

Vejledning til medarbejdere om retningslinjer for brug af generativ kunstig intelligens i departementet

Indledning

Generativ kunstig intelligens (AI) har mange nyttige anvendelsesmuligheder, men måden de er opbygget på medfører samtidig en række faldgruber, som medarbejdere i departementet skal være opmærksomme på. Formålet med retningslinjerne for brug af generativ AI i departementet er at sørge for, at værktøjerne anvendes på en ansvarlig måde over for borgere, virksomheder og medarbejdere. Både ift. hvordan vi bruger generativ AI og til hvilke formål. Samtidig skal vi sikre, at anvendelsen af generativ AI sker inden for rammerne af eksisterende lovgivning, herunder databeskyttelsesretten, forvaltningsretten samt immaterialretten.

Retningslinjerne er udarbejdet på baggrund af Digitaliseringsstyrelsens interne retningslinjer på området samt Digitaliseringsstyrelsen Guide til offentlige myndigheder om ansvarlig brug af generativ AI. Retningslinjerne er godkendt af departementets Cyber-, Informationssikkerhed og Databeskyttelsesudvalg (CID-udvalget) den 13. december 2023 og vil løbende blive opdateret i takt med udviklingen inden for generativ AI og departementets arbejde med området.

Hvis du er i tvivl, om brugen af generativ AI er i overensstemmelse med retningslinjerne, skal du lade være med at benytte værktøjerne.

Hvad er generativ AI

Generativ AI er værktøjer, der udnytter meget store datamængder sammen med maskinlæringsteknikker til at producere indhold, baseret på input fra brugerne kaldet prompts. Prompts er de spørgsmål eller kommandoer, brugeren giver til et generativt værktøj for at få et output.

Generativ AI er ikke intelligens i menneskelig forstand. I stedet er det avancerede modeller, der forsøger at forudsige, hvilken tekst, billede, lyd, video eller andet, der med størst sandsynlighed tilfredsstiller den givne prompt. Det er med andre ord (meget) avanceret statistik.

Eksempler på generative AI-værktøjer er ChatGPT og Google Bard.

Retningslinjer

1. Sensitive oplysninger må ikke deles med AI-værktøjer, som er gratis tilgængelige via en offentlig hjemmeside.

Årsag: Risiko for brud på f.eks. databeskyttelsesretten, immaterialretten (ophavsret mv.) eller tavshedspligten.

Sensitive oplysninger er en samlebetegnelse, som omfatter to overordnede kategorier:

- Alle personoplysninger omfattet af databeskyttelsesforordningen og databeskyttelsesloven (uanset om de er fortrolige eller ej)
- Andre fortrolige oplysninger omfattet af tavshedsbestemmelserne i forvaltningsloven og anden lovgivning.

Personoplysninger omfattet af databeskyttelsesforordningen/databeskyttelsesloven er enhver oplysning, der direkte eller indirekte kan henføres til en fysisk person. Det kan f.eks. være en persons fulde navn, CPR-nummer, en bils

registreringsnummer eller et fingeraftryk. Det kan også være f.eks. en adresse, et telefonnummer, en arbejds-mailadresse, en IP-adresse, eller et kontonummer.

Personoplysninger, der er blevet erstattet af en kode, fx B26482, og hvor det er muligt, ved hjælp af supplerende oplysninger, at føre koden tilbage til en bestemt person, er "pseudonymiserede" personoplysninger.

Pseudonymiserede personoplysninger er også personoplysninger.

Det er desuden vigtigt at være opmærksom på, at der stadig kan være tale om personoplysninger, selvom man f.eks. fjerner navn fra høringssvar eller rapporter, der behandles af det generative AI-værktøj. Det vil f.eks. være tilfældet, hvis der er andre oplysninger i dokumentet, der – selvom navnet ikke fremgår – kan henføres til en fysisk person.

Andre fortrolige oplysninger kan f.eks. være forretningshemmeligheder eller oplysninger som er nødvendige at hemmeligholde af hensyn til det offentlige interesser.

Hvis du er det mindste i tvivl om, hvorvidt en oplysning er sensitiv, skal du kontakte ISDB-enheden isdb@sum.dk.

Hvorfor?

I departementet arbejder vi med mange sensitive oplysninger, som skal holdes inden for vores interne systemer. Det er f.eks. personoplysninger, politiske sager og forretningsoplysninger.

Hvis nogle af disse deles med en tredjepart, kan der ske brud på f.eks. databeskyttelsesretten, immaterialretten (ophavsret mv.) eller tavshedspligten.

Persondatabeskyttelse. Departementet har ikke på nuværende tidspunkt godkendt generative AI-værktøjer til brug til personoplysninger. Personoplysninger må derfor ikke indtastes i generative AI-værktøjer, som er gratis tilgængelige via offentlige hjemmesider.

Immaterialretten. Når der anvendes generative AI-værktøjer, som er gratis tilgængelige via en offentlig hjemmeside, afgives rettighederne til, hvordan værktøjet anvender både prompts, metadata og svar. Udbyderne kan således bruge materialet i andre sammenhænge, som de finder relevant. F.eks. som træningsmateriale til videreudvikling af værktøjet.

Tavshedspligt. Offentligt ansatte har tavshedspligt. Det indebærer, at fortrolige oplysninger, som der opnås adgang til i forbindelse med arbejdet, ikke må videregives til uvedkommende. Departementet har ikke på nuværende tidspunkt godkendt generative AI-værktøjer til brug for fortrolige oplysninger. Fortrolige oplysninger må derfor ikke indtastes i generative AI-værktøjer, som er gratis tilgængelige via offentlige hjemmesider.

2. Tjek og gennemgå det indhold, der genereres af AI

Årsag: Værktøjerne har flere faldgruber såsom bias, opdigtede svar og potentielle brud på immaterialretten.

Hvorfor?

Når du bruger generative AI-værktøjer, er du ansvarlig for, hvordan du anvender det indhold, som værktøjet producerer. Det kan sammenlignes med, når autokorrektoren på din telefon ændrer et ord. Her er du som forfatter af teksten fortsat ansvarlig for, at teksten er korrekt, selvom det er teknologien, der har lavet ændringen. Teknologi er i mange tilfælde en stor del af vores arbejde, men den må ikke tilsidesætte vores dømmekraft, og den fjerner ikke vores ansvar. Derfor skal indhold, som produceres af generativ AI, altid gennemtjekkes, inden det bliver brugt. Du skal tjekke, om indholdet stemmer overens med det resultat, du efterspurgt i din prompt. Du skal især være opmærksom på, om indholdet er korrekt, om det er fri for bias, og om det krænker immaterialretten.

Generativ AI kan hurtigt producere indhold, men der er også risiko for unøjagtigt, forældet eller forkert indhold. F.eks. gør Digitaliseringsstyrelsen opmærksom på, at ChatGPT 3.5 har begrænset viden om verden efter september

2021, da datagrundlaget for versionen stopper her. Samtidig er modellerne ofte "overmodige", hvilket betyder, at de altid vil give dig et svar, selv hvis de ikke har et. F.eks. kan det opleves, at generativ AI opdigter informationer om offentligt kendte personer. Det er dit ansvar at kontrollere, at oplysningerne er korrekte ved selv at faktatjekke oplysningerne på anden vis.

Hvad kan jeg kigge efter?

- **Unøjagtige oplysninger**, herunder henvisninger til begivenheder, organisationer eller andre faktuelle oplysninger. Det kan være svært at gennemskue, hvorvidt der er tale om opdigtet indhold, da AI-værktøjer er gode til at få det til at fremstå troværdigt. Den skriftlige fremstilling afspejler altså ikke altid, at noget er galt.
- **Bias i indholdet**: Vær opmærksom på om indholdet har bias over for f.eks. specifikke køn, etniciteter, handikappede eller andre. Udbyderen af AI-værktøjet kan selv have indlagt regler, som styrer dens output. F.eks. kan der være anvendt sprogbrug, som nedtoner særlige problemer, eller der kan være emner, modellen undgår at udtale sig om. Du skal derfor være kritisk over for, om der er afledte ideologiske budskaber eller politiske holdninger i teksten.
- **Brud på immaterialretten**: Generativ AI bygger på eksisterende indhold fra bl.a. internettet og kan derfor give resultater, der sammenstykker elementer fra indhold, som kan være beskyttet af ophavsret, varemærker, patenter mv. Det kan f.eks. være logoer, slogans, tekststykker, lyd eller kildekode til patenteret software. Som altid, skal vi være opmærksomme på, at vi ikke publicerer indhold, som bryder immaterialretten.

Området er endnu nyt, og der er derfor usikkerhed, om hvor grænsen går. Der pågår flere retlige stridigheder, som på sigt vil kunne danne præcedens vedrørende immaterialret og generativ AI. Det er dog en god ide ikke at prompte generativ AI til at drage inspiration fra indhold, som er beskyttet af ophavsret, varemærker, patenter mv., da svarene derved vil kunne krænke rettighederne.

3. Brug ikke generativ AI til automatisk at træffe afgørelser

Årsag: Forvaltnings- og databeskyttelsesrettens principper om bl.a. gennemsigtighed og retten til ikke at være genstand for automatisk afgørelse.

Hvorfor?

Generative AI-værktøjer er trænet til at kunne generere nyt indhold. Værktøjerne er bygget på store mængder data, der især stammer fra internettet. Det betyder, at det er svært at gennemskue baggrunden for svarene på prompts, og derved om det bygger på et korrekt grundlag for afgørelser. Brug af generativ AI til at træffe afgørelser vil derfor være problematisk ift. forvaltnings- og databeskyttelsesrettens principper om bl.a. gennemsigtighed og retten til ikke at være genstand for automatisk afgørelse.

Gennemsigtighed i forvaltningen. Borgere og virksomheder skal ifølge forvaltningsloven kunne oplyses om, hvad der ligger til grund for en afgørelse. Da generative AI-værktøjer bygger på komplicerede matematiske modeller, er det på nuværende tidspunkt svært eller umuligt at forklare, hvordan de når frem til deres svar.

Retten til ikke at være til genstand for en fuldautomatiseret afgørelse. Som hovedregel er der ret til ikke at blive underlagt en afgørelse, der alene er baseret på automatisk behandling, som har retsvirkning eller på tilsvarende vis betydeligt påvirker vedkommende.

Det betyder, at det ikke er tilladt at bruge generativ AI til automatisk at træffe afgørelser i departementet. Der skal altid være en medarbejder/leder inde over konkrete vurderinger, beslutninger og godkendelser.

4. Brug af ikke personhenførbare e-mailadresser

Det er et krav, at du benytter en ikke-personhenførbare e-mailadresse, hvis du skal anvende generative AI-værktøjer som led i løsning af dine opgaver. Du får en ikke-personhenførbare e-mailadresse ved at skrive til InternIT@sum.dk. Adressen ser ud på denne måde "xxxx@chatgpt-test.dk". Mail-adressen fungerer som et "alias" til din personlige mailbox i Outlook, så du kan registrere dig hos f.eks. Chat-GPT med denne e-mailadresse og modtage bekræftelsesmail i din egen postkasse. Bestilling af en ikke personhenførbare e-mailadresse skal godkendes af din nærmeste chef, jf. punkt 5.

5. Aftale med din nærmeste chef, hvordan du bruger generativ AI til at løse dine opgaver

Årsag: Gennemsigtighed, som både giver mulighed for erfaringsdeling, dokumentation og opmærksomhed på evt. faldgruber.

Hvorfor?

Gennemsigtighed om brugen af generativ AI er både vigtig i forhold til dokumentation af myndighedsopgaver og for videndeling. Samtidig er ansvarlig brug af generativ AI et ledelsesansvar. Derfor skal du aftale med din nærmeste chef, hvordan du evt. bruger generativ AI i opgaveløsningen.

Ved at gøre opmærksom på og aftale, hvad generativ AI bliver brugt til, kan erfaringerne blive delt. Det giver dine kollegaer og chef mulighed for at være opmærksom på fordele såvel som muligt bias eller øvrige faldgruber.

Versioner

Version	Godkendt af	Gyldig fra dato	Status	Bemærkninger/ændringer
1.0	AC	15-01-2024	Godkendt	
1.1	CID-udvalg 10-9-2024	10-09-2024	Vedtaget	Krav om brug af ikke-personhenførbare e-mail og udbygget med afsnit om fortrolige oplysninger.