av virkningen «Negative miljøkonsekvenser»

Konsept	Kvantum	Enhetsverdi	Samfunns- økonomisk verdi
K2, K3	<p><i>Hvor mange blir berørt av virkningen?</i></p> <p>Alle innbyggere i Norge blir berørt av klima og miljø og således av klimaendringer. Antall personer som blir berørt er derfor vurdert til høy.</p> <p><i>Hvor stor er påvirkningen på hver enkelt berørt?</i></p> <p>Utbygging og bruk av en infrastruktur for tungregning vil få negative miljøkonsekvenser. Den største påvirkningen skjer gjennom innsatsfaktorene strøm og fysisk utbygging av datasentre. Særlig er strøm en viktig faktor for bruk av tungregning. I Norge er strømproduksjonen hovedsakelig basert på fornybar energi med vannkraft som den dominerende kilden, supplert av vindkraft og solenergi. Strømproduksjonen i Norge kan likevel ha negative miljøkonsekvenser som påvirkning på økosystemer og biologisk mangfold fra vannkraftutbygging, samt støyrelaterte konsekvenser fra vindkraftanlegg. I tillegg krever strømproduksjon og utbygging av datasentre m.m. en fysisk infrastruktur som medfører miljøutslipp ved utbygging og vedlikehold. Alle innbyggere i Norge vil kollektivt bli berørt av disse klima- og miljøendringene, samtidig som påvirkningen på den enkelte vil være begrenset i omfang. Det vil ta tid før virkningene gir utslag, men de vil til gjengjeld vare over lengre tid.</p> <p>Samlet vurderer vi kvantum i K2 og K3 til middels positivt.</p>	<p><i>Hva er verdien av virkningen?</i></p> <p>Klima- og miljømessig bærekraft er viktig for samfunnet og for verden samlet, og klima og miljø har direkte og indirekte innvirkning på alle innbyggere i Norge. Klima- og miljømessig bærekraft er sentral i FNs bærekraftsmål, og i Norge skal all statlig planlegging rette seg etter de globale bærekraftsmålene.</p> <p>Samlet vurderer vi enhetsverdien av negative miljøkonsekvenser i K2 og K3 som høy.</p>	<p><i>Samlet vurdering av virkningen</i></p> <p>Middels negativ virkning i K2 og K3 basert på verdimatrisen.</p>

9.6.7 Oppsummering av ikke-prissatte virkninger

I den samfunnsøkonomiske analysen har vi identifisert og vurdert seks ikke-prissatte virkninger. Tabell 9.12 gir en oversikt over verdsettingen av disse virkningene ved bruk av verdimatrisemetoden.

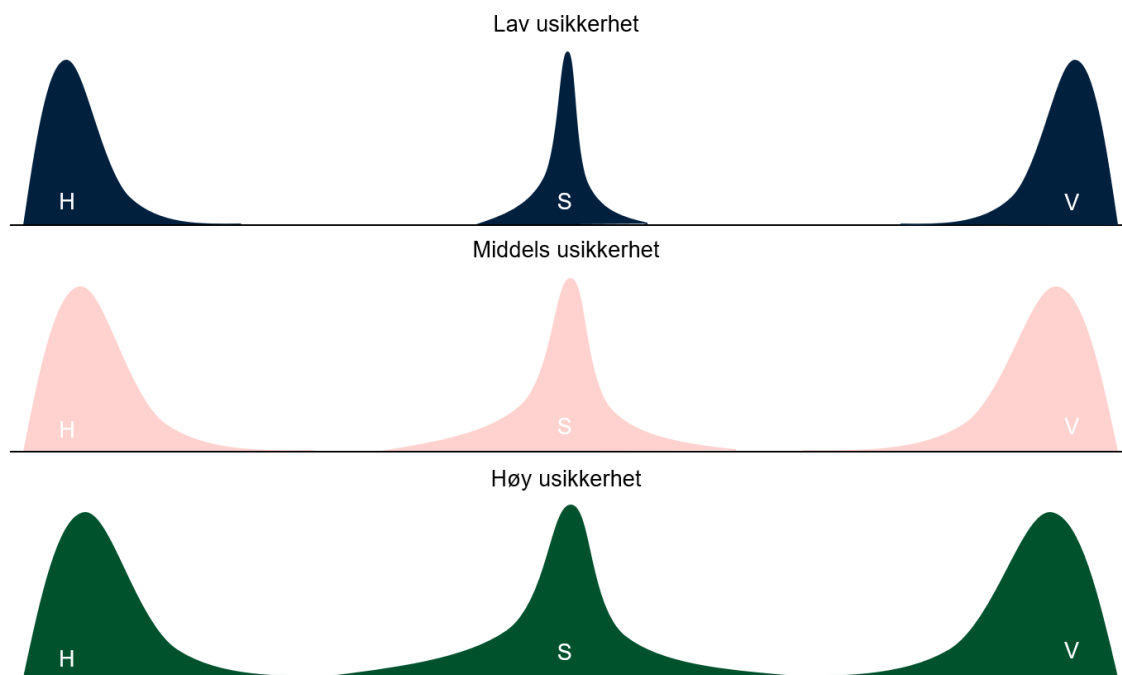
Tabell 9.12 Oversikt over verdsetting av ikke-prissatte virkninger

Virkning	KS «Sigma2 pluss»	K3 «Tverrsektoriell Sigma3»
Mer forskning i internasjonal forskningsfront	Middels positiv	Middels positiv
Økt innovasjon og konkurransekraft	Meget stor positiv	Meget stor positiv
Forbedret samfunnskritiske tjenester	Stor positiv	Meget stor positiv

Forbedret øvrig offentlig tjenesteyting	Liten positiv	Liten positiv
Økt verdiskaping	Liten positiv	Liten positiv
Miljøkonsekvenser	Middels negativ	Middels negativ

9.7 Usikkerhetsanalyse


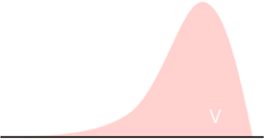

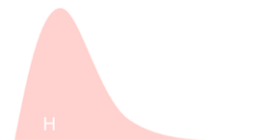
Vi har gjennomført en kvalitativ usikkerhetsanalyse for samfunnsøkonomiske virkningene ved å beskrive de viktigste usikkerhetsfaktorene for hver virkning. Usikkerheten vurderes på en trepunktsskala fra lav til høy usikkerhet, som vist i Figur 9.17. Lav usikkerhet (grønt) betyr at virkningens estimerte verdi vurderes som relativt sikker. Differansen mellom en optimistisk og pessimistisk verdi vil dermed være mindre, og nærmere den forventede estimerte verdien. Høy usikkerhet (lilla) betyr lav sannsynlighet for at forventet verdi er riktig, og differansen mellom en pessimistisk og optimistisk verdi vil være stor med et intervall langt unna den forventede estimerte verdien. Det er dermed et stort spenn av mulige verdier for virkningen. Middels usikkerhet brukes når virkningen vurderes til en usikkerhet mellom lav og høy. Videre skiller usikkerhetsanalysen mellom ulik grad av skjevhet, som illustrert ved de ulike formene i figuren. «H» står for høyreskjev, «S» for symmetrisk og «V» for venstreskjev. Høyreskjev vil si at usikkerheten er sterkere knyttet til optimistisk verdi. Venstreskjev vil si at usikkerheten er sterkere knyttet til pessimistisk verdi. Symmetrisk vil si at usikkerheten er likt knyttet til pessimistisk og optimistisk verdi.



Figur 9.17 Oversikt over ulik grad av usikkerhet og skjevhet for gjennomført usikkerhetsanalyse

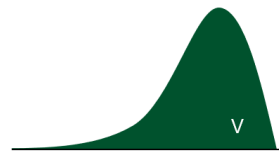
Tabell 9.13 fremstiller den kvalitative usikkerhetsanalysen for den prissatte virkningen «Økte offentlige utgifter». Analysen er delt i komponenter av kostnadsberegningene.

Tabell 9.13 Usikkerhetsanalyse av prissatt virkning «Økte offentlige utgifter»

Komponent	Beskrivelse av usikkerhet	Illustrasjon
Investering Sigma2	<p>Usikkerheten knyttet til investeringene i Sigma2 er høy og venstreskjev. Dette skyldes at kostnadene er basert på erfaringstall, men kan påvirkes av prisøkninger og etterspørsel etter teknologi. Det er sannsynlig at kostnadene kan bli høyere enn estimert, spesielt hvis det oppstår forsinkelser eller behov for tilleggsinvesteringer. Teknologisk utvikling og endringer i markedet kan også påvirke kostnadene, noe som gjør det viktig å ha fleksibilitet i budsjettet.</p>	
Drift Sigma2	<p>Usikkerheten for driftskostnadene i Sigma2 er middels og venstreskjev. Estimaterne er basert på prognoser som er hentet direkte fra Sigma2 sitt langtidsbudsjett, og baserer seg på dagens driftsmodell., men det er usikkerhet knyttet til om dimensjoneringen vil være tilstrekkelig for å håndtere tildeling av kapasitet utenfor forskningssektoren. Kostnadene kan øke hvis det kreves mer avansert brukerstøtte eller hvis det oppstår uforutsette driftsutfordringer. Det er viktig å overvåke driftskostnadene nøye og justere ressursallokeringen etter behov.</p>	
GPU-kapasitet til KI	<p>Usikkerheten for GPU-kapasitet til KI er høy og venstreskjev. Behovet for tungregnekapasitet til KI-modeller er vanskelig å forutsi, med estimater som varierer betydelig. Estimater fra ledende KI miljøer i Norge er i størrelsesorden 10-100 ganger kapasiteten som er tilgjengelig i dag. Som prissatt virkning er det tatt utgangspunkt et initelt behov på 2500 GPUer. Det er sannsynlig at behovet kan være mye høyere enn forventet, noe som kan føre til økte kostnader. Teknologiske fremskritt og endringer i KI-forskning kan også påvirke behovet for GPU-kapasitet, noe som krever fleksibilitet i planleggingen.</p>	
Investering gjennom EuroHPC/LUMI	<p>Vurderes til middels og høyreskjev usikkerhet. Sigma2 estimerer et investeringsnivå for den neste KI-relaterte LUMI-maskinen som utgjør 10-20 prosent av det nasjonale investeringsnivået på tungregning, for å få ønsket andel kapasitet. Det er knyttet usikkerhet til kvanteberegninger, men anbefales å reservere opptil 10 MNOK for tilgang på ressurser. I tillegg er det behov for medfinansiering for å utnytte muligheter som er i DIGITAL og EuroHPC, som estimeres å være 58,8 MNOK til 80 MNOK i perioden 2025-2028. Som prissatt virkning er 80 MNOK per år lagt inn i kostnadsanalysen. Det er vurdert som sannsynlig at kostnadene kan være lavere enn forventet hvis samarbeidet gir stordriftsfordeler.</p>	

Det er viktig å følge med på utviklingen i EuroHPC-samarbeidet og justere investeringene basert på tilgjengelig kapasitet og behov.

Annen sektorfinansiering Usikkerheten for annen sektorfinansiering er høy og venstreskjev. Det er utfordrende å estimere kostnadene for sektorspesifikke prosjekter, og behovet for tungregning kan variere betydelig mellom sektorer. I kostnadsanalysen er det dermed lagt et fokus på eksempler på tungregneprosjekter fra utvalgte sektorer i K2 og K3. Kapasitet som andre sektorer har behov for til tungregning er basert på nytt initielt behov på 2500 GPUer i K2 og K3, noe som beskrevet i kostnadselementet «GPU-kapasitet til KI» er sannsynlig et lavt estimat. Kostnadene kan bli høyere enn forventet hvis det oppstår uforutsette behov eller hvis sektorer må konkurrere om begrensede ressurser. Det er viktig å ha en fleksibel finansieringsmodell som kan tilpasses endrede behov.



Tabell 9.14 fremstiller den kvalitative usikkerhetsanalysen for den prissatte virkningen «Økte offentlige utgifter». Analysen er delt i komponenter av kostnadsberegningene.

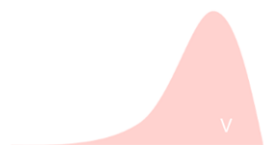
Tabell 9.14 Usikkerhetsanalyse av ikke-prissatte virkninger

Virkning	Beskrivelse av usikkerhet	Illustrasjon
Mer forskning i internasjonal forskningsfront	Usikkerheten er vurdert til middels og symmetrisk. Tilgang til tungregning er avgjørende for å fremme banebrytende forskning, men realiseringen avhenger også av forskernes evne til å utnytte disse ressursene effektivt. Usikkerheten er symmetrisk fordi både optimistiske og pessimistiske utfall er like sannsynlige, avhengig av hvordan ressursene allokeres og blir anvendt. Videre vil internasjonalt samarbeid og deling av kunnskap være viktige faktorer for å maksimere forskningsutbyttet.	
Økt innovasjon og konkurransekraft	Usikkerheten her er vurdert til middels og høyreskjev. Forskningsorganisasjonenes tilgang til tungregning kan potensielt stimulere til betydelig innovasjon, men dette avhenger av flere faktorer, inkludert samarbeid mellom forskningsmiljøer, næringsliv og offentlig forvaltning. Evnen til å omsette forskningsresultater til praktiske løsninger er avgjørende. Vi vurderer at det er større sannsynlighet for optimistiske utfall, spesielt hvis det legges til rette for effektiv utnyttelse av ressursene og gode samarbeidsmodeller. Innovasjon og konkurransekraft kan også påvirkes av	

politiske beslutninger og økonomiske insentiver som fremmer teknologisk utvikling.

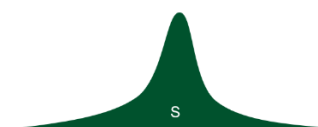
Forbedret samfunnskritiske tjenester

Usikkerheten er vurdert til middels og venstreskjev. Offentlige virksomheter vil kunne forbedre samfunnskritiske tjenester med økt tilgang til tungregning. Utfordringer kan imidlertid oppstå i implementeringsfasen, spesielt når det gjelder integrasjon av ny teknologi i eksisterende systemer og utnyttelse av teknologien. Risikoen ligger derfor mer i pessimistiske utfall, mens den generelle forventningen er positiv. Effektiv opplæring og tilpasning av teknologi til spesifikke behov og tilgang på rett kompetanse vil være avgjørende for å realisere potensialet.



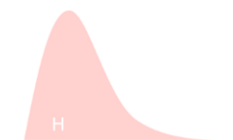
Forbedret øvrig offentlig tjenesteyting

Usikkerheten er høy og symmetrisk. Selv om tilgang på tungregning kan forbedre øvrige offentlige tjenester, avhenger det av hvordan ressursene blir prioritert og anvendt. Det er nødvendig med tydelige strategier og gjennomføringsevne for å sikre at teknologien brukes effektivt. Det er like sannsynlig med både positive og negative utfall, avhengig av implementeringsstrategier og ressursallokering. Forbedringer i tjenesteyting kan også kreve endringer i organisasjonskultur og prosesser, som kan være ressurs- og tidkrevende.



Økt verdiskaping

Usikkerheten er vurdert til middels og høyreskjev. Mens tungregning kan føre til økt verdiskaping, er dette avhengig av næringslivets evne til å integrere teknologi i sine prosesser og utvikle nye produkter og tjenester. Vi vurderer at det er større potensial for positive utfall hvis næringslivet klarer å utnytte teknologien effektivt og tilpasse seg raskt til endringer i markedet. Innovasjon i næringslivet kan også påvirkes av tilgang til kapital og kompetanse.



Miljøkonsekvenser

Usikkerheten er vurdert til høy og venstreskjev. Miljøkonsekvensene av økt tungregning er usikre, spesielt med tanke på energiforbruk og ressursbruk. Selv om Norge har tilgang til fornybar energi, kan økt etterspørsel etter strøm og utbygging av datasentre ha negative miljøpåvirkninger. Det er større sannsynlighet for negative utfall, spesielt hvis bærekraftige løsninger ikke implementeres effektivt. Langsiktig planlegging og investering i grønn teknologi vil være viktig for å minimere miljøpåvirkningen.



9.8 Realopsjoner

Verdien av realopsjoner består i de muligheter for fleksibilitet som ligger i konseptene hvis ny informasjon kommer til etter beslutningstidspunktet for tiltaket. Et eksempel på ny informasjon er at behovet for tungregnekapasitet endrer seg.

Et konsept som er mer fleksibelt enn et annet vil ofte være å foretrekke, og det finnes flere typer fleksibilitet. Fleksibilitet som vurderes er gjennomføringstidspunkt og tiltaksomfang.

Fleksibilitet i gjennomføringstidspunkt

Ved å utsette beslutningen for eventuell investering i tungregneinfrastruktur kan man tenke seg at ny informasjon kan utnyttes som vil påvirke valg av investering og organisering. Samtidig viser utfordringsbildet (kapittel 4) og behovene (kapittel 5) at det er allerede nå et betydelig behov for økt tungregnekapasitet, kompetanse m.m. Dermed vil en utsettelse av gjennomføringstidspunkt føre til at man ikke får nytten av raskere økt kapasitet og kompetanse ved en utbygging av en nasjonal infrastruktur for tungregning. Når det gjelder teknologi, blant annet innenfor kunstig intelligens, skjer utviklingen i et meget raskt tempo. Dersom man utsetter en eventuell investering i tungregneinfrastruktur, vil man således har mer teknologisk innsikt som man kan basere utbyggingen på. Samtidig er det ingenting per nå som tilsier at den teknologiske utviklingen vil stagnere og at man på et tidspunkt kan bedre forutsi hvordan teknologien vil se ut i fremtiden. Enhver investering i digital infrastruktur må heller etterstrebe å ha fleksibilitet og fremtidig handlingsrom for teknologiske nyvinninger når investeringer foretas.

Samlet sett er det indikasjoner på at det ikke vil lønne seg å utsette en beslutning om investeringer i tungregning. Ettersom K2 og K3 legger opp til en investering i tungregneinfrastruktur med økt tungregnekapasitet fra 2026, vurderer vi at begge konseptene har lav realopsjon mht. gjennomføringstidspunkt.

Fleksibilitet i tiltaksomfang

K2 gir en del fleksibilitet ved å tillate sektorer å velge hvordan de vil dekke sine behov for tungregning, enten gjennom Sigma2, egne løsninger, eller ved å benytte kommersielle aktører. Denne tilnærmingen gir sektorer mulighet til å tilpasse investeringene etter hvert som behovene endrer seg, og gir rom for å justere omfanget basert på spesifikke krav. Fleksibiliteten gjør det mulig å reagere raskt på teknologiske endringer og varierende etterspørsel. Selv om sektorer kan tilpasse seg endringer, vurderes realopsjonen vurderes som middels siden det fortsatt kreves koordinering og ressurser for å implementere større justeringer.

K3 gir noe fleksibilitet gjennom en stegvis utvikling. I første steg fokuseres det på å dekke behovet for tungregning til utvikling og testing av språkmodeller og generative nasjonale KI-modeller. I steg to gjennomføres sektorvise kost/nytte-vurderinger for å bestemme om sektorer skal få dekket sine behov hos «Sigma3» eller gjennom andre løsninger. Denne tilnærmingen gir mulighet til å tilpasse omfanget basert på spesifikke sektorbehov og endringer over tid. Til tross for denne fleksibiliteten, vurderes realopsjonen til lav fordi den sentraliserte strukturen kan begrense muligheten til å raskt justere kapasiteten i stor skala. Endringer krever omfattende koordinering og kan være tidkrevende, noe som reduserer den totale fleksibiliteten i tiltaksomfanget.

9.9 Fordelingsvirkninger

Selv om den samfunnsøkonomiske analysen viser at et tiltak totalt sett er lønnsomt for samfunnet, kan enkelte grupper komme dårligere ut som følge av tiltaket. Det er ofte slik at noen grupper kommer bedre ut og noen grupper kommer dårligere ut på at et offentlig tiltak iverksettes. Hvordan nytte- og kostnadsvirkningene fordeler seg mellom ulike grupper i samfunnet, kalles i en samfunnsøkonomisk analyse for en «fordelingsvirkning» og skal ikke inngå som en samfunnsøkonomisk verdi.

Realiseringen av nullalternativet, K2 eller K3 har fordelings effekter knyttet til at tiltakene krever skattefinansiering. Årsaken er at noen betaler mer skatt enn andre, samtidig som de som

betaler mest skatt ikke nødvendigvis er de som har størst nytte av de kollektive godene som samfunnet tilbyr. Skattefinansiering er imidlertid en akseptert form for finansiering av kollektive goder.

Det er ikke identifisert noen ytterligere fordelingsvirkninger av tiltaket.

9.10 Gjennomføringsrisiko

Gjennomføringsrisiko beskriver ulike risikoer som er knyttet til realisering av et konsept, og varierer mellom konseptene som inngår i alternativanalysen.

Vurdering av K2:

- **Koordinering og samarbeid:** K2 krever at sektorer selv finansierer og organiserer sine behov for tungregning. Dette kan føre til varierende grad av samarbeid og koordinering mellom sektorer, noe som kan skape utfordringer i å sikre en helhetlig tilnærming.
- **Eksisterende strukturer:** Konseptet bygger videre på eksisterende strukturer i Sigma2, noe som reduserer risikoen ved å unngå store organisatoriske endringer. Dette gir en viss stabilitet og forutsigbarhet.
- **Fleksibilitet:** Sektorene har frihet til å velge hvordan de vil dekke sine behov, enten gjennom Sigma2 (om det er kapasitet), egne løsninger eller kommersielle aktører. Denne fleksibiliteten kan imidlertid også føre til fragmentering og ineffektiv ressursbruk hvis den ikke er godt nok koordinert.
- **Usikkerhet i prioriteringer:** Det kan være utfordrende å sikre at alle sektorer prioriterer investeringer i tungregning, spesielt hvis de har andre presserende behov.

Vurdering av K3:

- **Omfattende omstrukturering:** K3 innebærer en betydelig omstrukturering av Sigma2 til en tverrsektoriell leverandør «Sigma3». Dette krever omfattende endringer i organisasjon og styringsmodeller, noe som kan være kompleks og tidkrevende.
- **Tverrsektoriell koordinering:** Å dekke behovene til flere sektorer krever effektiv koordinering og samarbeid på tvers av ulike offentlige aktører. Dette kan være utfordrende, spesielt når det gjelder å balansere ulike sektors behov og prioriteringer.
- **Endringsledelse:** Implementeringen krever endringsledelse for å sikre at alle involverte parter er engasjert og forplikter seg til den nye strukturen. Dette inkluderer å bygge tillit og sikre at sektorer som skal få tjenester av leverandøren blir ivaretatt.
- **Teknologisk kompleksitet:** Å integrere ulike teknologiske plattformer og sikre at de fungerer sømløst på tvers av sektorer kan være teknisk utfordrende og kreve spesialisert kompetanse.
- **Erfaringer fra lignende prosjekter:** Historisk sett har tverrsektorielle digitaliseringsprosjekter i offentlig sektor ofte møtt på utfordringer knyttet til kompleksitet og gjennomføring, noe som øker risikoen for forsinkelser og kostnadsoverskridelser.

Samlet sett vurderer vi at K3 mer ambisiøs og kompleks enn K2, noe som gir høyere gjennomføringsrisiko. Leverandøren skal levere tungregnetjenester til flere sektorer enn i K2, som innebærer at K3 er et større tverrsektorielt digitaliseringsprosjekt. Erfaring tilsier at slike prosjekter kan være utfordrende å gjennomføre i offentlig sektor. I K2 er gjennomføringsrisikoen vurdert til middels og i K3 er gjennomføringsrisikoen vurdert til stor.

10 Anbefalt konsept

Sammenstilling av samfunnsøkonomisk analyse

Vi har gjennomført en samfunnsøkonomisk analyse av de utvalgte konseptene K2 «Sigma2 pluss» og K3 «Tverrsektoriell Sigma3». Det vil si at vi har vurdert kostnads- og nyttevirkninger ved utbyggingen av en nasjonal infrastruktur for tungregning, samt en kvalitativ vurdering av gjennomføringsrisiko og realopsjoner ved konseptene. Basert på en samlet vurdering av de samfunnsøkonomiske virkningene, gjennomføringsrisiko og realopsjoner er K3 «Tverrsektoriell Sigma3» rangert best av konseptene, jf. Tabell 10.1.

Tabell 10.1 Resultater fra samfunnsøkonomisk analyse av utvalgte konsepter

	Null- alternativet	K2 «Sigma2 pluss»	K3 «Tverrsektoriell Sigma3»
Prissatte virkninger (MNOK, netto nåverdi)			
Investeringskostnad	-170	-2 050	-2 310
Driftskostnad	-610	-1 330	-840
Skattefinansieringskostnad	-160	-680	-640
Sum prissatte virkninger	-940	-4 050	-3 780
Ikke prissatte virkninger			
Mer forskning i internasjonal forskningsfront	-	Middels positiv	Middels positiv
Økt innovasjon og konkurransekraft	-	Meget stor positiv	Meget stor positiv
Forbedret samfunnskritiske tjenester	-	Stor positiv	Meget stor positiv
Forbedret øvrig offentlig tjenesteyting	-	Liten positiv	Liten positiv
Økt verdiskapning	-	Liten positiv	Liten positiv
Negative miljøkonsekvenser	-	Middels negativ	Middels negativ
Risiko og realopsjoner			
Gjennomføringsrisiko	-	Middels	Høy
Realopsjoner	-	Middels	Lav
Rangering basert på samlet vurdering		2	1

Anbefalt konsept for nasjonal infrastruktur for tungregning

Basert på den samfunnsøkonomiske analysen i utredningen, anbefaler vi at man på sikt beveger seg mot konsept K3 «Tverrsektoriell Sigma3» som et fremtidig mål bilde for nasjonal tungregneinfrastruktur.

I konsept K3 vil både behovene for tungregning innen forskningssektoren og sentrale deler av offentlig sektor dekkes av én leverandør. Leverandøren er en stegvis videreutvikling av dagens Sigma2 slik at leverandøren kan å bli en tverrsektoriell leverandør av tungregning («Sigma3»). Leverandøren vil dekke behovene for tungregning til forskningsdrevet og forvaltningsdrevet FoU, samt tungregning for utvikling, tilpasning og forbedringer av samfunnskritiske tjenester. I tillegg kan leverandøren levere noe tungregnekapasitet, støtte- og kompetansetjenester til næringslivet innenfor statsstøttereguleringen. Samtidig må hver offentlig virksomhet selv skaffe til

veie tungregning til å drifte samfunnskritiske tjenester (herunder 24/7-oppetid) og tungregning til bruk og tilpasning av øvrige offentlige tjenester, inkludert kunstig intelligens. Leverandør står fritt til å inngå og endre avtale om kjøp av kapasitet hos EuroHPC og andre samarbeidspartnere, og til å utforske muligheten til å inngå samarbeid med det private leverandørmarkedet for tungregning.

Den nye leverandøren vil motta finansiering til investeringer og grunnleggende driftsstøtte på vegne av departementsfellesskapet som koordineres gjennom Kunnskapsdepartementet, mens bruk av tungregning er helt eller delvis brukerfinansiert. Vi legger til grunn at differensierte modeller for brukerfinansiering må utvikles. Midler til bruken av tungregning i leveranse av samfunnskritiske tjenester kanaliseres til den enkelte forvaltningsområde i årlige budsjettprosesser basert på tversektorielle vurderinger av fordeling av midler.

Nyttevirkninger ved implementering av anbefalt konsept

Det anbefalte konseptet K3 kan bidra til en bærekraftig samfunnsutvikling ved å fremme datadrevet innovasjon og effektivisering i både forskning, offentlig og privat sektor, og bidra til at Norge blir bedre godt rustet til å møte fremtidens teknologiske og samfunnsmessige utfordringer.

K3 skal gi betydelige samfunnsøkonomiske gevinster ved å tilby en robust og fleksibel infrastruktur for tungregning som dekker både forsknings- og forvaltningsdrevet FoU. Ved å sikre forskningsmiljøer tilgang til nødvendig regnekraft, kan Norge styrke sin posisjon som en ledende kunnskapsnasjon. Dette er avgjørende for å flytte forskningsfronten, fremme innovasjon og tiltrekke seg internasjonale talenter. Forskningsmiljøene vil kunne utføre mer avanserte simuleringer og analyser, noe som vil bidra til å utvikle banebrytende løsninger på komplekse samfunnsutfordringer. Økt tilgang til tungregning kan også muliggjøre mer tverrfaglig forskning og samarbeid, noe som kan føre til nye oppdagelser og teknologiske fremskritt. Dette vil ikke bare styrke Norges konkurransevne, men også bidra til å løse globale utfordringer innen områder som klima, helse og energi.

Offentlige virksomheter vil dra nytte av denne infrastrukturen ved å kunne utvikle og forbedre både samfunnskritiske og øvrige tjenester. For samfunnskritiske tjenester kan det innebære økt sikkerhet og effektivitet i tjenesteleveransen, inkludert bedre beredskap, optimal ressursallokering og raskere respons i krisesituasjoner. Innenfor andre offentlige tjenester kan tungregning muliggjøre mer presise analyser og bedre beslutningsstøtte, noe som kan føre til mer effektiv forvaltning og høyere kvalitet på tjenestene. Dette inkluderer alt fra automatisering av administrative prosesser til personalisering av tjenester for innbyggerne. Ved å utnytte tungregning kan offentlige virksomheter også forbedre koordineringen mellom ulike sektorer, noe som gir en mer helhetlig og effektiv offentlig tjenesteyting.

Videre vil konseptet legge til rette for økt verdiskaping i næringslivet ved å tilby støtte- og kompetansetjenester innenfor statsstøttereguleringen. Dette kan stimulere til utvikling av nye teknologier og tjenester, og gi små og mellomstore bedrifter muligheten til å konkurrere på et globalt nivå.

K3 kan også fremme samarbeid på tvers av forskning, offentlig forvaltning og næringslivet, noe som gir bedre ressursutnyttelse og styrker Norges teknologiske suverenitet. Ved å integrere data fra ulike kilder kan offentlige og private aktører få et mer helhetlig bilde av utfordringer og muligheter, noe som forbedrer beslutningsgrunnlaget og effektiviteten i tjenesteleveranser og produktutvikling.

Kostnader ved implementering av anbefalt konsept

Konseptet innebærer en total investering på 3,4 mrd. kroner i perioden 2026 til 2030. Dette inkluderer et årlig driftsbudsjett mellom 150 og 185 mill. kroner, beregnet til totalt 850 mill. kroner i analyseperioden. I tillegg til investerings- og driftskostnader vil realiseringen av konseptet ha negative miljøkonsekvenser, særlig i form av økt strømforbruk og utbygging av fysisk infrastruktur.

Vi vil fremheve at det er svært stor usikkerhet knyttet til kostnadsberegningene og forutsetningene som ligger til grunn, herunder behovet for tungregning (både GPUer og CPUer). Usikkerheten er vurdert til å være høy og «venstreskjev», dvs. det kan potensielt bli vesentlig høyere kostnader knyttet til realiseringen av alle konsepter i analysen.

Gjennomføringsrisiko ved implementering av anbefalt konsept

Vi har vurdert gjennomføringsrisikoen i K3 som høy. Årsaken er at leverandøren skal levere tungregnetjenester til flere sektorer, som innebærer at dette er et større tverrsektorielt digitaliseringsprosjekt i offentlig sektor. Erfaringer fra flere andre utviklingsprosjekter tilsier at dette kan være utfordrende. Dette underbygger behovet for en tverrsektoriell tilnærming i de kommende fasene av prosjektet.

Det er avgjørende å sikre at alle brukergrupper over tid får dekket sine behov. En videreutvikling av Sigma2 til «tverrsektoriell Sigma3» hvor tungregnebehovet til flere sektorer dekkes, må ikke gå på bekostning av behovene til forskningsorganisasjonene som får dekket sine behov gjennom Sigma2 i dag. Behovene er ulike mellom brukergruppene og de vil kunne endre seg over tid. Det er derfor en kritisk suksessfaktor at risikoreducerende tiltak identifiseres og adresseres i forprosjektfasen.

Utvikling av leverandøren i anbefalt konsept

Infrastrukturen for tungregning skal så langt som mulig bygge på eksisterende systemer og strukturer. Dette er sentralt for å sikre at nasjonal kompetanse innenfor tungregning ikke forvirrer, men utnyttes på en god måte, og at man kan redusere behovet for nye investeringer og minimere implementeringstiden som bidrar til en kostnadseffektiv utbygging av en nasjonal tungregneinfrastruktur. I konsept K3 innebærer dette en utvidelse og omstrukturering av Sigma2, slik at leverandøren kan bli en tverrsektoriell leverandør av tungregning. Siden en konseptvalgutredning er en strategisk utredning som setter retning for framtiden, må organiseringen og struktureringen av leverandøren detaljeres i forprosjektet og dokumenteres i det sentrale styringsdokumentet (SSD).

Stegvis utvikling i leveranse av tungregnetjenester i anbefalt konsept

Utviklingen av leverandøren for tungregning må skje gradvis over tid for å redusere gjennomføringsrisiko og ta hensyn til den raske teknologiske utviklingen innenfor tungregning og kunstig intelligens. Den raske utviklingen gjør at behovene for tungregning fort kan endres, og en gradvis utbygging over tid vil sikre at man benytter seg av siste teknologiske nyvinninger og oppdatert kunnskap om behovene. Det kan bidra til å redusere usikkerhet og hindre feilinvesteringer i infrastruktur. Det er også et behov for å gå opp grenseskiller mot andre utviklingsløp, herunder behandling av gradert informasjon og håndtering av sensitive data.

Overordnet har vi vurdert følgende steg i utviklingen av tjenesteleveranse til leverandøren:

- Steg 1: Leverandøren må først dekke behovet for tungregning til utvikling og testing av språkmodeller og generative nasjonale KI-modeller.
- Steg 2: Gjennomføre kost/nytte-vurderinger for hvert forvaltningsområde og ta stilling til om det enkelte område skal få dekket sine behov for tungregning hos leverandøren eller gjennom andre løsninger/leverandører. Helsesektoren, som er en omfattende og kompleks sektor med særskilte krav til personvern, vil kreve en særskilt utredning.

Disse stegene må detaljeres ytterligere i forprosjektfasen av prosjektet.

11 Føringer for forprosjekt

Det anbefalte konseptet representerer en utviklingsretning og består av en rekke tiltak som samlet sett skal bringe oss mot et ønsket målbilde. Forprosjektet skal etablere et styringsdokument som beskriver hvordan Forskningsrådet og samarbeidspartnere, i og utenfor kunnskapssektoren, skal arbeide sammen for å realisere valgt konsept.

Dette kapitlet peker fremover mot neste fase som er planlegging og gjennomføring av anbefalt konsept (K3). Beskrivelsen legger til grunn at anbefalt konsept vedtas av regjeringen. Kapitlet beskriver særlig viktige utvalgte temaer som bør ivaretas i planleggings- og gjennomføringsfasen:

- Videreføring av utredningsprosjektet
- Planlegging og gjennomføring
- Prosjektspesifikke suksessfaktorer og fallgruver
- Risikoreduserende tiltak

11.1 Videreføring av utredningsprosjektet

Arbeidet til nå har vært ledet av Forskningsrådet, med Kunnskapsdepartementet som oppdragsgiver. Arbeidet har bygget videre på rapporten som Forskningsrådet leverte i august 2024⁵⁹ og har involvert ressurser fra instituttsektoren, universitetene og fra Sigma2. I tillegg er det gjennomført en rekke intervjuer med direktorater og utvalgte aktører i næringslivet.

I oppdragsbrevet fra Kunnskapsdepartementet fremgår det at arbeidet med forprosjektfasen skal startes opp før KS1 er gjennomført med tanke på en ferdig gjennomført KS1 til april 2023. Sentralt styringsdokumentet skal foreligge primo mai. Det innebærer at eksisterende prosjektorganisasjon bør videreføres, både for å sikre god samhandling med ekstern kvalitetssikrer og for å kunne utarbeide sentralt styringsdokument.

KVU-arbeidet har vist at det er av stor betydning å arbeide videre med:

- **Fireårig langtidspan for infrastruktur.** Igangsette arbeidet med å utarbeide en fireårig langtidspan for infrastruktur. I arbeidet med konseptvalgutredningen har behovet for tungregning blitt kartlagt. Imidlertid er det fortsatt knyttet sto usikkerhet til kapasitetsbehovet for GPU-er til KI-utvikling og trening. Gjennom forprosjektet skal det utarbeides en modell og en prosess for å kartlegge fremtidige behov, samt utarbeides terskelverdier for når det skal investeres i økt kapasitet. Langtidsplanen blir en sentral del av virksomhetsstyringen av «Sigma3» og bør oppdateres hvert annet år.
- **Bruksforpliktelser.** Gjennom arbeidet med sentralt styringsdokument skal man beskrive hvordan sikre at brukergrupper forplikter seg til bruk av kapasitet hos plattformleverandøren.
- **Finansieringsmodell.** Det anbefalte konseptet innebærer at investeringene finansieres gjennom grunnbevilgning over statsbudsjettet. Driftskostnadene skal dekkes gjennom en kombinasjon av grunnleggende driftsstøtte og brukerbetaling. Det sentrale styringsdokumentet skal beskrive hvordan ulik bruk av ulike tjenester skal belastes brukerne.
- **Organisering av plattformleverandør.** Etablere de sentrale prinsippene for hvordan plattformleverandøren skal organiseres. Skissere hvilke oppgaver plattformleverandøren skal ha med faste ansatte, hvordan skal drift- og generell brukerstøtte anskaffes (egne ansatte vs. eksterne leverandører).

⁵⁹ Forskningsrådet, «Behov for tungregnekraft for forskning og kunstig intelligens».

- **Organisering av avansert brukerstøtte.** Sentralt styringsdokument skal beskrive hvordan skal avansert brukerstøtte skal organiseres fremover og hvordan denne kompetansen skal tilgjengeliggjøres ulike brukergrupper (utenfor forskningssektoren). Beskrivelsen skal inneholde en dimensjonering av tjenesten, samt hvordan man skal sikre tverrsektoriell kompetanse.
- **Styrings- og tildelingsmodell.** Det anbefalte konseptet innebærer at omfanget av brukergrupper utvides, fra dagens situasjon, der BOTT-universitetene er styrende, til en situasjon der flere brukergrupper fra andre institusjoner må sikres nødvendig infrastruktur uten at det går på bekostning av dagens brukere. Sentralt styringsdokument skal beskrive hvordan denne balansen skal ivaretas gjennom styringsmodellen til plattformleverandøren. I tillegg skal man beskrive hvordan kapasitet skal tildeles ulike brukergrupper.

11.2 Planlegging og gjennomføring

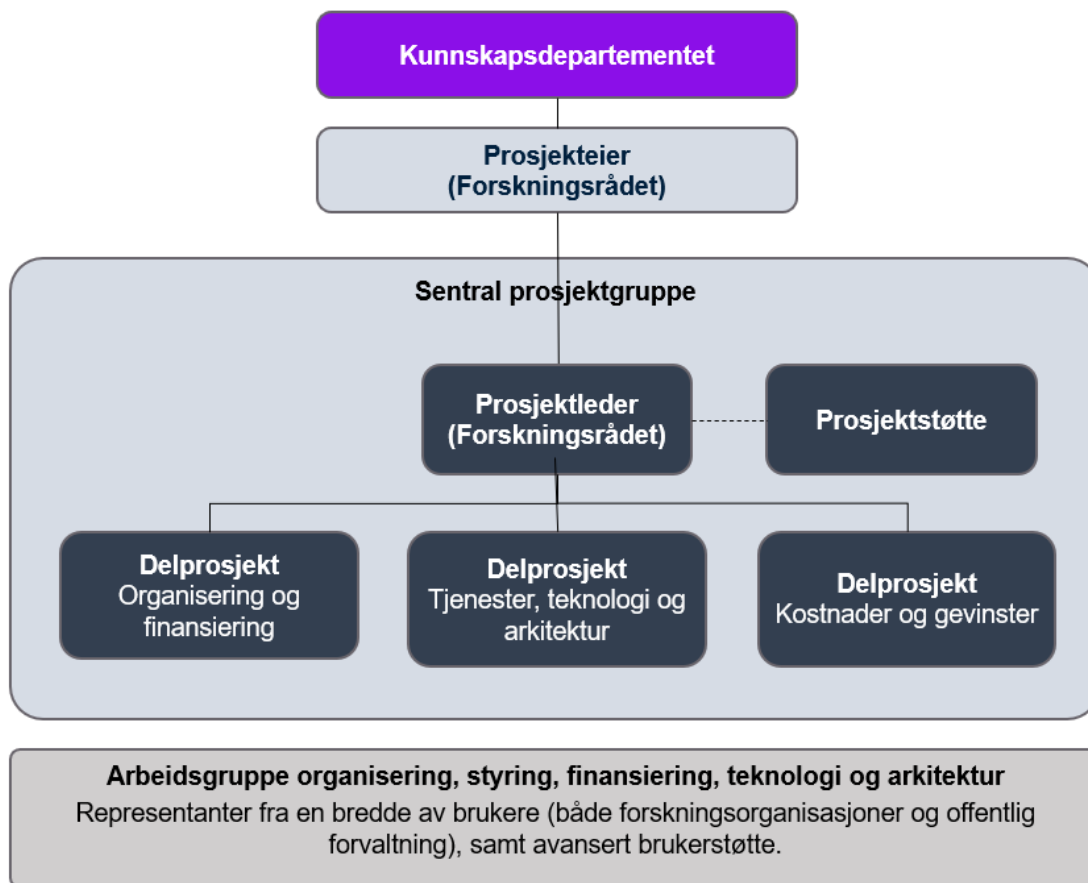
11.2.1 Organisering og finansiering

Det anbefales at forprosjektet blir gjennomført som et prosjekt i regi av Forskningsrådet, og styringsdokumentet bør foreslå en styringsstruktur for gjennomføringsfasen som blant annet ivaretar de viktigste interessentene.

Forskningsrådet er ansvarlig for finansiering av egne ressurser forprosjektet, men er helt avhengige av at andre virksomheter bidrar med ressurser i denne fasen.

Implementeringen av valgt konsept innebærer en stegvis videreutvikling av Sigma2 og eksisterende systemer og strukturer. Derfor bør Sigma2 være representert i den sentrale prosjektgruppen. I tillegg er det essensielt å involvere fremtidige brukere i design av tjenesten, valg av arkitektur, samt i utforming av styringsmodell og finansieringsmodell.

Figur 11.1 viser hvordan forprosjektet av valgt konsept kan organiseres.



Figur 11.1 Eksempel på hvordan prosjektet for forprosjektet av valgt konsept kan organiseres

Forprosjektet krever et sett med roller og ansvar. De ulike rollene i programmet er beskrevet i Tabell 11.1.

Tabell 11.1. Roller og ansvar i forprosjekt for implementering av valgt konsept

Rolle	Ansvarsområder
Kunnskapsdepartementet	<p>En videreføring av tiltaket og implementeringen av anbefalt konsept stiller en rekke krav til eierstyringen. Det vil innebære en gjennomføring som berører flere sektorer og departementer, og som potensielt vil ha grenseflater mot prosjekter og aktiviteter i flere av disse. Gjennomføringen av tiltaket vil således dekke et stort aktørbilde og behovet for involveringen av disse kan være ulikt i ulike faser av gjennomføringen. Riktig involveringen av virksomheten er en viktig forutsetning for å lykkes.</p> <p>Departementet gir oppdraget for forprosjektet og har ansvar for koordinering og forankring mot andre departementer.</p>
Prosjekteier	<p>Prosjekteier er en del av ledelsen i Forskningsrådet som skal sørge for at prosjektet når sine mål. Prosjekteier må ha erfaring, kompetanse og myndighet til å lede og motivere prosjektleder og delprosjektene. Prosjekteier har også ansvar for nødvendig forankring av prosjektet, samt sikre nødvendig gjennomføringskraft.</p>
Prosjektleder	<p>Prosjektleder har i forprosjektet ansvar for igangsetting og overordnet styring av forprosjektets delprosjekter. Prosjektleder har ansvar for å tilrettelegge for god koordinering mellom delprosjekter og mål, ivaretar avhengigheter, koordinering</p>

og ressurskonflikter mellom delprosjektene og oppgavene som gjennomføres av prosjektet.

Prosjektstøtte	Omfang og kompleksitet i prosjektet vurderes som så store at det er nødvendig å etablere et administrativt apparat til støtte for spesielt prosjekteier og prosjektleder. Det kan også være hensiktsmessig å støtte delprosjektene. Her kan eksempelvis ressurser på prosjektledelse, organisering/finansiering, arkitektur og løsningsdesign, gevinstrealisering, kommunikasjon, endringsledelse e.l. organiseres for å støtte delprosjektene med å designe og forprosjekttere tiltenkte endringer.
Delprosjektledere	Delprosjektledere har ansvar for å lede delprosjektet som er organisert under prosjektleder. Delprosjektleder har ansvar for å levere avtalte leveranser innenfor de rammene som er definert av prosjektet. Delprosjektleder har bl.a. en sentral rolle i å støtte gevinstarbeidet gjennom identifisering av gevinster.
Arbeidsgruppe	Arbeidsgruppen består av personer med kompetanse og ekspertise på endringer prosjektet skal utvikle. Gruppen har ikke ansvar for prosjektets suksess eller måloppnåelse, men skal gi støtte og veiledning til prosjektleder som berører prosjektets leveranser, og er viktig for å forankre prosjektet hos berørte parter. Gruppen settes sammen av en bredde av brukere og eksperter for å sikre god representativitet både når det gjelder behov og erfaring. Det legges opp til at arbeidsgruppen møtes jevnlig gjennom forprosjektfasen og bidrar med innspill til alle fasene av prosjektet.

11.2.2 Tentativ tidsplan

Gjennomføringen av forprosjektet vil følge strukturen til det sentrale styringsdokumentet.

Faser	Mobilisering	Overordnede rammer	Prosjektstrategi	Prosjektstyringsbasis
Hoved-aktiviteter	<ul style="list-style-type: none"> • Utnevne prosjekteier og prosjektleder. • Opprette prosjektkontor • Etablere arbeidsgruppe • Utarbeide plan for forprosjekt 	<ul style="list-style-type: none"> • Slutføre utredninger (ref. kapittel 11.1) • Beskrive overordnede rammer • Identifisere gevinstområder og planlegg arbeid med gevinstrealiseringsplan • Utarbeide utkast til prosjektnedbrytningsstruktur • Planlegg prosess for innhenting av enhetskostnader 	<ul style="list-style-type: none"> • Definere gjennomføringsstrategien, herunder detaljere ut: <ul style="list-style-type: none"> ○ Tjenester som skal leveres ○ Hvilke verktøy/ software skal tilgjengeliggjøres ○ Dimensjonering av plattform-leverandørens organisasjon ○ I hvilken rekkefølge skal dette bygges • Utarbeide kontraktsstrategi for: <ul style="list-style-type: none"> ○ Avansert brukerstøtte ○ Drift og generell brukerstøtte ○ Utstysleveranser ○ Entreprenørtjenester • Beskrive modell for styring av usikkerhet • Beskrive organisasjon og ansvarsdeling for gjennomføring • Utarbeide utkast til gevinstrealiseringsplan • Innhente oppdatert data for enhetskostnader • Beskrive prosjektstrategi 	<ul style="list-style-type: none"> • Utarbeide prosjektnedbrytningsstruktur • Beskrive arbeidsomfang • Utarbeide kostnadsoverslag, budsjett og investeringsplan • Gjennomføre usikkerhetsanalyse • Slutføre gevinstrealiseringsplan • Utarbeide fremdriftsplan for steg 1 i implementeringen • Slutføre dokumentasjon av sentralt styringsdokument
Leveranser	<ul style="list-style-type: none"> • Prosjektmandat • Gjennomføringsplan 	<ul style="list-style-type: none"> • Utkast til fireårig langtidspan for infrastruktur • Modell for bruksforpliktelser • Finansieringsmodell • Organisasjonsmodell for avansert brukerstøtte • Organisasjonsmodell for plattformleverandør • Utkast til styrings- og tildelingsmodell • Utkast til overordnede rammer 	<ul style="list-style-type: none"> • Utkast til prosjektstrategi • Utkast gevinstrealiseringsplan • Utkast prosjektnedbrytningsstruktur og innsamlede data for enhetskost 	<ul style="list-style-type: none"> • Fireårig langtidspan for infrastruktur • Sentralt styringsdokument
Arbeidsgruppens deltakelse		<ul style="list-style-type: none"> • Deltakelse på heldags arbeidsgruppemøte for å forankre og få innspill på overordnede rammer. 	<ul style="list-style-type: none"> • Deltakelse på heldags arbeidsgruppemøte for å forankre og få innspill på prosjektstrategi 	<ul style="list-style-type: none"> • Deltakelse på heldags arbeidsgruppemøte for å forankre og få innspill på prosjektstyringsbasis

11.3 Suksessfaktorer og fallgruver

Kritiske suksessfaktorer er forhold som planleggings- og gjennomføringsfasen må lykkes med for å oppnå gode resultater med utgangspunkt i målsettingene. Suksessfaktorene er således spesifikke for den omstillingen som skal gjennomføres, og bygger på det overordnede usikkerhetsbildet fra prosjektets usikkerhetsanalyse. Suksessfaktorene gjelder i utgangspunktet alle prosjektets aktører og gir et bilde over hva som kreves av aktørene for at prosjektet skal bli en suksess. Ut ifra suksessfaktorene vil en vurdere de mest relevante tiltak som optimaliserer mulighetene for å oppnå prosjektets målsettinger.

Det er identifisert en rekke kritiske suksessfaktorer, som i ulik grad vil treffe målsettingene med det anbefalte konseptet (K3):

- Tydelig innsats på, og krav til, kompetanseutvikling hos eier, forvalter og bruker for å kunne ivareta sine nye roller i modellen, samt etablering av gode rutiner for evaluering og læring.
- Godt forankrede brukerbehov med alle brukergrupper for å sikre riktig innretning av tjenesteleveransene fra en videreutviklet Sigma2, herunder evnen til å håndtere differensierte tildelingsmodeller, ulike plattformbehov (verktøy, software), samt evnen til å gi avansert brukerstøtte.
- God prosjekteierstyring for å sikre effektive beslutningsprosesser og nødvendig ressurstilgang.
- God forankring og kommunikasjon av andre pågående tiltak, utredninger og analyser som vil kunne påvirke finansiering, organisering, roller og ansvar med mer.
- Prosjektorganisasjonen må bemannes med tilstrekkelige ressurser med nødvendig kompetanse. Prosjektets egenart og risiko tilsier at prosjektet krever betydelige ressurser med hensyn til planlegging, styring, ledelse og gjennomføring.
- God og presis kommunikasjon med alle relevante interessenter og aktører for å sikre gode beslutningsprosesser gjennom hele planleggings- og gjennomføringsfasen.
- Oppmerksomhet på tekniske løsninger knyttet til det tverrsektorielle konseptets sin egenart og nødvendige integrasjoner med andre plattformer og systemstøtteverktøy.
- Anbefalt konsept K3 må ha tilstrekkelig fleksibilitet i innretningen slik at den er levedyktig over tid. Dette innebærer fleksible løsninger for omgjøring knyttet til endringer i teknologi, deltakelse i internasjonalt samarbeid, organisatoriske endringer med mer.

Fallgruver er det motsatte av suksessfaktorer og viser derfor til forhold som kan bidra til at prosjektets målsettinger ikke innfris. Slik sett vurderes ikke fallgruver ut over at de vil uttrykkes som det motsatte av nevnte suksessfaktorer. Slike forhold søkes unngått gjennom relevante tiltak som inngår i prosjektkontorets arbeid med risikostyring.

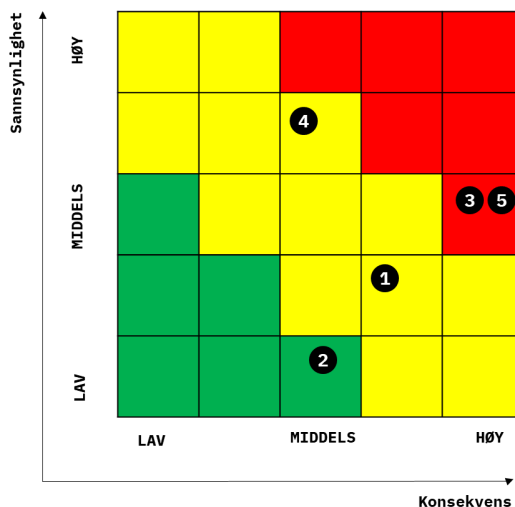
11.4 Forslag til risikoreduserende tiltak

Nedenfor beskrives hvordan prosjektrisiko kan håndteres i planlegging og gjennomføring av implementeringen av valgt konsept (K3). I forbindelse med arbeidet med konseptvalgutredningen er det identifisert noen risikofaktorer som har vesentlig konsekvens. Disse risikofaktorene vil kunne redusere kvaliteten på implementeringen av valgt konsept (K3), både i form av hvordan implementeringen planlegges og gjennomføres. Implementering av konseptet er kompleks med hensyn til at målbildet skal omfatte flere brukere med ulike behov for kompetanse, og dette setter høye krav til prosjektstyringen, herunder løpende vurdering av prosjektets risiko.

Risikofaktorene nedenfor er ikke fullstendig utfyllende og bør oppdateres gjennom hele gjennomføringsperioden. Ansvar for oppdateringen bør ligge hos sentral prosjektgruppe, som bør utarbeide en oppdatert analyse av prosjektrisikoen i forbindelse med mobilisering av

prosjektet. Gjennom planleggingen og gjennomføringen av prosjektet vil man søke å identifisere de risikofaktorer som kan påvirke prosjektgjennomføringen. Underveis er god kommunikasjon mellom eier, bruker og forvalter den viktigste aktiviteten knyttet til risikohåndtering. Årsaken til risikoen beskrives som del av identifiseringen. Risikofaktorene med tilhørende forslag til tiltak er nærmere beskrevet nedenfor.

Prosjektrisiko	
1	Samarbeidet mellom eier, brukere og plattformleverandør er ikke tilfredsstillende
2	Kapasitetsutfordringer hos eier, brukere og plattformleverandør
3	Kompetanseutfordringer hos eier, bruker og plattformleverandør i forhold til økt myndighet og endrede roller.
4	Usikkerhet knyttet til kapasitetsbehov medfører under-/overinvestering
5	Endringer i anbefalt konsept har kan medføre redusert kvalitet på tjenesten til dagens brukere (forskningsdrevet FoU)



Figur 11.2. Risikofaktorer i risikomatrixen

Tabell 11.2. Utvalgte risikomomenter ved oppdraget

#	Prosjektrisiko	Tiltak
1	Samarbeidet mellom eier, brukere og plattformleverandør er ikke tilfredsstillende.	Planleggingsfasen bør omfatte en bred involvering og forankring av endringsbehovet som følge av valgt konsept. Hensikten er å etablere kunnskap og forståelse av hvilke tiltak som må gjøres og hva som må gjøres for å gjennomføre implementeringen.
2	Kapasitetsutfordringer hos eier, brukere og plattformleverandør.	I planleggingsfasen bør det gjøres et grundig arbeid med å identifisere kapasitetsbehov hos eier, bruker og forvalter. Samtidig bør det etableres en detaljert beskrivelse av forventet kapasitet i gjennomføringsfasen. Sentral prosjektgruppe følger opp kapasitetsbehovet i første omgang i forprosjektfasen og melder avvik til prosjekteier.
3	Kompetanseutfordringer hos eier, bruker og plattformleverandør i forhold til myndighet og endrede roller.	Planleggingsfasen inkluderer en kartlegging av nødvendig kompetanse for eier, bruker og plattformleverandør.
4	Usikkerhet knyttet til kapasitetsbehov medfører under-/overinvestering.	Planleggingsfasen inkluderer utarbeidelse av en fireårig investeringsplan. Det er plattformleverandørens styre sitt ansvar å sikre at denne blir årlig oppdatert, basert på en strukturert og dyptgående kartlegging av brukerbehov.

#	Prosjektrisiko	Tiltak
5	Endringer i anbefalt konsept kan medføre redusert kvalitet på tjenesten til dagens brukere (forskningsdrevet FoU).	<p>I planleggingsfasen er det avgjørende at BOTT-universitetene er tett involvert i design av ny organisasjons- og styringsmodell.</p> <p>Gjennomføringen skjer stegvis:</p> <ul style="list-style-type: none">• Steg 1: Leverandøren må i tillegg til å dekke dagens brukere behov, dekke behovet for tungregning til utvikling og testing av språkmodeller og generative nasjonale KI-modeller.• Steg 2: Gjennomføre sektorvise kost/nyttevurderinger og ta stilling til om sektorer (forvaltningsområder) få dekket sine behov for tungregning hos leverandør eller sikre behov gjennom andre løsninger/leverandører.• Helsesektoren krever særskilt vurdering, kan potensielt være mulig å 'dele opp' sektorens behov.

12 Referanser

Digitaliserings- og forvaltningsdepartementet. «Fremtidens digitale Norge. Nasjonal digitaliseringsstrategi 2024-2030». https://www.regjeringen.no/contentassets/c499c3b6c93740bd989c43d886f65924/no/pdfs/nasjonal-digitaliseringsstrategi_ny.pdf

Digitaliserings- og forvaltningsdepartementet. «Strategi for norsk deltakelse i Programmet for et digitalt Europa – DIGITAL». 2024. <https://www.regjeringen.no/contentassets/d73845895fdb4225b75bc6eb60d220a7/no/pdfs/norsk-deltakelse-i-programmet-digital.pdf>

Forskningsrådet. Rapport til Kunnskapsdepartementet fra Norges Forskningsråd: «Behov for tungregnekraft for forskning og kunstig intelligens». 2024. <https://www.forskningsradet.no/siteassets/publikasjoner/2024/rapport-med-vedlegg---behov-for-tungregnekraft-for-forskning-og-ki.pdf>

Forskningsrådet. «Et helhetlig forsyningssystem for åpen, skjermet og gradert forskning». 2024. https://www.forskningsradet.no/siteassets/publikasjoner/2024/sluttrapport_helhetlig-forskningssystem.pdf

Forskningsrådet. Rapport til Kunnskapsdepartementet fra Norges Forskningsråd: «Forslag til rammeverk for rask oppskalering av analysekapasitet og -infrastruktur i kriser». 2023. [231108-oppskalering-av-analysekapasitet-i-kriser---rapport-endelig.pdf](https://www.forskningsradet.no/siteassets/publikasjoner/2023/231108-oppskalering-av-analysekapasitet-i-kriser---rapport-endelig.pdf)

Forskningsrådet. Rapport fra Datainfrastrukturutvalget: «Investering i infrastrukturer for FAIR forskningsdata og særlig relevante forvaltningsdata for forskning». 2022. <https://www.forskningsradet.no/siteassets/publikasjoner/2022/rapport-fra-datainfrastrukturutvalget-2022.pdf>

Forskningsrådet. «Evaluation of the INFRASTRUKTUR initiative as a funding instrument. 2021». [infrastruktur-evalueringsrapport.pdf](https://www.forskningsradet.no/siteassets/publikasjoner/2021/infrastruktur-evalueringsrapport.pdf)

Helsedirektoratet. «Felles KI-plan for trygg og effektiv bruk av KI i helse og omsorgstjenesten 2024 – 2025». 2024. [Felles KI-plan for trygg og effektiv bruk av KI i helse og omsorgstjenesten 2024 - 2025 - Helsedirektoratet](https://www.helsedirektoratet.no/globalassets/publikasjoner/2024/felles-ki-plan-for-trygg-og-effektiv-bruk-av-ki-i-helse-og-omsorgstjenesten-2024-2025.pdf)

Innst. 170 S (2022–2023). *Innstilling til Stortinget fra utdannings- og forskningskomiteen*. Stortinget. <https://www.stortinget.no/globalassets/pdf/innstillinger/stortinget/2022-2023/inns-202223-170s.pdf>

Kommunal- og moderniseringsdepartementet. «Nasjonal strategi for kunstig intelligens». 2020. <https://www.regjeringen.no/contentassets/1feb3bb2c4fd4b7d92c67ddd353b6ae8/no/pdfs/ki-strategi.pdf>

Meld. St. 5 (2022–2023). *Langtidsplan for forskning og høyere utdanning 2023 – 2032*. Kunnskapsdepartementet. <https://www.regjeringen.no/contentassets/9531df97616e4d8eabd7a820ba5380a9/no/pdfs/stm202220230005000dddpdfs.pdf>

Meld St. 9 (2023-2024). *Nasjonal helse- og samhandlingsplan 2024–2027 — Vår felles helsetjeneste*. Helse og omsorgsdepartementet. [Meld. St. 9 \(2023–2024\) - regjeringen.no](https://www.regjeringen.no/globalassets/publikasjoner/2023/meld-st-9-2023-2024.pdf)

Nærings- og fiskeridepartementet/Kunnskapsdepartementet. «Strategi for å øke næringslivets investeringer i forskning og utvikling». 2024. [Strategi for å øke næringslivets investeringer i forskning og utvikling](https://www.regjeringen.no/globalassets/publikasjoner/2024/strategi-for-a-oke-naeringslivets-investeringer-i-forskning-og-utvikling.pdf)

Ot.prp. nr. 62 (2006-2007). *Om lov om endringer i skattelovgivningen (internprising)*. Finansdepartementet. [Ot.prp. nr. 61 \(2006-2007\)](https://www.regjeringen.no/globalassets/publikasjoner/2006/ot-prp-nr-62-2006-2007.pdf)

Prop. 1 S (2024-2025). Proposisjon til Stortinget (forslag til stortingsvedtak). Kunnskapsdepartementet. [Prop. 1 S \(2024–2025\)](https://www.regjeringen.no/globalassets/publikasjoner/2024/prop-1-s-2024-2025.pdf)

Prop. 87 S (2023-2024). *Forsvarsløftet – for Norges trygghet. Langtidsplan for forsvarssektoren 2025–2036*. Forsvarsdepartementet.

<https://www.regjeringen.no/contentassets/27e00e5acc014c5ba741aacff235d99/no/pdfs/prp202320240087000dddpdfs.pdf>

Regjeringen. «Liste over virksomheter med kritisk samfunnsfunksjon og nøkkelpersonell.» https://www.regjeringen.no/contentassets/8da70b8196a24296ae730eaf99056c1b/liste-over-kritiske-samfunnsfunksjoner_endelig-versjon-22.12.pdf

Samfunnsøkonomisk analyse. «Kunstig intelligens i Norge – nytte, muligheter og barrierer». Rapport 35-2023. [Kunstig intelligens - nytte, muligheter og barrierer | Samfunnsøkonomisk Analyse](#)

Sigma2. «Project description E-INFRA 2023 – A National e-infrastructure for science». Søknad om midler til forskningsinfrastruktur av nasjonal viktighet. 2023.

Skate. «Rammeverk for utforming av finansieringsmodeller for nasjonale og sektorielle fellesløsninger». 2023. [Rammeverk for utforming av finansieringsmodeller for nasjonale og sektorielle fellesløsninger](#)

Technavio. «High Performance Computing (HPC) Market Analysis APAC, North America, Europe, South America, Middle East and Africa - China, US, Japan, Germany, France - Size and Forecast 2024-2028». 2024. [High Performance Computing \(HPC\) Market Growth Analysis - Size and Forecast 2024-2028](#)

Teknologirådet. «Generativ kunstig intelligens i Norge». 2024. [Generativ kunstig intelligens i Norge - Teknologirådet](#)

13 Vedleggsoversikt

1. Spørreundersøkelse – Bruk og behov for tungregnekraft i offentlig forvaltning

Norges forskningsråd

Besøksadresse: Drammensveien 288

Postboks 564

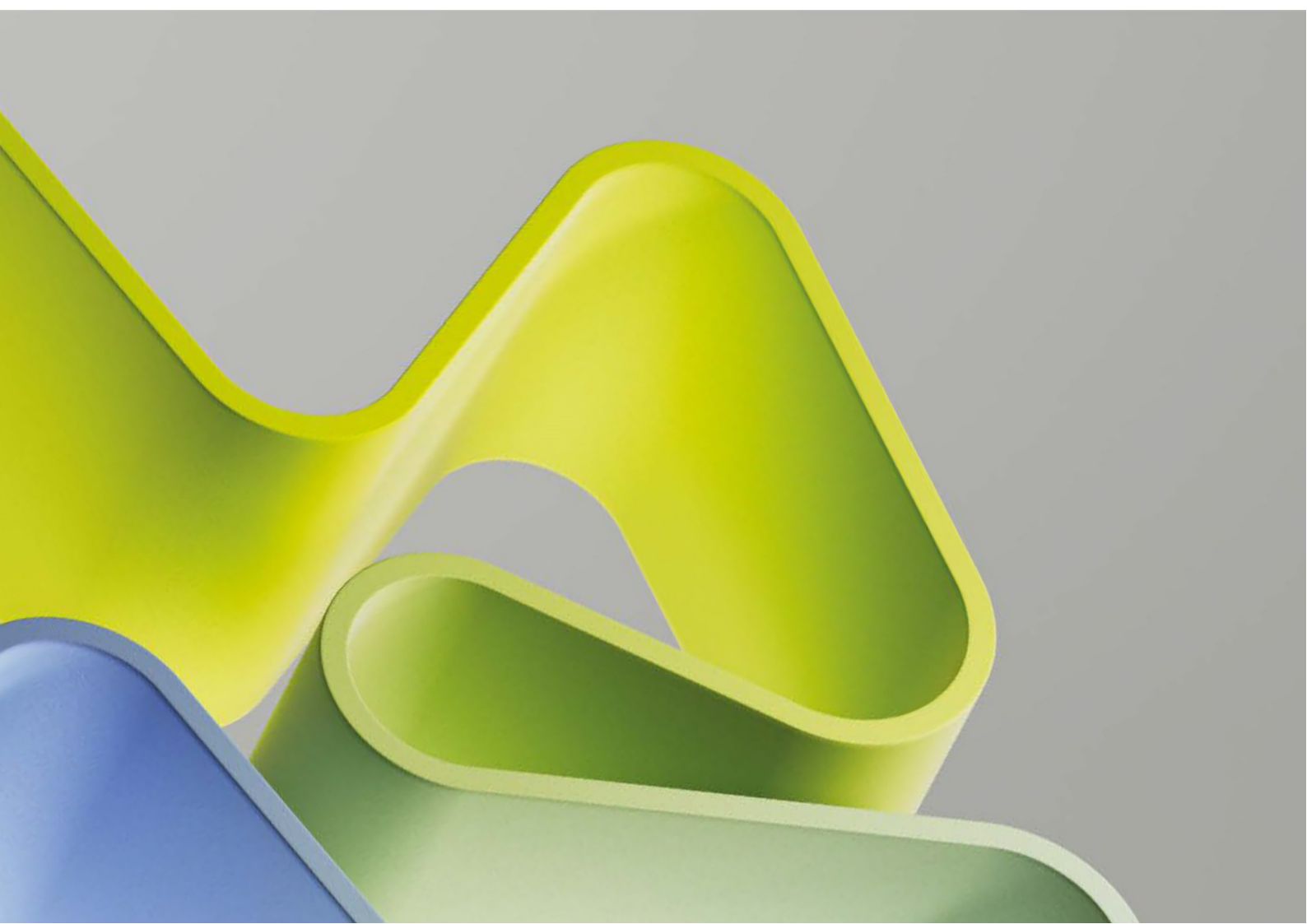
1327 Lysaker

Telefon: 22 03 70 00

Telefaks: 22 03 70 01

post@forskningsradet.no

www.forskningsradet.no



Vedlegg 1: Spørreundersøkelse - Bruk og behov for tungregnekraft i offentlig forvaltning

Spørreskjema

Forskningsrådet kartlegger behovet for tungregnekraft i offentlig forvaltning, særlig i tilknytning til utvikling av KI.

I utarbeidelsen av konseptvalgutredningen for tungregnekraft er det behov for økt innsikt i behovene for tungregnekraft i offentlig forvaltning, særlig i tilknytning til utvikling av KI. Vi håper derfor at dere kan svare på denne spørreundersøkelsen om deres bruk og behov for tungregnekraft for å løse forvaltningsoppdraget. Svarfrist: 3. desember 2024.

* Obligatorisk

Bakgrunnsinformasjon

1. Hvilken institusjon representerer du? *
2. Vennligst oppgi kontaktinfo i tilfelle vi trenger ytterligere avklaringer (Navn, epost, telefon)*

Dagens situasjon og bruk av tungregnekraft for tradisjonell og KI-relatert tungregning

3. Bruker dere tungregneressurser for å løse forvaltningsoppdraget deres i dag? *
 - a. Ja
 - b. Nei
4. Ser dere behov for å bruke tungregnekraft for å løse forvaltningsoppdraget de nærmeste årene? *
 - a. Ja
 - b. Nei
5. Utdyp hvilke behov
6. Har dere egne, interne tungregneressurser i dag? *
 - a. Ja
 - b. Nei
7. Hva har dere av egne tungregneressurser i dag?
Vi er ute etter kapasitet av en viss størrelse/mengde (større og nyere enn (>V100) 16GPUer eller 3000 CPUer) og som kan brukes sammen slik at de utgjør en betydelig ressurs for et forskermiljø/senter eller andre, enten interne brukere eller eksterne.
8. Hvis aktuelt, kan dere si noe om fordelingen mellom tradisjonell tungregning og KI-relatert kapasitet?
Fordelingen kan oppgis i prosent.
9. Hvem er dagens brukere av disse tungregneressursene?
Med interne brukere mener vi brukere på egen institusjon. Med eksterne brukere mener vi brukere fra andre FoU-institusjoner, offentlig sektor og/eller næringsliv.
 - a. Interne brukere fra et bestemt forskningsmiljø/senter/institutt
 - b. Interne brukere fra hele organisasjonen
 - c. Både interne og eksterne brukere
 - d. Annet
10. Hvordan anskaffer dere eksterne tjenester/kapasitet for tungregning? *
 - a. Vi bruker kun egne, interne ressurser

- b. Vi anskaffer tjenester/kapasitet hos vår skyleverandør
 - c. Vi anskaffer tjenester/kapasitet hos nasjonal, offentlig leverandør av tungregnekraft (som Sigma2)
 - d. Vi anskaffer tjenester/kapasitet hos internasjonal, offentlig leverandør av tungregnekraft (som EuroHPC)
 - e. Annet
11. Utdyp gjerne hvilke leverandører dere bruker

Fremtidige behov for tungregnekraft i et femårsperspektiv

Her ønsker vi at dere vurderer behovet de nærmeste fem arene.

- 12. Kan dere anslå framtidig behov for tungregnekraft i deres organisasjon når det gjelder kapasitet til tradisjonelle beregningsoppgaver?
- 13. Kan dere anslå framtidig behov for tungregnekraft i deres organisasjon når det gjelder kapasitet til KI-relaterte beregninger?
- 14. Har dere noen synspunkter på hvordan Norge skal lykkes med innovasjon og verdiskapning med KI i Norge, innen 5 år?
Si gjerne noe om muligheter som ligger i samarbeid mellom FoU-institusjoner, offentlig sektor og næringsliv.
- 15. Har dere noen synspunkter på hvor stor andel av fremtidige, nasjonale tungregneressurser som bør ligge i Norge, med tanke på nasjonal sikkerhet og beredskap?
- 16. Leverer dere tjenester av samfunnskritisk betydning som krever bruk av tungregneressurser?
- 17. I hvilken grad er disse tungregneressursene norskeide og/eller lokalisert i Norge?
- 18. Er det andre krav til tungregneressursene som er viktig for dere?
Mulig å håndtere sensitive data, mulig å håndtere gradert informasjon, kontinuerlig tilgang hele døgnet, m.m.

Respondenter

Arkivverket	Avinor - AI og Data
Brønnøysundregistrene	Datatilsynet
Departementenes servicesenter	Direktoratet for byggkvalitet
Direktoratet for Forvaltning og Økonomistyring	Direktoratet for høyere utdanning og kompetanse (HK-dir)
Direktoratet for medisinske produkter	Entur AS
Fiskeridirektoratet	Helse Nord RHF
Jernbanedirektoratet	Kartverket
Kriminalomsorgsdirektoratet	Kripos
Kulturdirektoratet	Kystverket
Landbruksdirektoratet	Luftfartstilsynet
Lånekassen	Miljødirektoratet

Vedlegg 1: Konseptvalgutredning for tungregning

Norsk helsenett SF	Norsk Romsenter
Norsk romsenter	Norsk Utenrikspolitisk Institutt
Politiets IT-enhet	Riksantikvaren
Ruter AS / TET Digital AS	Skatteetaten
Sokkeldirektoratet	Sokkeldirektoratet
St. Olavs hospital HF	Statens arbeidsmiljøinstitutt
Statens vegvesen	Statens vegvesen
Statkraft	Økokrim