

Vejledning til danske institutioner og virksomheder

Om anvendelse af ophavsretsbeskyttet indhold til udvikling af kunstig intelligens

Behov for skærpet opmærksomhed

Der har vist sig et behov for at skærpe opmærksomheden på de rettighedsmæssige forhold, når forskere, institutioner, organisationer og virksomheder udvikler kunstig intelligens (AI) på baggrund af indhold skabt af den kreative branche i Danmark.

RettighedsAlliancen, der arbejder for at beskytte danske rettighedshavere mod ulovlig brug af deres kreative indhold såsom litteratur, nyhedsartikler, film, serier og musik, har fulgt udviklingen af kunstig intelligens tæt.

Vi har igennem det seneste halvandet år undersøgt, hvordan dansk indhold bruges i forbindelse med træning af generative AI-modeller, og hvordan indhold uploades til eller bruges til at promte AI-løsninger med. Vi har desværre måtte konstatere, at denne brug af ophavsretsbeskyttet indhold sker uden de påkrævede aftaler med rettighedshaverne.

RettighedsAlliancen har på denne baggrund iværksat en række håndhævelsesaktiviteter for at stoppe den ulovlige brug af vores medlemmers indhold.

Ophavsretten og AI

Denne vejledning giver et kortfattet overblik over, hvordan ophavsretsbeskyttet indhold kan anvendes lovligt til at træne kunstig intelligens. Vejledning beror på RettighedsAlliancens praktiske erfaringer med at beskytte rettigheder mod ulovlig brug til udvikling af AI-modeller. Samtidig beror vejledningen på nogle principielle standpunkter fra rettighedshavernes side.

Udgangspunktet i ophavsretslovgivningen er, at rettighedshaverne til ophavsretsbeskyttet indhold skal give samtykke til, at deres indhold anvendes i forskellige sammenhænge, fx kopiering og tilrådighedsstillelse for almenheden. Dette udgangspunkt er fraveget i nogle særlige situationer af relevans for træning af kunstig intelligens med reglerne om tekst- og datamining (TDM) i ophavsretslovens § 11 b (generel adgang) og 11 c (forskning), som trådte i kraft den 7. juni 2023.

TDM-begrebet kan omfatte analyseprocesser i forbindelse med træning af generativ AI. Det er dog uklart, om TDM med henblik på træning af generativ AI altid er lovligt, da det kan vise sig at være uforeneligt med grundlæggende ophavsretlige principper, herunder 3-trins testen, men også EU-retslige principper om snæver fortolkning af undtagelser.

Det står klart af udmeldingerne fra de store AI virksomheder, at træning af både generativ som ikke-generativ AI kræver kopiering af træningsdata.¹ Derudover står det klart fra en række retssager, at færdigtrænede AI modeller er i stand til at reproducere ophavsretsbeskyttet indhold.²

Rettighedshaverne lægger i denne forbindelse til grund, at der ved træning af AI-modeller sker en reproduktion af evt. værker fra modellens træningsdataset, samt at de færdigtrænede modeller indeholder kopier af træningsdata og dermed evt. beskyttede værker. Deraf følger også, at der ved udbud og brug af AI-tjenester sker en overføring til almenheden af det beskyttede indhold, der er trænet på.

Vejledning om anvendelse af beskyttet indhold i udvikling af AI

På baggrund af ovenstående og relevante bestemmelser i ophavsretsloven kan RettighedsAlliancen give følgende generelle vejledning:

1. Der må ikke laves kopier af og trænes AI modeller på beskyttet indhold, der lå på offentligt tilgængelige hjemmesider, og som blev indsamlet før den 7. juni 2023, medmindre der er givet udtrykkeligt samtykke hertil fra de relevante rettighedshavere til indholdet.
2. Forudsat at rettighedshaverne ikke har taget forbehold jfr. pkt. 3, må der gerne laves kopier af beskyttet indhold, der ligger på offentligt tilgængelige hjemmesider, og som er indsamlet efter den 7. juni 2023, hvis kopieringen sker med henblik på tekst- og datamining (TDM). Bemærk usikkerheden omtalt ovenfor omkring anvendelsesområdet for TDM med henblik på træning af generativ AI.
3. Der må ikke laves kopier med henblik på tekst- og datamining, hvis rettighedshaverne har taget et forbehold på passende vis mod, at deres indhold må blive kopieret til brug for tekst- og datamining.
 - Hvis indholdet ligger til rådighed for almenheden online, så skal forbeholdet tages i maskinlæsbar form, herunder i metadata og i vilkår og betingelser for brug af et websted eller en tjeneste, hvor indholdet ligger tilgængeligt.
 - Hvis indholdet ikke ligger på internettet, så kan forbeholdet tages på anden vis, herunder ensidigt eller via kontraktmæssige aftaler.

¹ Fx skriver Google: "Fundamentally, and as explained above, the training of ML models, including those underlying both generative and non-generative AI systems, captures the statistical relationships among training data, such as, in the case of an LLM, the relationships between words as they are used in writing. If training could be accomplished without the creation of copies, there would be no copyright question here." side 9 <https://www.documentcloud.org/documents/24117935-google>.

² Fx New York Times v OpenAI: https://nytc0-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf

4. Forskningsorganisationer og kulturarvsinstitutioner må – uanset om rettighedshaverne har taget forbehold – lave kopier af beskyttet indhold, på betingelse af organisationen eller institutionen har en lovlige adgang til indholdet, og det sker med henblik på tekst- og datamining i en forskningsmæssig sammenhæng. Se pkt. 5 og 6 om lovlige adgang.
5. Det er ikke tilladt at omgå effektive tekniske foranstaltninger, såsom DRM-beskyttelse og krav om tegning af abonnement for at få adgang til og lave kopier af beskyttet indhold.
 - Eksempelvis må der ikke laves kopier af videoklip fra YouTube, da YouTube har implementeret tekniske beskyttelsesforanstaltninger mod download af indhold på platformen.
 - Hvis en platform for fx videnskabelige artikler eller e-bøger kræver, at man har et abonnement for at få adgang til indholdet, så er en virksomhed eller forsknings- og kulturarvsinstitution nødt til at have et abonnement for at få adgang til at kopiere indholdet med henblik på tekst- og datamining. For virksomheders vedkommende skal forbehold overholdes jf. pkt. 3 ovenfor.
6. Der må aldrig foretages kopiering med henblik på tekst- og datamining af indhold, der ligger på internettet uden rettighedshavernes samtykke.
 - Derfor er det fx ulovligt at bruge det meget omtalte "Books3-datasæt" og lignende datasæt med piratkopierede bøger og øvrigt beskyttet indhold.
 - Det er fx heller ikke tilladt at kopiere indhold fra sociale medier, som er lagt op uden samtykke fra rettighedshaverne.
7. Det er vedkommende, som laver kopier til brug for tekst- og datamining, herunder træning af AI, der har ansvaret for at sikre, at der er en lovlige adgang til indholdet, herunder at det ligger lovligt tilgængeligt på internettet.
8. Såfremt en virksomhed eller institution allerede er i besiddelse af kopier af beskyttet indhold, eller får udleveret indhold fra en tredjepart, må indholdet alene anvendes til at træne kunstig intelligens, såfremt de relevante rettighedshavere har givet udtrykkelig tilladelse hertil.
9. Pligtafleverede og -indsamlede værker må ikke uden en konkret aftale med de relevante rettighedshavere anvendes til tekst- og datamining.
10. Såfremt der trænes videre på allerede eksisterende AI-modeller, er det videreudviklernes ansvar at sikre, at der ikke trænes videre på modeller, der oprindeligt er trænet uden lovlige hjemmel eller på ulovligt indhold.
11. Det er ikke tilladt at promte en AI tjeneste med kopier af ophavsretsbeskyttet indhold. Det kan fx være et prompt med hele eller dele af teksten fra en bog i

ChatGPT. I dette eksempel skabes der digitale kopier af teksten på servere tilhørende OpenAI.

12. Det er ikke tilladt at uploade kopier af ophavsretsbeskyttet indhold til en AI-tjeneste uden forudgående tilladelse fra de relevante rettighedshavere. Det kan fx være i forbindelse med skabelsen af en GPT med OpenAI's GPT-service, hvor der uploades én eller flere bøger som "ekstra viden" til GPT'en. I dette eksempel skabes der digitale kopier af teksten på servere tilhørende OpenAI. Derudover gøres kopier af bøgerne tilgængelige for almenheden, hvis GPT'en offentliggøres.
13. Konsekvensen af manglende overholdelse af ophavsretslovgivningen i forbindelse med udviklingen af AI-modeller og -tjenester er, at rettighedshaverne til det beskyttede indhold i træningsdata kan kræve modellerne eller tjenesterne fjernet fra internettet og destrueret, ligesom rettighedshaverne kan kræve erstatning og vederlag for uberettiget anvendelse af deres indhold.

Vejledningen er med forbehold for anden lovgivning, eksempelvis persondata- og konkurrenceretlige regler.

/RettighedsAlliancen, maj 2024

[Se oversigt over RettighedsAlliancens medlemmer her](#)