

Kultur- og likestillingsdepartementet
Postboks 8030 Dep
0030 OSLO

Deres ref.

Vår ref.
24/521-1/MAJE

Dato
21.03.2024

Høringssvar fra Språkrådet om: Høring – endringer i åndsverkloven mv. (gjennomføring av digitalmarkedsdirektivet og nett- og videresendingsdirektivet)

I dette høringssvaret vil Språkrådet peke på noen aspekter ved forslaget til implementering av digitalmarkedsdirektivet fra EU inn i norsk rett. Alle merknadene er knyttet til tekst- og datautvinning (TDM):

- Definisjonen av TDM i § 50 d bør vise tydelig at trening av maskinlæringsmodeller er inkludert.
- Vi støtter implementeringen av artikkel 3 fra digitalmarkedsdirektivet slik den er foreslått av departementet. Merknaden vår går ut på at bestemmelsen kan ha noe begrenset nytteverdi for forskningsmiljøet for språkteknologi i Norge.
- Vi støtter implementeringen av artikkel 4, «fribruksretten», fra digitalmarkedsdirektivet slik den er foreslått av departementet. Omfanget av bruken av reservasjonsretten bør evalueres etter en tid for å sikre at hensynet til utvikling av språkteknologi på norsk er ivaretatt.
- Språkrådet har en liten bemerkning til departementets forespørsel i avsnitt 3.6.9, om videreføring av forskrift til åndsverkloven § 4.

Bakgrunn for merknadene

Språkrådet er statens forvaltningsorgan i språkspørsmål og følger opp språkpolitikken som er nedfelt i Prop. 108 L (2019–2020) *Lov om språk (språkløva)*. Et av hovedmålene i språkpolitikken er at norsk skal kunne brukes, og også blir brukt, i alle deler av samfunnet. Språkteknologi er et prioritert område for Språkrådet fordi vi ser at automatisk tekstbehandling og -produksjon er, og kommer til å være, en viktig arena for norsk språkbruk. Dersom Norge skal nå de språkpolitiske målene som Stortinget har stilt seg bak, vil språkteknologien og store språkmodeller som nasjonal infrastruktur stå sentralt.

For at språkteknologi skal utvikles og fungere på norsk, trengs det store mengder språkdata, altså tekst og tale som kan brukes til å trene maskinlæringsmodeller. Norsk er et lite språk i verdenssammenheng, og aktørene på språkteknologifeltet opplever at tilgangen til språkdata er en flaskehals for utviklingen av språkteknologi på norsk. Tekst- og datautvinning

(TDM) er et viktig element i tilgangen til språkdata, og endringene i regelverket for TDM som nå er på høring, er derfor svært viktige for norsk språkteknologi. Uten god tilrettelegging for TDM vil gapet mellom norsk og engelsk språkteknologi kunne bli enda større enn det er i dag, og det vil bli mer attraktivt for sluttbrukeren å bruke engelskspråklige tjenester.

Både offentlig forvaltning og privat næringsliv er avhengig av god språkteknologi på bokmål og nynorsk for å kunne utnytte effektiviseringspotensialet som ligger i kunstig intelligens. Her vil vi også påpeke hvor viktig det er at modellene blir lært opp på godt norsk fagspråk, slik at alle samfunnsområder kan dra nytte av teknologien. I et lite språksamfunn som Norge, der den totale mengden tekst på hvert fagfelt er ekstra liten, er det viktig med tilgang til det som finnes av fagtekst.

Departementet skriver at digitalmarkedsdirektiver fra EU har som formål å «legge til rette for teknologisk utvikling». De nye bestemmelsene i åndsverkloven bør støtte oppunder dette formålet ved at de gjør det mulig å benytte TDM for språkteknologiske formål.

Nasjonalbiblioteket arbeider for tiden med et prosjekt på oppdrag fra KUD: *Mimir*. Prosjektet skal undersøke hvilken effekt tilgang på opphavsrettsbeskyttet materiale har på kvaliteten til generative språkmodeller. Det er uheldig at resultatene fra *Mimir*-prosjektet ikke blir klare før endringene i åndsverkloven er ferdig utredet. Disse resultatene vil kunne gi en indikasjon på hvor tungt språkteknologiske behov bør veie i møtet med andre hensyn.

Språkrådet ser positivt på endringene som er foreslått av departementet. Vi har kommentarer til enkelte aspekter ved lovforslaget og notatet med utgangspunkt i hensynet til tilrettelegging for utvikling av norsk språkteknologi.

Definisjonen av TDM (§ 50 d)

Departementets foreslåtte definisjon:

- *Med tekst- og datautvinning menes i §§ 50 e og 50 f enhver automatisert metode som brukes for å analysere tekst og data i digital form for å fremskaffe informasjon.*

Språkrådet forstår det slik at adgang til TDM for (språk)modellutvikling gjennom maskinlæring er en del av formålet med lovendringene. Slik ordlyden er i forslaget, fremstår det noe uklart for oss hvorvidt trening av språkmodeller er inkludert. Hvis formålet med ordlyden er å gi adgang til TDM (både artikkel 3 og 4 i direktivet) for språkmodellutvikling, bør det vurderes om dette kan uttrykkes mer eksplisitt.

Gjennomføringen av artikkel 3 i direktivet: TDM for forskningsformål (§ 50 f)

Språkrådet støtter implementeringen av tilgangen til TDM for forskningsformål (§ 50 f) uten reservasjonsrett for rettighetshavere. Resultatene av slik forskning vil ikke nødvendigvis direkte øke kvaliteten av språkteknologiske produkter på norsk for sluttbrukerne. Det vil likevel være snakk om grunnforskning som vil gi det norske fagmiljøet mer kunnskap og erfaring som senere vil kunne nyttiggjøres i ferdige, språkteknologiske tjenester med hjemmel i § 50 e.

Vi forstår lovforslaget slik at de relevante forskningsaktørene med denne bestemmelsen kan gjennomføre forskningsprosjekter, men **ikke** trene modeller som skal publiseres for allmennheten, finjusteres av andre eller brukes kommersielt. De forskningsaktørene vi har kontakt med, utvikler og kommer til å utvikle modeller også for sistnevnte formål. Dermed er det grunn til å tro at disse heller vil benytte den generelle «fribruksretten» i § 50 e, som innebærer reservasjonsrett for rettighetshavere.

Gjennomføringen av artikkel 4 i direktivet: generell adgang til TDM: «fribruksretten» (§ 50 e)

Implementeringen av denne artikkelen i § 50 e vil ha stor betydning for utviklingen av norsk språkteknologi og for statusen til og bruken av det norske språket. Vi vurderer det slik at dette vil avhenge av i hvor stor grad rettighetshaverne benytter seg av reservasjonsretten i tredje ledd.

Språkrådet vil oppfordre til at omfanget av bruken av reservasjonsretten blir evaluert etter at loven trer i kraft. Hvis reservasjonsretten blir mye brukt, kan dette sette sterke begrensninger for tilgangen på språkdata for utviklere av norsk språkteknologi. Det bør i så fall gjøres en ny avveining mellom reservasjonsretten og tilrettelegging for norsk språkteknologi.

Departementets forespørsel knyttet til forskrift til åndsverkloven § 4 (avsnitt 3.6.9 i notatet)

Departementet spør om forskriftens § 4 blir overflødig (side 41 i høringsnotatet) ved gjennomføringen av artikkel 3 (forskningsparagrafen). Vi kommenterer her kun andre ledd i bestemmelsen.

Språkrådet ønsker ikke å ta stilling til om lovverket (inkludert andre lover enn åndsverkloven) tar hensyn til Nasjonalbibliotekets anledning til å lage og tilgjengeliggjøre korpus i Språkbanken også for kommersiell bruk. Vi vil derimot påpeke at forskriften slik den er formulert i dag, med en presisering om at korpus kun kan brukes til «forskningsformål», uansett ikke tillater Språkbanken å dele korpus til fri språkteknologisk bruk i privat og offentlig sektor. Dermed er ikke paragrafens andre ledd avgjørende for om Språkbanken kan eller ikke kan utøve denne delen av sitt mandat. Hvordan Nasjonalbiblioteket løser denne problemstillingen i praksis, er nærmere belyst i Prop. 108 L (2019–2020) *Lov om språk*, kapittel 6.

Avslutning

Avslutningsvis vil Språkrådet understreke at tilgang på språkdata er essensielt for utviklingen av språkteknologi på norsk, noe som igjen er viktig for norsk språks status og bruk i fremtiden. Det er viktig å ta hensyn til dette i avveiningene av brukersidens og rettighetshavernes interesser. Innarbeidelsen av EUs direktiv i norsk rett er i denne sammenhengen positiv.

Vennlig hilsen
Lars Ivar Nordal
avdelingsdirektør

Matias Jentoft
Rådgiver

Brevet er elektronisk godkjent og sendes uten underskrifter.

Mottakere:

Kultur- og
likestillingsdepartementet

Postboks 8030 Dep 0030

OSLO

Kopi til:

Constance Ursin
Espen Arneberg Børset