NESH (2026):
Ansvarlig bruk av generativ kunstig intelligens i forskning:
en forskningsetisk veileder

Engelsk:
Responsible use of generative artificial intelligence in research:
a guide to research ethics

UTKAST

Innhold

FORORD 3
INNLEDNING 4
Normer og verdier 4
Risiko og forsvarlighet 5
Roller og ansvar 7
Relevante ressurser 7
1. FORSKNINGENS INTEGRITET 10
1.1. Pålitelighet 10
1.2. Ansvarlighet 12
1.3. Åpenhet 13
1.4. Institusjonenes ansvar 14
2. FORSKNING PÅ MENNESKER 16
2.1. Respekt 16
2.2. Beskyttelse 17
2.3. Rettferdighet 18
2.4. Institusjonenes ansvar 19
3. FORSKNING OG SAMFUNN 20
3.1. Offentlighet 21
3.2. Usikkerhet 22
3.3. Forsvarlighet 23
3.4. Institusjonenes ansvar 24

UTKAST

FORORD

Veilederen er laget av Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH) og supplerer *Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora* (2023) og *Forskningsetisk veileder for internettforskning* (2019).[1]

Veilederen drøfter forskningsetiske spørsmål og dilemmaer knyttet til bruk av generativ kunstig intelligens i forskning, med henvisning til grunnleggende forskningsetiske verdier som sannhet, menneskeverd og demokrati. Den fremhever at generativ kunstig intelligens reiser nye utfordringer, spesielt knyttet til risiko, usikkerhet og forsvarlighet. Formålet med veilederen er å fremme god vitenskapelig praksis, bidra til kritisk refleksjon over forskningsetiske spørsmål og dilemmaer og gi råd og veiledning om ansvarlig bruk.[2]

Veilederen er utviklet for forskning, men den kan også være relevant for høyere utdanning. Den er utviklet for samfunnsvitenskap og humaniora, men den kan også brukes på andre fagområder. Arbeidet ble påbegynt i 2025 av den forrige NESH-komiteen (2022–2025), og sluttføres av den nåværende (2026–2029). NESH bistår gjerne med råd og veiledning om ansvarlig bruk av generativ KI i forskning, og komiteen kan behandle saker som reiser prinsipielle spørsmål.

Oslo, 29.09.2026

NESH (2026–2029): Kari Steen-Johnsen (leder), Anette Bringedal Houge, Anne Line Wittek, Annette Alstadsæter, Hanne Hagtvedt Vik, Hans Jakob Walnum, Jakob Mattias Emanuel Maliks, Jesper Aagaard Petersen, Karl Harald Søvig, Kristian Sandbekk Norsted, Lars Frers, Monika Abels, Per Henning Uppstad, Rani Lill Anjum, Shaher Ahmmad Ibrahim Shalfawi, Steinar Sneås Skauge og Vidar Enebakk (sekretariatsleder).

[1] NESH (2023); NESH (2019). Det vises fortløpende til relevante deler (A–E) eller punkter (1–50) i retningslinjene, for eksempel (NESH, 2023: 4), samt relevante uttalelser, for eksempel (NESH, 2026/140).
[2] NENT (2019); FEK (2020).

INNLEDNING

Kunstig intelligens (KI) er i rask utvikling og tilbyr store muligheter. Samtidig medfører KI betydelig risiko og usikkerhet for forskningen, for mennesker og for samfunnet. KI kan svekke forskningens integritet, vernet av forskningsdeltakerne og tilliten til forskning. Derfor er det viktig å fremheve og sikre grunnleggende forskningsetiske verdier som sannhet, menneskeverd og demokrati.

KI er et bredt begrep, som omfatter ulike teknologier, systemer og verktøy. Veilederen fokuserer spesielt på bruk av *generativ* KI i forskning. I praksis dreier det seg om ulike verktøy som kan generere nytt innhold i form av tekst, bilder, video eller annet medieinnhold, basert på bestemte instruksjoner eller prompter. Generativ KI kan inngå i alle faser av forskningen, og brukes for eksempel til datainnhenting, koding, analyse, litteraturarbeid og skriving.

Det er viktig å skille mellom bruk av ulike typer generativ KI i forskning. Noen KI-verktøy er åpent tilgjengelige for allmenn bruk. Samtidig utvikles det stadig nye verktøy spesifikt for forskning, med tanke på å sikre normer om pålitelighet, ansvarlighet og åpenhet. Forskjellige KI-verktøy reiser ulike forskningsetiske spørsmål og dilemmaer. Derfor er det nødvendig med konkrete vurderinger og forsvarlige begrunnelser ved all bruk av generativ KI i forskning.

Normer og verdier

Selv om utviklingen går fort, og de mulige konsekvensene for mennesker og samfunn er store, endres ikke de grunnleggende forskningsetiske normene og verdiene. Snarere er det behov for å formulere nye råd og retningslinjer i møte med nye utfordringer.

NESHs *Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora* (2023) bygger på tre ulike sett med normer og verdier som er utviklet over tid og forankret i forskerfellesskapet. Det første er den ufravikelige forpliktelsen til *sannhet*, knyttet til normer om pålitelighet og åpenhet. Det omfatter også metodologiske normer om saklighet og etterrettelighet, samt institusjonelle normer som skal sikre at forskning er uavhengig og kritisk. Disse normene skal sikre forskingens integritet. Det andre handler om beskyttelse av mennesker, knyttet til *menneskeverdet* og normer om respekt, beskyttelse og rettferdighet. Det tredje er mer allmenne normer, avledet av samfunnets krav og forventninger og knyttet til et bredere samfunnsansvar. Alle mennesker har rett til å få del i den vitenskapelige fremgangen og dens goder, og et *demokrati* er avhengig av forskning som kilde til pålitelig kunnskap, til

myndiggjøring og til utvikling av ansvar, autonomi og integritet. Forskningsetisk ansvarlighet kan bidra til å sikre forskningens tillit og troverdighet. (NESH, 2023: Innledning).

Veilederen er organisert i tre deler, og bygger på inndelingen i NESHs retningslinjer. Del 1 omhandler forskningens integritet, med utgangspunkt i sannhet som grunnleggende verdi. Bruk av generativ KI kan utfordre forskningens pålitelighet og svekke tilliten til forskningens resultater. Derfor er det viktig å reflektere over hvordan bruk av KI kan utfordre grunnleggende normer som *pålitelighet, ansvarlighet* og *åpenhet* (NESH, 2023: A).

Del 2 omhandler forskning på mennesker, og her er menneskeverdet den grunnleggende verdien, knyttet til personvern og privatlivets fred. Bruk av generativ KI kan utfordre frihet og selvbestemmelse, konfidensialitet og anonymitet, samt vernet mot skade og urimelige belastninger. I denne sammenheng er det viktig å reflektere over grunnleggende normer som *respekt*, *beskyttelse* og *rettferdighet* (NESH, 2023: B og C).

Del 3 omhandler forskning og samfunn, med vekt på demokrati som grunnleggende verdi. Bruk av generativ KI påvirker både forskningen og samfunnet, samtidig som utviklingen av teknologien er dominert av store selskaper og preget av kommersielle interesser, skarp konkurranse, uoversiktlig regulering og fravær av åpenhet. Det reiser nye spørsmål om vitenskap og samfunn knyttet til grunnleggende normer knyttet til håndtering av *risiko, usikkerhet* og *forsvarlighet* (NESH, 2023: D og E).

Risiko og forsvarlighet

Forskere bør være åpne og nysgjerrige på de mulighetene generativ KI tilbyr. KI kan effektivisere forskningens prosesser og gi nye resultater og innsikter, som ellers ikke kan komme samfunnet til gode. KI åpner for enklere tilgang til data og kilder, noe som kan redusere skjevheter i det globale forskerfellesskapet. Og KI kan bidra i selve skriveprosessen og til å teste ut argumenter og ideer, noe som kan utjevne forskjeller mellom forskere med ulik språkbakgrunn eller med lese- og skrivevansker. Det er viktig at forskere møter og utforsker disse mulighetene og utforsker dem på en ansvarlig måte.

Forskere bør også være ydmyke og opptre med varsomhet. Ansvarlig bruk av generativ KI i forskning innebærer å vurdere risiko og usikkerhet. Veilederen anlegger derfor en *risikobasert tilnærming*. NESH mener det er behov for å tenke bredt og helhetlig om risiko i forbindelse med generativ KI, og har formulert en liste med fem punkter som forskere bør forholde seg til i slike forskningsetiske risikovurderinger:

UTKAST

a) ***Forskningens integritet:*** Generativ KI kan bidra til å endre grunnleggende vitenskapelige praksiser som tenkning, skriving, tolkning, analyse og formidling. Bruk av KI i forskning kan også utfordre grunnleggende vitenskapelige normer og verdier, noe som svekker forskningens tillit i samfunnet. Derfor må forskere vurdere risiko for forskningens integritet (Del 1 i denne veilederen).

b) ***Personvern:*** KI kan utfordre menneskeverd og personvern, herunder beskyttelse av privatliv og personlig integritet. Ulike former for innsamling og gjenbruk av informasjon kan bidra til reidentifisering, bakveisidentifisering og urimelig belastning. Derfor må forskere vurdere risiko for personvern i bred forstand (Del 2).

c) ***Svakstilte og sårbare grupper:*** KI-generert kunnskap har systematiske skjevheter kan bidra til å forsterke sosiale stereotypier og diskriminering, selv om hensynet til den enkeltes personvern er ivaretatt. Derfor må forskere vurdere risiko for svakstilte og sårbare grupper, spesielt ved forskning på barn (Del 2).

d) ***Natur, miljø og klima:*** KI har store og uforutsigbare konsekvenser for naturen, knyttet til enormt forbruk av energi og vann til drift av datasentre, men også i form av arealbeslag, nedbygging av naturområder, utvinning av mineraler og økt elektronisk avfall. Derfor må forskere vurdere risiko for natur, klima og miljø (Del 3).

e) ***Samfunn, offentlighet og demokrati:*** KI har store og uforutsigbare konsekvenser for samfunnet, knyttet til spørsmål om makt, interesser, eierskap, tilgang og sikkerhet. Det kan svekke forskningens frihet, åpenhet og uavhengighet, noe som kan få konsekvenser for forskningens tillit og troverdighet. Derfor må forskere vurdere risiko for samfunn, offentlighet og demokrati (Del 3).

NESH presiserer at formålet med en vurdering av risiko og usikkerhet ikke er å begrense eller hindre forskning. Tvert imot handler det om å bidra til å begrunne at forskningen er etisk forsvarlig i tråd med prinsipper om akademisk frihet og ansvar.[3] Ansvarlig bruk av KI i forskning innebærer også at forskere må vurdere om det ikke er forsvarlig å bruke bestemte KI-verktøy, spesielt i prosjekter med uakseptabel eller høy risiko.

[3] Universitets- og høyskoleloven, § 2–2.

Roller og ansvar

Ansvaret for forskningsetikken er fordelt mellom forskere, forskningsinstitusjoner og andre forskningsaktører, for eksempel finansiører, oppdragsgivere og samarbeidspartnere i offentlig eller privat regi (NESH, 2023: Introduksjon).

Forskere har et selvstendig ansvar for å sikre forskningens integritet og forsvarlighet.[4] Ved bruk av generativ KI i forskning må forskere vurdere etiske spørsmål og dilemmaer, og de må gjøre rede for sine valg og prioriteringer. Det er spesielt viktig med pålitelighet, ansvarlighet og åpenhet i alle faser av prosessen, fra planlegging til publisering, for å sikre forskningens integritet. Forskere må ivareta respekt, beskyttelse og rettferdighet, for å sikre forskningsdeltakernes menneskeverd. Og forskere må vurdere risiko og kommunisere usikkerhet, for å sikre forskningen forsvarlighet.

Forskningsinstitusjoner skal sikre at forskningen ved institusjonen skjer i henhold til anerkjente forskningsetiske normer.[5] De må legge til rette for at forskere får opplæring og veiledning om ansvarlig bruk av generativ KI, og de må ha rutiner for å vurdere risiko ved bruk av ulike KI-verktøy og -systemer for personvern og informasjonssikkerhet. Mange av disse konkrete utfordringene blir i praksis håndtert av studieadministrasjon, bibliotekarer og databehandlere. Derfor må institusjonene ha en helhetlig tilnærming i arbeidet med å utvikle rutiner og retningslinjer for ansvarlig bruk av KI i forskning.[6]

Andre aktører som finansiører, oppdragsgivere og samarbeidspartnere, har et medansvar for i sikre at forskningsetikken blir respektert og ivaretatt. I dag er det stor press både i forskning og i offentlig og privat sektor på utvikling og bruk av generativ KI, men andre aktører må unngå å gi føringer og insentiver som truer forskningens integritet og forsvarlighet, knyttet til data, metode, analyse, konklusjoner og resultater (NESH 2023: D).

Relevante ressurser

Veilederen er utformet i samspill med ressurser som er utviklet internasjonalt, særlig innenfor EU-området. NESH vil spesielt kommentere fire relevante ressurser som veilederen bygger på, i tillegg til retningslinjene og forskningsetikkloven.

ALLEAs *European Code of Conduct for Research Integrity* (2023) er et felles rammeverk for forskningsintegritet i Europa, basert på grunnleggende prinsipper som pålitelighet, ærlighet, respekt og ansvarlighet. Retningslinjene omfatter etiske, juridiske og

[4] Forskningsetikkloven, § 4.
[5] Forskningsetikkloven, § 5.
[6] FEK (2023).

vitenskapelige forpliktelser, og de henvender seg til både forskere og forskningsinstitusjoner. Her fremgår det at god forskningspraksis innebærer åpenhet og redegjørelse for bruk av generativ KI både i forskning og i vurdering av andres arbeid. Forsøk på å skjule bruk av KI er en uakseptabel praksis som i alvorlige tilfeller kan innebære uredelighet.[7]

Disse grunnleggende prinsippene er videreutviklet i EU-kommisjonens *Living Guidelines on the Responsible Use of Generative AI in Research* (2026), som er mer juridisk forankret og har en bredere innretning som også omfatter personvern, datalagring, sikkerhet og immaterielle rettigheter. Det henvender seg til forskere, forskningsinstitusjoner og forskningsfinansiører i offentlig og privat sektor. Her formuleres en rekke konkrete råd om ansvarlig bruk av generativ KI i forskning.[8]

Living Guidelines bygger på EUs *AI Act* (2024), som er et overordnet juridisk regelverk i Europa. Loven er primært rettet mot utviklere som ønsker å introdusere KI-produkter i det europeiske markedet, og den omfatter ikke forskning direkte. Loven har imidlertid en risikobasert tilnærming som er relevant også for forskning, basert på en inndeling i fire risikonivåer med økende krav og regulering: Minimal, begrenset, høy og uakseptabel risiko.[9]

NESH mener det er nødvendig å formulere et tydeligere rammeverk internasjonalt for vurdering av risiko, usikkerhet og forsvarlighet knyttet til bruk av generativ KI i forskning. Europarådets *Rammekonvensjon om kunstig intelligens og menneskerettigheter, demokrati og rettsstat* (2024), som også inngår i forslaget til ny norsk KI-lov sammen med *AI Act*, er spesielt relevant i denne sammenhengen[10] Rammekonvensjonen formulerer et bredt perspektiv på risiko i en samfunnsmessig kontekst og drøfter mulige negative konsekvenser for menneskerettigheter, demokrati og rettstaten knyttet til grunnleggende normer som pålitelighet, ansvarlighet, åpenhet, menneskeverd, personvern, rettferdighet, risiko, usikkerhet og forsvarlighet.

Veilederen er formulert innenfor rammen av de fire relevante ressursene, og med de juridiske regelverkene som referansepunkter, men den formulerer *ikke* retningslinjer og regler som skal følges punkt for punkt. Den forsøker snarere å avgrense et rom for refleksjon hvor forskere kan gjøre sine egne forskningsetiske vurderinger. Veilederen inviterer til en selvstendig forskningsetisk refleksjon med utgangspunkt i NESHs retningslinjer og i tråd med

[7] ALLEA (2023), 2.3, 2.8 og 3.1.
[8] EU-kommisjonen (2026).
[9] EU (2024), *AI Act*, art. 25–27.
[10] Utkast til ny lov om kunstig intelligens (30.09.2025).

prinsipper om individuell akademisk frihet. Samtidig fremheves institusjonenes ansvar i tråd med prinsipper om institusjonell autonomi og selvregulering.

I tillegg til de nevnte finnes det flere relevante juridiske regelverk, både i Norge og i Europa, for eksempel om behandling av personopplysninger og helseopplysninger.[11] Spørsmål om opphavsrett og immaterielle rettigheter, som kan være regulert i kontrakter med oppdragsgivere, samarbeidspartnere, finansiører eller forlag.[12] Det kan også være nødvendig å vurdere risiko og usikkerhet i lys av regelverk for bærekraft, sikkerhetspolitikk og ansvarlig internasjonalt samarbeid.[13] Forskningsinstitusjonene må sikre etterlevelse av lovverk, og forskere må sette seg inn i rutiner og retningslinjer ved egen institusjon og på eget fagområde.

UTKAST

[11] Personopplysningsloven; Helseforskningsloven.
[12] Åndsverksloven; Patentloven.
[13] FNs bærekraftsmål; Sikkerhetsloven.

1. FORSKNINGENS INTEGRITET

Bruk av generativ KI i forskning gir nye muligheter, men kan også endre forskningens karakter og kvalitet. Det kan bidra til å svekke forskningens integritet og tillit, både i forskerfelleskapet og i samfunnet. Derfor er det nødvendig å sikre forsvarlig bruk av KI i forskning, både i forskningens resultater og prosesser og i forskerfellesskapet.

Den grunnleggende normen i forskning handler om etterlevelse av vitenskapens sannhetsforpliktelse. Den er en forutsetning for forskningens kvalitet og pålitelighet, og for at forskere kan sikre tillit og troverdighet (NESH 2023: Innledning). Å overholde sannhetsforpliktelsen kan imidlertid være utfordrende ved bruk av generativ KI i forskning. Økt akselerasjon av forskningsprosesser kan endre selve forståelsen av hva som er god vitenskapelig praksis. Og økt press fra myndigheter og andre eksterne aktører kan sette forskningen under press. Derfor er det viktig å fremheve grunnleggende forskningsetiske normer om *pålitelighet, ansvarlighet* og *åpenhet*.[14]

Det er vanlig å skille mellom *vesentlig* og *ikke vesentlig* bruk av generativ KI i forskning.[15] Bidraget er vesentlig dersom det påvirker forskningens resultater, for eksempel i formulering av formål, utvikling av hypoteser, valg av metoder, tolkning av materialet, manipulasjon av illustrasjoner eller utledning av konklusjoner, prediksjoner og konsekvenser. Generelt sett er bidraget ikke vesentlig dersom generativ KI bare blir brukt til språkvask eller for å finne bakgrunnsinformasjon. Her er det imidlertid nyanser og forskjeller mellom fagområder. I humaniora og samfunnsvitenskap, hvor utvalg av litteratur og nyanser i språk åpenbart påvirker både analyse og resultat, kan også slik bruk være vesentlig. Derfor må all bruk av generativ KI drøftes åpent, og bidraget fra KI-verktøy må diskuteres kritisk, på en måte som også gjør det forståelig for ikke-eksperter og eventuelle forskningsdeltakere. Det er viktig for å sikre pålitelighet og tillit i forskningen og i samfunnet.

1.1. Pålitelighet

Forskerens tillitsverdighet er en forutsetning for forskningens pålitelighet. Slik tillitsverdighet hviler i stor grad på kildekritikk og etterlevelse av metodologiske normer for å sikre forskningens kvalitet (NESH, 2023: Innledning). Ansvarlig bruk av KI forutsetter derfor

[14] ALLEA (2023); EU (2026).
[15] EU-kommisjonen (2026) skiller mellom «*substantial use*» og «*not a substantial use*», n. 21, 28 og 33.

kritisk refleksjon gjennom hele forskningsprosessen, fra forskningsdesign og metodevalg til analyse, skriving og publisering.[16]

Forskere må reflektere over mulige feilkilder ved bruk av generativ KI i forskning. Forskere må så langt som mulig, sette seg inn i og ta hensyn til teknologiske begrensinger og systematiske skjevheter ved ulike KI-verktøy. KI-modeller angir bare hva som er sannsynlige formuleringer, de har ingen kriterier for å skille mellom sant og falskt. Det henger sammen med «svart boks»-problematikken, altså at det ikke er mulig å avdekke nøyaktig hvordan KI-modeller kommer fram til et gitt resultat.

Forskere må etterprøve hvilke kilder den KI-genererte kunnskapen bygger på. Kritisk refleksjon forutsetter kildekritikk, og pålitelighet sikres med god henvisningsskikk (NESH 2023: 8). Kildehenvisninger må vise til eksisterende faglitteratur, som faktisk er konsultert.

Det er viktig å være oppmerksom på *hallusinering* som en teknologisk begrensning, altså at KI-modeller viser til innhold, referanser, kilder eller sitater som ikke er virkelige.. Slike fiktive kilder påvirker ikke nødvendigvis forskningens resultater, slik Granskningsutvalget nylig har påpekt.[17] Men de *kan* også påvirke selve analysen. Her er det også relevante forskjeller mellom ulike fagområder. I humanistisk og samfunnsvitenskapelig forskning er ofte tekster og historiske kilder både forskningsmateriale og studieobjekt. I slike tilfeller kan bruk av fiktive kilder påvirke selve resultatet og, i alvorlige tilfeller, innebære fabrikkering eller fordreining (NESH 2023: 11–12). Ved bruk av generativ KI må forskere derfor etterprøve analyser og resultater i henhold til god vitenskapelig praksis.

Forskere må håndtere og korrigere systematiske skjevheter, som kan forekomme i KI-generert kunnskap. Systematiske skjevheter kan skyldes begrensninger i materialet som KI-modeller er trent på, noe som kan føre til misvisende fremstillinger av virkeligheten og forsterkning av sosiale stereotypier (*training data bias*). KI-modeller kan også gjøre tilpasninger når de svarer på forespørsler for å imøtekomme oppfattede preferanser og holdninger hos den som bruker dem *(prompt bias*), noe som kan forsterke feilaktige fremstillinger og gi fordreide svar. Skjevheter kan dessuten være intenderte eller programmerte, innebygd i selve algoritmene, for å ramme bestemte sårbare eller svakstilte grupper eller fremme bestemte oppfatninger og ideologier (*programmed bias*). Kritisk refleksjon over slike skjevheter omfatter også refleksivitet knyttet til egen kontekst, perspektiv og posisjon.

[16] ALLEA (2023), s. 5; EU-kommisjonen (2026), s. 5 (om *reliability*).
[17] Granskingsutvalget (2026/6; 2026/7); Jf. Resnik & Hosseini (2026a).

Forskere må formidle vitenskapelig usikkerhet og faglige begrensninger på en klar og tydelig måte. Det er viktig i alle faser av forskningen, i dialogen med forskningsdeltakerne, ved vitenskapelig publisering, i annen formidling og ved senere bruk (NESH, 2023: 47). Forskere bør bare bruke generativ KI i kontekster der de selv har tilstrekkelig kompetanse til å identifisere og håndtere tekniske begrensninger og systematiske skjevheter.

Bruk av generativ KI i forskning kan få utilsiktede og uønskede konsekvenser for vitenskapelige normer som originalitet og kreativitet. Det kan svekke muligheten for selvstendig tenkning og kritisk refleksjon, og gjøre det vanskelig å vedlikeholde forskerens egen kompetanse. Overdrevet bruk av generativ KI kan bidra til at forskere ikke utvikler nødvendige ferdigheter for å ivareta god vitenskapelig praksis. Derfor er det nødvendig med opplæring i forsvarlig bruk av generativ KI med tanke på både muligheter og begrensninger, herunder også risiko for forskningens integritet.[18]

1.2. Ansvarlighet

Forskere har et selvstendig ansvar for å ivareta forskningsetikken (NESH, 2023, Innledning). Ansvarlighet innebærer at den enkelte forsker *har* ansvar, *tar* ansvar og kan *holdes* ansvarlig ved bruk av generativ KI i forskning. Derfor må forskere gjøre seg kjent med, og ta hensyn til, teknologiens muligheter og begrensninger. Dette ansvaret gjelder i alle faser av forskningen, fra idé til publikasjon. Ansvarlig bruk av generativ KI i forskning forutsetter *menneskelig tilsyn*.[19]

Forskere er ansvarlige for forskningens resultater og for publikasjoner hvor de står oppført som (med)forfatter. KI-verktøy kan *ikke* stå oppført som (med)forfatter, fordi forfatterskap må utøves aktivt, med ansvarlighet, noe som bare tilligger mennesker. KI-verktøy kan ikke holdes ansvarlig, både fordi de ikke kan gi troverdige forklaringer på de valg som er gjort og fordi de ikke kan holdes juridisk ansvarlige eller blir sanksjonert dersom noe har gått galt.[20]

Forskere må være klar over at uansvarlig bruk av generativ KI kan resultere i vitenskapelig uredelighet, forstått som plagiat, fabrikkering og forfalskning (FFP), eller andre alvorlige brudd som fordreining og fortielse (NESH, 2023: 10–12). Forskere må aktivt begrense mulige brudd på anerkjente forskningsetiske normer forårsaket av algoritmer, datasett og prompter eller av den konteksten eller sammenhengen hvor generativ KI blir

[18] UK RIO (2025), s. 19.
[19] ALLEA (2023), s. 5; EU-kommisjonen (2026), s. 5 (om *accountability*).
[20] EU-kommisjonen (2026), 2.1.1; UK RiO (2025).

brukt. KI må uansett ikke brukes på måter som skader forskningens integritet eller forskerfellesskapets praksiser.[21]

Forskere må vurdere om det er forsvarlig å bruke KI-verktøy i forbindelse med fagfellevurdering, søknadsevaluering eller ved vurdering av ansettelser og opprykk. Vesentlig bruk av KI-verktøy kan føre til feilaktig vurdering og urettferdig behandling. Det kan bryte med prinsipper om personvern og konfidensialitet. Og det kan undergrave prinsipper om originalitet, opphavsrett og immaterielle rettigheter (IPR). Det er også en risiko for at man utsetter kollegers upubliserte arbeid for uønsket offentliggjøring og mulig plagiering av originale ideer.[22] Forskere må derfor vurdere om det er forsvarlig å laste opp andres arbeid, for eksempel upubliserte manus, publiserte artikler eller utgitte bøker. Selv om personvernet blir ivaretatt, er det ikke gitt at det er mulig å beskytte originale ideer og immaterielle rettigheter. Forskere må derfor sette seg inn i hvordan ulike KI-verktøy virker. De står selv ansvarlig for den vurderingen som blir gjort. Ansvarlighet innebærer uansett begrenset bruk av KI-verktøy.

Forskere må sette seg inn i retningslinjer for ansvarlig bruk av generativ KI ved egen institusjon, hos finansiører og hos tidsskrifter og forlag. De må avklare hva som skjer med informasjon som blir lastet opp eller generert ved bruk av spesifikke KI-verktøy.[23] Forskere bør også kunne veilede studenter og bistå med forskningsetiske vurderinger om bruk av KI-verktøy. Dette forutsetter at institusjonene har gode rutiner og retningslinjer for ansvarlig bruk av KI i forskning. NESH vil oppfordre forskere til å delta aktivt i utviklingen av slike retningslinjer, slik at disse er best mulig tilpasset forskningens karakter og formål.

1.3. Åpenhet

Åpenhet er en grunnleggende forskningsetisk forpliktelse, og forskere må være åpne om roller og interesser (NESH, 2023: 2–3). Ansvarlig bruk av generativ KI i forskning forutsetter også åpenhet om anvendelse. Forskere må redegjøre for hvordan de har brukt KI-verktøy, i både forskningsprosessen og skriveprosessen, på en ærlig, fullstendig og upartisk måte. Det er spesielt viktig ved publisering og formidling av resultatene.[24]

Forskere bør oppgi *hvilke* KI-verktøy som er benyttet og *hvordan* de er brukt. Det er spesielt relevant i samfunnsvitenskap og humaniora, hvor skriveprosessen ofte er en sentral

[21] EU-kommisjonen (2026), 2.1.6.
[22] Resnik & Hosseini (2026b).
[23] NESH (2024/121).
[24] ALLEA (2023), s. 5; EU-kommisjonen (2026), s. 5 (om *honesty*).

del av forskningsprosessen. I samarbeidsprosjekter kan det også være relevant å oppgi *hvem* som har brukt KI-verktøy. Forskere bør drøfte systematiske skjevheter og teknologiske begrensninger, og de bør redegjøre for om bruken er vesentlig eller ikke. Det kan være relevant å tilgjengeliggjøre prompt-logger eller tidligere versjoner av en tekst, i tråd med prinsippene for åpen forskning. Dette er relevant også ved en eventuell granskning ved mistanke om vitenskapelig uredelighet.

Åpen vitenskap og åpen publisering må være ansvarlig, slik at det bidrar til å sikre forskningens integritet og tillit.[25] Derfor er det viktig å vurdere konsekvenser av KI for publiseringsetikk (NESH, 2023: 6–8, 44). Forskere må være oppmerksomme på at ulike tidsskrift og forlag eller finansiører kan ha forskjellige standarder for åpenhet om bruk, ulike krav for hva som skal oppgis, og ulik praksis på om deklarasjon er obligatorisk eller ikke. Forskere må uansett kvalitetssikre og sjekke innhold og resultater ved bruk av KI-verktøy.

Generativ KI har skapt store utfordringer for vitenskapelig publisering og åpen forskning. Såkalte artikkelfabrikker eller papirmøller (*paper mills*) masseproduserer og selger KI-genererte artikler til forskere, som er under stor press i et forskningssystem som belønner publiseringer og siteringer. Samtidig finnes det sterke insentiver for åpen publisering, hvor kvaliteten ikke alltid er ivaretatt. Det undergraver troverdigheten til redelige forskere, og det svekker tilliten til forskning i samfunnet. Det bidrar også til å at den vitenskapelige kunnskapsbeholdningen blir forurenset av fabrikkerte, forfalskete og fordreide funn.[26] Mange tidsskrifter og forlag har derfor gått sammen om å utvikle felles standarder og betingelser for ansvarlig bruk av generativ KI i vitenskapelig publisering.[27]

Forskere, forskningsinstitusjoner og andre aktører, for eksempel finansiører, har et felles ansvar for å motvirke denne utviklingen, for eksempel ved å oppdatere det nasjonale kanalregisteret, utvikle nye former for kvalitetssikring av vitenskapelig publisering, og vurdere andre kriterier for kvalitet ved vurdering av karriereløp og ansettelser.[28]

1.4. Institusjonenes ansvar

Forskningsinstitusjoner må utvikle egne retningslinjer for ansvarlig bruk av generativ KI i forskning og sikre at forskerne har tilgang på relevant opplæring og veiledning. De må utvikle plattformer for godkjente og sikre KI-verktøy og ha oversikt over utvikling og bruk av

[25] Bernabe, R. mfl. (2024).
[26] The Royal Society (2025); Cambridge University Press (2025).
[27] STM (2023); Frisch, K. (2025).
[28] NOR-CAM (2021); COARA (2022).

generative KI-systemer i hele organisasjonen. Institusjonene må også sikre ansvarlig bruk av generativ KI i ledelse, administrasjon og organisering, i opplæring, utdanning og veiledning, ved vurdering av kandidater, publikasjoner og ansettelser, og med tanke på forskningens bredere samfunnsmessige konsekvenser. [29]

Det forutsetter en rekke institusjonelle ordninger, blant annet for databehandling, beskyttelse av opphavsrett og intellektuelle rettigheter, risikohåndtering, samfunnssikkerhet, bærekraft og teknisk dokumentasjon.[30] Ved granskning av vitenskapelig uredelighet må institusjonen også ha klare kriterier for vurdering av bruk av generativ KI.

UTKAST

[29] EU-kommisjonen (2026), 2.2.
[30] EU (2024), *AI Act*, art. 9–11.

2. FORSKNING PÅ MENNESKER

Forskningsetikken bygger på respekt for menneskeverdet. Det enkelte individ har interesser og integritet, som ikke uten videre kan settes til side i forskningen. Forskere skal respektere grunnleggende verdier som menneskeverd og individuell autonomi (NESH, 2023: Innledning). Respekt for menneskeverdet handler ikke bare om behandling av personopplysninger i juridisk forstand, men også om personvern i en bredere forstand, knyttet til privatlivets fred, konfidensialitet og anonymisering.[31]

I fortsettelsen fremheves betydningen av tre grunnleggende prinsipper som skal ivareta hensynet til mennesker som deltar eller inngår i forskning, og som forskere må forholde seg til ved bruk av generativ KI: *Respekt* for likeverd, frihet og selvbestemmelse*; beskyttelse* mot risiko for skade og urimelig belastning*; og rettferdighet* i prosedyrer og fordeling av goder og byrder

2.1. Respekt

Forskere skal vise respekt for likeverd, frihet og selvbestemmelse. Derfor må forskere vurdere om det er nødvendig å gi informasjon eller å få samtykke fra mennesker som deltar eller inngår i forskning. Informasjon og samtykke er to ulike forskningsetiske hensyn, som må vurderes hver for seg (NESH, 2023: 15–16).[32]

Hvorvidt det er nødvendig å gi informasjon, avhenger av forskningens art, kildemateriale og datagrunnlag. Informasjonen må være klar og tydelig, tilpasset mottakerens alder og bakgrunn og formidlet på et språk og på en måte de forstår. Forskere må uansett sikre at informasjonen er tilstrekkelig for å respektere deres menneskeverd og deres integritet, sikkerhet og velferd. Følgende spørsmål er relevante: Hva er formålet med prosjektet? Hva innebærer prosjektet? Hva er fordeler og ulemper? Hvilke interesser er involvert? Hvem er ansvarlig? (NESH, 2023/156).

Når generativ KI brukes enten til datainnsamling eller dataanalyse, kan dette ha implikasjon for hensynet til informasjon og samtykke. Forskere bør informere om bruk av generativ KI i ulike faser av forskningen, for eksempel til datainnsamling, databehandling og analyse. Forskere må også informere om risiko og usikkerhet og redegjøre for hvordan dette

[31] Europarådet (2024), art. 7.
[32] ALLEA (2023), s. 5; EU-kommisjonen (2026), s. 5 (om *respect*).

er håndtert i prosjektet.[33] Det handler om risiko for informasjonssikkerhet, for personopplysningsvern og for personvern i bredere forstand (NESH, 2024/121).

Når generativ KI brukes for å få innhente data fra åpne kilder, for eksempel i sosiale medier, er fire faktorer relevante for å vurdere om det er nødvendig å be om samtykke i tillegg til å informere: Grad av åpenhet, informasjonens sensitivitet, forskningens interaksjon og de berørtes sårbarhet. De fire faktorene kan kombineres på ulike måter og vektes forskjellig i ulike prosjekt. Her finnes det ingen enkle regler eller entydige svar, og det er alltid behov for skjønnsmessige vurderinger og tydelige begrunnelser. Forskere bør også ta hensyn til *forventet offentlighet* og *kontekstuell integritet*. Disse hensynene er utdypet i NESHs veileder for internettforskning (2019).

2.2. Beskyttelse

Forskere må beskytte personvern og privatliv og vurdere behovet for anonymisering og konfidensialitet, samt risiko for reidentifisering. Det stiller krav til hva slags informasjon som kan lastes opp i KI-verktøy, og til hvordan man kan bruke KI-generert informasjon.[34] Kravet om beskyttelse gjelder spesielt ved behandling av personopplysninger, også når informasjonen ikke er samlet inn direkte fra enkeltindivider, for eksempel i tekster eller bilder generert av KI.

Forskere må gjøre en selvstendig vurdering av behovet for anonymisering, både i analyse og ved publisering (NESH, 2023: 20). Videre må de behandle innsamlet informasjon konfidensielt hvis det er avtalt eller hvis det er nødvendig av andre grunner (NESH, 2023: 21). Dersom KI-generert informasjon inneholder private eller sensitive opplysninger, knyttet til privatliv, familieliv, hjem og kommunikasjon, må forskeren vurdere om de må behandles konfidensielt. Det kan handle om mennesker som er i en sårbar eller svakstilt posisjon, som har krav på beskyttelse, eller som må beskyttes mot seg selv.

Forskere må være oppmerksomme på at enkeltpersoner kan bli indirekte identifisert ved sammenstilling på stordata og bruk av generativ KI. Her kan det være hensiktsmessig å skille mellom reidentifisering og bakveisidentifisering. *Reidentifisering* vil si at opplysninger som i utgangspunktet er pseudonymiserte, men hvor forskeren har en koblingsnøkkel, likevel kan bidra til å indentifisere et individ. *Bakveisidentifisering* vil si at opplysninger som i utgangspunktet er anonymiserte, det vil si fullstendig avidentifiserte, likevel kan bidra til å

[33] NESH (2024/121).
[34] EU-kommisjonen (2026).

identifisere et individ.[35] Ulike opplysninger kan altså, ved sammenstilling i KI-verktøy, *bli* personopplysninger, i betydningen identifiserende, selv om de ikke var det i utgangspunktet. Forskere må derfor vurdere om innsamling, analyse og bruk av slike opplysninger faktisk er ansvarlig, ikke minst fordi jussen på området er i utvikling.

Forskere må foreta en helhetlig vurdering av risiko og sårbarhet for å avgjøre om behandlingen av slike opplysninger er forsvarlig (NESH 2025/140, NESH 2023: 28). Det er spesielt viktig å vurdere risiko for reidentifisering og bakveisidentifisering ved bruk av personopplysninger, helseopplysninger, genetiske opplysninger og lokasjonsdata, som gir stor mulighet for identifikasjon. Det er også viktig å vurdere risiko for direkte og indirekte berørte, for svakstilte og sårbare grupper og for barns menneskeverd og integritet (NESH 2023: 18, 26 og 31). Ved å laste opp informasjon i KI-systemer som ikke er sikre, eller godkjent av egen institusjon, kan informasjonen i prinsippet bli offentliggjort og formidlet til andre på måter som bryter med prinsippet om konfidensialitet. Det kan bidra til å svekke tilliten til prosjektet og til forskning mer generelt.

2.3. Rettferdighet

Forskere må vurdere risiko for enkeltpersoner som deltar eller inngår i forskningen, og avklare alvorlighetsgraden og sannsynligheten for mulige negative konsekvenser, herunder også samfunnsmessige konsekvenser knyttet til likeverd og rettferdighet.[36] KI-verktøy kan gjøre det enklere å få tilgang til kunnskap om medlemmer av svakstilte og sårbare grupper. Men selv om informasjonen er lett er tilgjengelig, må forskere uansett sikre at forskningen skjer i tråd med anerkjente forskningsetiske normer om respekt, beskyttelse og rettferdighet (NESH, 2023: 31).

KI-modeller har innebygde systematiske skjevheter, som kan føre til misvisende fremstillinger av virkeligheten og forsterke sosiale stereotypier. Det kan få urettferdige konsekvenser for medlemmer av svakstilte eller stigmatiserte grupper: Svakstilte grupper kan bli enda mer underrepresentert, og sårbare grupper kan bli enda mer stigmatisert. Slike former for strukturell diskriminering kan medføre risiko for menneskerettigheter, demokrati og rettstaten. Ved bruk av generativ KI må forskere være oppmerksomme på systematiske skjevheter, identifisere feilaktige fremstillinger og korrigere analysen med forskningsbaserte data og vitenskapelige metoder.

---

[35] FEK (2020), s. 17.
[36] Europarådet (2024), art. 10; Diskrimineringsombudet (2023).

KI-modeller kan også skape syntetiske data, det vil si kunstig genererte opplysninger, som kan brukes til å trene modeller eller å simulere virkeligheten. Slike data kan brukes til å begrense skjevheter i KI-modellene, men de kan også føre til misvisende fremstillinger av virkeligheten og forsterke sosiale stereotypier. Forskere må være åpne om bruk av syntetiske data og redegjøre for risiko og usikkerhet knyttet til forskningens tillit og troverdighet.[37]

2.4. Institusjonenes ansvar

Forskningsinstitusjonene må sikre personvern i bred forstand, knyttet til menneskeverd, autonomi og personlig integritet. Det er viktig å ha sikre systemer for behandling av personopplysninger, men også å beskytte selve individet mot krenkelse av personlige forhold eller privatlivets fred (NESH, 2025/156). Dersom bruk av generativ KI i forskning medfører høy risiko for menneskers friheter og rettigheter, må det gjennomføres en vurdering av personvernkonsekvenser (DPIA). Institusjonene må ha gode systemer for informasjonssikkerhet og rutiner for å vurdere og godkjenne hvilke KI-verktøy som er forsvarlige å bruke, til hvilke typer data og i ulike faser av forskningen. Institusjonene må også bistå forskerne med nødvendig dokumentasjon av godkjente KI-verktøy til relevante tredjeparter, for eksempel oppdragsgivere, finansiører og samarbeidspartnere.

[37] Jf. kap. i NENT-antologi.

3. FORSKNING OG SAMFUNN

Generativ KI kan få store og uforutsigbare konsekvenser for samfunnet og utfordre menneskeverd, demokrati og rettstaten. Generativ KI kan også utfordre og endre vitenskapens rolle i samfunnet. Derfor er det viktig at forskningsetiske prinsipper som pålitelighet, ansvarlighet og åpenhet (Del 1) og respekt, beskyttelse og rettferdighet (Del 2) ses i en bredere samfunnsmessig kontekst. Både offentlige og private institusjoner må sikre at forskningsetikken blir ivaretatt ved utvikling og bruk av generativ KI.

NESH mener det er behov for å utvikle nye rammer for vurdering av risiko, usikkerhet og forsvarlighet. Forskere må foreta en selvstendig forskningsetisk vurdering av risiko og forsvarlighet, herunder alvorlighetsgraden og sannsynligheten for potensiell skade og belastning for samfunnet. Og forskningsinstitusjonene må utvikle rutiner for systematisk risikohåndtering, om nødvendig med en uavhengig forskningsetisk vurdering.

Europarådets *Rammekonvensjon om kunstig intelligens og menneskerettigheter, demokrati og rettsstat* (2024) har et bredt perspektiv på risiko i en samfunnsmessig betydning og drøfter mulige negative konsekvenser for menneskerettigheter, demokrati og rettstaten knyttet til grunnleggende normer som pålitelighet, ansvarlighet, åpenhet, menneskeverd, personvern, rettferdighet, risiko, usikkerhet og forsvarlighet.[38] NESH mener at dette rammeverket åpner for relevante forskningsetiske refleksjoner innenfor samfunnsvitenskap og humaniora, som samtidig kan være relevante for andre fagområder.

EUs *AI Act* (og utkast til norsk KI-lov) har en risikobasert tilnærming og stiller krav om en systematisk tilnærming til håndtering av risiko i fire ledd: Risikovurdering, klassifisering, dokumentasjon og kommunikasjon. Basert på en risikovurdering er det nødvendig å vurdere alvorlighetsgrad og sannsynlighet for negative konsekvenser for å skille mellom fire ulike risikonivåer: Minimal, begrenset, høy og uakseptabel risiko. Gitt at minimal og begrenset risiko er mindre alvorlig, og uakseptabel risiko kan være forbudt, handler loven primært om behovet for å regulere tilfeller av høy risiko.[39] NESH mener at dette regelverket er hensiktsmessig utgangspunkt for å vurdere risiko ved bruk av generativ KI i forskning, men det må tilpasses forskningens verdier og normer på ulike fagområder.

[38] Europarådet (2024).
[39] EU AI Act.

3.1. Offentlighet

Et velfungerende, kunnskapsbasert og demokratisk samfunn er avhengig av forskning som kilde til pålitelig kunnskap, og et viktig formål med forskningsformidling er å styrke individuell autonomi og ytringsfrihet samt å bidra til en åpen og opplyst offentlig samtale (NESH, 2023: 45).[40] Utvikling og bruk av generativ KI utfordrer forskningens rolle i samfunnet på måter som kan innebære høy og uakseptabel risiko både for forskning og samfunn. Derfor er det viktig å beskytte integriteten i demokratiske prosesser og respekten for rettsstaten. For å sikre individers tilgang til og deltakelse i offentlig debatt, samt deres evne til fritt å danne meninger, er spesielt det viktig å beskytte grunnleggende prinsipper som offentlighet og uavhengighet.[41]

Forskningens bidrag til en velfungerende *offentlighet* blir konstituert og vedlikeholdt gjennom ulike praksiser, for eksempel tilgang til offentlige data, åpenhet om roller og interesser, offentliggjøring gjennom publisering og formidling, samt deltakelse i det offentlige ordskiftet. Det er spesielt viktig å beskytte disse praksisene ved bruk av generativ KI i forskning, hvor potensialet for misbruk er stort. Forskerfellesskapet skal sikre at det åpne offentlige ordskiftet er opplyst i den forstand at samfunnet får tilgang til pålitelig kunnskap. Forskere skal også være et kritisk korrektiv til myndigheter og andre mektige aktører i samfunnet, og de skal imøtegå misbruk av forskning i form av pseudovitenskap, politisering og propaganda (NESH, 2023: E).

Uavhengighet er en forutsetning for forskningens troverdighet og for samfunnets tillit til forskning (NESH 2023: 1, 36). Den kan komme i press i en digital kunnskapsøkonomi hvor også selve forskningssystemet er i endring. Bruk av generativ KI kan utfordre normer om åpenhet og uavhengighet, for eksempel på grunn av i) privatisering og monopolisering av eierskap til KI-systemer, ii) utvisking av skillet mellom vitenskapelig forskning og andre former for kunnskapsproduksjon, og iii) avmakt i møte med tech-giganter som kontrollerer utviklingen av KI.[42] Forskere må bidra til å beskytte integriteten i forskningen og i demokratiske institusjoner og prosesser, inkludert prinsippet om maktfordeling, respekt for domstolenes uavhengighet og tilgang til rettferdighet.[43]

Ved bruk, og spesielt ved utvikling, av generative KI-systemer og KI-verktøy må forskere være åpne om finansiering, roller og interesser: Åpenhet om finansiering gjør det

[40] Grunnloven, § 100.
[41] Europarådet (2024), art. 5 (2).
[42] Steen-Johnsen, K. og Enjolras, B. (2015).
[43] Europarådet (2024), art. 5 (1).

lettere for forskere å beskytte seg mot utilbørlig press og dermed sikre forskningens uavhengighet (NESH, 2023: 41). Forskere må også være åpne om eventuelle føringer eller bindinger fra samarbeidspartnere: Det er viktig alltid å være tydelig på ulike roller og mulige interessekonflikter, både for å fremme god forskning og for å sikre forskningens tillit i samfunnet (NESH, 2023: 39–40).

Vurdering av risiko for forskningens integritet er en nødvendig for å avgjøre om et prosjekt er forsvarlig å gjennomføre: Et grunnleggende forskningsetisk premiss er at forskning skal være åpen og uavhengig. Hvis forsvarlige avtaler ikke kan inngås eller opprettholdes, bør ikke forskningen gjennomføres (NESH, 2023: 37). Slike vurderinger, for eksempel av bestemte KI-verktøy, forutsetter at forskningsinstitusjonene kan bistå med relevant støtte og veiledning.

3.2. Usikkerhet

Forskere må kommunisere både vitenskapelig usikkerhet og faglige begrensninger (NESH, 2023: 47). Det er spesielt viktig ved bruk av generativ KI i forskning, i dialogen med mennesker som deltar i forskning, ved vitenskapelig publisering og i annen formidling til samfunnet. NESH vil fremheve at vitenskapelig usikkerhet er noe kvalitativt annerledes enn uvitenhet, skepsis eller sikkerhetskontroll.

Det er viktig å skille mellom vitenskapelig usikkerhet og uvitenhet eller mangel på kunnskap. Ved bruk at generativ KI i forskning kan det være vanskelig for forskeren å vite hva som er mulige negative konsekvenser. I slike tilfeller er det viktig med ydmykhet i stedet for skråsikkerhet. Forske bør for eksempel ikke si til deltakere at bruken av generativ KI er trygg hvis de faktisk ikke vet det. Forskere bør heller ikke fremstille forskning basert på generativ KI som sikker og pålitelig med mindre de på en etterrettelig måte har etterprøvd kilder, data, analyse og resultater. Det samme gjelder ved formidling. Forskere bør være åpne om mulige føringer og begrensninger ved bruk av generativ KI i forskning (NESH, 2023: 48).

Det er også viktig å skille mellom det å formidle vitenskapelig usikkerhet og å bidra til ulike former for skepsis eller mistro til forskning. Det er relevant for eksempel når forskere skal gi informasjon eller få samtykke fra mennesker som deltar i forskning: Ærlighet og åpenhet om vitenskapelig usikkerhet kan bidra til å redusere uberettiget frykt og usikkerhet hos deltakerne, og det er en forutsetning for at et eventuelt samtykke er tilstrekkelig informert og dermed gyldig. Det er viktig i forbindelse med vitenskapelig publisering: Forskere må være åpne og mulig risiko og usikkerhet ved bruk av generativ KI i forskning. Det er også viktig i formidling: Forskere bør tydeliggjøre de faglige kriteriene for den vitenskapelige

usikkerheten, slik at de kan være gjenstand for offentlig debatt i tråd med normer om etterrettelighet om saklighet i formidling (NESH, 2023: 47).

Samtidig finner forskning sted i et samfunn som i økende grad er preget av sikkerhetspolitiske føringer og geopolitiske spenninger. Ulike former for sikkerhetsstyring kan legge begrensninger på forskningen, spesielt der det kan være stor grad av usikkerhet, for eksempel ved bruk av generativ KI. Det er nødvendig å sikre forskningen mot press fra politikken. Derfor er det viktig å skille mellom forskningssikkerhet og sikkerhetspolitikk, og å finne en forsvarlig balanse mellom åpenhet og aktsomhet: Forskningssikkerhet er nært knyttet til den nasjonale sikkerheten, men handler også om å ivareta akademiske verdier, normer og prinsipper som kan komme under press. Det forutsetter at forskere er ansvarlige og åpne og risiko og usikkerhet. I tillegg må institusjonene utvikle rutiner og systemer for sikkerhetsstyring som faktisk ivaretar forskningssikkerheten.

3.3. Forsvarlighet

Åpenhet er nødvendig, men ikke tilstrekkelig for å håndtere risiko på en forsvarlig måte. Forskere må også gjøre en selvstendig forskningsetisk vurdering av risiko og forsvarlighet, hvor de har identifisert og vurdert mulige uønskete konsekvenser og formulert tiltak for å begrense og forebygge dem. NESH presiserer at det ikke er tilstrekkelig med DPIA og/eller ROS-analyse for å sikre at forskningen er etisk forsvarlig. Det er nødvendig å vurdere risiko for skade og belastning i bred forstand for å kunne vurdere om prosjektet er forsvarlig i etisk forstand (NESH, 2025/140).

I vurderingen av risiko, usikkerhet og forsvarlighet mener NESH det er hensiktsmessig å trekke inn *føre-var prinsippet*, som ble kodifisert i Rio-erklæringen i 1992. Føre-var prinsippet formidler mellom forskning og politikk, og det fokuserer på forebygging fremfor forbud. I utgangspunktet ble det formulert for beslutningstakere i forvaltning og politikk, men det kan også forstås som et forebyggende etisk prinsipp som utgangspunkt for håndtering og regulering av risiko. Føre-var prinsippet er spesielt relevant i situasjoner med stor vitenskapelig usikkerhet, mulighet for skade, behov for forebygging og viktig å plassere ansvar. Prinsippet er mindre relevant i tilfeller der vitenskapen er entydig og sikker.[44]

Selv om føre-var prinsippet opprinnelig ble utviklet med utgangspunkt spørsmål om natur, klima og miljø, finnes det en omfattende faglitteratur innenfor samfunnsvitenskap og humaniora som har videreutviklet dette perspektivet for eksempel knyttet til svakstilte og

[44] NENT (1997).

sårbare grupper og til det bredere begrepet om risikosamfunnet. NESH mener at slike perspektiver er relevante for en forskningsetisk vurdering av risiko, usikkerhet og forsvarlighet i forbindelse med bruk av generativ KI i forskning. Disse bredere perspektivene på samfunn, offentlighet og demokrati er ikke bare relevante i humanistisk og samfunnsfaglig forskning, men er også relevante i tverrfaglig samarbeid og på andre fagområder.

3.4. Institusjonenes ansvar

Forskningsinstitusjoner må sikre at bruk av generativ KI i forskning er forsvarlig, basert på en forskningsetisk vurdering av risiko, usikkerhet og forsvarlighet. Dersom bruk av generativ KI i forskning innebærer høy risiko, bør institusjonene utvikle egne retningslinjer og rutiner for å sikre at forskningen er forsvarlig. Slike systemer kan virke forebyggende og kan erstatte mer inngripende former for kontroll.

Institusjonene bør også vurdere om det er behov for ytterligere tiltak i tilfeller hvor det er snakk om uakseptabel risiko. For å sikre tillit til kontroversiell forskning kan det være nødvendig med en uavhengig forskningsetisk vurdering, for eksempel fra et forskningsetisk utvalg. Institusjonene står fritt til å etablere slike ordninger ved behov.[45]

Bredere spørsmål om risiko for demokrati, offentlighet og rettsstat tilsier at institusjonene bør se arbeidet med ansvarlig bruk av KI i sammenheng med andre rammeverk knyttet til samfunnssikkerhet og ansvarlig internasjonalt samarbeid.[46] Bredere spørsmål om risiko for natur, klima og miljø tilsier at institusjonene også bør se arbeidet med ansvarlig bruk av KI i sammenheng med andre rammeverk knyttet til bærekraft.

[45] FEK (2023), kap. 6; UK RIO (2025), kap. 10; TENK (2026).
[46] HK-dir (2026).

Litteraturliste

ALLEA (2023). *The European Code of Conduct for Research Integrity – Revised Edition 2023*. Berlin. DOI: 10.26356/ECOC

Bernabe, R. m.fl. (2024). *The ROSiE General Guidelines on Responsible Open Science.*

Cambridge University Press (2025). *Publishing futures: Working together to deliver radical change in academic publishing*.

COARA (2022). *Agreement on Reforming Research Assessment*.

Diskrimineringsombudet (2023). *Innebygd diskrimineringsvern*.

EU-kommisjonen (2026). *Living guidelines on the responsible use of generative AI in research. An ERA Forum stakeholders' document.*

Europarådet (2024), *Rammekonvensjon om kunstig intelligens og menneskerettigheter, demokrati og rettsstat.*

FEK (2020). *Stordata i forskning: store muligheter, store utfordringer*. De nasjonale forskningsetiske komiteene.

FEK (2023). *Veileder om institusjonenes ansvar for forskningsetikk*. De nasjonale forskningsetiske komiteene.

Fossheim, H. og Østerhaug, T. (2026, kommer). *Research Ethics and Artificial Intelligence*. Cappelen Damm Forskning / Den nasjonale forskningsetiske komité for naturvitenskap og teknologi (NENT).

Frisch, Katrin (2025). *FAQ Artificial Intelligence and Research Integrity.* The Ombuds Committee for Research Integrity in Germany (OWID), Version 2. doi.org/10.5281/zenodo.17349995

Granskingsutvalget (2026/6). *Om mulig vitenskapelig uredelighet i forbindelse med fiktive referanser*.

Granskingsutvalget (2026/7). *Om mulig vitenskapelig uredelighet i forbindelse med fiktive referanser*.

HK-dir (lastet 11.09.2026). *Retningslinjer for ansvarlig internasjonalt samarbeid*. Direktoratet for høyere utdanning og kompetanse.

NENT (1997). *Før-var prinsippet: Mellom forskning og politikk*. Den nasjonale forskningsetiske komité for naturvitenskap og teknologi.

NENT (2019). *Statement on research ethics and artificial intelligence*. Den nasjonale forskningsetiske komité for naturvitenskap og teknologi.

NESH (2019). *Forskningsetisk veileder for internettforskning*. Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora.

UTKAST

NESH (2023). *Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora*. Femte reviderte utgave. Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora.

NESH (2024/121). *Uttalelse om prosjektet KidAd*. Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora.

NESH (2025/140). *Vurdering av risiko og forsvarlighet ved forskning på skoler og barnehager*. Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora.

NOR-CAM (2021). *A toolbox for recognition and rewards in academic careers*. Universitets- og høgskolerådet.

Resnik, D.B. og Hosseini, M. (2026a). «Hallucinated citations produced by generative artificial intelligence may constitute research misconduct when citations function as data in scholarly papers». *Accountability in Research*. doi.org/10.1080/08989621.2026.2645390

Resnik, D.B. og Hosseini, M. (2026b). «Plagiarism of ideas in the age of generative artificial intelligence». *Nature Machine Intelligence* 8, 860–861. doi.org/10.1038/s42256-026-01247-3

Steen-Johnsen, K. og Enjolras, B. (2015). «Social research and Big Data – the tension between opportunities and realities», i Fossheim, H. og Ingierd, H. (red.). *Internet Research Ethics*. De nasjonale forskningsetiske komiteene. Cappelen Damm Akademisk: Oslo, s. 122–140.

STM (2023). *Generative AI in Scholarly Communications*. The International Association of Scientific, Technical & Medical Publishers.

TENK (2026). *Artificial Intelligence in Research: Research Integrity and Ethical Principles*. Recommendation of the Finnish National Board on Research Integrity.

The Royal Society (2025). *Reformation of science publishing: the Stockholm Declaration*.

UK RIO (2025). *Embracing AI with integrity. A practical guide for researchers*.

Relevante lover:

- Forskningsetikkloven
- Personopplysningsloven
- Helseforskningsloven
- Åndsverksloven

- Patentloven
- Sikkerhetsloven
- EU AI Act
- Utkast til ny lov om kunstig intelligens (30.09.2025).

UTKAST