Oslo

Byrådsavdeling for finans

Veileder for kunstig intelligens

Versjon 10.10.2023

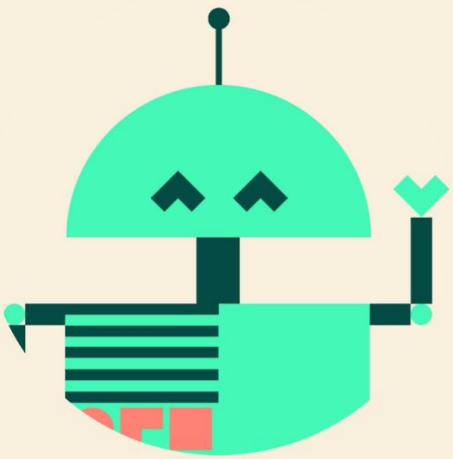

Foto: Anette Brekke-Bjørkedal

Innhold

Veileder for kunstig intelligens i Oslo kommune

1 DEL 1: Introduksjon

Kunstig intelligens, KI, har eksistert i mange år – bare tenk på spam-filteret i innboksen din. Det som er nytt er regnekraften og den åpne tilgjengeliggjøringen av generativ KI, som ChatGPT. EU jobber med ny lovgivning for kunstig intelligens, som etter hvert vil sette rammer også for norsk lov. Uten felles retning vil alle bruke KI på forskjellig måte.

Dette dokumentet gir deg veiledning i hvordan du i dag kan ta ansvar for, og prøve ut, kunstig intelligens. KI-veilederen baserer seg blant annet på rådene til Digitaliseringsdirektoratet og på Oslo kommunes strategi for kunnskapsbasert utvikling, innovasjon og digital teknologi, samt nasjonale og europeiske retningslinjer og anbefalinger – med mål om at bruk av kunstig intelligens i Oslo kommune skal være lovlig, etisk og sikker.

1.1 Definisjon

«Kunstig intelligente systemer utfører handlinger, fysisk eller digitalt, basert på tolkning og behandling av strukturerte eller ustrukturerte data, i den hensikt å oppnå et gitt mål.» (Regjeringen, Nasjonal strategi for kunstig intelligens, 2020).

1.2 Hva finner du i denne utgaven av veilederen?

Denne veilederen skal bidra til etisk og ansvarlig bruk av kunstig intelligens i Oslo kommune. KI skal forbedre kommunens tjenester og sørge for at KI blir brukt på en kostnadseffektiv og målbar måte - når KI er hensiktsmessig. Den er bevisst laget som en *veileder* og gir kun *forslag* til retningslinjer for virksomhetene. Det er virksomhetene selv som er ansvarlig for å innføre disse. Veilederen retter seg mot deg som er ansatt og skal bruke KI, og mot deg som er ansvarlig i virksomhets- eller tjenesteutvikling og vil bruke KI til dette.

Hva finner du *ikke* i denne utgaven av veilederen?
Veilederen vil ikke gi et totalt bilde av hva som skal til for å ta i bruk KI i Oslo kommune, men den er et første steg på veien.

1.3 Tre dimensjoner KI: Integrert KI, egenutviklet KI og generativ KI

1. Med *integrert KI* mener vi for eksempel synonymforslag i Word, Workplace-algoritmer, bakgrunnsfilteret på Teams og Excel smart fill. Dette er KI som «siver inn» i vår arbeidshverdag gjennom kontorstøtteprogrammene vi bruker.
2. Det er forskjell på å bruke KI og utvikle KI. Veilederen vil ikke gå i dybden på *utvikling av KI*, som handler om å løse et problem ved å bruke data, metoder, teknikker og teknologi for å utvikle et KI-system, men veilederen peker på ressurser som tar for seg dette.

3. Veilederen vil gå mer i dybden på *generativ KI* og gi forslag til retningslinjer for bruk i virksomhetene. Dette er KI som mange ansatte vil ta i bruk, bevisst eller ubevisst.

Generativ KI er en ny generasjon KI. Den bruker maskinlæringsteknikker for å *generere* nytt innhold som ligner på menneskeskapt innhold. At KI er generativ betyr at den ikke bare tolker informasjon, men også skaper nytt, originalt innhold. Husk at denne type KI er språkmodeller, ikke faktasjekkmodeller eller søkemotorer. Generativ KI lager tekst – som for eksempel epost, informasjonsartikler, CV-er, søknader, dikt, kodelinjer, musikk og brev, basert på den tekstlige inputen den får fra deg. Generativ KI oversetter også mellom språk, til og med mellom programmeringsspråk. Slik kunstig intelligens har stort potensial for å påvirke måten vi jobber på, og er et nytt fenomen som virksomhetene i Oslo kommune må forholde seg til. Noen av de mest utbredte og kjente verktøyene som finnes i dag er ChatGPT, Dall-E, Midjourney, Bard AI, LaMDA, Aleph Alpha, Bloom og Stable Diffusion. Generative KI-modeller er systemer som er designet for å virke «menneskeaktige».

1.4 Regelverk for KI er fortsatt i støpeskjeen, men det finnes likevel rammer

Kunstig intelligens er et fagområde som beveger seg raskt. Vi har allerede mange lover som gjelder ved bruk av KI. Utfordringen ligger i evnen til å anvende lovene, som Grunnloven, menneskerettighetsloven, forvaltningsloven, likestillings- og diskrimineringsloven og personopplysningsloven med forordning. Disse lovene har per i dag ikke egne kapitler om KI, men du som skal bruke og utvikle KI må forholde deg til dem.

I kommunen har vi egne instrukser for hvordan vi skal bruke IKT, men dagens IKT-retningslinjer omfatter ikke generativ kunstig intelligens. Gjeldende IKT-reglement er fra 2010, og det nærmeste vi kommer å beskrive hvem som har ansvar for generativ kunstig intelligens er i paragraf 10:
«Den virksomhet som er systemeier har ansvaret for at systemet dekker tjenestens behov, at det er i samsvar med gjeldende lover og regler og at det blir tilpasset ved behov for endringer. Systemeier har videre ansvar for at systemet implementeres og brukes på en optimal måte, slik at systemets gevinstpotensial realiseres.»

Virksomhetsleder har ansvar for informasjonssikkerhet, informasjonsforvaltning og etterfølgelse av personvernregelverket. En utfordring virksomhetsleder møter er at siden 2010 har verden endret seg. Google er for eksempel nå en integrert del av ansattes hverdag uten at Google har en systemeier i kommunen. Det samme skjer med ChatGPT. Hvem som helst kan nå opprette en konto og bruke generativ KI som ChatGPT og Dall-E. Det må vi håndtere.

1.4.1 Kommende regelverk

Den kommende KI-forordningen fra EU vil ha betydning for den videre utviklingen av rammer for KI også i Oslo kommune.

1.5 Nyttige begreper

Du kan klikke på lenkene.

Dette sier Store norske leksikon om ordbruken rundt KI. «Man sier ofte at en datamaskin som er i stand til å løse oppgaver uten å få instruksjoner fra et menneske om hvordan den skal gjøre det, har kunstig intelligens. For eksempel foreslår «intelligente» søkemotorer treff på grunnlag av data om tidligere søk og annen brukeradferd. Dette kalles *maskinlæring*, og har en lang rekke bruksområder, fra enkle programmer i smarttelefoner til chatbots og selvkjørende biler. Det finnes også kunstig intelligens som ikke lærer, for eksempel såkalte *regelbaserte systemer*, hvor komplekse regler for «intelligent adferd» er spesifisert av mennesker på forhånd.
Det arbeides spesielt mye med kunstig intelligens innen språkteknologi, talegjenkjenning, bildegjenkjenning, brukerinteraksjon og styring av fysiske prosesser. Når det snakkes om kunstig intelligens, refereres det som regel til såkalt *dype nevrale nettverk* eller *dyp læring*.*
Som fagfelt er kunstig intelligens en sammensmelting av datateknikk, logikk, matematikk, psykologi og nevrovitenskap. Begrepet har vært en del av datateknikken siden 1950-årene, men det er særlig siden 2010-tallet at teknologien har fått stor utbredelse. I dag er kunstig intelligens en vital del av utviklingen innen informasjonsteknologi, og introduseres på stadig nye felt.» **Ett av flere felt innen kunstig intelligens.*

1.6 Risikovurdering av kunstig intelligens

Per oktober 2023 finnes ennå ikke et nasjonalt regelverk som fastsetter hvordan risikovurdering av kunstig intelligens skal gjennomføres, i motsetning til for eksempel på informasjonssikkerhetsområdet.

Digitaliseringsdirektoratet foreslår ved KI-risikovurdering å ta utgangspunkt i:

- Eksisterende sikkerhetsarbeid
- Følge standarder som gir god veiledning til risikovurderinger
- Være oppmerksom på at regelverk kan påvirke risikovurderingen
- Og gir noen råd til hvordan du kan gjennomføre risikovurdering av et KI-system (lenke)

Oslo kommune har i tillegg egne rutiner og retningslinjer for risikovurderinger, utarbeidet av Byrådsavdeling for finans og publisert på Felles intranett.

2 DEL 2: Til deg som er ansatt og som skal bruke KI

Du skal være trygg på at du når du bruker Oslo kommunes IKT-løsninger, så har arbeidsgiver ansvaret for at programmene du bruker er innenfor regelverket som gjelder for KI.

Integrert KI i Oslo kommune sine fellessystemer er det UKE som har ansvar for, mens integrert KI i IKT-systemer som virksomhetene har anskaffet er det virksomheten selv som har ansvar for.

Vær veldig varsom med å bruke IKT-verktøy i ditt arbeid som ikke er arbeidsgivers. For eksempel: Vit at når du googler noe, så bruker du KI. Og når du taster noe inn i et eksternt KI-verktøy for å lage et bilde eller få forslag til en tekst, så deler du informasjon som blir lagret et sted som du ikke vet noe om, og som det kan være tilnærmet umulig å få innsyn i. Husk også at det du skriver inn blir en del av språkmodellens treningsgrunnlag. For å gjøre det enklere for ansatte å ta i bruk generativ KI er det utarbeidet Vær varsom-prinsipper samt retningslinjer som virksomhetene kan ta i bruk:

2.1.1 Oslo kommunes Vær varsom-prinsipper for bruk av generativ KI

1. KI er en maskin, ikke et menneske - selv om det føles slik.
2. KI kan oppgi opplysninger som ser riktige ut, men som kan være helt feil.
3. Bruker du KI-tjenester i jobbsammenheng, vær klar over at informasjon om deg og det du skriver inn blir brukt til å trene opp KI-modellen. Det glemmes ikke.

2.1.2 Oslo kommunes retningslinjer for bruk av generativ KI

Retningslinjene under er basert på Digitaliseringsdirektoratets retningslinjer for generativ KI i offentlig sektor, retningslinjer for bruk for EU-kommisjonens ansatte, og for Harvard-universitetets studenter.

Du skal aldri

- dele personlig informasjon, som personopplysninger om ansatte og innbyggere
- dele sensitiv informasjon, som virksomhetskritiske opplysninger
- dele brukerinformasjon, som påloggingsinformasjon og passord
- lime inn interne eller sensitive dokumenter, eller utklipp fra disse
- stole på at en KI sier sannheten – sjekk alltid fakta
- bruke kun KI til å ta beslutninger som i betydelig grad påvirker personer

ChatGPT og annen generativ KI: Kan jeg skrive inn personopplysninger?
KI-veilederen anbefaler å vente med å bruke generativ KI til personopplysninger. Det er usikkert om personopplysningsloven kan ivaretas i møte med dagens utgave av generativ KI. Vi vet lite om hvordan disse språkmodellene fungerer. Du må være forberedt på betydelig juridisk forarbeid dersom du ønsker å bruke personopplysninger i generative KI-tjenester, per i dag. Bruk av personopplysninger må holde seg innenfor GDPR, og slik bruk er derfor noe ledelsen i din virksomhet må ta stilling til.

Du skal alltid

- være kritisk til det du leser – se opp for unøyaktigheter, skjevheter og feilaktig informasjon
- være kritisk til om KI bryter opphavsrett, særlig for tredjepart
- tydelig merke innhold som du har brukt generativ KI til å lage. For eksempel "Denne illustrasjonen er laget ved hjelp av KI", i tillegg til kreditering av KI-tjenesten du har brukt
- huske at du er ansvarlig for KI-generert tekst du står som avsender av, eller publiserer
- være transparent og si fra om når du bruker en KI-generert tekst, bilde, kode eller annet
- være oppmerksom på at KI har gjort det lettere for trusselaktører å lage svindel. Kontakt alltid kommunens team for datasikkerhetshendelser om du er i tvil om noe er ekte eller fake: csirt@oslo.kommune.no

3 DEL 3: Til deg som er øverste ansvarlig i virksomhetene

Øverste leder i virksomhetene er ansvarlig for virksomhetens bruk av kunstig intelligens. Opplæring av ansatte og bevisstgjøring om prinsipper og retningslinjer for bruk er også virksomhetsleders ansvar.

Transformativ teknologi som KI krever at virksomhetsledelsen tar bevisste og informerte valg knyttet til informasjonsforvaltning, informasjonssikkerhet, personvern, virksomhetskritisk informasjon og dataverdier. *Hva kan virksomheten bruke KI til* og *Hva bør virksomheten bruke KI til?* Står du overfor mulig KI-bruk med lav risiko, som dekker et stort behov? Eller omvendt?

- Når det gjelder *integrert KI* der det er snakk om behandling av personopplysninger, må du sørge for at databehandleravtaler dekker behandlingen av personopplysninger til KI-formål.
- Når det gjelder *generativ KI* anbefaler vi at virksomheten tar i bruk retningslinjer, for eksempel med utgangspunkt i Digitaliseringsdirektoratets veiledningsmateriale for ansvarlig utvikling og bruk KI i offentlig sektor.
- Når det gjelder å *utvikle KI selv* i virksomheten kan du ta utgangspunkt i Metier AI Training Model, samt Oslo kommunes strategi for kunnskapsbasert utvikling, innovasjon og digital teknologi (lenke) som gir nyttige retninger om du vurderer utvikling av helt nye KI-løsninger.
- Når det gjelder KI-strategi kan du for eksempel ta utgangspunkt i Metiers 5-stegs modell for AI- og digitaliseringsstrategi.

4 DEL 4: Til deg som skal utvikle KI

Nyutvikling av KI bør ta utgangspunkt i personvern og transparens. Det er viktig at virksomheter som utvikler eller tar i bruk løsninger med kunstig intelligens har fokus på innebygget personvern fra starten av, for eksempel ved at virksomheten sørger for at innhenting og bruk av treningsdata skjer på en lovlig og åpen måte. For å finne ut av hvordan du kan balansere risiko og muligheter bør du sette deg inn i Digitaliseringsdirektoratets veiledning for utvikling av KI.

KI-gevinster som effektivisering og bedre kvalitet på tjenester er effekter som kommunen ønsker. På den annen side innebærer utvikling av nye KI-løsninger stor risiko, som du både skal kunne vurdere og håndtere - som kvalitet på datagrunnlaget, datasikkerhet, personvern, og etiske og juridiske implikasjoner. For å ivareta disse forholdene kreves tverrfaglige, kompetente team.

Kommunens utviklingsstrategi skal legges til grunn for vurderingene, og du må være nøye med å holde deg innenfor alle lover og regelverk.

4.1 Finn mer veiledning for utvikling av KI her

- Digitaliseringsdirektoratets veiledning for ansvarlig bruk og utvikling av kunstig intelligens (lenke)
- Oslo kommunes strategi for kunnskapsbasert utvikling, innovasjon og digital teknologi (lenke)
- Datatilsynets regulatoriske sandkasse (lenke)