

Motion till riksdagen 2024/25:2021

av **Susanne Nordström m.fl. (M)**

Tillgängliggörande av textmaterial i Kungliga biblioteket

Förslag till riksdagsbeslut

Riksdagen ställer sig bakom det som anförs i motionen om att överväga att utreda hur och med vilka begränsningar man kan tillgängliggöra textdata från Kungliga biblioteket och Riksarkivet för att träna AI-modeller och tillkännager detta för regeringen.

Motivering

För att träna stora språkmodeller inom AI krävs tillgång till massiva mängder högkvalitativ textdata. Internationellt sett domineras denna utveckling av stora teknikbolag, främst i USA och Kina, som under de senaste tre decennierna har byggt upp infrastruktur och samlat in enorma mängder data. Det finns ett direkt samband mellan kvaliteten och mängden av träningsdata och prestandan hos de språkmodeller som utvecklas. Detta har lett till att de mest framstående språkmodellerna främst har sitt ursprung i dessa länder. En följd av detta är att modellerna speglar kulturella värderingar från andra länder än Sverige.

På senare tid har det blivit allt tydligare att data från internet, som tidigare varit lättillgänglig för träning av AI-modeller, blir alltmer svårtillgänglig då fler och fler aktörer ser till att hävda sin upphovsrätt. Många aktörer motsätter sig användningen av deras material för AI-träning, och den globala tillgången på kvalitativ data minskar

därmed. Dessutom ser vi nu en ökande andel av materialet online som genereras av redan existerande språkmodeller, vilket försämrar kvaliteten på den data som används för att träna nya modeller. Denna trend förväntas accelerera under kommande år.

Utmaningar i Europa och möjligheter i Sverige

I Europa är situationen än mer komplicerad på grund av AI-förordningens krav på transparens och efterlevnad av upphovsrättslagstiftningen, vilket ytterligare begränsar tillgången till högkvalitativ träningsdata. Även om detta skyddar rättighetshavare, skapar det hinder för utvecklingen av konkurrenskraftiga AI-modeller. Samtidigt har vi i Europa, och i Sverige särskilt, en unik fördel: en lång tradition av att noggrant bevara och katalogisera textmaterial. Kungliga biblioteket och Riksarkivet har arkiverat en betydande mängd historisk och kulturellt viktig text som kan utgöra en ovärderlig resurs för att träna AI-modeller.

Stora språkmodeller spelar redan idag en avgörande roll i att effektivisera offentlig förvaltning och myndighetsutövning. De möjliggör snabbare och mer enhetliga beslut och förbättrar medborgarnas interaktion med det offentliga genom verktyg som starkt förstärker tjänstemäns förmågor att skyndsamt utföra sina sysslor. Genom att integrera dessa modeller i offentlig verksamhet kan Sverige både öka produktiviteten och stärka förtroendet för offentliga institutioner.

Strategisk betydelse för träningsmaterialet

Det är viktigt att förstå att AI-modeller inte bara är tekniska verktyg utan också speglar de värderingar och den kunskap som finns i det material de tränas på. Därför är det av strategisk betydelse att de modeller som används för svensk offentlig sektor tränas på data som reflekterar svenska kulturella och juridiska normer. Genom att använda material från Kungliga biblioteket och Riksarkivet kan vi säkerställa att AI-modeller som utvecklas för svensk myndighetsutövning bygger på mänskligt genererat material som är relevant och kontextuellt korrekt för vårt samhälle.

Möjligheter till internationellt samarbete

Europa anklagas ofta för att ligga efter i AI-utvecklingen, men en styrka vi har ligger i våra väldokumenterade och omfattande text- och mediearkiv, som representerar århundraden av mänsklig kunskap och kulturell utveckling. Genom att tillgängliggöra dessa arkiv för träning av språkmodeller skulle Sverige kunna ta ledningen i

utvecklingen av europeiska AI-modeller, antingen på egen hand eller i samarbete med våra nordiska grannar och/eller EU-medlemmar. Detta skulle inte bara bidra till att stärka svensk innovation utan även förbättra vår förmåga att utveckla AI-modeller som kan användas inom offentlig sektor med hög kvalitet och effektivitet. För att nå detta mål måste dock ansträngningarna att digitalisera materialet i arkiven mångdubblas och resurser anslås.

Utredning om tillgängliggörande av textdata

Kungliga biblioteket har redan ett eget lagutrymme som tillåter viss rörlighet i hur man använder upphovsskyddat material men det är oprövat vad gäller i synnerhet generativa AI-modeller. Denna motion syftar till att se över möjligheten att tillsätta en utredning som undersöker hur Sverige kan tillgängliggöra de enorma textresurser som finns i Kungliga biblioteket och Riksarkivet för att träna AI-modeller för nyttjande på annan plats än ex Kungliga biblioteket. Utredningen bör också överväga vilka juridiska begränsningar som behöver beaktas, särskilt i förhållande till upphovsrättslagstiftningen, och hur vi kan säkerställa att den offentligt finansierade verksamheten får tillgång till de resurser som behövs för att uppnå maximalt värde av den moderna tekniken. Detta kommer såklart att hänga samman med i vilken och vems infrastruktur som datan sen kommer att användas vilket behöver beaktas i besluten.

Susanne Nordström (M)

Marie Nicholson (M)

Katarina Tolgfors (M)

Anna af Sillén (M)