**NTNU S-sak /20**

**Norges teknisk-naturvitenskapelige**

**universitet**

01.02.2024
Saksansvarlig: Håkon Alstad
Saksbehandler: Heine Skipenes, Morten Sørlie
Arkiv: 2024/5301

**N O T A T**

Til: Rektor

Fra: IT-avdelingen og Avdeling for utdanning

**Om:** **Saksfremlegg – Innføring av KI-chat verktøy for studenter**

**Tilråding:**

*«IT-avdelingen anbefaler at rektor godkjenner innføring av verktøyet Microsoft Copilot (tidligere Bing Chat Enterprise) for alle ansatte og studenter på NTNU. Rektor godkjenner at verktøyet kan brukes til å behandle åpen og intern informasjon jf Retningslinje for klassifisering av informasjon. Verktøyet skal ikke benyttes til å behandle personopplysninger»*

**Bakgrunn**

IT-avdelingen skrudde i september 2023 på Microsoft sitt verktøy «Bing Chat Enterprise» (<https://www.bing.com/chat>) for alle ansatte på NTNU. Bing Chat Enterprise er en samtalerobot og en språkmodell som ligner på andre kjente verktøy som for eksempel ChatGPT. I motsetning til mange andre verktøy var sikkerhetsnivået høyere og løsningen var lett tilgjengelig for NTNU. Verktøyet er integrert i nettleseren og innlogging skjer automatisk for alle som er pålogget (Microsoftkonto). Informasjon om verktøyet finnes på innsida: [https://i.ntnu.no/wiki/-/wiki/Norsk/Bing+Chat+Enterprise](https://i.ntnu.no/wiki/-/wiki/Norsk/Bing%2BChat%2BEnterprise)

I september 2023 var Bing Chat Enterprise ikke gjort tilgjengelig for studentlisensene, og IT-avdelingen har siden dette jobbet med å tilby et KI-verktøy for studentene. IT-avdelingen har vurdert både Sikt sin Sikker KI-Chat, GPT.UIO og utvikling av en egen NTNU-løsning. IT-avdelingens ledergruppe falt til slutt ned på å jobbe videre med UiO sin løsning. For å kunne tilby UiO sin løsning på NTNU, var det nødvendig med noe utvikling som tidligst kunne være ferdig i februar 2024. Like før jul meldte Microsoft at de kommer til å tilgjengeliggjøre verktøyet Bing Chat Enterprise også for studentlisensene under et nytt navn: «Microsoft Copilot». Verktøyet tilbys som en såkalt «opt out»-løsning, det vil si at Microsoft skrur det på og NTNU må i så fall aktivt skru det av for at det ikke skal bli gjort tilgjengelig for brukerne. Det medfører ingen kostnad for NTNU å ta løsningen i bruk, men institusjonen må sørge for at alle vurderinger av informasjonssikkerhet og personvern er ivaretatt. Når et samtalerobotverktøy generativ kunstig intelligens skal tilbys alle brukere anbefales det å gjennomføre en personvernkonsekvensvurdering (DPIA) jf artikkel 35 i Personvernforordningen.

**Hvorfor samtalerobot/språkmodell for NTNUs ansatte og studenter?**

Hovedformålet med å tilgjengeliggjøre Bing Chat Enterprise var å kunne tilby et språkmodellverktøy for behandling av informasjon med litt høyere [informasjonssikkerhetsklassifiseringsnivå](https://i.ntnu.no/wiki/-/wiki/Norsk/Informasjonsklassifisering%2B-%2Binformasjonssikkerhet) (intern) enn det som anbefales for åpne verktøy på nett (åpen). Ved å tilby verktøy hvor ansatte kan legge inn interne data så reduserer vi risiko for at åpne plattformer som ChatGPT brukes til å behandle informasjon som NTNU ønsker å holde internt i virksomheten. IT-avdelingen ønsket også å tilby en lett tilgjengelig og trygg språkmodell der ansatte kan gjøre seg kjent med fordeler og ulemper med denne typen verktøy og teknologi. Språkmodeller er ny teknologi for mange, og mange har aldri turt å prøve. Ved å tilby en trygg «sandkasse» for testing ønsket IT-avdelingen å bidra til at alle ansatte skal få større kjennskap om hva denne typen verktøy kan bidra med av positive og negative effekter på måten ansatte kan jobbe på. I tillegg har IT-avdelingen ønsket å tilby et verktøy hvor vi er åpne om hvordan vi har tenkt risikovurdering, informasjonssikkerhet og personvern. IT-avdelingen håper at dette skal kunne bidra til økt forståelse for hvorfor og hvordan vi alle bør tenke på informasjonssikkerhet i hverdagen.

IT-avdelingen har tatt utgangspunkt i følgende behandlingsformål i vurderingen av om det skal tilbys et felles KI-verktøy for studenter og ansatte (samtalerobot / språkmodell):

1. Tilby en språkmodell for behandling av gule/interne data (ikke personopplysninger)
	1. Å tilby og anbefale et sikrere alternativ til studentene enn åpne tjenester på nett (for eksempel ChatGPT)
2. Tilby en lett tilgjengelig og trygg språkmodell der våre ansatte og studenter kan gjøre seg kjent med fordeler og ulemper med ny teknologi.
	1. Å tilby det samme verktøyet til både studenter og ansatte slik at fagspesifikk opplæring kan gis i ordinær undervisning
	2. Å tilby et verktøy raskt nok til at det kan inngå i undervisningsplanleggingen for våren 2024
3. Tilby en teknisk løsning hvor vi samtidig tilgjengeliggjør informasjon om hvordan vi har tenkt risikovurdering og personvern

IT-avdelingen ønsker ikke at betegnelsen «sikker» skal knyttes til dette verktøyet. Dette kan gi et feilaktig inntrykk av at løsningen kan brukes til å behandle data med et høyere informasjonssikkerhetsnivå enn åpen og intern. I tillegg anbefales det ikke å behandle personopplysninger i verktøyet.

IT-avdelingen har 31.01.2024 skrudd av tilgangen for studentene (det vil si at når Microsoft gjør verktøyet tilgjengelig for studentlisenser i verden så blir det ikke tilgjengelig for våre studenter). Verktøyet blir ikke gjort tilgjengelig for studentene før den blir aktivt skrudd på, men kan skrus på og vil være tilgjengelig for alle noen timer etterpå.

**Vurdering av konsekvenser – informasjonssikkerhet og personvern**

Forslag til personvernkonsekvensvurdering (DPIA) er utarbeidet i tråd med NTNUs retningslinjer ([https://i.ntnu.no/wiki/-/wiki/Norsk/Vurdere+personvernkonsekvenser](https://i.ntnu.no/wiki/-/wiki/Norsk/Vurdere%2Bpersonvernkonsekvenser)) og ligger vedlagt. Det er dette dokumentet med forslag til tiltaksplan som foreslås godkjent av rektor. DPIAen utarbeidet i samarbeid med Personvernklyngen i Fellesadministrasjonen, Personvernombud og i tråd med [Digdirs Veiledning for ansvarlig utvikling og bruk av kunstig intelligens i offentlig sektor.](https://www.digdir.no/kunstig-intelligens/kunstig-intelligens/4132) I tillegg til DPIAen gjøres det en egen teknisk risiko og sårbarhetsvurdering i risikovurderingssystemet Diri.

DPIA konkluderer med følgende overordnet risiko- og tiltaksliste:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Risiko**  | **Tiltak**  | **Effekt** **på** **risiko**  | **Restrisiko**  | **Tiltak** **godkjent**  |
| Manglende samsvar mellom den registrertes rettigheter og den behandlingsansvarliges plikter etter Personvernforordningen (GDPR) For eksempel manglende rett til innsyn, sletting osv | Må gjennomgå NTNUs personvernerklæring for å kontrollere at informasjon til den registrerte er tydelig nok kommunisert | *Redusert* | *Lav* | *Må gjennomgås på nytt* |
| Manglende opplæring eller lav kompetanse hos brukerOgEn bruker legger inn noen andre sine personopplysninger som kommer på avveie | Lage informasjonssider på nett norsk og engelsk om hva verktøyet er, og hva det kan brukes til Det blir opprettet en wikiside som det er mulig å finne informasjon om tekniske vurdering, risikovurderinger, personvernkonsekvensvurdering osv. | *Redusert* | *Lav* | *Dagens wikisider må oppdateres* |
| Manglende opplæring eller lav kompetanse hos bruker | Legge ut melding til alle ansatte og studenter om at verktøyet nå er på plass med informasjon om hva det kan brukes til. | *Redusert* | *Lav* | *Må gjennomgås* |
| Manglende opplæring eller lav kompetanse hos bruker | Tilby «å komme på besøk» til brukergrupper for å informere om verktøyet og bidra til økt bevissthet om bruk av kunstig intelligens  | *Redusert* | *Lav* | *Gjennomføres kontinuerlig på forespørsel*  |
| Manglende opplæring eller lav kompetanse hos bruker | I informasjonsmateriellet er det gitt informasjon om hvordan det er mulig å komme med tilbakemeldinger til oss som IT-avdeling, og komme med generelle tilbakemeldinger og refleksjoner om positive og negative ting med løsningen.  | *Redusert*  | *Lav* | *Gjennomføres kontinuerlig.* *Kun sporadiske tilbakemeldinger så langt* |
| Manglende opplæring eller lav kompetanse hos bruker | NTNU bør lage retningslinjer for bruk av generativ kunstig intelligens  | *Redusert*  | *Lav* | *Under gjennomføring, endelig høringsfrist 1/2-24.* |
| Manglende opplæring eller lav kompetanse hos bruker | Tema «hvordan skal vi forholde oss til Kunstig intelligens?», «hva slags kjøreregler bør NTNU ha?» osv har blitt behandlet i til medvirknings- og medbestemmelsesorganer: * SESAM (06.11.2023)
* Administrativt lederutvalg (24.11.2023)
* Formøte Studenttinget (27.11.23)
* Dekanmøtet (16.01.24)
 | *Redusert*  | *Lav* | *Under gjennomføring,*  |
| Tjenesten er ikke bra nok for våre brukere. Brukere ønsker seg verktøy som tar vare på data over tid og som kan «lære deg å kjenne». Denne funksjonaliteten kan ikke tilfredsstilles med krav til sletting, og det er risiko for at andre verktøy tas i bruk uten tilstrekkelig kunnskap om hva dette innebærer | Alle tiltak på lista over vil bidra til risikoreduksjon: * Informasjonsmateriell på innsida om hva tjenesten kan brukes til, og hva det ikke kan brukes til
* Retningslinjer
* Opplæring

I tillegg må det jobbes med videreutvikling og vurdere om NTNU også skal tilby verktøy som tilfredsstiller behovet for å ta vare på data over tid.  | *Redusert*  | *Lav* | *Under gjennomføring,* *Ansvar for oppfølging Heine Skipenes* |
| Risiko for at studenter blir tatt i fjusk, (plagiat, tekstlikhet osv) ved bruk av dette verktøyet. | Avdeling for utdanning har satt i gang et arbeid hvor en arbeidsgruppe med representanter fra alle fakultet skal se nærmere på klargjøring av rektorvedtaket om fjusk og skjema for redegjørelse for bruk av KI i arbeidet med besvarelser (deklarasjonsskjema). Gruppa vil også se på behovet for kompetanseutvikling knyttet til KI i utdanningsvirksomheten. | *Redusert*  | *Lav* | *Under gjennomføring,* *Ansvar for oppfølging Morten Sørlie* |

Sammendrag av personvernombudets råd:

*«Innføring av Microsoft Copilot gjelder utprøving av ny teknologi, og det er viktig å ha fortsatt trykk på det som gjelder rettighetene til den registrere. Fortsette utvikling av informasjon, opplæringsmateriell, svare på spørsmål osv, og ha tilstrekkelig med ressurser for å kunne «være der» for brukeren. En stor utfordring er hva de brukerne faktisk kommer til å bruke løsningen til, og det vil ikke være mulig å kontrollere eventuell feil bruk. God kontinuerlig informasjon blir derfor viktig.*

*Personvernombudet anbefaler ikke at verktøyet brukes til å behandle personopplysninger. Dette må komme tydelig frem i informasjonsmateriellet som gjøres tilgjengelig for brukerne.*

*Personvernombudet ønsker at det blir gjennomført kontroll av at tjenesten blir benyttet som tenkt, systematisk gjennomgang av hva brukerne spør om og revidert tiltaksliste. Personvernombudet ber om å få delta i denne prosessen.»*

**Fremtidige verktøy, muligheter og utfordringer**

Utviklingen av generativ kunstig intelligens feltet er det vanskelig å spå hvilke verktøy som blir «den neste ChatGPT» og hvilken retning teknologiutviklingen tar. IT-avdelingen følger kontinuerlig med på utviklingen for å se hvilke muligheter og utfordringer som dukker opp. Temaer det jobbes med:

* KI-chat verktøyet «Copilot» for studenter og ansatte er bare første skritt. Det er viktig at NTNU jobber videre med utvikling og tilgjengeliggjøring av nye og bedre verktøy, for eksempel verktøy som i større grad tilfredsstiller pedagogiske forventninger og tar vare på input/output over lenger tid.
* NTNU er det i dag mange som jobber med problemstillinger knyttet til kunstig intelligens, og Det er viktig med god samordning internt. For eksempel pågår det et viktig arbeid på utdanningsområdet, hvor vurderingene som gjøres her kan påvirke vurderinger som må gjøres innenfor andre områder. En arbeidsgruppe med representanter fra alle fakultet skal se nærmere på klargjøring av rektorvedtaket om fjusk og skjema for redegjørelse for bruk av KI i arbeidet med besvarelser. Gruppa vil også se på behovet for kompetanseutvikling knyttet til KI i utdanningsvirksomheten.
* Det arbeides med «Retningslinjer for ansvarlig bruk av (generativ) kunstig intelligens ved NTNU». Retningslinjene er i høringsprosess med endelig høringsfrist 1. februar.
* IT-avdelingen har søkt og fått opptak i Datatilsynets regulatoriske sandkasse for personvernvennlig innovasjon og digitalisering våren 2024 med prosjektet «Pilotere Microsoft 365 Copilot». Microsoft 365 Copilot er neste generasjons KI-verktøy, og NTNU gjennomfører et pilotprosjekt for å teste om Microsofts kunstige intelligente assistent kan bli tatt i bruk i en stor offentlig organisasjon. Det som skiller de to forskjellige «copilotene» ligger i at Microsoft 365 Copilot blir integrert i allerede eksisterende Microsoft-tjenester som Word, Excel, Powerpoint osv, og at verktøyet kan potensielt få tilgang til lokale filer og organisasjonsintern informasjon. Dette vil innebære et «teknologisk taktskifte» som kan gi NTNU mange muligheter, men også en del nye utfordringer. Microsoft bygger en kraftig merkevare i begrepet «Copilot», og det ser ut til at de fleste tastaturer på Windows11 PCer for eksempel utstyres med en egen Copilot-knapp.
* Kostnaden for neste generasjons KI-verktøy er høy og det er viktig at det settes i gang en prinsippdiskusjon om alle brukere skal få tilgang, hvilke finansieringsmekanismer som skal legges til grunn osv. IT-avdelingen kommer til å forberede og løfte en egen sak om dette i løpet av våren 2024.

Vedlegg: Personvernkonsekvensvurdering «Tilgjengelig språkmodell for NTNU / implementere Microsoft Copilot som verktøy ved NTNU»