Kopinors høringsinnspill til Utdannings- og forskningskomiteen – Prop. 1 S for budsjettåret 2027

Budsjettkapitler Prop. 1 S, kap. 337, post 71:

Ansvarlig KI-bruk og digital suverenitet er blitt viktige temaer i utdannings- og forskningssektorene. Malthe-Sørenssen-utvalgets rapport, som kommer i løpet av et par uker, vil sette KI i høyere utdanning høyt på dagsordenen. Regjeringen har pekt på utvikling av norske språkmodeller som en sentral brikke i å bygge norsk KI-infrastruktur som er konkurransedyktig. Norske språkmodeller må prestere på høyt nivå for å kunne være nyttige verktøy i forskning og høyere utdanning. Dette krever store mengder kvalitetsdata. Regjeringens budsjettforslag setter disse ambisjonene i fare.

En avtale om trening av norske språkmodeller på boktekster er ferdig framforhandlet mellom Kopinor og Nasjonalbiblioteket. Trening på boktekster er avgjørende for å gjøre modellene bedre på bokmål, nynorsk og samisk. I forslag til statsbudsjett for 2027 bevilget ikke regjeringen midler til denne. Finansiering må sikres slik at avtalen kan iverksettes og ambisjonene om bedre språkmodeller kan realiseres.

Hittil har Norge vært et foregangsland i byggingen av en nasjonal KI-infrastruktur. Måten regjeringen og Nasjonalbiblioteket har gått fram for å sikre lovlig tilgang til opphavsrettsbeskyttet materiale har blitt sett på som et eksempel til etterfølgelse for andre nasjonale KI-satsinger.

For 2026 bevilget regjeringen 45 millioner kroner til en lignende avtale om trening på avisinnhold mellom Nasjonalbiblioteket og Kopinor. Avtalen løper videre for 2027 og regjeringen har lagt inn 47,8 mill. kr[1] til denne avtalen i sitt budsjettforslag. Dette var den første avtalen av sitt slag i verden – en offentlig finansiert vederlagsavtale som muliggjør trening av åpne språkmodeller på opphavsrettsbeskyttet materiale der rettighetshaverne blir kompensert. De første språkmodellene trent på norske avistekster ble lansert i august.

Daværende Digitaliserings- og forvaltningsminister Karianne Tung kalte lanseringen «en sensasjon i Norge og en sensasjon i verden» og framhevet betydningen av å kompensere rettighetshaverne når språkmodellene trenes på opphavsrettsbeskyttet materiale. Kultur– og likestillingsminister Lubna Jaffery sa: «For oss er det et viktig premiss i seg selv: Vi mener at den raske utviklingen innen KI ikke skal gå på bekostning av rettigheter til dem som produserer innhold som språkmodellene trenes på for å bli gode».

[1] Vederlaget KPI-justeres årlig.

Da statsbudsjettet for 2027 ble presentert, ventet rettighetshaverorganisasjoner både i Norge og internasjonalt spent på neste steg. Forskerforbundet og NFFO er blant de organisasjonene som tydeligst har uttalt behovet for KI-løsninger som respekterer opphavsrett.

KI-teknologien er i rask endring, og det er viktig å holde tempoet oppe for å utvikle modeller som holder en konkurransedyktig standard. Denne dynamikken kjenner regjeringen godt til: Da kulturministeren presenterte bevilgningen til trening på avistekster i fjorårets statsbudsjett sa hun at regjeringen «øker tempoet i satsingen på språkmodeller og kunstig intelligens. Det er viktig for å ivareta norsk språk og kultur».

I sitt svar til skriftlig spørsmål fra SV-representant Mirell Høyer-Berntsens om de manglende midlene, forsøker kulturministeren å gå tilbake på denne uttalelsen ved å påpeke at «den framforhandlede avtalen er uten utløpsdato». Men dette hjelper lite dersom de norske språkmodellene ikke klarer å holde tritt med den lynraske utviklingen til de utenlandske modellene. I KI-utvikling er ett år faktisk svært lang tid. Det kan derfor ha langvarige konsekvenser for norske språkmodeller at satsingen stanser halvveis.

Kulturministerens svar på representantens skriftlige spørsmål understreker også mange av grunnene til at trening på bokinnhold er viktig for de norske språkmodellenes konkurransedyktighet og nytteverdi. Avisinnhold gir tilgang til dagsaktuelt språk, samfunnsdebatt, nyheter, politikk og journalistikk. Bøkene tilfører noe annet: lange sammenhengende tekster, komplekse resonnementer, fagkunnskap, litterære uttrykk, kulturhistorie, sjangermangfold og større språklig variasjon. Aviskorpus og bokkorpus er komplementære, ikke alternative datakilder. Det framforhandlede bok-korpuset inneholder like store mengder tekstdata som avis-korpuset, og en dobling av rettighetsbeskyttet materiale vil gi langt bedre modeller og øke sannsynligheten for at modellene faktisk blir tatt i bruk. Minoritetsspråkene er særlig utsatte i møte med internasjonal KI-påvirkning: For å oppnå tilstrekkelig kvalitet på språkmodeller som behersker nynorsk og samisk, er bokinnholdet helt avgjørende.

Det er bred politisk oppslutning om behovet for en sterk og bærekraftig norsk KI-infrastruktur.
Regjeringen har selv sagt i sin «Plan for Norge» at de skal skape muligheter og trygge rammer for kunstig intelligens. De har også lagt vekt på at de skal videreutvikle norske språkmodeller. Om denne budsjettposten mangler også i det vedtatte budsjettet, vil Norge sende et tydelig signal om at norsk KI-infrastruktur ikke lenger er en prioritet. Det ville være svært uheldig og ville svekke troverdigheten til Norges KI-ambisjoner både nasjonalt og internasjonalt.

Vi ber komiteens medlemmer om å sikre at satsingen på gode norske språkmodeller ikke bremses. Avtalen mellom Kopinor og Nasjonalbiblioteket tilbyr en konkret løsning som gir tilgang til norsk kvalitetsinnhold og sikrer at

rettighetshaverne får betalt. Finansiering av avtalen må sikres slik at ambisjonene om bedre språkmodeller kan realiseres.

Forslag til merknad:
Komiteens medlemmer understreker betydningen av å ha gode norske språkmodeller på bokmål, nynorsk og samisk. For at språkmodellene skal bli best mulig for norsk og samisk bruk, må de trenes på store mengder kvalitetsdata. Disse medlemmene mener Norge trenger språkmodeller som er basert på norsk og samisk språk og innhold som er kvalitetssikret. I en tid preget av globale tek-giganter bidrar satsingen til å styrke Norges digitale suverenitet og handlingsrom. Satsingen er viktig for at KI trygt skal kunne tas i bruk i norske utdanningsinstitusjoner og for at norske språkmodeller skal være tilstrekkelig gode til å brukes til forskningsformål. Avtalen mellom Kopinor og Nasjonalbiblioteket gir en dobling av rettighetsbeskyttet materiale modellene kan trenes på. Modellene vil dermed få langt bedre kvalitet, noe som øker sannsynligheten for at modellene tas i bruk. For å oppnå gode norske språkmodeller som behersker bokmål, nynorsk og samisk er bokinnholdet helt avgjørende. Disse medlemmene mener at Kopinor og Nasjonalbiblioteket bør få en bevilgning til bokavtalen på 47,8 millioner kroner slik at tempoet i arbeidet med å utvikle kvalitetssikrede språkmodeller opprettholdes.

Bakgrunn:
I 2023 ga regjeringen Nasjonalbiblioteket (NB) i oppdrag å undersøke om trening med opphavsrettsbeskyttet innhold ga bedre norske språkmodeller, og å foreslå modeller for betaling til rettighetshaverne som har skapt og utgitt innholdet. NB etablerte Mímir-prosjektet, som fastslo at språkmodellene fikk bedre kvalitet dersom de ble trent på innhold fra norske aviser og bøker.

Om Kopinor:
Kopinor sikrer at rettighetshavere får betalt når verkene deres kopieres, brukes og deles.
Vi tilbyr avtaler som forenkler brukernes lovlige tilgang til attraktivt innhold. Slik sørger vi for store verdier som gjør at ny kultur og kunnskap kan skapes og utgis til nytte for samfunnet.
Kopinor ble etablert i 1980 og er en organisasjon av opphavere og utgivere i til sammen 23 medlemsorganisasjoner.