UNIVERSITETET
I OSLO

Norges forskningsråd
1324 LYSAKER

Dato: 10.09.2026
Deres ref.:
Vår ref.: 26/20158-12

Forskningsrådet - UiOs innspill retningslinjer for bruk av KI i forskningssøknader og søknadsevalueringer

Universitetet i Oslo (UiO) viser til Forskningsrådets arbeid med retningslinjer for bruk av kunstig intelligens (KI) i forskningssøknader og søknadsbehandling, og setter pris på muligheten for å gi innspill i tidlig fase.

Innledningsvis for orientering, UiO har egne retningslinjer for KI og disse har som overordnet prinsipp at KI er IT-tjenester og underlagt de samme rammer som gjelder for universitetets øvrig IT-virksomhet når det gjelder IT-sikkerhet, personvern og andre juridiske og regulatoriske bestemmelser[1]. UiO har også utarbeidet en utprøvingsnettside om agentisk KI, og har en hjelpe-nettside for Kunstig intelligens (KI) ved UiO ment for alle ved UiO som skal ta i bruk KI.

Metodikk og kunnskapsgrunnlag
Dette innspillet er forankret i en analyse av retningslinjer fra sentrale aktører som EUs ERA Forum, European Research Council (ERC), den finske nemnda for forskningsetikk (TENK), Swiss National Science Foundation (SNSF), det nederlandske forskningsrådet (NWO) og skandinaviske forskningsråd Vetenskapsrådet, Finlands akademi og NordForsk. Vi har også orientert oss i de nasjonale dokumentene NTNU og UiT har på dette området.

Analysen har særlig fokusert på:
1. deklarasjonsplikt for KI-bruk
2. forbud mot KI i evaluering/fagfellevurdering
3. konfidensialitet og immaterielle rettigheter (IP)
4. regulering av agentisk KI og miljøansvar

UiO mener det er avgjørende at norske retningslinjer harmoniseres med eksisterende retningslinjer i land vi naturlig sammenlikner oss med. Innspillene har lagt seg inntil de internasjonale standardene i Horizon Europe og ERA-retningslinjene, og et noe strengere grep for deklarasjonsplikt og konfidensialitet. Men det er spørsmål om man i Norge trenger å være strengere enn de skandinaviske landene. En mellomting vil kunne være å legge seg på linje med Danmark når det gjelder deklarasjonsplikt, der man må angi hvordan man har brukt KI, men at denne informasjonen ikke (automatisk, i hvert fall) videreformidles til evaluator.

Vedlagt finnes svar på de tre områdene som Forskningsrådet etterspør innspill på:
- søkernes bruk av KI i søknader

[1] Generelt om kunstig intelligens (KI) ved Universitetet i Oslo - Universitetet i Oslo

Universitetsdirektør

Postadresse: Postboks 1072 Blindern 0316 Oslo NO
Besøksadresse: Problemveien 7,Lucy Smiths hus 2.etg
E-post: postmottak@uio.no
Nettside: www.uio.no
Org.nr.: 971 035 854

- eksterne evaluatorers bruk av KI
- interne saksbehandleres bruk av KI

For hvert område nevnes prinsippene først, etterfulgt av hva vi foreslår at Forskningsrådet kan bidra med. Vi har her tatt utgangspunkt i følgende ansvarsfordeling:

> 1) Forskningsrådet: Tydeliggjør kravene til søknadsutforming og – evaluering rådet har for å bidra til at forskeres og institusjoner kan følge opp disse, sikrer sikker infrastruktur for evaluering og etablerer tekniske barrierer mot manipulasjon.
> 2) Forskere: Personlig ansvar for integritet, deklarasjonsplikt for bruk av generativ KI (med unntak av enkel språkvask og fomratering), verifiseringsplikt, beskyttelse av konfidensialitet/IP og tilstrekkelig KI-kompetanse.
> 3) Institusjoner: Opplæring og støtte i KI-kompetanse, sikker KI-infrastruktur (enterprise/lukkede systemer), interne retningslinjer og rutiner for overvåking/risikovurdering, samt miljøansvar ved KI-tung forskningsbruk

Vi håper at disse innspillene kan være nyttige i Forskningsrådets videre arbeid med å utvikle retningslinjer for ansvarlig bruk av KI i norsk forskningsfinansiering. Vi bidrar gjerne videre i dialogen, og ser fram til å se Forskningsrådets forslag.

Med hilsen

Arne Benjaminsen	Ingrid Sogner
Universitetsdirektør	Avdelingsdirektør

Dokumentet er elektronisk godkjent og har ingen underskrifter.

Vedlegg:

Saksbehandler: Vibeke Alm

Vedlegg 1- Forskningsrådets arbeid med retningslinjer for bruk av KI i forskningssøknader og søknadsevaluering–UiOs innspill

a. Søkernes bruk av KI i utforming av prosjektsøknader

- Menneskelig ansvar og forfatterskap
  Forskerne har alltid det fulle ansvaret for innholdet i søknaden inkludert den faglige kvaliteten, originaliteten, korrektheten og påliteligheten, uavhengig av om KI er benyttet. KI kan aldri oppføres som medforfatter, da forfatterskap krever evne til å ta både vitenskapelig og etisk ansvar.

  For å bidra til dette, oppfordrer vi Forskningsrådet til å:

  Tydeliggjøre krav i utlysninger: Forskningsrådet må klargjøre kravene til søknadsutforming slik at det kommer tydelig frem at KI ikke kan krediteres som forfatter, og at forskeren selv står ansvarlig for alt innhold.

  Innføre deklarasjonsplikt i søknadsskjemaet: Det anbefales at Forskningsrådet etablerer et eget felt i søknadsskjemaet (alternativt i de administrative skjemaene), der søkere må deklarere all bruk av generativ KI. Dette bidrar til transparens og ansvarliggjøring.

  Implementere standardiserte advarsler: UiO foreslår at Forskningsrådet inkluderer en kort, standardisert advarsel i alle utlysninger og veiledningstekster. Denne bør presisere at:
  - *Forskeren har det fulle ansvaret for søknadens korrekthet og pålitelighet (reliabilitet).*
  - *Brudd på retningslinjene for KI-bruk kan vurderes som brudd på forskningsetiske retningslinjer og forskningsintegritet.*
  - *Det foreligger en risiko knyttet til å laste opp søknadsmateriale til eksterne KI-tjenester, spesielt med tanke på konfidensialitet, personvern og immaterielle rettigheter.*

  Veiledning om verifisering: Forskningsrådet bør understreke forskerens plikt til manuelt å validere alt KI-generert innhold, inkludert faktaopplysninger og referanser, for å unngå «hallusinasjoner».

  Sikre menneskelig skjønn i evalueringen: For å speile ansvaret som kreves av søkerne, må Forskningsrådet også stille krav til sine eksterne evaluatorer. Dette innebærer å innføre en standarderklæring der evaluatorer bekrefter at de har vurdert innholdet basert på eget faglig skjønn og ikke delegert dette til KI-verktøy.

  Sperre mot skjulte instrukser («hidden promots»). Det bør innføres filtre for å fange opp skjulte instrukser (f.eks. instrukser om å gi positive evalueringer) som er lagt inn i søknadsteksten for å manipulere administrative KI-systemer. Administrative KI-verktøy i lukkede systemer bør nettopp brukes til å avdekke slike manipulasjonsforsøk.

  Oppsummert er Forskningsrådets ansvar å tydeliggjøre rammene, legge til rette for deklarering og advare om konsekvensene ved manglende overholdelse av det personlige forskeransvaret, samt å påse at retningslinjene som etableres, faktisk håndheves.

Agentisk KI – risikoklassifisering
Retningslinjene bør ta høyde for agentisk KI – systemer som utfører kjedede autonome funksjoner og komplekse forskningsoppgaver selvstendig.

For slike systemer anbefales det at Forskningsrådet innfører en egen risikokategori hvor prosjekter må:

- *beskrive kontrollmekanismer og menneskelig overoppsyn («human oversight»)*
- *redegjøre for hvordan irreversible handlinger mot kritiske data forhindres eller loggføres.*
- *imøtekomme krav om spesifikke kontrollmekanismer («kill-switches») som sikrer at menneskelig overstyring kan skje i sanntid ved kritiske operasjoner.*

Deklarasjon av all generativ KI-bruk

Det bør innføres deklarasjonsplikt for all bruk av generativ KI. Siden KI-teknologien utvikler seg raskt, er det avgjørende at utformingen av søknadsskjemaet og deklarasjonsfeltet revideres jevnlig for å fange opp denne dynamikken.

Vi anbefaler at Forskningsrådet stiller som et absolutt krav at deklarasjonsfeltet i søknadsskjemaet gir mulighet til å spesifisere tre punkter:

1. *Hvilke deler av søknaden generativ KI er brukt i.*
2. *Hva teknologien konkret er brukt til.*
3. *Hvilket KI-verktøy som er benyttet.*

Det må klargjøres tydelig at deklarasjonsplikten utelukkende gjelder generativ KI, mens etablerte maskinlæringsmetoder og statistiske analyser som utgjør standard faglig praksis (for eksempel innenfor mednat-miljøene), faller utenfor denne spesifikke plikten for å unngå overrapportering. Manglende deklarasjon ved generativ KI-bruk kan vurderes som brudd på forskningsetikk og forskningsredelighet.

Syntetiske data, verifiseringsplikt og språk

Dersom KI brukes til å generere syntetiske data, bør disse skilles klart fra empiriske data, og mulige skjevheter (*biaser)* rapporteres transparent. Søkere må validere alt KI-generert innhold (faktaopplysninger, referanser, kode, modeller) for å unngå «hallusinasjoner». UiO oppfordrer til bruk av verktøy som ivaretar norsk og samisk som fagspråk, slik at språklig støtte ikke svekker nasjonale fag- og forskningsspråk.

Forskningsrådet bør stille krav til transparens og validering ved bruk av KI-genererte data og innhold.

Miljøansvar

Bruk av tunge KI-modeller og agenter bør stå i rimelig og proporsjonalt forhold til det vitenskapelige utbyttet. I tråd med anbefalingene i ERA 2026 må ressurskrevende KI-bruk minimeres for å redusere miljøavtrykket.

Forskningsrådet kan eksplisitt inkludere miljø- og ressursbruk som et vurderingskriterium ved KI-tung metodebruk.

Konfidensialitet, immaterielle rettigheter (IP) og standardisert advarsel

Det bør slås eksplisitt fast at det er strengt forbudt å laste opp upubliserte ideer og forskningsplaner, konfidensielle data eller søknadstekster til åpne eller eksterne KI-tjenester uten eksplisitt tillatelse. Dette gjelder også forskernes egne produksjoner, da slik opplasting kan medføre datalekkasjer, tredjepartsbruk til modelltrening, tap eller svekkelse av IP, samt brudd på personvern og forskningsetiske normer.

UiO anbefaler at Forskningsrådet inkluderer en kort, standardisert advarsel i alle utlysninger og veiledningstekster, som:

- *Tydelig advarer mot opplasting av søknads- og forskningsdata til åpne KI-systemer*
- *Forklarer konsekvenser for personvern, konfidensialitet og IP*
- *Presiserer at brudd kan vurderes som brudd på forskningsetiske retningslinjer og forskningsintegritet*

b. Eksterne eksperter/evaluatorers bruk av KI

For å ivareta rettferdighet og konfidensialitet bør evalueringen preges av en restriktiv linje basert på prinsippet om *ingen delegasjon* (*«non-delegation»).* Bruk av KI i vurderinger vil øke faren for konformitet og tap av originalitet, da generative verktøy baserer seg på eksisterende mønstre og etablert kunnskap.

KI-bruk i både skriving og evaluering medfører en stor risiko for en systematisk tiltrekning mot det konvensjonelle – mot de vanligste argumentasjonstypene, de mest etablerte teoretiske rammeverkene og de dominerende forskningstradisjonene innenfor et felt. Et KI-assistert vurderingssystem vil systematisk favorisere søknader som samsvarer med mønstre for tidligere finansiert forskning, og dermed selektere mot originalitet og nyskaping.

Ingen delegasjon av faglig skjønn
Den menneskelige fagfellevurderingen er en grunnpilar som må bestå, og faglige vurderinger må bygge på ekspertenes egne kvalifiserte faglige skjønn. KI skal ikke brukes til å evaluere innhold, gi poengsummer, eller oppsummere søknader for å spare tid. Dette innebærer en uakseptabel delegasjon av evaluators ansvar og kan svekke kvalitet og individuell vurdering.

IP-beskyttelse og konfidensialitet
Det må være et absolutt forbud mot å laste opp søknader eller upublisert materiale i åpne KI-modeller for å beskytte søkernes immaterielle rettigheter og konfidensialitet. UiO anbefaler en restriktiv praksis der bruk av KI i søknadsvurderinger i utgangspunktet ikke tillates. Dersom Forskningsrådet likevel velger å åpne for slik bruk, må dette utelukkende skje innenfor en lukket og sikker infrastruktur som Forskningsrådet selv kontrollerer, og som evaluatorene forpliktes til å bruke.

UiO anbefaler at Forskningsrådet vurderer tekniske sikkerhetstiltak som begrenser evaluatorers mulighet til å kopiere tekst direkte ut av søknadene, for å beskytte søkernes IP.

Oppmerksomhet mot «kunstig autoritet»
Evaluatorer må gjøres oppmerksomme på risikoen for tekst som fremstår som troverdig og velskrevet, men som mangler reell vitenskapelig substans – såkalt «kunstig autoritet». *Forskningsrådet bør utvikle en kort veiledning om hvordan slik tekst kan gjenkjennes og vurderes kritisk.*

Intervjufase: *Det bør innføres et absolutt forbud mot å bruke KI-verktøy under selve søknadsintervjuene for å sikre en genuin menneskelig vurdering i samtid.*

Transparens fra finansiør: *Forskningsrådet bør selv være åpne om sin egen eventuelle bruk av KI i evalueringsprosessen for å sikre tillit.*

Standarderklæring: UiO anbefaler at *alle evaluatorer må godta en erklæring der de bekrefter at de ikke har brukt KI til oppsummering eller poengsetting, ikke har lastet opp tekst til åpne verktøy, og har vurdert innholdet basert på eget faglig skjønn.*

Algoritmisk rettferdighet: *Retningslinjene bør adressere risikoen for at evaluatorer ubevisst favoriserer søknader som er språklig «perfeksjonert» av KI, slik at originale ideer ikke nedvurderes til fordel for mer strømlinjeformede, men mindre nyskapende prosjekter.*

c. Interne saksbehandleres bruk av KI i søknadsbehandlingen

UiO ser positivt på at administrasjonen tar i bruk interne KI-verktøy i sikre, lukkede systemer, men understreker at denne bruken strengt må begrenses til administrative støtteoppgaver. KI må aldri benyttes til faglige vurderinger, innholdsmessige oppsummeringer, prioriteringer eller beslutninger om tildeling av forskningsmidler. Saksbehandling skal følge prinsipper om transparens, etterprøvbarhet og sikkerhet. Saksbehandlere som bruker KI-verktøy i søknadsbehandling, må ha tilstrekkelig forståelse av verktøyenes systematiske tendenser og begrensninger.

Forståelse av metodiske svakheter og strukturell bias

Saksbehandlingen må ta inn over seg at store språkmodeller har en dokumentert systematisk bias mot deduktiv resonnering – mot argumenter som beveger seg fra etablerte premisser til konklusjoner. Abduktiv resonnering – som beveger seg fra observasjon til det beste tilgjengelige forklaringsrammeverket – er imidlertid den dominerende argumentasjonsformen i store deler av humanistisk forskning, fortolkende samfunnsvitenskap og tidligfase empirisk forskning. En hypotese eller prosjektdesign som holder flere levende tolkninger åpne, vil fremstå som metodologisk svakt formulert for et KI-system.

Søknader fra disipliner der fortolkende, abduktiv eller genuint åpen faglig tilnærming er normen, er systematisk utsatt for å bli feilrepresentert i KI-assistert behandling. Dette kan føre til en *strukturell bias mot hele fagtradisjoner.* Uten tilstrekkelig kompetanse til å identifisere og kompensere for denne skjevheten *(biasen)*, risikerer saksbehandlere å produsere vurderingsgrunnlag som systematisk favoriserer søknader med konvensjonell deduktiv struktur, og som undervurderer søknader fra der den faglige verdien ligger i en annen type argumentasjonslogikk.

Tydelig oppdeling av administrative oppgaver

UiO anbefaler derfor at Forskningsrådet, i sitt arbeid med retningslinjer for saksbehandlere, tydelig skiller mellom tillatt og forbudt bruk:

1. *Operasjoner der KI-assistanse er pålitelig og forsvarlig (tillatt bruk):*
    - *Kontroll av formelle krav, sjekklisteverifisering, terminologisk konsistens, formatering og kontroll av tegngrenser,*
    - *Bruk av lukkede administrative verktøy til å avdekke udeklarert KI-bruk, manipulasjonsforsøk eller 'skjulte instrukser' (hidden prompts).*
2. *Operasjoner der KI-assistanse innebærer strukturell risiko (forbudt bruk):*
    - *Innholdsmessig oppsummering og komprimering av søknadsinnhold, vurdering av argumentasjonsstyrke, og identifisering av faglig originalitet.*
    - *KI skal aldri brukes til faglige vurderinger eller prioriteringer av søknader, og det må understrekes at administrasjonen ikke skal fatte beslutninger om tildeling av forskningsmidler på bakgrunn av KI-genererte analyser.*

Lukkede systemer og tekniske grep

Saksbehandling må skje i sikre, lukkede, risikovurderte KI-systemer som Forskningsrådet selv kontrollerer. Som en operasjonell anbefaling bør Forskningsrådet implementere filtre og tekniske løsninger for å fange opp skjulte instrukser (*«hidden prompts»)*, for eksempel gjennom skanning av metadata, konvertering til ren tekst, og implementering av tekniske filtre.

Etterprøvbarhet og arkivering

Relevante samtalelogger knyttet til saksbehandling bør arkiveres sikkert, slik at KI-bruk kan dokumenteres og etterprøves. Dette sikrer at man kan dokumentere at KI ikke har påvirket konklusjoner på en måte som bryter med likebehandling, konfidensialitet eller faglig skjønn.

KI-kompetanse og revisjon

Forskningsrådet bør definere minimumskrav til KI-kompetanse («*AI literacy*») for saksbehandlere og sikre regelmessig opplæring. Retningslinjene bør revideres jevnlig i lys av den teknologiske utviklingen og kravene i AI Act, som understreker behovet for solid KI-kompetanse hos ansatte og beslutningstakere.