**NENT** • Den nasjonale forskningsetiske komité for naturvitenskap og teknologi

Kongens gate 14
0153 Oslo
Telefon 23 31 83 00

post@etikkom.no
www.etikkom.no

Org.nr. 999 148 603

christine.hafskjold@kmd.dep.no).

Vår ref.:2018/213

Dato: 01.07.2019

# Innspill til nasjonal strategi for kunstig intelligens

## 1. Innledning

Den nasjonale forskningsetiske komité for naturvitenskap og teknologi (NENT) gir med dette innspill til regjeringens strategi for kunstig intelligens. NENT ønsker en nasjonal strategi for kunstig intelligens velkommen, og vil gjerne understreke betydningen av at etiske, inkludert forskningsetiske, aspekter ivaretas i en nasjonal strategi. Kunstig intelligens påvirker allerede i dag mange områder av samfunnet. Den videre utvikling av kunstig intelligens-teknologi og usikkerheten knyttet til konsekvenser for mennesker og samfunn, krever en refleksjon, som også forskere må forholde seg til. Dette er også bakgrunnen for NENTs pågående arbeid med en forskningsetisk betenkning om kunstig intelligens, som skal ferdigstilles i høst.

NENT skal gi uavhengige råd til forskere og myndigheter om forskningsetiske spørsmål, og NENT ønsker med dette innspillet å bidra til god og ansvarlig kunstig intelligens-forskning i Norge. Innspillene fra NENT oppsummerer de særegne forskningsetiske spørsmål som oppstår som en følge av kunstig intelligens-forskning. Innspillene gjelder selve utviklingen av teknologien, men også anvendelsen av allerede utviklet kunstig intelligens-teknologi i forskning. *Forskningsetiske retningslinjer for naturvitenskap og teknologi* vektlegger forskningens selvstendige ansvar for egen rolle i samfunnsutviklingen, og innspillene utdyper spesielt hvordan forskningens samfunnsansvar bør forstås i lys av utfordringer som reises.

## 2. Kjennetegn ved kunstig intelligens

Vi vil her sammenfatte hva som vektlegges som særegne karakteristiske trekk ved kunstig intelligens i tilbakemeldingene NENT har fått fra fagmiljøene i Norge og i rapporter komiteen har gått gjennom. Kunstig intelligens er kjennetegnet ved at teknologien:

- *Etterligner, erstatter og utvider menneskelig intelligent handling, og menneskelig beslutningstaking og vurdering.* Teknologien kan også forsøke å simulere menneskelige følelser. De menneskelige handlinger som erstattes eller automatiseres har ulik grad av kompleksitet.

- Har utallige *anvendelsesområder og enormt endringspotensiale.* Utviklingen skjer dels i samhandling med oss og dels i interaksjon mellom de tekniske systemene. Teknologien er også kjennetegnet ved *uforutsigbarhet*; det er utfordrende eller umulig å forutse hvilke virkninger teknologien vil ha for enkeltindivider, samfunn og miljø.

- *Generering av stordata* har muliggjort de siste års kunstig intelligens-utvikling, samtidig som kunstig intelligens igjen genererer stordata, som kan inneholde personopplysninger.

I det følgende skisserer NENT de forskningsetiske implikasjonene av disse kjennetegnene. Dette er forskningsetiske aspekter NENT mener det er viktig å ha oppmerksomhet om – både blant forskere, forskningsinstitusjoner og andre aktører som finansierer forskning og utvikler rammevilkårene for forskning i Norge.

## 3. Forskningsetiske aspekter: 9 innspill

### *1. Uforutsigbarhet*

Forskningsetisk sett er det vesentlig å vurdere og kommunisere usikkerhet i forskningen. Kunstig intelligens-utviklingen er kjennetegnet ved grunnleggende usikkerhet og uforutsigbarhet. Usikkerheten i kunstig intelligens-forskning er knyttet til a) utviklingen og tilpasningen av systemene, inkludert kvaliteten på data; b) bruken av systemene og konsekvenser for enkeltpersoner, dyr, miljø og samfunn; c) verdiene som (eksplisitt eller implisitt) bygges inn i systemene og hvordan disse påvirker utfall

eller, i en større sammenheng, hvordan disse påvirker enkeltpersoner, dyr, miljø og samfunn; og d) konsekvensene ved ikke å utvikle teknologien.

NENT ser et behov for systematiske studier av risikoene forbundet med utviklingen av kunstig intelligens. Det er av betydning at forskere, så vel som politiske beslutningstakere, anerkjenner usikre mulige konsekvenser og *ukjente* ukjente, dvs. fremtidige konsekvenser vi ikke kjenner til. Myndigheter og forskningsfinans bør legges til rette for tverrfaglighet i forskning for å anerkjenne uforutsigbarhet, og minimere usikkerhet, der det er mulig. Etikk bør inn i utdanningen av fremtidige utviklere av kunstig intelligens.

## 2. *Bred involvering*

Forskere har et ansvar for å kommunisere risiko som følger av forskningsfunn. Hvilke risikoer og muligheter ved teknologien som vektlegges vil imidlertid være ulikt avhengig av hvilket etisk perspektiv og hvilke verdier og interesser som legges til grunn. De som blir mest påvirket av beslutninger som tas, har ofte ikke en stemme i beslutningsprosessene.

Myndigheter og forskningsinstitusjoner bør legge til rette for bred involvering av innbyggerne i diskusjoner om hva som er formålet med forskningen, innretningen på forskningssatsninger og bruk av forskningen.

## 3. *Ansvar*

a) Spørsmålet om ansvar er essensielt i kunstig intelligens-sammenheng og viktigheten av dette spørsmålet forsterkes ved utvikling og bruk av adaptive og autonome systemer. Generelt kan det hevdes at jo mer adaptivt og autonomt et kunstig intelligens-system er, jo vanskelige vil det være å kontrollere det og ansvaret er vanskeligere å lokalisere.

   NENT mener myndighetene bør stille krav om å kunne lokalisere ansvar om det skjer en feilvurdering som konsekvens av en beslutning i et kunstig intelligens-system.

b) Det må skilles mellom forskerens ansvar for kunstig intelligens-forskning og forskerens ansvar for videre bruk av forskningsresultatet. Forskere har et medansvar for bruk av forskningen.

   Ved oppdragsforskning eller planlagt kommersialisering av forskningsresultater bør forskere samarbeide med eksterne aktører for å vurdere risiko ved videre bruk av forskningen.

## *4. Menneskeverd*

Forskningen må ikke bryte med de rettighetene som er nedfelt i anerkjente internasjonale konvensjoner om sivile, politiske, økonomiske, sosiale og kulturelle menneskerettigheter. Kunstig intelligens-forskning kan bidra til å fremme menneskeverdet, men også true det. Kunstig intelligens må utvikles og brukes på måter som har respekt for menneskeverdet. Dette fordrer en innebygd etikk. Innebygd etikk (*Ethics by design*) er et begrep som viser til behovet for en proaktiv tilnærming for å sikre god og ansvarlig kunstig intelligens-forskning. Det bygger på det mer veletablerte begrepet «innebygd personvern», som vi blant annet finner i personvernlovgivningen.

Forskere og forskningsinstitusjoner må sikre at systemene er bygd opp på en måte som ivaretar personvern, samt forventet eller mulig innvirkning på enkeltpersoner, dyr, miljø og samfunn, og krav til rettferdig og etisk bruk av slike systemer.

## *5. Personvern og hensyn til enkeltmennesker*

Grunnleggende personvernprinsipper, nedfelt i personvernlovgivningen, om lovlighet, rettferdighet, gjennomsiktighet, formålsbestemthet, dataminimering, riktighet, lagringsbegrensning, integritet og fortrolighet og ansvarlighet skal følges. Forskningsetisk sett er samtykke en hovedregel når personopplysninger brukes i forskning.

Forskere har et ansvar for å vurdere opplysningens offentlighet, informasjonens sensitivitet, de berørtes sårbarhet og forskningens

interaksjon og konsekvenser i den forskningsetiske vurderingen knyttet til informasjon og samtykke (NESH 2018).

Men samtykke er ikke alltid mulig å innhente, og selv når samtykke er mulig, vil det ofte ha sine begrensninger fordi informasjonen som gis ofte er et biprodukt av samhandling med ulike produkter og tjenester. Utfordringer oppstår også fordi opplysninger som i utgangspunktet er anonymiserte kan bli personopplysninger ved nye sammenstillinger.

Derfor er det en økt erkjennelse av at forskere også trenger nye tilnærminger for å sikre respekten til dem som deltar med sine personlige opplysninger i forskning.

### 6. *Kvalitet og skjevhet*

Det kan være særlig grunn til å stille kritiske spørsmål ved dataenes kvalitet, sannferdighet og relevans i kunstig intelligens-forskning fordi vi ikke alltid vet hvor dataene kommer fra, og metadata kan mangle eller være usikre. Mulighetene for slutningsfeil øker pga. skjevheter i materialet, analyseverktøyet og grunnet menneskelig fortolkning. Dette gir grunnlag for usikkerhet knyttet til tolkninger og beslutninger basert på kunstig intelligens.

NENT mener det er vesentlig at forskere og forskningsinstitusjoner legger til rette for åpne og allment tilgjengelige data for å sikre etterprøvbarhet og kvalitet. Samtidig bør usikkerhet og begrensninger ved forskningen anerkjennes og kommuniseres.

### 7. *Eierskap, tilgang til og deling av data*

Forskningsetisk sett er det vesentlig å legge til rette for at forskningen, inkludert data og resultater, som hovedregel gjøres tilgjengelig for alle. NENT ser en risiko for at store deler av forskningsinnsatsen innen kunstig intelligens unndrar seg de krav til åpenhet som gjelder for forskning ellers (bl.a. nedfelt i FAIR-prinsippene), for eksempel med henvisning til nødvendigheten av hemmelighold for å sikre konkurransefortrinn. Deling av data er viktig for etterprøving og etterbruk av forskningsmateriale. Åpenhet er også en forutsetning for samfunnets tillit til forskning

NENT mener myndigheter og forskningsinstitusjoner bør sikre åpenhet om hvem som skal ha eierskap til teknologi, infrastruktur og data, hvilke forskningsområder som prioriteres, hvorfor de prioriteres og hvem som kan dra nytte av forskningsinnsatsen.

### *8. Forskningsformidling*

Forskere skal bidra til informert samfunnsdebatt, slik at samfunnets vurderinger funderes i realisme. Det er imidlertid utfordrende å sikre en balansert diskusjon om risikoer og muligheter ved kunstig intelligens. Fremstillingene av kunstig intelligens i offentligheten kan tendere mot en slags «moralsk panikk», der scenarier knyttet til muligheten for superintelligens fremheves. På den annen side kan muligheter overdrives og farer som konsekvens av teknologi underkommuniseres for å sikre finansiering av utvikling og forskning. Vi bør som samfunn unngå naivitet og være klar over mulige risikoer og muligheter, for eksempel at kunstig intelligens kan komme i gale hender.

Forskere har et særlig ansvar for å formidle risikoer og muligheter på edruelig vis, da de best vet hvor langt utviklingen har kommet.

### *9. Inspiserbarhet*

Åpenhet, forskningsetisk sett, innebærer å være åpen og eksplisitt om valg av datakilder, utviklingsprosesser og interessenter. *Inspiserbarhet* betegner en evne til å beskrive, inspisere og reprodusere hvordan beslutninger tas av systemene, samt opprinnelsen til data som brukes og lages av systemet. Inspiserbarhet er helt avgjørende for å sikre rettferdighet og tillit ved automatiserte avgjørelser, som "profilering", der personopplysninger blir brukt til automatisk å analysere eller forutsi forhold ved enkeltpersoner eller grupper (f.eks. knyttet til økonomi, helse og atferd). Sort-boks-problematikk betegner ulike utfordringer knyttet til at kunstig intelligens-systemene og algoritmene kan være så kompliserte at vi ikke forstår hvordan systemene har kommet frem til et svar eller en beslutning.

Et system bestående av en sort boks vil ofte kunne ha bedre ytelse enn et mer inspiserbart system, så forskere må i enkelte tilfeller gjøre en avveining mellom kvalitet og åpenhet. Forskere bør synliggjøre og begrunne slike avveininger, og det bør

være et mål å oppnå «glassbokser» i kunstig intelligens-forskning, dvs. systemer som lar seg inspisere.

NENT stiller seg til rådighet for departementet, og ønsker lykke til i arbeidet med en nasjonal strategi. Komiteen holder gjerne departementet orientert om arbeidet med en forskningsetisk betenkning om kunstig intelligens, og vil oversende denne når den er klar.

Med vennlig hilsen,

| | |
|---|---|
| Øyvind Mikkelsen | Helene Ingierd |
| Leder, NENT | Sekretariatsleder, NENT |