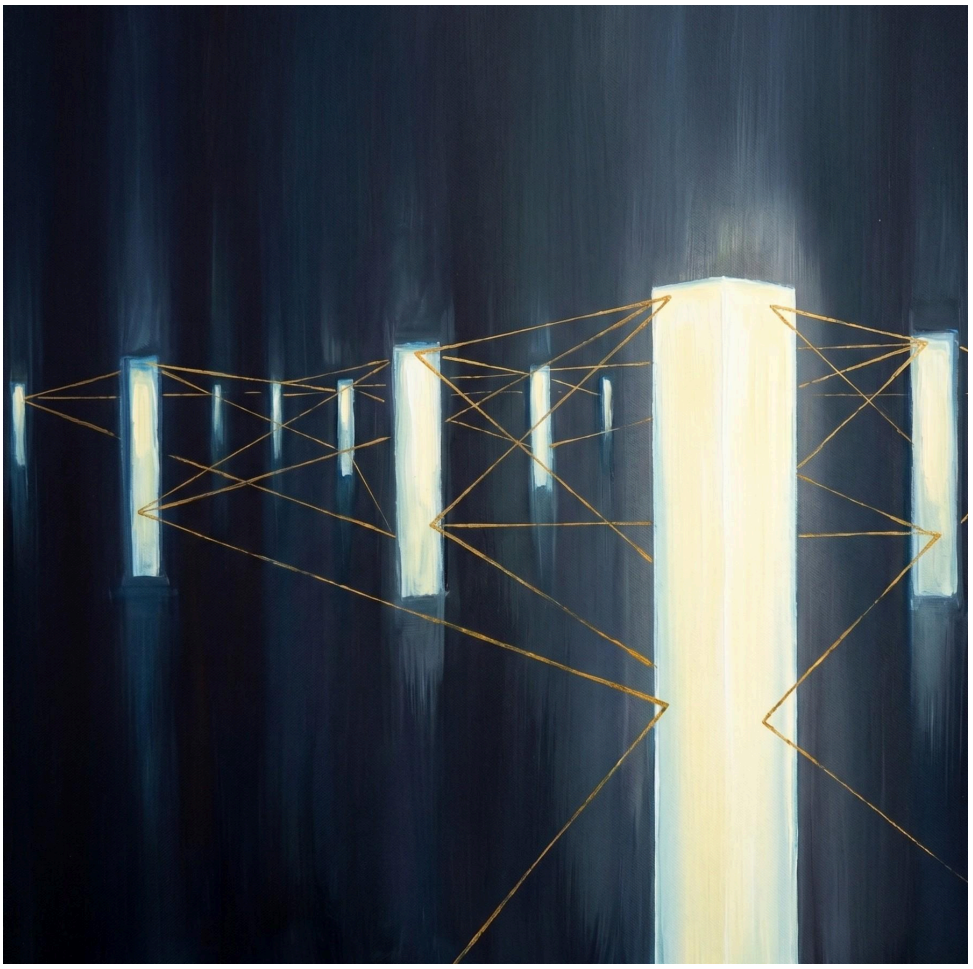
Langsikt

07.09.2026

Konseptnotat for et norsk KI-sikkerhetsorgan

LOTTE SKOLEM (fungerende daglig leder, Langsikt)
AKSEL BRAANEN STERRI (fagsjef, Langsikt, PhD)
RAM EIRIK GLOMSETH (rådgiver, Langsikt)
TELLEF RAABE (*tidligere seniorrådgiver*, Langsikt, PhD)
PEDER SKJELBRED (*tidligere rådgiver*, Langsikt)



"KI er nå motoren bak økonomisk makt og hard makt. Og fremtiden kommer imot oss med stor hastighet – ikke i løpet av de neste tiårene, men de neste årene."

- Rt. Hon. Liz Kendall MP, Storbritannias minister for Forskning, innovasjon og teknologi, 28. april 2026

Sammendrag

Verdens ledende teknologer og forskere, støttet av skyhøye kapitalinvesteringer, utvikler i dag systemer som løser oppgaver bedre og raskere enn mennesker. De neste årene kan dette innebære den største teknologiske revolusjonen verden har sett. Utviklingen vil også medføre et endret trusselbilde som krever økt institusjonell kapasitet og kompetanse innen kunstig intelligens (KI). Derfor har flere av våre allierte etablert nasjonale KI-sikkerhetsinstitutter (AI Safety Institutes, forkortet AISI-er).

Norge har etablert, eller er i ferd med å etablere, flere organer for å håndtere KI-fremveksten. Det er likevel flere kritiske funksjoner Norge ikke har dekket. Dette notatet identifiserer tre kjernebehov som kan og bør dekkes av et nasjonalt KI-sikkerhetsorgan:

1. Koordinering av internasjonalt KI-sikkerhetsarbeid gjennom deltakelse i et viktig nettverk av likesinnede stater.
2. Teknisk og strategisk rådgivning av norske myndigheter.
3. Sikre at KI-modeller brukt i staten, kommuner og kritisk infrastruktur er trygge gjennom koordinering av test- og evalueringsarbeidet.

I notatet gjennomgår vi hvordan andre land har organisert sine KI-sikkerhetsinstitutter og skisserer forslag til hvordan et norsk organ kan organiseres. Vi peker ikke på én konkret organisering, men anbefaler at Norge oppretter et nasjonalt KI-sikkerhetsorgan.

1. Hvorfor trenger Norge et KI-sikkerhetsorgan?

Norge har allerede flere forvaltnings- og kompetansemiljøer på KI-området. Rollene er fordelt, men KI-loven er ennå ikke vedtatt: den skal på ny høring høsten 2026, med sikte på proposisjon til Stortinget våren 2027. Rollefordelingen som er besluttet, dekker:

- koordinering av implementeringen av EUs AI Act og tilsynsvirksomheten (Nkom),
- veiledning om legal og ansvarlig KI-bruk og kapasitetsbygging (Digdir/KI Norge),
- tilsyn og håndheving (Datatilsynet med ansvar for å drive algoritmetilsyn).

Det er viktig å understreke at et KI-sikkerhetsorgan ikke skal ta over disse funksjonene. Et NAISI skal være en kunnskapsformidlende og koordinerende aktør, samt et organ som formelt representerer norske interesser på KI-sikkerhetsfeltet. Mer konkret skal organet produsere teknisk evidens om hvordan spesifikke modeller faktisk oppfører seg i norske kontekster og holde sentrale beslutningstakere løpende oppdatert om KI-utviklingen og risikobildet.

Et nasjonalt KI-sikkerhetsorgan er heller ikke et akademisk forskningssenter, som vi har flere av allerede. Det er naturlig at NAISI trekker på kompetansen som allerede ligger i de sterke forskningsmiljøene som arbeider med KI-sikkerhet, blant annet TRUST ved UiO, aiD ved SINTEF/NTNU og Centre for AI Security and Safety ved Simula, slik det også gjøres i andre land. Men disse eller nye forskningssentre kan ikke fylle forvaltningsoppgavene NAISI er ment å fylle, som rådgivning av politiske myndigheter og å representere Norges interesser internasjonalt.

Et KI-sikkerhetsorgan er altså tenkt å dekke tre kjernefunksjoner som ikke ivaretas tilstrekkelig i dag:

Funksjon	Beskrivelse
1. Internasjonal koordinering	Koordinere og samarbeide med andre lands AISIer i International Network for Advanced AI Measurement, Evaluation and Science (tidl. International Network of AI Safety Institutes). Delta på AI Summit-serien og andre multilaterale arenaer. Sikre en norsk stemme i internasjonale prosesser. Tilrettelegge for tettere nordisk og europeisk samarbeid om KI-sikkerhet.
2. Strategisk rådgivning	Holde sentrale beslutningstakere, spesielt statsministerens kontor (SMK) og politisk ledelse i relevante departementer, løpende oppdatert om KI-utviklingen og risikobildet. Være rådgiver i KI-sikkerhetsspørsmål for departementer og tilsynsmyndigheter.
3. Evaluering og testing	Sørge for testing og evaluering av KI-modeller brukt i staten, kommuner og kritisk infrastruktur i tett samarbeid med utviklingsmiljøene. For å unngå dobbeltarbeid bør organet samarbeide tett med AISI-er i allierte land.

1.1 Internasjonal koordinering

De ledende KI-modellene utvikles ikke i Norge. Vår mulighet til å påvirke utviklingen skjer derfor hovedsakelig gjennom internasjonale fora. I juli 2026 ble det første møtet til Global

Dialogue on AI Governance arrangert under FNs mandat, i Genève. AI Summit-serien (Bletchley 2023, Seoul 2024, Paris 2025, New Delhi 2026) har til nå vært den mest sentrale arenaen for internasjonal KI-sikkerhetspolitikk. Her møtes regjeringsledere, teknologiselskaper og forskere for å diskutere felles retningslinjer og tiltak.

I Plan for Norge står det at regjeringen skal "fremme ansvarlig og etisk KI-utvikling internasjonalt". Norges evne til å ta en konstruktiv rolle begrenses uten et eget KI-sikkerhetsorgan. Norge risikerer å bli en passiv mottaker av retningslinjer og regelverk utviklet av andre, med begrenset mulighet til å ivareta norske verdier og interesser i prosessen.

Mye av det internasjonale KI-sikkerhetsarbeidet har opphav i, og koordineres innenfor, et nettverk av de nasjonale KI-sikkerhetsinstituttene. Nettverket, som inntil desember 2025 het International Network of AI Safety Institutes, går nå under navnet International Network for Advanced AI Measurement, Evaluation and Science (NAAIMES). Medlemslandene deler evalueringsmetodikk, forskningsresultater og beste praksis. Norge er per i dag ikke representert. Tyskland etablerte AISI Deutschland 31. august 2026, etter et vedtak i det nasjonale sikkerhetsrådet i juni, som en virtuell struktur bygget på BSI og Bundesnetzagentur.

Globale fora er viktige, men for en liten aktør som Norge vil regionalt samarbeid være vel så avgjørende – både for å samle ressurser og for å sikre at egne interesser faktisk blir hørt. Et nordisk eller nordisk-baltisk samarbeid peker seg ut som den mest aktuelle rammen: landene deler verdigrunnlag, regulatorisk tradisjon, og småspråk-utfordringer, og står overfor mange av de samme problemstillingene når store internasjonale KI-modeller skal integreres i offentlig forvaltning og kritisk infrastruktur. Regionen er allerede i bevegelse. Den svenske AI-kommisjonen har foreslått å opprette et AISI, Estland bygger opp Estonian Center for Safe and Trustworthy AI (ECSTAI) og organisasjonen New Nordics AI har siden 2025 samlet nordiske og baltiske aktører om utbredelse av KI, med støtte fra Nordisk ministerråd. Sikkerhet og evaluering ligger imidlertid utenfor mandatet, og der er regionen fortsatt uten felles struktur. Et norsk NAISI vil kunne sette fart i tilsvarende prosesser hos nabolandene, og posisjonere Norge til å være med å forme arkitekturen for regionalt KI-sikkerhetssamarbeid. Vi vet at det er interesse for dette hos sentrale stakeholdere i disse landene.

I tillegg til de multilaterale og regionale sporene er bilaterale partnerskap viktige. Senest 13. mars 2026 undertegnet Canada og Norge en felleserklæring om digital suverenitet og KI. Canada er et av flere allierte land som har opprettet et KI-sikkerhetsinstitutt.

Et AISI gir altså tilgang til viktige allianser med likesinnede land med en naturlig motpart til eksisterende organisasjoner, samt tilgang til kompetanse og testing av frontier-kapabiliteter gjennom de amerikanske og britiske AISI-ene sin tilgang. Et norsk KI-sikkerhetsorgan, et NAISI, vil bygge kompetanse i Norge og styrke våre allianser i disse foraene.

1.2. Nasjonal strategisk rådgivning

KI-feltet utvikler seg så raskt at tradisjonelle utredningsformater ikke holder tritt. Dette er ikke et område hvor man kan lene seg på en NOU hvert fjerde år eller periodiske rapporter fra Teknologirådet (hvis oppgave er å dekke all teknologi) for å holde seg oppdatert. Det er heller ikke tilstrekkelig å lene seg på enkeltforskere eller forskningsmiljøer uten rapporteringsplikt. Det er behov for dedikerte eksperter tett knyttet opp mot sentrale beslutningstakere, med mandat

til å kontinuerlig overvåke utviklingen, rådgi om implikasjonene for Norge og holde beslutningstakere orientert.

KI-sikkerhetsorganets mandat bør inkludere en eksplisitt foresight-funksjon: strukturert analyse av utviklingstrender og tidlig identifisering av fremvoksende risikoer, som et supplement til løpende evaluering og overvåkning. Uten dette risikerer organet å bli reaktivt i et felt der tiden fra teknologisk gjennombrudd til samfunnsmessig konsekvens kan være svært kort (som illustrert av Claude Mythos).

1.3. Evaluering og testing
I dag stoler vi på teknologiselskapenes egne evalueringer når KI-systemer tas i bruk i Norge. Dette er uholdbart, som Inga Strümke (NTNU) og Michael Riegler (SimulaMet) skriver i Aftenposten. Det som trengs er evalueringer av hvordan systemene fungerer i norske brukskontekster, i norske offentlige systemer, på norsk og med særegne forventninger til helseråd og juridisk veiledning. Dette forvaltningsområdet mangler. NorEval har ingen safety-komponent og LinguaSafe inkluderer ikke norsk språk. Analyse, evaluering og testing er et pågående arbeid. Hver modelloppdatering krever ny testing. Hver ny utrullingskontekst – om det er innenfor NAV, Helsenorge eller kommunal saksbehandling – krever nye evalueringer. Skal dette gjøres tilfredsstillende, krever det et fagmiljø som jobber kontinuerlig med disse oppgavene.

2. Hvordan er andre lands AISIer organisert?

Mange land har etablert eller er i ferd med å etablere KI-sikkerhetsinstitutter etter at Storbritannia og USA ledet an i 2023. Internasjonalt kan vi identifisere tre hovedmodeller for organisering. Det er også presentert en interessant fjerde modell. Det de har til felles er at de tilfredsstiller de tre funksjonene vi nevnte over. De produserer kunnskap gjennom evaluering og testing av KI-modeller og regimer for slik testing. De rådgir sentrale beslutningstakere og deltar i internasjonalt KI-sikkerhetsarbeid, blant annet gjennom alliansen av KI-sikkerhetsinstitutter.

Modell 1: Fagetat under departement

Storbritannias AI Security Institute (AISI) ble etablert i 2023. Det er plassert i Department for Science, Innovation and Technology (DSIT). Instituttet er i en særklasse med et årlig budsjett på omtrent 900 millioner kroner. Det ble etablert etter initiativ fra statsministeren og har en uvanlig direkte kobling til politisk toppnivå. UK AISI gir regelmessige briefinger til både statsråder og statsministerens stab, særlig i saker knyttet til frontiermodeller, nasjonal beredskap og KI-risiko. Instituttet er tett integrert med det britiske sikkerhetsapparatet, som Government Communications Headquarters (GCHQ) og National Cyber Security Centre (NCSC). UK AISI fremstår som en "startup i forvaltningen", kjennetegnet av høy operasjonell fleksibilitet og sterk sikkerhetspolitisk forankring. Instituttet har over 100 tekniske ansatte, rundt 150 policyeksperter og mulighet til å tilby ekstraordinært høy lønn for å tiltrekke seg nødvendig kompetanse. Les mer i denne artikkelen i New York Times.

USAs Center for AI Standards and Innovation (CAISI) ble etablert i 2023. Det er plassert under National Institute of Standards and Technology (NIST) i Department of Commerce. Organet

har et budsjett på ca. 100 millioner kroner for 2024/2025. Under Trump-administrasjonen skiftet instituttet navn fra US AI Safety Institute til dagens navn, med fokus på "standards and innovation". Dette er bakgrunnen for at det internasjonale AISI-nettverket byttet navn i desember 2025. USA og UK samarbeider svært tett, og deres AISI-er er per nå de eneste med formaliserte avtaler om å teste nye modeller fra de amerikanske KI-selskapene før de lanseres, såkalt "pre-deployment testing". CAISI utvidet ordningen i mai 2026 med avtaler om pre-deployment-evaluering med Google DeepMind, Microsoft og xAI, i tillegg til de reforhandlede avtalene med OpenAI og Anthropic. Testingen skjer fortsatt på frivillig basis. Trump-administrasjonens executive order fra 2. juni 2026 om frontier-modeller viderefører frivilligheten: den etablerer graderte benchmarks administrert av NSA og en frivillig kanal for gjennomgang før lansering, men ingen lisensordning og ingen krav om forhåndsgodkjenning.

Canada lanserte Canadian AI Safety Institute (CAISI) i november 2024. I likhet med det britiske AISI er CAISI organisert under departementet Innovation, Science and Economic Development Canada (ISED). Instituttet har et budsjett på 345 millioner kroner over 5 år. Instituttet samarbeider tett med de ledende kanadiske forskningsmiljøene, gjennom Canadian Institute for Advanced Research, de nasjonale KI-instituttene Mila, Amii og Vector Institute og det kanadiske forskningsrådet. CAISI har også et Safe and Secure AI Advisory Group, ledet av professor Yoshua Bengio.

Japan sitt AISI ble etablert i februar 2024. J-AISI er plassert under Information-technology Promotion Agency (IPA), som ligger under Ministry of Economy, Trade and Industry. Instituttet samarbeider med andre departementer og etater og har omtrent 25-30 ansatte. Budsjettet er ikke kjent.

Australia har nylig etablert sitt AI Safety Institute. Instituttet ligger under Department of Industry, Science and Resources og samarbeider tett med eksisterende tilsynsmyndigheter og forskningsinstitusjoner som CSIRO, Data61 og UNSW AI Institute. Det er bevilget rundt 200 millioner kroner til etableringen av instituttet.

Styrken ved denne modellen er direkte politisk påvirkning og institusjonell forankring. Når evalueringsresultater skal inn i anskaffelsesprosesser eller briefes til statsråder, er den korteste veien fra et organ som allerede sitter i forvaltningsstrukturen. Det gir også legitimitet i internasjonale fora; andre lands AISI-er vet hvem de forholder seg til. Svakheten her er at faglig uavhengighet er strukturelt krevende når organet rapporterer til et departement som samtidig kan ha andre politiske prioriteringer.

Modell 2: Todelt struktur med policy i etat og testing i forskningsinstitusjon

Singapore har valgt en todelt struktur. Digital Trust Centre (DTC) ble opprettet i 2022, da Infocomm Media Development Authority (IMDA) ga Nanyang Technological University (NTU) oppdraget med å etablere et nasjonalt senter for tillitsteknologi, finansiert med 50 millioner singaporske dollar fra IMDA og forskningsrådet NRF. I mai 2024 ble DTC utpekt som Singapores AI Safety Institute. AISI-funksjonen er altså lagt til et forskningssenter som allerede fantes, og drives av DTC i samarbeid med IMDA. IMDA står for finansiering, har policyansvar og representerer Singapore i internasjonale fora. Det tekniske evaluerings- og

forskningsarbeidet utføres av DTC ved NTU. DTC har betydelig autonomi i å bedrive det evalueringsarbeidet de mener er klokt. Det er altså ikke oppdragsforskning i streng forstand.

Styrken ved denne modellen er at den løser rollekonflikten ved at staten beholder policyansvar, internasjonal representasjon og finansieringskontroll, mens den tekniske evalueringskapasiteten legges der kompetansen er. Svakheten er koordineringskostnader og prinsipal-agent-problematikk. Forskningsmiljøet har egne prioriteringer, publiseringslogikk og finansieringskilder som kan trekke i retninger som avviker fra statens behov. Det krever et sterkt og kompetent bestillerledd.

Modell 3: Koordinert nettverk av eksisterende organisasjoner

Frankrike valgte med "Nasjonalt institutt for evaluering og sikkerhet av kunstig intelligens" (INESIA, lansert januar 2025) å koordinere fire eksisterende statlige organisasjoner: ANSSI (cybersikkerhet), Inria (IT-forskning), LNE (standarder) og PEReN (digital reguleringsekspertise), fremfor å opprette en ny juridisk enhet. Koordineringen skjer, på vegne av statsministeren, fra Generalsekretariatet for nasjonalt forsvar og nasjonal sikkerhet sammen med Generaldirektoratet for næringslivet.

Tyskland har valgt en tilsvarende nettverksmodell. Nasjonalt sikkerhetsråd besluttet 9. juni 2026 å opprette et KI-sikkerhetsinstitutt, som åpnet i Berlin 31. august 2026. I første fase utgjør IT-sikkerhetsmyndigheten BSI og telemyndigheten BNetzA en «virtuell» kjerne under digitaliserings- og innenriksdepartementene. Instituttet gjennomfører tekniske evalueringer, utarbeider situasjonsbilder og rådgir regjeringen, men er ikke tilsynsmyndighet. Allerede i oppbyggingsfasen er faglig utveksling med AI Office og med instituttene i Storbritannia og Frankrike i gang. Tyskland danner et godt sammenligningsgrunnlag for Norge: en mellomstor stat uten egne frontier-laboratorier som bygger på eksisterende etater fremfor en ny enhet.

Styrken ved nettverksmodellen er at den unngår å bygge ny byråkratisk kapasitet og heller henter inn eksisterende fagmiljøer ved behov. Den franske forankringen i SMK gir også politisk tyngde, mens den tyske ligger i fagdepartementene. Svakheten er at ingen tar reelt eierskap. Begge er også så nye at det er vanskelig å si om modellen faktisk fungerer i praksis.

Modell 4: Statlig lisensiert privat regulering

En fjerde modell er foreslått i artikkelen "Regulatory Markets: The Future of AI Governance". Modellen flytter det operative reguleringsarbeidet til et lisensiert marked. Fremfor at et statlig AISI selv gjennomfører evalueringer og red-teaming, fastsetter myndighetene ønskede utfall og lar private aktører kontrollere teknologiselskapene. KI-selskapene pålegges å kjøpe "regulatoriske tjenester" fra en lisensiert aktør. Sikkerhetsinstituttets rolle blir å utforme utfallskrav og lisenskriterier og føre tilsyn med markedet. Et utfallskrav kan være at frontiermodeller ikke skal øke ikke-statlige aktørers evne til å produsere biologiske våpen.

Styrken ved modellen er at privat kapital og teknisk talent kan mobiliseres i et tempo staten ikke kan matche, og at den er skalerbar på tvers av land gjennom gjensidig anerkjennelse av lisenser. En utfordring er at staten må bygge betydelig egen kapasitet for å føre meningsfullt tilsyn. Det andre er at lisensierte aktører som konkurrerer om oppdrag fra KI-selskaper har insentiver til å ikke finne for mange problemer.

3. Forslag: Et norsk KI-sikkerhetsorgan

Grunnleggende designvalg

Tre sentrale spørsmål knyttet til et et norsk KI-sikkerhetssenter (NAISI) er:

1. Hvilken teknisk kompetansemodell skal sikkerhetsorganet ha?
    a. bestilling fra forskningsmiljøer/kommersielle aktører
    b. egen teknisk kapasitet,
    c. en kombinasjon.
2. Hvor skal enheten plasseres organisatorisk?
    a. DFD/Digdir/KI Norge
    b. Forsvarsdepartementet
    c. Justisdepartementet
    d. Forskningsinstitutt
3. Hvor stort skal budsjettet være?
    a. Mindre enn 30 millioner kr
    b. 30-100 millioner kr
    c. Mer enn 100 millioner kr
    d. Oppdragsforskning

Disse henger delvis sammen, siden plasseringen påvirker hvilke kompetansemodeller som er realistiske og hensiktsmessige. Tabellen på neste side tar for seg det første spørsmålet.

Teknisk kompetansemodell

Et viktig spørsmål er hvordan senteret sikrer teknisk evalueringskapasitet. Vi vurderer tre modeller:

	Bestillermodell	**In-house kapasitet**	**Hospitering**
Beskrivelse	Kjerneenhet bestiller evalueringer fra forskningsmiljøer eller kommersielle aktører	Egne tekniske forskere og ingeniører gjennomfører evalueringer internt	Hospitering av forskere som gjennomfører tekniske evaluering i deltidsstillinger
Fordeler	– Relativt rask oppstart[1] – Lavere faste kostnader – Fleksibilitet	– Større faglig uavhengighet – Potensielt raskere responstid – Varig institusjonell kompetanse – Normen i AISI-nettverket	– Forskere beholder sin akademiske tilknytning – Billig
Ulemper	– Avhengig av at de tekniske miljøene svarer på anbud og prioriterer arbeidet over andre oppgaver de har. – Svakere egenkontroll	– Vesentlig dyrere – Vanskelig å rekruttere – Kan tappe forskningsmiljøer for kompetanse	– Verken dypt nok eller fleksibelt nok – Svak institusjonell hukommelse – Ingen får eierskap

In-house-kapasitet er det ideelle alternativet, men det kan ta lang tid å bygge opp og i en del plasseringer være urealistisk selv på lang sikt. En bestillermodell er enklere å få på plass. Forslagene kan i noen grad også kombineres. For eksempel kan en tenke seg at en starter med en bestillermodell for å sikre rask oppstart, men med et mål om å tilegne seg et in-house teknisk miljø på sikt.

Bestillermodell

Hvordan en eventuell bestillermodell organiseres, er avgjørende. Vi skiller mellom tre ulike varianter som varierer i grad av formalisering og i hvor mye autonomi det tekniske fagmiljøet har.

[1] Et forbehold er at Anskaffelsesloven som regel krever åpne anbudskonkurranser med betydelig ledetid.

Anbudsmodell

NAISI legger konkrete evalueringsoppdrag ut på anbud og forskningsmiljøer og kommersielle aktører konkurrerer om oppdragene. Modellen gir fleksibilitet og lar organet hente kompetanse fra ulike fagmiljøer. Svakheten er at hver bestilling krever ny prosess, at faglig kontinuitet og kompetanseoverføring er krevende å bygge opp, og at de mest relevante miljøene ikke nødvendigvis prioriterer å svare på utlysningene.

Fast partner med spesifikke oppdrag

NAISI inngår én eller flere langsiktige rammeavtaler med utvalgte forskningsmiljøer. Organet definerer oppdragene konkret, og partneren leverer på bestilling. Modellen sikrer faglig kontinuitet og bygger gjensidig forståelse mellom bestiller og utfører over tid, samtidig som NAISI beholder styringen over hva som faktisk evalueres. Den krever imidlertid at NAISI har tilstrekkelig egen kompetanse til å spesifisere meningsfulle oppdrag og kvalitetssikre leveranser. Canadas CAISI representerer en variant der det formelle ansvaret ligger hos et statlig organ, men der det meste av det tekniske arbeidet utføres av etablerte forskningsmiljøer som Mila, Amii og Vector Institute.

Fast partner med høy autonomi

NAISI knytter til seg ett eller flere uavhengige forskningsmiljøer som tildeles et bredt mandat og betydelig frihet til selv å definere hva som skal undersøkes. Singapores AISI bruker denne modellen, der Digital Trust Centre ved Nanyang Technological University i stor grad selv prioriterer hvilke evalueringer som er klokest å gjennomføre. Styrken er at sterke akademiske miljøer kan forventes å komme opp med flere nyvinninger på sikkerhetsfeltet enn hvis bestillingen spesifiseres av noen med mindre kompetanse. Faglig uavhengighet er også mer attraktivt for flinke forskere. Svakheten er mindre mulighet for politisk styring og prinsipal-agent-problematikk: at forskningsmiljøet kan prioritere publiserbare problemstillinger fremfor det NAISI selv anser som mest samfunnsnyttig. Modellen forutsetter høy gjensidig tillit og en felles forståelse av oppdraget.

Organisatorisk plassering

Vi vurderer tre alternativer for plassering av et NAISI.

Alternativ A: Under Digitaliserings- og forvaltningsdepartementet (DFD)

Dedikert enhet i DFD. Kan eventuelt legges under Digdir eller KI Norge. Kan samarbeide med KD.

- Fordeler: DFD er ansvarlig for å gjennomføre EUs AI Act. Datatilsynet og KI Norge ligger under samme departement. Nærmest det sivile KI-økosystemet.
- Ulemper: DFD er Norges minste departement. Det er mange andre nystartede prosjekter som kjemper om oppmerksomhet og ressurser, som KI Norges foreløpige mandat. Uten et eksisterende teknisk-vitenskapelig miljø må evalueringskapasitet bygges fra grunnen av eller bestilles. Mindre erfaring med gradert informasjon.

Alternativ B: Under Forsvarsdepartementet (FD)

Plassert i tilknytning til Forsvarets eksisterende KI-miljø, som f.eks. det nyetablerte KI-senteret.

- Fordeler: Et eksisterende miljø som jobber med gradert informasjon og sikkerhetsspørsmål. Det er et eksisterende teknisk KI-miljø ved Forsvarets KI-senter som kan utvides.
- Ulemper: Risiko for å bli et rent forsvarsprosjekt. Ingen AISI internasjonalt er forsvarsplassert.

Alternativ C: Under Justis- og beredskapsdepartementet (JD)

Plassert under JD, med kobling til beredskapsapparatet, inkludert PST, NSM og DSB.

- Fordeler: Nærhet til beredskapsapparatet samt NSM og PSTs arbeid med teknologiske trusler.
- Ulemper: JD mangler tradisjon for tekniske forskningsmiljøer. Ingen AISI internasjonalt har justisplassering.

Størrelsen på budsjettet

Under 30 millioner kroner per år: Koordinerings- og bestillerorgan

Et budsjett under 30 millioner gir rom for omtrent 8–12 årsverk, primært innen policy og koordinering. Dette er tilstrekkelig til å ivareta internasjonal koordinering og nasjonal strategisk rådgivning. Compute-behovet på dette nivået er begrenset og gjelder primært API-tilgang til modeller og sporadiske evalueringskjøringer som kan dekkes av eksisterende nasjonal regnekraftstruktur (Sigma2) eller skybudsjetter.

Det som ikke kan ivaretas er selvstendig evaluering og testing. Et organ på denne størrelsen kan bestille evalueringer fra forskningsmiljøer, men mangler kapasitet til å faglig kvalitetssikre bestillingene, følge opp kontinuerlig ved modelloppdateringer og opprettholde institusjonell hukommelse. Resultatet er et organ som er synlig internasjonalt, men svakt operativt, omtrent på nivå med Japans J-AISI.

50-100 millioner kroner per år: Reell bestillermodell med begrenset in-house kapasitet

50-100 millioner tilsvarer et AISI på linje med det kanadiske. Det gjør 20-35 årsverk mulig, med rom for en kjerne av tekniske ansatte. Deployment-kontekst-evaluering blir mulig: organet kan finansiere og koordinere norskspråklige safety-evalueringer, utvikle scenarioer for offentlig sektor og bygge et evalueringsrammeverk. Et dedikert compute-budsjett på minst fem millioner kroner per år gir kapasitet til systematisk inferens-testing og red-teaming ved modelloppdateringer.

Organet kan ikke gjennomføre frontier modell-evalueringer selvstendig, men kan bidra meningsfullt til felles evalueringsarbeid i AISI-nettverket. Tilgang til nordisk compute-infrastruktur kan løfte den reelle evalueringskapasiteten utover hva budsjettet alene skulle tilsi.

Over 100 millioner kroner per år: In-house teknisk kapasitet

Over 100 millioner åpner for 35–50 årsverk og gjør det mulig å bygge reell in-house evalueringskapasitet. Et compute-budsjett på minst 15 millioner kroner per år gir rom for både systematisk deployment-evaluering og begrenset deltagelse i frontier model-testarbeid gjennom AISI-nettverket. På dette nivået er det også realistisk å inngå formelle avtaler om compute-tilgang med Sigma2 eller tilsvarende nordisk infrastruktur, noe som reduserer avhengigheten av kommersiell sky.

Dette er nivået der norsk deltakelse i AISI-nettverket kan gå fra politisk representasjon til faglige bidrag og der den strategiske foresight-funksjonen blir reell fremfor å være sammenfatning av andres rapporter.

En praktisk anbefaling

Et minimumsbudsjett som faktisk oppfyller de tre kjernefunksjonene ligger trolig på rundt 50–70 millioner kroner per år, inkludert et compute-budsjett på 5–10 millioner. Under dette nivået risikerer man å opprette et organ som er synlig nok til å skape forventninger, men for svakt til å innfri dem. Det er bedre å starte med et tydelig avgrenset mandat og tilstrekkelig finansiering enn å spre et utilstrekkelig budsjett over alle tre funksjoner.

Samlet vurdering

Plassering / Funksjon	**DFD**	**FD**	**JD**
Internasjonal koordinering	Middels	Svak	Svak
Nasjonal rådgivning og policy	Middels	Middels	Middels
Teknisk evaluering	Svak	Middels	Svak
Forvaltningskapasitet	Sterk	Middels	Middels
Tidslinje	Raskest	Middels	Langsom

Plasseringsuavhengige fallgruver

- **Politisk sårbarhet.** USA omgjorde sitt AISI til "Center for AI Standards and Innovation" under Trump-administrasjonen i 2025. Opprettelsen av et NAISI bør ideelt få tverrpolitisk støtte, flerårig finansiering og en klar plan for oppskalering.
- **Manglende kontakt med SMK.** En viktig funksjon for et AISI er rådgivning til politiske myndigheter. KI er ikke bare et digitaliseringsanliggende. Det er derfor viktig at rådgivningen er bredere og spesielt går til SMK og Nasjonal sikkerhetsrådgiver. Det bør derfor opprettes klare linjer mellom et AISI og SMK.
- **For lite eller ikke skalerbart.** Minimumsbudsjettet må sikre at funksjonene til et NAISI faktisk blir oppfylt. At det starter smått, må heller ikke bli en hemsko for å kunne

oppskalere hvis KI-utviklingen tilsier det. Plassering må derfor være robust for potensiell framtidig oppskalering.

- **Motstand fra etablerte aktører.** Modellen må ha bred nok oppslutning for å sikre at nødvendige partnere er villige til å samarbeide.
- **Kulturell innflytelse.** En vertsinstitusjon preger senteret med sin kultur. Et NAISI må klare å bevare sitt fokus som distinkt fra eksisterende mandater.

Vurdering

Gjennomgangen peker ikke entydig mot én organisasjonsmodell, men den utelukker noen og gjør ett alternativ mer plausibelt enn de andre. Vi presenterer her vår tentative vurdering som grunnlag for videre politisk avklaring.

Vi tar til orde for en todelt struktur med policyfunksjon i departement og testing i en forskningsinstitusjon, inspirert av Singapore. DFD er trolig best egnet til å eie policyfunksjonen, og det kan være hensiktsmessig å gi mandatet til KI Norge. Singapore-modellen demonstrerer at denne kombinasjonen er mulig i praksis: et statlig bestillerledd med policyansvar og internasjonal representasjon, kombinert med teknisk evalueringskapasitet forankret i et eksisterende forskningsmiljø.

For Norge betyr dette i praksis: et lite statlig organ under DFD med klart mandat, direkte linje til SMK og nasjonal sikkerhetsrådgiver, og et formalisert samarbeid med eksisterende tekniske miljøer om gjennomføring av evalueringsarbeidet. Det statlige leddet er bestiller og ansvarlig; forskningsmiljøene er utførere med faglig autonomi innenfor definerte oppdrag. Oppdraget bør koordineres med KD. Høringen på KI-loven høsten 2026 er naturlig for å spille inn dette, og lovproposisjonen våren 2027 den naturlige anledningen til å avklare mandatet.

To forbehold er viktige. For det første forutsetter denne modellen at bestillerleddet har reell faglig kompetanse – ikke bare administrativ kapasitet – til å spesifisere og kvalitetssikre det de bestiller. For det andre må DFD-plasseringen håndtere den strukturelle interessekonflikten mellom to ulike oppdrag: KI-adopsjon (KI Norge) og kritisk evaluering (NAISI). Dette krever tydelig mandatseparasjon, ikke bare organisatorisk avstand.

Budsjettet bør være på 50-70 millioner kroner per år som minimum. Under dette nivået kan ikke organet ivareta alle tre funksjoner på en måte som innfrir forventningene opprettelsen vil skape. Et organ som er synlig nok til å representere Norge i NAAIMES, men for svakt til å produsere evalueringsresultater av faglig verdi, tjener verken norske interesser eller nettverkets. Budsjettet bør inkludere et dedikert compute-budsjett på minst 5 millioner kroner per år, og plasseringen bør være skalerbar dersom KI-utviklingen tilsier økt kapasitet.

OM FORFATTERNE

Lotte Skolem er fungerende daglig leder i Langsikt, og leder for fremvoksende teknologier. Hun er utdannet sivilingeniør i nanoteknologi fra NTNU, og har arbeidet i skjæringspunktet mellom teknologi, innovasjon og strategi som direktør i Aker BioMarine og konsulent i Boston Consulting Group. Som styreleder for næringsklyngen The Life Science Cluster har hun også erfaring med samspillet mellom akademia, industri og offentlige myndigheter.

Aksel Braanen Sterri er fagsjef i Langsikt og ansvarlig for tankesmiens analyser og politikkutvikling. Han er statsviter og filosof og har lenge arbeidet i skjæringspunktet mellom økonomi, politikk og etikk. Han leder et forskningsprosjekt om verdsetting av dyr på Universitetet i Oslo og har undervist i KI-etikk ved Harvard University, University of Oxford og UiO. Sterri har vært kommentator i Dagbladet og skrevet bøker om blant annet sosialdemokrati og Arbeiderpartiet.

Ram Eirik Glomseth er rådgiver på kunstig intelligens, bioteknologisikkerhet og sikkerhetspolitikk i Langsikt. Han har erfaring fra internasjonal KI-styring som Governance Fellow hos Simon Institute, og tar en mastergrad i internasjonal sikkerhetspolitikk ved Sciences Po Paris. Han har tidligere arbeidet med nedrustning i FN og forsket på atomvinter-scenarier.

Tellef Solbakk Raabe er tidligere seniorrådgiver i Langsikt. Han har en master- og doktorgrad i sosiologi fra Universitetet i Cambridge og har de siste årene jobbet som forsker ved SNF på Norges Handelshøyskole. Han har forsket på hvordan medier og teknologi virker i samfunnet og er særlig opptatt av politisk økonomi, regulering, strategi og innovasjon.

Peder Skjelbred er tidligere rådgiver på kunstig intelligens og bioteknologisikkerhet i Langsikt. Han er stipendiat i etikk ved Universitetet i Oslo, og er fra høsten 2026 Visiting Fellow ved Department of Philosophy ved Harvard. Han har en mastergrad i filosofi fra University of Oxford og har også studert samfunnsøkonomi ved UiO.

OM LANGSIKT

Langsikt er en uavhengig, ideell tankesmie for mer langsiktig politikk. Vi forener forskning og politikkutvikling for å bidra til at samfunnets ressurser brukes der de gjør mest nytte. Vårt fokus er på store globale utfordringer som kunstig intelligens, pandemier og ekstrem fattigdom. Vi utvikler ny politikk, bidrar i den offentlige samtalen og er en møteplass for alle som er interessert i langsiktig og kunnskapsbasert politikk. Vår finansiering kommer fra stiftelser og privatpersoner, uten at økonomiske bidrag gir noen innflytelse på hva vi skal mene. Les mer om oss og vårt arbeid på www.langsikt.no.

Konseptnotat for et norsk KI-sikkerhetsorgan
Dato: 07.09.2026
Redaksjonelt ansvar: Aksel Braanen Sterri, fagsjef

Har du innspill til notatet, ta gjerne kontakt på kontakt@langsikt.no.