

Analys och förberedelse för en bredare användning av Kungl. bibliotekets språkmodeller i svensk offentlig sektor

Slutredovisning av regeringsuppdrag enligt
regeringsbeslut U2025/01837



National Library
of Sweden

Analys och förberedelse för en bredare användning av Kungl. bibliotekets språkmodeller i svensk offentlig sektor



Denna rapport är licenserad med [Creative Commons Erkännande 4.0 Internationell, CC BY](https://creativecommons.org/licenses/by/4.0/)

Kungliga biblioteket

www.kb.se

Diarienummer: KB 2025–930

Handläggare: Christopher Natzén

Analys och förberedelse för en bredare användning av Kungl. bibliotekets språkmodeller i svensk offentlig sektor

**Slutredovisning av regeringsuppdrag enligt
regeringsbeslut U2025/01837**

Innehållsförteckning

Förord	6
Sammanfattning	8
Förslag på hur KB:s verksamhet kan bidra till att möta behoven av att använda KB:s språkmodeller även från andra relevanta verksamheter i offentlig sektor.....	9
Summary	11
Inledning.....	12
Regeringsuppdraget.....	13
Läsanvisning	14
Sammanfattning av delredovisningen	14
Genomförandet av slutredovisningen	15
Arbetsformer	15
Genomförande av dialoger	16
Arbetsgrupp	16
Avgränsningar i genomförandet av slutredovisningen	16
Utgångspunkter för slutredovisningen.....	17
Betydelsen av nationella språkresurser.....	18
Offentlig sektors behov av språkmodellsstöd.....	18
Utveckling av svenska språkmodeller	22
Data- och/eller modellportabilitet	23
Flerspråkighetens förbannelse	25
Digitalt kulturarv som del av nationell AI-infrastruktur för offentlig sektor	29
Betydelsen av digitiserade kulturarvssamlingar.....	32
Betydelsen av reviderad plikttagstiftning.....	34
Andra nationalbiblioteks utveckling av språkmodeller	38
Bibliothèque nationale de France, Frankrike.....	39
Det Kongelige Bibliotek, Danmark	41
Nasjonalbiblioteket, Norge	42

Ekonomiska aspekter vid KB:s utveckling av språkmodeller.....	45
Kostnader för utveckling och förvaltning av språkmodeller	45
Utökad teknisk driftorganisation för lagring och beräkning	46
Utökad laborativ kapacitet för långsiktigt konkurrenskraftig utveckling....	47
Utökad AI-understödjande digitisering	47
Effektiviseringspotential i offentlig sektor	48
Upphovsrättsliga aspekter vid KB:s utveckling av språkmodeller	51
Rättsliga utlåtanden av sakkunniga	51
Rättsläget.....	52
Vägen framåt för KB	53
Redogöra för KB:s roll i utvecklingen av språkmodeller inom offentlig sektor i förhållande till andra aktörers arbete med att utveckla språkmodeller inom offentlig sektor.....	54
Redogörelse för KB:s roll i relation till forskningsinfrastrukturer och excellenta miljöer	54
Redogörelse för KB:s roll i relation till andra kulturarvsmyndigheter.....	56
Redogörelse för KB:s roll i relation till rätts- och säkerhetsvårdande myndigheter	58
KB:s samverkansroll	62
Redogörelse för KB:s roll till andra relevanta myndigheter och KB:s medverkan i en svensk AI-verkstad för offentlig sektor	64
Bilaga 1: Genomförda dialoger och inhämtad information	68
Bilaga 2: Källor	71
Bilaga 3: Advokatfirman Delphis utlåtande	77
Bilaga 4: Professor emeritus Jan Roséns utlåtande.....	87
Bilaga 5: Uppdragsdirektiv	98

Förord

Utvecklingen av artificiell intelligens (AI) och språkmodeller innebär ett teknikskifte som redan påverkar forskning, informationshantering och offentlig verksamhet i grunden. På relativt kort tid har möjligheterna att analysera, strukturera och tillgängliggöra mycket stora informationsmängder förändrats på ett sätt som aktualiserar nya möjligheter men också nya vägval för samhället. För offentlig sektor innebär utvecklingen en potential att stärka kvalitet, effektivitet och tillgång till information, samtidigt som frågor om transparens, rättssäkerhet, kontroll över data och långsiktig förvaltning får ökad betydelse. För ett mindre språkområde som det svenska blir dessa frågor särskilt angelägna. Om svenska språkliga och institutionella förhållanden inte ges tillräcklig representation i utvecklingen av AI riskerar det att påverka både kvaliteten i framtida digitala tjänster och den långsiktiga tillgången till svenska språkresurser.

Kungliga biblioteket (KB) har som Sveriges nationalbibliotek ett särskilt ansvar för att samla in, bevara och tillgängliggöra det svenska publicerade kulturarvet. Samtidigt är KB en nationell forskningsinfrastruktur med uppdrag att stödja forskningens långsiktiga kunskapsutveckling. Under de senaste åren har myndigheten, genom forskningsnära utveckling av språkmodeller och språkteknologiska resurser, successivt byggt upp kompetens och kapacitet inom ett område som snabbt har fått ökad strategisk betydelse. Detta arbete har ytterst handlat om att skapa bättre förutsättningar för forskning på svenska textmängder, men också om att säkerställa att det svenska språket, svenska samhällskontexter och det digitiserade kulturarvet finns representerade i utvecklingen av ny teknik.

Regeringens uppdrag till KB att analysera och förbereda för en bredare användning av KB:s språkmodeller i offentlig sektor har gett myndigheten möjlighet att närmare belysa hur denna verksamhet kan bidra till att möta behov även utanför forskningen. Utredningen visar att intresset för språkmodeller är stort inom offentlig sektor, men också att nyttiggörandet förutsätter långsiktighet, samverkan och tydliga strukturer för ansvar, transparens och tillit. Parallellt har det varit viktigt att analysera de rättsliga förutsättningarna för fortsatt utveckling. KB har därför tagit in utlåtande från externa sakkunniga med expertis inom upphovsrätt och EU-rätt avseende de upphovsrättsliga förutsättningarna för text- och datautvinning (TDM) för forskningsändamål. De externas bedömningar ger stöd för att KB:s forskningsnära utveckling av språkmodeller kan fortsätta inom ramen för gällande rätt. Detta skapar stabila förutsättningar för att språkmodeller utvecklade på vetenskaplig grund ska kunna komma forskning och offentlig sektor till del.

I arbetet med detta uppdrag har det blivit tydligt att språkmodeller är del av en bredare nationell infrastruktur där tillgång till data, digitiserat kulturarv, kompetens och säker hantering av information är avgörande. KB:s roll är detta sammanhang i första hand möjliggörande och infrastrukturell, och inte en tjänsteleverantör av verksamhetsspecifika AI-tjänster. En förutsättning för detta arbete är att utvecklingen ska ske med utgångspunkt i forskningens behov, principerna för öppen vetenskap och ett långsiktigt ansvarstagande för

de resurser som förvaltas av det offentliga. Genom KB:s språkresurser, språkmodeller, metodstöd och långsiktig dataförvaltning, kan KB bidra till att stärka offentlig sektors möjligheter att använda AI på ett transparent, rättssäkert och språkligt välanpassat sätt.

Denna slutredovisning utgör ett underlag för fortsatt dialog om hur Sverige kan stärka sin långsiktiga förmåga inom språkteknologi och AI. Det är vår förhoppning att rapporten kan bidra till en mer samlad förståelse för hur nationella språkresurser, kulturarvssamlingar och forskningsinfrastruktur kan utgöra en del av de förutsättningar som krävs för att AI ska komma forskning och offentlig sektor till nytta på ett sätt som stärker tillit, kvalitet och demokratisk förankring.

Karin Grönvall

Riksbibliotekarie

Sammanfattning

Den snabba utvecklingen av AI och språkmodeller skapar nya möjligheter för forskning, informationshantering och offentlig verksamhet. För svensk offentlig sektor innebär utvecklingen potential att effektivisera arbetsprocesser, förbättra informationssökning och stärka analysförmågan. Samtidigt aktualiseras frågor om språklig precision, transparens, rättssäkerhet, långsiktig kontroll över data och beroenden av internationella kommersiella aktörer. För mindre språkområden såsom den svenska är dessa frågor särskilt betydelsefulla.

KB har som Sveriges nationalbibliotek och nationell forskningsinfrastruktur särskilda förutsättningar att bidra till utvecklingen av svenska språkmodeller. Tack vare pliktinsamling, digitisering och långsiktig förvaltning av omfattande kulturarvs- och textsamlingar förvaltar KB resurser av stor betydelse för språkmodeller anpassade till svenska förhållanden. KB har inom sin forsknings- och utvecklingsverksamhet (FoU) utvecklat språkmodeller och språkresurser för forskningsändamål i nära samverkan med forskarsamhället. Arbetet har stärkt forskningens möjligheter att analysera stora textmängder och samtidigt skapat resurser som fått betydande användning även utanför akademien. KB är i dag en viktig del av Sveriges AI-infrastruktur för forskning och denna roll kan vidareutvecklas för att möta offentlig sektors behov.

I detta sammanhang, som redovisningen visar, är digitiserade kulturarvssamlingar hos KB och andra kulturarvsmyndigheter en strategisk nationell resurs för utveckling av språkmodeller som kan stödja offentlig sektors behov. Däremot är endast en begränsad del av dessa samlingar hittills digitiserade, där dessutom nuvarande pliktlagstiftning bara delvis speglar dagens digitala publiceringslandskap.

Slutredovisningen av uppdraget innehåller även en fördjupad analys av de upphovsrättsliga förutsättningarna för KB:s utveckling av språkmodeller för forskningsändamål. KB har till denna redovisning inhämtat rättsliga utlåtanden från sakkunniga med särskilt hög akademisk och praktisk kompetens inom upphovsrätt och EU-rätt. Dessa utlåtanden ger samstämmt och tydligt stöd för KB:s tillämpning av 15 b § lagen (1960:729) om upphovsrätt till litterära och konstnärliga verk (URL).

Utlåtandena, vilka biläggs denna utredning, visar att utveckling av språkmodeller på vetenskaplig grund kan bedrivas inom KB:s forskningsinfrastruktur och att bestämmelsen är teknikneutral, vilket innebär att tillämpningen inte är beroende av vilken typ av språkmodell som utvecklas. Utlåtandena ger även stöd för att uppdragsforskning kan omfattas av samma rättsliga grund när verksamheten bedrivs inom ramen för forskningsändamål. Sammantaget finns det ett väl underbyggt rättsligt stöd för att KB även fortsatt ska kunna utveckla språkmodeller för forskningsändamål på ett sätt som även kan möta offentlig sektors behov.

Utredningens slutsats är att KB:s verksamhet med språkmodeller har potential att bidra till forskning, offentlig förvaltning och nationell AI-kapacitet på ett sätt som stärker svenska språkliga förutsättningar, minskar beroenden av externa aktörer och förbättrar förutsättningarna för transparent och rättssäker användning av AI.

Förslag på hur KB:s verksamhet kan bidra till att möta behoven av att använda KB:s språkmodeller även från andra relevanta verksamheter i offentlig sektor

Förslagen redovisas på en övergripande nivå. De överväganden och analyser som ligger till grund för bedömningarna utvecklas löpande genom hela redovisningen. Hänvisade sidor avser de avsnitt där respektive förslag behandlas mer specifikt.

Se sidorna 23-25.

Bedömning: Modellportabilitet är en förutsättning för att KB ska kunna möta forskningens och offentlig sektors behov av språkmodeller. Öppna modeller möjliggör en kumulativ språkmodellsutveckling, där tidigare investeringar och vetenskapliga framsteg tas till vara. Användning av språkmodeller i såväl forskning som offentlig sektor förutsätter transparens och granskningsbarhet. Eftersom träningsdata i många fall inte kan delas direkt mellan aktörer av rättsliga och institutionella skäl, är för ett mindre språkområde som Sveriges öppna modeller med öppna vikter särskilt betydelsefulla. Det möjliggör återanvändning och vidareträning utan att känsliga eller rättsligt skyddade datamängder behöver flyttas mellan organisationer. Bristande modellportabilitet riskerar att skapa tekniska inlåsningar, dubbelarbete och ökat beroende av externa aktörer, vilket begränsar både offentlig sektors handlingsutrymme och den långsiktiga utvecklingen av svenska språkmodeller.

Förslag: För att skapa fortsatt rätt förutsättningar, och för att KB ska kunna möta offentlig sektors behov utan att investeringar behöver återupprepas, bör nationella strategier eller riktlinjer inom språkmodellområdet ha som grundprincip att offentligt finansierade svenska språkmodeller i möjligaste mån ska utvecklas med öppna vikter och dokumenterade gränssnitt för vidareträning och interoperabilitet.

Se sidorna 32-34.

Bedömning: KB:s utveckling av språkmodeller, och därmed möjligheten att bidra till att möta behov i offentlig sektor, är beroenden av tillgång till stora, representativa och väldokumenterade digitiserade kulturarvssamlingar. Digitisering av kulturarv bör därför förstås som en strategisk infrastrukturell investering med betydelse för både forskning och offentlig sektor. En moderniserad plikttagstiftning som speglar dagens samhälle skulle också på sikt stödja tillgången till representativa svenska datamängder som grund för nationell språkmodellsutveckling.

Förslag: För att möta det växande behovet av anpassade språkmodeller inom offentlig sektor tillförs succesivt resurser för en nationell digitiseringssatsning för digitisering av samlingar med särskild relevans för språkmodellutveckling. En moderniserad plikttagstiftning antas för att på sikt säkerställa datamängder för modellutveckling inom samtliga modaliteter.

<https://www.regeringen.se/rattsliga-dokument/statens-offentliga-utredningar/2021/04/sou-202132/>

Se sidorna 62-63.

Bedömning: KB:s språkmodellutveckling, och möjligheten att stödja offentlig sektors behov, skulle stärkas om denna verksamhet i ökande grad skedde i samverkan med andra myndigheter. Det skulle öka modellernas språkliga representativitet till förmån för offentlig sektors övergripande behov. För att kunna hantera skyddsvärda och känsliga data krävs även en säker beräkningsmiljö som flera myndigheter, inklusive KB, kan ges tillgång till. På en sådan plats kan träning/finjustering ske samt data anonymiseras och/eller aggregeras.

Förslag: För att möta särskilda behov inom offentlig sektor kopplade till nationell säkerhet och beredskap, etableras en säker nationell beräkningsmiljö på lämplig myndighet där KB:s roll är rådgivande under uppbyggnaden samt användare av miljön vid samarbeten. KB ges utökat anslag för att vid sådana samarbeten hantera datamängder med viss nivå av skydd från andra myndigheter.

Se sidorna 64-67.

Bedömning: KB ska bidra till att möta behoven inom offentlig sektor bland annat genom medverkan i Försäkringskassans och Skatteverkets AI-verkstad. KB:s roll bör vara möjliggörande och infrastrukturell, medan utveckling och drift av verksamhetsnära AI-tjänster bör ligga hos respektive myndighet.

Förslag: För att möta behov i offentlig sektor kan KB via uppdragsforskning tillhandahålla svenska språkmodeller och språkteknologiska komponenter för användning i AI-verkstadens utvecklingsmiljöer eller direkt till specifika myndigheter. KB kan också bidra med metod- och kompetensstöd till myndigheter vid användandet av KB:s språkmodeller.

Summary

The rapid development of AI and language models is creating new opportunities for research, information management, and public administration. For the Swedish public sector, these developments offer the potential to streamline work processes, improve information retrieval, and strengthen analytical capabilities. At the same time, they raise important questions regarding linguistic precision, transparency, legal certainty, long-term control over data, and dependencies on international commercial actors. For smaller language communities such as Swedish, these issues are particularly significant.

As Sweden's national library and a national research infrastructure, (Kungliga biblioteket, KB) is uniquely positioned to contribute to the development of Swedish language models. Through legal deposit collection, digitisation, and the long-term stewardship of extensive cultural heritage and text collections, KB manages resources of major importance for language models adapted to Swedish conditions. Within its research and development activities, KB has developed language models and language resources for research purposes in close collaboration with the research community. This work has bolstered the research possibilities involving the analysis of large volumes of text while also creating resources that have seen substantial use beyond academia.

At the same time, digitised cultural heritage collections held by KB and other cultural heritage authorities constitute a strategic national resource for the development of language models capable of supporting the public sector. However, only a limited proportion of these collections has so far been digitised. Furthermore, the current legal deposit framework reflects today's digital publishing landscape only in part.

The final report includes an in-depth analysis of the copyright conditions governing KB's development of language models for research purposes. For this final report, KB commissioned legal opinions from experts with particularly high academic and practical expertise in copyright and EU law. These opinions provide consistent and clear support for KB's application of Section 15 b of the Swedish Copyright Act (Act 1960:729 on Copyright in Literary and Artistic Works) concerning text and data mining (TDM). The opinions, which are appended to this inquiry, demonstrate that the development of language models on a scientific basis can be conducted within KB's research infrastructure and that the provision is technology-neutral, meaning that its application does not depend on the type of language model being developed. The opinions further support the view that commissioned research may fall under the same legal basis when carried out within a research-purpose framework.

Altogether, the inquiry finds that there is a well-founded legal basis for KB to continue developing language models for research purposes in ways that can also meet the needs of the public sector. Overall, KB's work on language models has the potential to contribute to research, public administration, and national AI capacity in ways that strengthen Swedish linguistic conditions, reduce dependencies on external actors, and improve the conditions for transparent and legally robust use of AI.

Inledning

Utvecklingen av AI och språkmodeller utgör ett betydande teknikskifte som har förändrat förutsättningarna för hur information kan analyseras, bearbetas och tillgängliggöras. Den snabba tekniska utvecklingen de senaste åren har möjliggjort automatisk bearbetning av mycket stora textmängder samt avancerade språkliga analyser i en omfattning som tidigare inte var möjligt. Tekniken påverkar i allt större utsträckning forskningens metoder, den offentliga förvaltningens arbetsprocesser samt förutsättningarna för samhällsanalys och beslutsfattande. För offentlig sektor innebär utvecklingen nya möjligheter att effektivisera verksamheter och förbättra tillgången till information. Samtidigt väcker teknikskiftet frågor om kontroll över data, transparens och långsiktig tillgång till språkliga resurser. Dessa frågor är särskilt betydelsefulla för mindre språkområden såsom det svenska. För svenska myndigheter innebär utvecklingen därför både möjligheter och strategiska utmaningar, där det finns ett betydande behov av att säkerställa att svenska språket och svenska informationsresurser är representerade i utvecklingen av AI-baserade system.

KB är Sveriges nationalbibliotek och en nationell forskningsinfrastruktur med uppdrag att samla in, bevara, beskriva och tillgängliggöra det publicerade kulturarvet i tryckt och digital form, inklusive audiovisuella media såsom radio, tv och film. Via pliktinsamling, digitisering och långsiktig förvaltning av samlingarna har KB en av Sveriges mest omfattande resurser av svenskspråkigt material. Samlingarna utgör därför en unik grund för utveckling av språkteknologiska resurser och språkmodeller för svenska språket samt de nationella minoritetsspråken. För att i linje med KB:s instruktion bidra till svensk forsknings kvalitet startade KB i samverkan med forskarsamhället verksamheten KB-labb som ett projekt för datadriven forskning inom humaniora och samhällsvetenskap. Syftet var att stärka KB:s roll som nationell forskningsinfrastruktur och möjliggöra forskning på mycket stora textmängder över långa tidsperioder.

Den nationella diskussionen har mer och mer kommit att handla om ”en stor svensk språkmodell” med bred användning i samhället. KB har inte i uppdrag att utveckla någon sådan språkmodell, utan KB:s uppdrag är att verka för kvaliteten på svensk forskning. Språkmodeller som tas fram inom myndigheten är därför utvecklade utifrån en forskningsfråga och inom ramen för forskningsprojekt. Resultaten från denna FoU-verksamhet har gett KB en strategisk betydelse för utvecklingen av språkteknologiska resurser i Sverige som går utöver nationalbibliotekets traditionella uppdrag. De öppna forskningsresultat i form av svenska språkmodeller detta har resulterat i har även bidragit till att lägga grunden för en bredare användning och vidareutveckling av AI-baserade lösningar inom offentlig sektor.

Regeringsuppdraget

KB fick den 18 september 2025 i uppdrag att analysera och förbereda för en bredare användning av myndighetens språkmodeller i svensk offentlig sektor.¹ I uppdraget ingår att:

- genomföra en fördjupad kartläggning av möjligheter och utmaningar med KB:s arbete med språkmodeller för forskning,
- belysa ekonomiska och upphovsrättsliga aspekter,
- ge förslag på hur verksamheten kan bidra till att möta behoven av att använda KB:s språkmodeller även från andra relevanta verksamheter i offentlig sektor,
- redogöra för KB:s roll i utvecklingen av språkmodeller i förhållande till andra aktörers arbete med att utveckla språkmodeller inom offentlig sektor.

Inom ramen för uppdraget ska KB föra nära dialog med och inhämta synpunkter från Vetenskapsrådet och lärosäten samt med andra relevanta myndigheter och aktörer, i synnerhet Myndigheten för digital förvaltning, Riksarkivet, Institutet för språk och folkminnen, Försäkringskassan och Skatteverket.

Skälen för regeringens uppdrag till KB är att det under de senaste åren har blivit allt tydligare att AI och maskininlärning kommit att revolutionera forskningen och avsevärt påverka samhället i stort. Sverige har goda förutsättningar att vara ledande inom detta område genom tillgång till olika typer av data, beräkningskapacitet och modeller som kan användas för att utveckla och använda AI som vetenskapligt verktyg. För att säkerställa en transparent och ansvarsfull AI-utveckling och för att underlätta och dra nytta av användningen av AI inom forskningen och offentlig sektor bedömde regeringen i forskningspropositionen att det fanns behov av nationell samordning och vägledning.²

Regeringen konstaterar i uppdragsdirektivet att KB idag utvecklar AI-modeller för forskningens behov med utgångspunkt i de omfattande samlingarna av historiskt och aktuellt material. Mängden tillgängliga data gör att dessa modeller kan spegla alla slags variationer av det svenska språket liksom minoritetsspråken. AI-modellerna görs öppet tillgängliga i enlighet med principerna för öppen vetenskap och kan därför användas även utanför specifika forskningsområden, vilket ger bred samhällsnytta. Regeringen såg därför anledning att förbereda för hur KB i än större utsträckning skulle kunna möta offentlig sektors behov med sina språkmodeller och därigenom stärka och bredda rollen som resurs för utvecklingen och användningen av AI i Sverige.

Uppdraget delredovisades den 15 januari 2026.³

¹ Regeringen, Utbildningsdepartementet, *Uppdrag till Kungl. biblioteket att analysera och förbereda för en bredare användning av myndighetens språkmodeller i svensk offentlig sektor* (U2025/01837).

² Regeringskansliet, *Forskning och innovation för framtid, nyfikenhet och nytta*, prop. 2024/25:60, sidan 67.

³ Kungl. biblioteket, *Analys för en bredare användning av Kungl. bibliotekets språkmodeller i svensk offentlig sektor: Delredovisning av regeringsuppdrag enligt regeringsbeslut U2025/01837* (KB 2025-930), 2026. <https://www.kb.se/om-oss/nyheter/nyhetsarkiv/2026-01-16-delredovisning-av-uppdrag-om-sprakmodeller-i-offentlig-sektor.html>, kontrollerad 2026-05-26.

Läsanvisning

Slutredovisningen kombinerar analyser av behov, tekniska och organisatoriska förutsättningar, internationella jämförelser samt ekonomiska, upphovsrättsliga och infrastrukturella perspektiv för att ge en samlad kontext av KB:s verksamhet med att utveckla språkmodeller för forskningsändamål. Syftet har varit att visa på de nationella förutsättningarna för en bredare användning av KB:s språkmodeller inom offentlig sektor och samtidigt belysa KB:s roll i utvecklingen inom området.

Redovisningen inleds med en genomgång av betydelsen av nationella språkresurser och offentlig sektors behov av språkmodellsstöd. Här behandlas användningsområden för språkmodeller, organisatoriska och kompetensmässiga hinder samt frågor om transparens, rättssäkerhet och långsiktig kontroll över data och modeller. Kapitlet omfattar även utvecklingen av svenska språkmodeller, data- och modellportabilitet samt betydelsen av digitaliserat kulturarv och pliktinsamling för svensk språkmodellutveckling.

Vidare redovisas erfarenheter från andra europeiska nationalbibliotek för att sätta KB:s arbete i ett internationellt sammanhang och synliggöra hur liknande satsningar organiserats i relation till forskning, kulturarv och offentlig verksamhet i Europa. Redovisningen innehåller även ett särskilt avsnitt om ekonomiska aspekter kopplade till KB:s verksamhet, där kostnader för utveckling och förvaltning vägs mot potentiella effektivitetsvinster inom offentlig sektor.

Därefter behandlar redovisningen de upphovsrättsliga frågor som aktualiseras vid utveckling och användning av språkmodeller. Här redovisas förutsättningarna för TDM för forskningsändamål enligt upphovsrättslagstiftningen, med utgångspunkt i externa sakkunnigutlåtanden från experter inom upphovsrätt och EU-rätt. Avsnittet belyser även rättsutvecklingen på EU-nivå, inklusive betydelsen av akademisk frihet och kulturarvsinstitutioners roll i utvecklingen av språkmodeller för forskning.

I de avslutande analyskapitlen behandlas KB:s roll i utvecklingen av språkmodeller inom offentlig sektor i relation till andra aktörers arbete. Här analyseras särskilt relationen till forskningsinfrastrukturer, excellenta forskningsmiljöer, kulturarvsmyndigheter samt rätts- och säkerhetsvårdande myndigheter, liksom KB:s möjliga bidrag inom ramen för en svensk AI-verkstad för offentlig sektor.

Sammanfattning av delredovisningen

Delredovisningen av regeringsuppdraget, som överlämnades till Regeringskansliet den 15 januari 2026,⁴ syftade till att ge en samlad bild av möjligheter och utmaningar kopplade till KB:s utveckling av språkmodeller för forskningsändamål. Redovisningen hade främst karaktären av en nulägesanalys och ett kunskapsunderlag för den fortsatta utredningen, med

⁴ Kungl. biblioteket, *Analys för en bredare användning av Kungl. bibliotekets språkmodeller i svensk offentlig sektor: Delredovisning av regeringsuppdrag enligt regeringsbeslut U2025/01837 (KB 2025-930)*, 2026. <https://www.kb.se/om-oss/nyheter/nyhetsarkiv/2026-01-16-delredovisning-av-uppdrag-om-sprakmodeller-i-offentlig-sektor.html>, kontrollerad 2026-05-26.

fokus på den verksamhet som redan bedrivs inom KB:s FoU-verksamhet. Några slutliga ställningstaganden eller konkreta förslag presenterades inte, utan avsikten var att skapa ett underlag för fortsatt analys. Det framhölls att KB:s språkmodeller bidrar till forskning genom att förbättra analysen av stora textsamlingar och möjliggöra nya metoder inom bland annat humaniora och samhällsvetenskap. Delredovisningen identifierade även ett växande intresse från offentlig sektor för språkteknologiska resurser inom områden som informationsanalys, dokumenthantering och kunskapsutvinning.

Delredovisningen visade hur KB:s språkmodeller successivt utvecklats via forskningssamarbeten och blivit en integrerad del av KB:s forskningsinfrastruktur. Modellerna har utvecklats för att stärka forskningens möjligheter att analysera mycket stora textmängder från nationalbibliotekets samlingar och därigenom identifiera språkliga och kulturella mönster över tid. Samtidigt aktualiserades frågor om hur KB:s resurser skulle kunna göras tillgängliga för en bredare användning inom offentlig sektor på ett sätt som beaktar juridiska, tekniska och organisatoriska förutsättningar. Delredovisningen identifierade därför behovet av fördjupade analyser av organisering, styrning och användning, särskilt mot bakgrund av frågor om rättssäkerhet, transparens, ansvarsfördelning och kompetensförsörjning. Ett annat strategiskt utvecklingsbehov som lyftes fram var ökad digitisering av KB:s samlingar, eftersom den begränsade graden av digitisering i dag hämmar forskning och utveckling av svenska språkmodeller.

Delredovisningen behandlade även upphovsrättsliga aspekter av utvecklingen av språkmodeller. Redovisningen redogjorde för de rättsliga ramarna som gäller för TDM enligt EU-rätten och svensk upphovsrättslagstiftning. I delredovisningen konstaterade KB att verksamheten bedrivs inom ramen för den inskränkning för TDM som gäller för forskningsändamål, vilket möjliggör exemplarframställan av material som KB har laglig tillgång till. För KB innebär detta bland annat att material som samlas in genom pliktlagstiftningen kan användas i forskningsrelaterad modellutveckling under de förutsättningar som lagstiftningen medger. Samtidigt betonades att utvecklingen av språkmodeller är ett relativt nytt och rättsligt delvis oprövat område, vilket innebär ett fortsatt behov av juridisk analys, rättsutveckling och dialog med rättighetsinnehavare.

Genomförandet av slutredovisningen

Arbetsformer

Arbetet med uppdraget har genomförts genom en kombination av analyser, dialoger och kartläggningar av befintliga initiativ inom myndigheter, forskningsmiljöer och andra relevanta aktörer. Detta har inkluderat studier av såväl nationella som internationella initiativ för utveckling och användning av språkmodeller och språkteknologiska resurser. I arbetet har även ingått analyser av de tekniska och organisatoriska förutsättningarna för utveckling av språkmodeller, där frågor om beräkningskapacitet, tillgång till träningsdata, dataförvaltning och kompetensförsörjning har varit centrala.

Genomförande av dialoger

Dialogerna har haft två huvudsakliga syften. Det första har varit att identifiera konkreta behov och användningsområden inom offentlig sektor. Det andra har varit att belysa vilka organisatoriska och tekniska förutsättningar som krävs för att en bredare användning av KB:s språkmodeller ska kunna ske på ett ändamålsenligt sätt.

Det har även omfattat dialog med Läromedelsförfattarna, Svenska Förläggareföreningen, Sveriges Författarförbund, Sveriges Tidskrifter och Tidningsutgivarna i syfte att redogöra för KB:s utveckling av språkmodeller för forskningsändamål, diskutera användningen av upphovsrättsligt skyddat material vid utveckling av språkmodeller samt ge de medverkande möjlighet att ställa frågor och framföra synpunkter.

För inhämtning av lärosätenas synpunkter kontaktades initialt UniDig (Universitet och högskolors organisation för digital infrastruktur), som avsåg att sammanställa ett samlat och övergripande underlag. Sent under utredningsarbetet meddelade dock UniDig att det inte hade varit möjligt att samla in tillräckligt underlag för att kunna lämna ett sådant bidrag. Utredningen riktade därför, med kort varsel, en bred förfrågan direkt till lärosätena. Den begränsade tidsramen för svar innebar dock att flera lärosäten inte hade förutsättningar att lämna mer fördjupade yttranden.

Nationell akademisk infrastruktur för superdatorer (NAISS) kontaktades i egenskap av en central nationell aktör för utveckling av AI och språkmodeller. Efter att NAISS bekräftat mottagandet av KB:s skrivelse inkom dock inga ytterligare svar.

Arbetsgrupp

Utredningsarbetet har genomförts av forskningsstrateg Christopher Natzén, enhetschef Love Börjeson, handläggare och bibliotekarie Kristin Halverson, chefsjurist och enhetschef Karin Lodin, IT-arkitekt Martin Malmsten, data scientist Faton Rekathati samt verksjurist Jerker Rydén. Arbetsgruppen har tillsammans bidragit med kompetens inom språkmodeller, data och digital infrastruktur, forskningsstöd, juridik samt strategisk verksamhetsutveckling. Uppdraget har bedrivits i nära samverkan mellan berörda funktioner inom KB för att säkerställa ett sammanhållet perspektiv på de frågor som aktualiseras vid utveckling och användning av språkmodeller för forskningsändamål.

Avgränsningar i genomförandet av slutredovisningen

Redovisningen inkluderar inga analyser av enskilda tekniska implementationer, eftersom sådana lösningar varierar mellan myndigheter och nära sammanhänger med respektive verksamhets operativa behov och tekniska förutsättningar. Kommersiella språkmodeller behandlas inte heller i detalj, även om deras framväxt och användning utgör en viktig kontext för analysen. Syftet har varit att belysa hur KB:s språkmodeller för forskningsändamål kan bidra till offentlig sektors behov samt vilka organisatoriska, kompetensmässiga och resursmässiga förutsättningar som krävs för att möjliggöra detta.

Utgångspunkter för slutredovisningen

KB är Sveriges nationalbibliotek och utgör samtidigt en nationell forskningsinfrastruktur. KB har därför tolkat uppdraget som att redovisningen primärt ska utgå från den verksamhet med språkmodeller som bedrivs inom myndighetens FoU-verksamhet, samt bidra till att tydliggöra relationen mellan nationalbibliotekets uppdrag och de externa aktörer som arbetar med eller har behov av språkteknologiska resurser. Eftersom flera myndigheter inom offentlig sektor har ansvar för digitisering, dataförvaltning och informationshantering aktualiseras också frågor om samordning och ansvarsfördelning. En väl fungerande samverkan mellan berörda aktörer är viktig för att offentliga resurser ska kunna användas effektivt och för att skapa sammanhållna förutsättningar för utveckling och användning av språkmodeller i offentlig verksamhet.

KB har bedömt att uppdraget behöver analyseras i relation till bredare nationella strategiska initiativ inom AI och digitisering. Detta inkluderar bland annat Sveriges nationella AI-strategi samt de förslag som presenterats i AI-kommissionens utredning *Färdplan för Sverige*.⁵ Utifrån denna strategi och dessa förslag har KB analyserat hur myndighetens verksamhet med språkmodeller kan bidra till att stödja nationella mål inom forskning, digitisering och offentlig förvaltning. Analysen har samtidigt omfattat identifiering av möjliga utmaningar och utvecklingsbehov i den nuvarande nationella strategiska inriktningen. En genomgående fråga i slutredovisningen är hur olika initiativ och aktörer kan samordnas för att skapa mer sammanhållna och långsiktiga förutsättningar för utveckling och användning av språkmodeller i Sverige.

En särskild utgångspunkt för utredningsarbetet har varit att analysera hur utvecklingen av språkmodeller kan bidra till att stärka tillgången till svenska språket i digitala miljöer. Detta omfattar frågor om kvaliteten i språkliga resurser samt förutsättningarna för att utveckla språkmodeller som fungerar väl i svenska språkliga, kulturella och samhälleliga kontexter. I Sveriges nationella AI-strategi framhålls att dessa frågor är av särskild betydelse för mindre språkområden, där tillgången till relevanta språkresurser och marknadsmässiga incitament för utveckling ofta är mer begränsade.⁶ I detta sammanhang bedöms KB ha särskilda förutsättningar att bidra genom sitt uppdrag att samla in, bevara och tillgängliggöra det svenska kulturarvet samt genom förvaltningen av omfattande text-, ljud- och videosamlingar. Dessa resurser utgör en viktig strategisk nationell tillgång för utveckling av språkmodeller anpassade till svenska förhållanden. Mot denna bakgrund har särskild vikt lagts vid att, utifrån KB:s perspektiv som forskningsstödande myndighet, identifiera strukturella faktorer som kan påverka utvecklingen på längre sikt. Detta omfattar bland annat frågor om tillgång till data, kompetensförsörjning, teknisk och beräkningsmässig infrastruktur samt långsiktiga organisatoriska förutsättningar. En central faktor i sammanhanget är även hur lagstiftning och reglering utvecklas i relation till AI, där EU:s rättsliga och strategiska initiativ i hög grad bedöms sätta ramarna för utvecklingen även för svenska förhållanden.

⁵ Regeringskansliet, *Sveriges AI-strategi*, 2026; SOU 2025:12, *AI-kommissionens Färdplan för Sverige*, 2025.

⁶ Regeringskansliet, *Sveriges AI-strategi*, 2026, sidan 16.

Betydelsen av nationella språkresurser

De stora kommersiella språkmodeller som utvecklas internationellt av stora teknikföretag är i första hand anpassade för den globala marknaden och de största språken. Mindre språk får i dessa modeller en svagare representation, där träningsdata huvudsakligen består av sådant som samlats in genom storskalig skrapning av internet. Om offentlig sektor enbart förlitar sig på dessa modeller kan det på sikt påverka både kvaliteten och legitimiteten i digitala tjänster och informationssystem genom försämrad språklig precision och begränsad transparens kring vilka data som använts vid modellutvecklingen. Från ett säkerhets- och beredskapsperspektiv innebär beroendet av externa kommersiella leverantörer dessutom potentiella risker och sårbarheter. Den teknik som möjliggör för analys av stora textmängder i myndigheters dokumentation kan samtidigt användas för informationspåverkan, desinformation och andra former av digital påverkan. Att utveckla nationell kapacitet inom området bidrar därför till att stärka Sveriges digitala suveränitet, där KB:s verksamhet för utveckling av språkmodeller stärker den långsiktiga nationella förmågan att hantera informationspåverkan och dessutom säkerställer att väsentliga språkliga resurser blir representerade.

Nyttor och konsekvenser av AI i offentlig sektor kan förstås i ljuset av utvecklingen inom det svenska forskningssystemet under de senaste decennierna. Internationella jämförelser visar att Sverige kännetecknas av höga investeringar i FoU samt en omfattande vetenskaplig produktion. Samtidigt har genomslaget för svensk forskning, mätt mot citeringsgrad, inte utvecklats i takt med investeringsnivån.⁷ Med andra ord är sambandet mellan resurstillförsel och faktisk effekt inte linjärt. Denna iakttagelse är överförbar till utvecklingen och användningen av språkmodeller i offentlig sektor, och som det har visat sig i KB:s verksamhet, specifikt för utvecklingen av språkmodeller för forskningsändamål. Ökade investeringar, fler projekt eller ytterligare regeringsuppdrag är i sig inte tillräckliga för att uppnå önskade samhälls- och verksamhetseffekter. Avgörande är i stället i vilken utsträckning tekniken kan omsättas i konkreta tillämpningar som skapar nytta, effektivitet och kvalitet i myndigheters verksamheter. Vid utveckling av språkmodeller innebär detta att fokus behöver ligga på användbarhet, språklig precision, robusthet och integration i befintliga arbetsprocesser. KB:s språkmodellverksamhet för forskningsändamål kan ses som exempel på hur relativt begränsade resurser, genom ett långsiktigt fokus på funktionalitet och praktisk forskaranvändning, har omsatts i betydande samhällsnytta.

Offentlig sektors behov av språkmodellsstöd

Offentlig sektors förutsättningar att använda AI påverkas av organisatorisk mognad, tillgång till kompetens och etablerade former för verksamhetsutveckling. Flera av de rapporter som har granskats till denna utredning indikerar att användningen av AI inte primärt begränsas av teknikens tillgänglighet, utan av verksamheters förmåga att omsätta tekniska möjligheter i praktisk och långsiktig verksamhetsutveckling. De kartläggningar som genomförts av Statistiska centralbyrån (SCB) visar exempelvis att organisationer som i större utsträckning

⁷ Vetenskapsrådet, *Forskningsbarometern 2025: Svensk forskning i internationell jämförelse* (dnr. 3.5-2025-07677), 2025, sidorna 8-10.

redan bedriver ett systematiskt innovations- och utvecklingsarbete också har visat sig ha förutsättningar för att införa och använda AI-stöd på ett mer genomgripande sätt.⁸ SCB:s statistik visar också att organisationer som i större omfattning redan använder AI även har väletablerade organisationsstrukturer för systematisk verksamhetsutveckling. Teknikens faktiska genomslag framstår med andra ord som nära sammanlänkad med verksamhetens kapacitet att identifiera nya arbetssätt och att integrera dessa i nya tekniska lösningar inom befintliga organisationsstrukturer.

När genomslag realiserar fungerar AI-stöd i ett flertal olika användningsområden inom offentlig sektor. Skatteverket pekar exempelvis på att AI och språkmodellers användning inom offentlig sektor i första hand aktualiseras som stöd i framtagandet av beslutsunderlag snarare än som verktyg för automatiserat beslutsfattande.⁹ Myndigheten framhåller att de mest tidskrävande momenten i handläggningsnära verksamhet ofta består av att söka, samla in och sammanställa relevant information från flera olika källor och system. Ekobrottsmyndighetens perspektiv förstärker denna bild av att språkmodeller inom offentlig sektor huvudsakligen behöver utformas som stöd för informationsinhämtning, sammanställning och analys i handläggnings- och utredningsprocesser. Myndigheten framhåller särskilt att rättssäkerhet och integration i befintliga systemmiljöer utgör avgörande aspekter vid införande av AI-stöd.¹⁰ Vad som framgår av båda myndigheternas skrivelser är att stödet som användandet av språkmodeller kan ge inte främst är som verktyg för automatiserat beslutsfattande, utan som stöd för professionella bedömningar inom tydligt definierade organisatoriska och rättsliga ramar. Skatteverket konstaterar dessutom att ett alltför snävt fokus på att automatisera enskilda delmoment riskerar att i stället leda till suboptimering.

I redovisningen av sitt regeringsuppdrag 2020¹¹ identifierade SCB flera hinder för offentlig sektors användning av AI för att effektivisera verksamheterna, däribland begränsad tillgång till data och varierande kvalitet på den data som fanns tillgänglig. Datautmaningarna är fortsatt relevanta, där många organisationer, trots ökad tillgång till digitala verktyg, fortfarande saknar tillräckliga förutsättningar att införa och nyttiggöra AI på ett sätt som är förenligt med verksamheternas behov och krav. Vid SCB:s kartläggning 2023, baserad på data från 2019 och 2021, framkom dessutom att användningen av AI var relativt begränsad och ojämnt fördelad mellan organisationer och sektorer.¹² Större privata organisationer hade i ökande utsträckning tagit tekniken i bruk, medan mindre aktörer och stora delar av offentlig sektor hade i lägre utsträckning integrerat AI i sina verksamheter. Liknande iakttagelser framkom i analyser genomförda av Diskrimineringsombudsmannen (DO) och

⁸ Statistiska centralbyrån, *Förhöjd innovation – En mikrodataanalys om AI och innovation*, 2024, sidorna 11-12.

⁹ Skatteverket, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

¹⁰ Ekobrottsmyndigheten, skrivelse via mejl, inkommen 2026-05-05 (KB 2025-930).

¹¹ Regeringen, Infrastrukturdepartementet, *Uppdrag att kartlägga användningen av artificiell intelligens respektive analys av stora datamängder i Sverige*, I2019/01964/D.

¹² Statistiska centralbyrån, *AI-användning i företag och offentlig sektor*, 2023, sidorna 11-13.

Statskontoret,¹³ där myndigheter lyfte fram att specialistkompetens och förmågan till dataanalys ofta var begränsad.

I och med att KB inom sin FoU-verksamhet utvecklar och tillgängliggör öppna och väldokumenterade språkmodeller för forskningsändamål som kan användas i verksamhetsnära tillämpningar kan vissa av de trösklar som identifierats i SCB:s, DO:s och Statskontorets kartläggningar minska. KB:s språkmodeller kan bidra till att sänka trösklarna för användning av AI i verksamheter där resursbegränsningar annars riskerar att utgöra betydande hinder. Försäkringskassans och Skatteverkets uppdrag att etablera en AI-verkstad för offentlig sektor är ytterligare ett led i en mer samordnad ansats för utveckling och användning av AI inom statsförvaltningen.¹⁴ Genom att samla kompetens och erfarenheter i en gemensam miljö kan förutsättningarna stärkas för att minska fragmentering och stödja myndigheter som saknar egen kapacitet att utveckla avancerade AI-lösningar. Här har KB:s erfarenheter och modellutveckling för forskningsändamål stor potential att bidra. AI-verkstaden kan därmed fungera som en plattform för metodutveckling och gemensamt lärande mellan myndigheter. De långsiktiga strategiska satsningarna som görs inom den nationella AI-strategin blir därmed kompletterade med verksamhetsnära och flexibla stödstrukturer som bidrar till ett mer praktiskt och brett genomslag för AI i offentlig sektor.

Liknande, för forskningen verksamhetsnära behov, uttrycker flera lärosäten i sina inspel till utredningen.¹⁵ Ett återkommande perspektiv är behovet av transparens, dokumentation och vetenskaplig reproducerbarhet i utvecklingen av språkmodeller. Flera lärosäten efterfrågar standardiserad dokumentation av träningsdata, modellernas begränsningar, utvärderingsmetoder och eventuella efterträningsteg för att möjliggöra granskning och ansvarsfull användning. Det framhålls också att fullständig teknisk reproducerbarhet inte alltid är realistisk för större modeller, varför fokus snarare bör ligga på spårbarhet, versionshantering och konsistenta resultat under likvärdiga förutsättningar. Det betonas även att transparens behöver omfatta information om modellernas kända svagheter. KB:s erfarenheter av att tillgodose forskningens behov av öppna och dokumenterade utvecklingsprocesser som stärker tilliten till språkmodellers användning kan därmed vidareutvecklas i ett breddat arbete gentemot offentlig sektor.

¹³ Diskrimineringsombudsmannen, *Transparens, träning och data: Myndigheters användning av AI och automatiserat beslutsfattande samt kunskap om risker för diskriminering*, Rapport 2022:1; Statskontoret, *Myndigheterna och AI: En studie om möjligheter och risker med att använda AI i statsförvaltningen*, skriftserien Om offentlig sektor, 2024.

¹⁴ Regeringen, Finansdepartementet, *Uppdrag till Försäkringskassan och Skatteverket att etablera en AI-verkstad för den offentliga förvaltningen* (Fi2026/00018).

¹⁵ Blekinge tekniska högskola, skrivelse via mejl, inkommen 2025-05-25 (KB 2025-930); Göteborgs universitet, skrivelse via mejl, inkommen 2025-05-20 (KB 2025-930); Högskolan Dalarna, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Högskolan i Borås, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Högskolan i Halmstad, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Högskolan Kristianstad, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Karlstads universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Mittuniversitetet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Stockholms universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Sveriges lantbruksuniversitet, skrivelse via mejl, inkommen 2025-05-20 (KB 2025-930), Södertörns högskola, skrivelse via mejl, inkommen 2025-05-26 (KB 2025-930).

Informationsinhämtningen visar därmed att behovet av forskningsnära språkmodeller som kan hantera dessa aspekter inte är begränsat till akademiska miljöer utan även har tydlig relevans för offentlig verksamhet. SCB uppgav exempelvis till utredningen att myndigheten redan använder språkmodeller utvecklade av KB, däribland KB-BERT, inom statistikproduktionen och analys av fritextsvar. SCB framhöll samtidigt ett intresse för samverkan kring mer specialiserade modeller anpassade till statistikproduktionens behov.¹⁶ Myndigheten betonade särskilt vikten av dokumenterade data, versionshanterad kod, reproducerbara miljöer och möjligheter att följa resultat över tid inom etablerade kvalitets- och styrningsramverk. Statskontoret framhöll samtidigt i dialog med utredningen behovet av språkmodeller med bättre förståelse för svensk förvaltning, svensk rätt och offentliga dokumentstrukturer samt modeller som kan stödja avancerad semantisk sökning i stora informationsmängder.¹⁷ Myndigheten aktualiserade även behovet av mer systematiska former för utvärdering av språkmodellers prestation i svenska språkliga och förvaltningsmässiga kontexter. Även E-hälsomyndigheten framhöll till utredningen behovet av språkmodeller som kan hantera komplex och domänspecifik information där precision, transparens och reproducerbarhet utgör centrala krav.¹⁸ Myndigheten betonade särskilt användningsområden kopplade till analysarbete, intern kunskapshantering och avancerad informationssökning i omfattande informationsmiljöer. Behoven illustrerar betydelsen av infrastrukturer såsom KB, där svenska språkmodeller kan utvecklas, testas och utvärderas mot sektorsspecifika användningsfall med höga krav på kvalitet och tillförlitlighet.

Det finns däremot en utmaning som alla myndigheter oavsett organisatorisk AI-mognad behöver hantera. AI-system fungerar inte enbart som tekniska verktyg för informationsbearbetning, utan påverkar i praktiken hur myndigheter organiserar arbetsprocesser, prioriterar information och utformar beslutsunderlag.¹⁹ Till skillnad från traditionella regelbaserade system, där beslut fattas utifrån explicit formulerade regler och rättsliga normer, bygger AI-system på statistiska samband och sannolikhetsbaserade modeller. Det här innebär att analyser, rekommendationer och sammanställda beslutsunderlag riskerar att grundas i mönster som återfinns i data. Principiellt innebär detta en viktig förändring i hur stöd för offentligt beslutsfattande utformas, där frågor om ansvar, kontroll och institutionell legitimitet blir centrala förutsättningar för ett långsiktigt och ändamålsenligt nyttiggörande av AI i offentlig verksamhet.

När AI används som stöd i processer som påverkar myndigheters prioriteringar eller utformningen av beslutsunderlag uppstår ökade krav på möjligheten att förstå hur resultat genereras och vilka begränsningar som följer av använda modeller. Nödvändigtvis förutsätter det inte fullständig teknisk insyn i varje komponent eller algoritmisk process, men det behöver finnas tydlig och tillgänglig dokumentation om exempelvis språkmodellernas syfte, datagrund, metodologiska utgångspunkter och begränsningar. Möjligheten att i efterhand

¹⁶ Statistiska centralbyrån, skrivelse via mejl, inkommen 2026-05-08 (KB 2025-930).

¹⁷ Dialog KB-Statskontoret, 2026-04-29.

¹⁸ E-hälsomyndigheten, skrivelse via mejl, inkommen 2026-04-22 (KB 2025-930).

¹⁹ Simon Vinge and Maja Fjaestad, "Introduction: From algocracy to democracy", *Algorithmic rule: AI and the Future of Democracy in Sweden and Beyond* (2025), sidorna 8-9.

förstå hur exempelvis ett beslutsunderlag har tagits fram behövs för att säkerställa rättssäkerhet, möjliggöra överprövning och upprätthålla principen om likabehandling.

DO och Statskontoret har i sina rapporter identifierat just att rättssäkerhet, transparens och risken för reproduktion av systematiska skevheter är avgörande frågor vid offentlig sektors användning av AI.²⁰ DO lyfter särskilt tre riskområden som offentlig sektor behöver hantera. Det är bristande datakvalitet, otillräcklig kompetens och träning samt begränsad transparens i AI-systemens funktion och användning.²¹ Av detta skäl vill KB framhålla vikten av att det offentliga Sverige inte är hänvisat till kommersiella aktörer utan även bygger och utnyttjar redan befintlig egen kapacitet. DO:s risker är särskilt relevanta i relation till användandet av språkmodeller, eftersom modellernas egenskaper påverkas av de data de tränats på. Snedvridningar i datamaterialet kan därigenom reproduceras eller förstärkas i modellernas svar, vilket kan påverka såväl kvalitet som rättvisa i informationsbearbetning och beslutsstöd. Ett realistiskt mål är däremot inte att helt eliminera risker eller bias, utan snarare att minska systematiska fel, synliggöra begränsningarna och säkerställa tydlig dokumentation om språkmodellernas egenskaper och användningsområden. Potentialen i den stora mängden kulturarvsdata, som till exempel KB och Riksarkivet (RA) förvaltar, har stor betydelse för representation och transparens (se vidare sidan 32).

Utveckling av svenska språkmodeller

Som beskrivits inledningsvis förespråkar KB öppna och väldokumenterade svenska språkmodeller i och med att de kan bidra till att stärka offentlig sektors förutsättningar att använda AI på ett transparent, kontrollerbart och rättssäkert sätt. Till exempel Högskolan i Borås lyfter problemen som följer med bristande insyn i modellträningen tillförlitligheten, spårbarheten och transparensen.²² När träningsdata, metodval och dokumentation däremot är kända förbättras förutsättningarna att förstå modellernas styrkor, begränsningar och tillämpliga användningsområden. Det skapar också bättre möjligheter att identifiera systematiska felkällor, analysera språkliga eller samhällseliga snedvridningar samt utveckla mer ändamålsenliga former för kvalitetssäkring och uppföljning. I och med att KB utvecklar och tillgängliggör öppna och väldokumenterade språkmodeller för forskningsändamål bidrar därför KB också till ökad spårbarhet, transparens och tillit i offentlig sektors användning av språkteknologi.

Organisation for Economic Co-operation and Development (OECD) har pekat på att de globala investeringarna i AI-infrastruktur och modellutveckling som görs i dag är av en sådan omfattning att länder med jämförelsevis mindre resurser har begränsade möjligheter att

²⁰ Diskrimineringsombudsmannen, *Transparens, träning och data: Myndigheters användning av AI och automatiserat beslutsfattande samt kunskap om risker för diskriminering*, Rapport 2022:1, sidan 30; Statskontoret, *Myndigheterna och AI: En studie om möjligheter och risker med att använda AI i statsförvaltningen*, skriftserien Om offentlig sektor, 2024, sidan 36.

²¹ Diskrimineringsombudsmannen, *Transparens, träning och data: Myndigheters användning av AI och automatiserat beslutsfattande samt kunskap om risker för diskriminering*, Rapport 2022:1, sidorna 23-25.

²² Högskolan i Borås, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

konkurrera på lika villkor.²³ Sverige är resursmässigt i denna jämförelse ett litet land och satsningar på svenska språkmodeller bör i första hand inte ses som ersättningar för redan utvecklade internationella modeller, utan snarare som ett kompletterande och strategiskt viktigt skikt. Från ett sådant perspektiv framstår nationella språkmodeller, så som de KB utvecklar för forskningsändamål, som ett sätt att kombinera internationell teknikutveckling med svenska krav på språklig precision, verksamhetsanpassning, transparens och långsiktig institutionell kontroll.

Data- och/eller modellportabilitet

En central fråga för språkmodeller på systemnivå, och därmed för Sverige på nationell nivå, är hur kunskap, kapacitet och investeringar i AI kan återanvändas och vidareutvecklas utan att behöva byggas upp på nytt eller låsas in hos enskilda aktörer. I stora ekonomier med omfattande privata och offentliga resurser hanteras detta genom ett överskott av resurser och konkurrerande modellutveckling. Det främsta exemplet är USA där de ledande företagen var för sig lägger betydande resurser på infrastruktur och utveckling av språkmodeller.²⁴ Kapacitetsackumuleringen drivs i den amerikanska kontexten i huvudsak av kommersiell och vetenskaplig konkurrens och kräver därmed inte central samordning.

Mindre länder som Sverige saknar förutsättningar för sådan redundans. Resurser behöver därför delas eller samutnyttjas. Samtidigt finns en risk att alltför långtgående samordning hämmar den innovativa förmågan, eftersom denna gynnas av incitament för spetsutveckling, självständiga prioriteringar, konkurrerande angreppssätt och möjligheten att snabbt fatta beslut och omfördela resurser. Delning och samutnyttjande av resurser bör därför utformas så att den stärker, snarare än begränsar, vetenskaplig och kommersiell konkurrens, eftersom denna konkurrens är central för systemets totala innovativa förmåga.

För att utveckla stora språkmodeller behövs, förutom de legala och förvaltningsmässiga förutsättningarna, kompetens, data och beräkningskraft. KB bedömer att Sverige som land står sig väl kompetensmässigt när det gäller utveckling av språkmodeller (Sverige har starka forskningsmiljöer och kompetenskluster i akademien, på myndigheter och i den privata sektorn). Nödvändig kapacitet har även uppnåtts när det gäller beräkningskraft för utveckling av små och medelstora språkmodeller i och med EU-kommissionens och medlemsländernas gemensamma satsning på superdatorer (EuroHPC JU) och förstärkningen av beräkningskapaciteten vid svenska lärosäten främst genom NAISS. Även i offentlig och privat sektor ökar investeringarna i beräkningskraft. KB:s bedömning är att beräkningskraft finns att tillgå för samlade tyngre initiativ. Den kvarvarande trånga sektorn för att utveckla varaktig svensk förmåga för att utveckla konkurrenskraftiga språkmodeller är därför tillgången till träningsdata.

²³ OECD, *Competition in Artificial Intelligence Infrastructure*, OECD Roundtables on Competition Policy Papers, No. 330 (2025), sidan 23.

²⁴ OECD, *Competition in Artificial Intelligence Infrastructure*, OECD Roundtables on Competition Policy Papers, No. 330 (2025), sidorna 17-18.

Det blir inte minst tydligt utifrån lärosätenas inspel till denna utredning där de överlag framhåller betydelsen av långsiktig tillgång till svenska språkresurser som grund för utveckling och vidareutveckling av språkmodeller.²⁵ Flera pekar särskilt på värdet av digitaliserade kulturarvssamlingar och myndighetsmaterial. Återkommande är synpunkter om behovet av bredare digitalisering, förbättrad textigenkänning av äldre material och tillgång till samtida språkdata, inklusive tal, multimodalt material och professionsspecifika datamängder.

KB och andra kulturarvsmyndigheter såsom RA förvaltar just den typ av data som efterfrågas och behövs. Dessa kulturarvsinstitutioners samlingar innehåller språkliga, historiska och kulturella representationer som är avgörande för att språkmodeller ska kunna utveckla hög kvalitet för svenska språket, svenska varieteter och svenska samhällskontexter. Samtidigt är stora delar av materialet rättsligt, etiskt eller institutionellt känsligt, vilket innebär att direkt datadelning endast i begränsad omfattning är möjlig. Därtill kommer myndigheternas uppdrag att garantera samlingarnas (datans) integritet och säkerhet, så att framtida generationer kan ta del av oförvanskat källmaterial. Från ett säkerhets- eller upphovsrättsperspektiv är det därför mycket svårt att dela data med andra institutioner. I vissa fall inom ramen för forskningsprojekt kan kopior upprättas till exempel för träning inom EuroHPC JU. Även i dessa fall kommer det vara administrativt och tekniskt komplext, med risker för samlingarnas integritet och kontroll. För kulturarvsinstitutioner är det dessutom centralt att digitala samlingar inte gradvis förs över till privata aktörer på ett sätt som urholkar det långsiktiga offentliga förvalterskapet över kulturarvet. Att göra data flyttbart (dataportabilitet) mellan språkmodellsutvecklare för att möjliggöra kumulativ förmågeutveckling är med andra ord betydligt svårare och dyrare och förknippat med större risker än att göra själva språkmodellerna flyttbara (modellportabilitet). Huvudspåret för att möjliggöra kapacitetsackumulering utan inlåsning eller monopolisering, samtidigt som kulturarvsinstitutionernas samlingar värnas, är därför att det primära delningsobjektet inte är data utan öppna modeller med öppna vikter.²⁶ Detta ligger i linje med principerna för öppen vetenskap och möjliggör att investeringar i träning, annotering och modellutveckling successivt kan återanvändas och vidareutvecklas av flera aktörer över tid.

Det är av central betydelse för KB att hantering av data sker i enlighet med upphovsrättsliga regler. KB ser att delning av data utanför myndigheten bland annat kan innebära ett tillgängliggörande för allmänheten av upphovsrättsligt skyddade verk. Möjligheten till

²⁵ Blekinge tekniska högskola, skrivelse via mejl, inkommen 2025-05-25 (KB 2025-930); Göteborgs universitet, skrivelse via mejl, inkommen 2025-05-20 (KB 2025-930); Högskolan Dalarna, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Högskolan Kristianstad, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Karlstads universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Linnéuniversitet, skrivelse via mejl, inkommen 2025-05-27 (KB 2025-930); Lunds universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Mittuniversitetet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Mälardalens universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Stockholms universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Sveriges lantbruksuniversitet, skrivelse via mejl, inkommen 2025-05-20 (KB 2025-930), Södertörns högskola, skrivelse via mejl, inkommen 2025-05-26 (KB 2025-930), Uppsala universitet, skrivelse via mejl, inkommen 2025-05-19 (KB 2025-930).

²⁶ Öppna vikter innebär endast att parametrarna, de tränade numeriska värdena eller "vikterna", i en AI-modell görs tillgängliga. Det betyder att användare kan använda modellen lokalt, vidareträna den på egen data eller anpassa den till specifika användningsområden.

dataportabilitet kan därför bli begränsad, eftersom ett sådant tillgängliggörande typiskt sett kräver upphovsrättsligt tillstånd eller licens. Dessa kan i praktiken vara svåra att erhålla, bland annat till följd av ett stort antal rättighetshavare samt att kollektiva licenslösningar inte nödvändigtvis omfattar alla relevanta materialtyper eller användningar.

När det gäller de modeller som KB har tagit fram gäller, såsom framgår av delredovisningen, att de bland annat i sig inte innehåller något upphovsrättsligt skyddat material och att de inte heller kan generera den data de tränats på vid användning. Under dessa förutsättningar kan portabilitet av modeller således inte medverka till, eller innebära, upphovsrättsintrång i ett användningsskede.

En central aspekt är att öppna modeller med öppna vikter skapar kumulativ kapacitet. Även om en enskild svensk modell inte når absolut högsta globala prestanda (State-of-the-Art, SOTA) kan dess träningsresultat, språkförmåga och domänkunskap återanvändas som grund för nästa generations modeller som i sin tur för specifika kontexter eller funktioner kan uppnå SOTA eller near-SOTA. Detta minskar behovet av att återupprepa mycket kostsamma träningsprocesser från början och gör att investeringar i språkresurser, beräkningskraft och expertis får långsiktigt värde på systemnivå. Öppna vikter innebär också att svenska aktörer kan behålla strategisk handlingsfrihet. Om modellutvecklingen i huvudsak sker genom slutna kommersiella API:er riskerar såväl offentlig sektor som forskning att bli beroende av externa leverantörer för grundläggande språkteknologisk infrastruktur. Ett ekosystem baserat på öppna modeller möjliggör däremot vidareträning och anpassning för svenska behov och regulatoriska krav. Detta innebär inte att all utveckling måste ske i offentlig regi eller att konkurrens ska begränsas. Tvärtom kan öppna grundmodeller fungera som gemensam infrastruktur ovanpå vilken kommersiella och akademiska aktörer konkurrerar genom specialisering, effektivisering och innovation.

Flerspråkighetens förbannelse

Utveckling och användning av språkmodeller inom offentlig sektor medför betydande möjligheter, men samtidigt aktualiseras ett flertal utmaningar. Stora språkmodeller har successivt förbättrat sin förmåga att hantera svenska språket men trots detta kvarstår återkommande begränsningar avseende språklig precision, terminologisk träffsäkerhet och förståelse för idiomatiska uttryck, särskilt i verksamhetskritiska eller kontextspecifika sammanhang. Förklaringen till dessa begränsningar är obalanser i träningsdata. Eftersom stora språkmodeller tränas på mycket omfattande textmängder där globala språk, framför allt engelska, dominerar får mindre språk generellt en svagare representation. Detta påverkar modellernas språkliga precision och medför att genererad text framstår som mindre naturlig eller präglas av formuleringar som närmast motsvarar direktöversättningar från större språk. Begränsningarna avser dock inte enbart språklig korrekthet, utan även förståelsen för kulturella och samhällsliga kontexter, lokala referenser och språkliga nyanser som är specifika för Sveriges förhållanden. Institutet för språk och folkminnen (Isof) betonar också att utvecklingen av språkmodeller för offentlig sektor behöver inkludera nationella

minoritetsspråk, dialekter och andra språkliga varieteter för att säkerställa bredare språklig representation och tillgänglighet.²⁷

Utmaningarna framträder särskilt tydligt inom specialiserade verksamhetsområden såsom forskning och offentlig förvaltning där krav på terminologisk precision och kontextuell förståelse är höga. Lärosätena betonar samstämmigt exempelvis att svenska språkmodeller inte enbart berör språkförståelse i snäv mening, utan om en förmåga att hantera professionsspråk, myndighetsterminologi, dialektala variationer och svenska samhällsförhållanden.²⁸ Flera pekar på att internationella modeller i vissa sammanhang riskerar att missa betydelsebärande nyanser eller skapa terminologiska förskjutningar som påverkar resultatens tillförlitlighet och rättssäkerhet.

I forskningen om språkmodellsutveckling har uttrycket ”flerspråkighetens förbannelse” (the curse of multilinguality) blivit ett etablerat koncept.²⁹ Den förklarar en observerad trend där språkmodellers förmågor i ett enskilt språk försämras när allt fler språk adderas i träningen, jämfört med när träningen i stället sker enspråkigt eller koncentrerat med färre språk. Den negativa trenden förstärks särskilt när obesläktade språk samsas under en och samma modells ändliga förmåga, medan träning tillsammans med närbesläktade språk tvärtom kan leda till bättre resultat bland samtliga inkluderade språk.³⁰

Språkmodeller som KB:s är specifikt tränade på större mängder data från ett mindre språkområde och presterar därför bättre eftersom de är tränade med ett tydligare fokus. Genom att använda data som huvudsakligen eller enbart är på målspråket får dessa modeller en djupare förståelse för grammatik, ordförråd och språkliga nyanser. Det här leder dels till att de kan producera mer naturliga och korrekt formulerade texter, dels till att de också har större möjlighet att inkludera kulturellt relevanta referenser och därmed ge mer träffsäkra svar i lokala sammanhang. En annan fördel är att de kan optimeras för specifika användningsområden exempelvis i en myndighetskontext. Dessutom kräver mindre modeller ofta färre resurser vilket gör dem mer effektiva att använda och att vidareutveckla för specifika ändamål.

När utvecklingsprocessen sker inom offentlig sektor, så som sker inom KB:s FoU-verksamhet, och närmare de myndigheter som faktiskt ska använda modellerna ökar också

²⁷ Institutet för språk och folkminnen, skrivelse via mejl, inkommen 2026-05-07 (KB 2025-930).

²⁸ Blekinge tekniska högskola, skrivelse via mejl, inkommen 2025-05-25 (KB 2025-930); Chalmers tekniska högskola, skrivelse via mejl, inkommen 2026-05-26 (KB 2025-930); Högskolan Dalarna, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Högskolan i Borås, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Högskolan i Halmstad, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Högskolan Kristianstad, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Karlstads universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Linnéuniversitet, skrivelse via mejl, inkommen 2025-05-27 (KB 2025-930); Mälardalens universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Stockholms universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Sveriges lantbruksuniversitet, skrivelse via mejl, inkommen 2025-05-20 (KB 2025-930), Södertörns högskola, skrivelse via mejl, inkommen 2025-05-26 (KB 2025-930).

²⁹ Chang et al., *When Is Multilinguality a Curse? Language Modeling for 250 High- and Low-Resource Languages*. EMNLP 2024.

³⁰ Longpre et al., *ATLAS: Adaptive Transfer Scaling Laws for Multilingual Pretraining, Finetuning, and Decoding the Curse of Multilinguality*. ArXiv preprint, 2025.

insynen. Det skapar bättre förutsättningar för spårbarhet eftersom varje steg i modellernas livscykel transparent kan dokumenteras och följas upp. På så sätt möjliggörs tydligare ansvarsstrukturer vilket står i kontrast till de stora internationella modellerna. Det blir enklare att implementera kontinuerlig uppföljning och kvalitetssäkring, eftersom modellen är anpassad till specifika behov och regelverk där kontrollmekanismer dessutom kan integreras direkt i utvecklingen snarare än läggas till i efterhand. Därutöver kan modellerna i utvecklingsstadiet även anpassas efter en myndighets administrativa processer, vilket minskar risken för missförstånd och felaktiga beslut.

I princip samtliga underlag som har delgetts utredningen pekar på vikten av reproducerbarhet, transparens och spårbarhet i träningsdata både från ett forskningsperspektiv och i relation till offentlig sektors krav på insyn och ansvar. Med det följer utmaningar för rättslig granskning, kvalitetssäkring och vetenskaplig reproducerbarhet. För offentlig sektor aktualiseras dessutom frågor om hur de datagrunder som påverkar modellernas funktion och resultat kan beskrivas, granskas och följas upp. KB bedriver därför inom ett forskningsprojekt med Språkbanken Text vid Göteborgs universitet under 2026 ett arbete för att stärka spårbarheten i träningsdata och öka transparensen i modellutvecklingen.³¹ Samarbetet syftar till att inventera, sammanställa, bearbeta, versionshantera, dokumentera och förädla relevanta datamängder från både egna och externa källor. Eftersom datamängderna härrör från många olika källor och förekommer i varierande format krävs ett omfattande arbete för att kvalitetssäkra, harmonisera och anpassa materialet till standarder som lämpar sig för utveckling av språkmodeller. Arbetet omfattar även metoder och utvärderingsdataset som kan användas för att bedöma modellernas kvalitet, analysera effekter av olika metodval och identifiera systematiska snedvridningar. Detta samarbete mellan två forskningsinfrastrukturer (KB och Språkbanken Text) har stor potential att bidra till utvecklingen inom offentlig sektor.

Som SCB pekar på berör krav på transparens, objektivitet och reproducerbarhet allt från rättstillämpning och säkerhetsrelaterade verksamheter till myndigheters analys- och statistikproduktion.³² Myndigheten framhåller bland annat att officiell statistik måste vara möjlig att granska, förstå och reproducera även när AI-baserade metoder används. Det innebär bland annat krav på dokumenterade träningsdata, versionshantering, tydligt beskrivna bearbetningsprocesser samt möjlighet att följa hur slutsatser och prediktioner har genererats. KB:s bedömning är att detta stärker behovet av språkmodeller som utvecklas inom ramar där transparens och granskningsbarhet kan säkerställas över tid. SCB lyfter samtidigt risker kopplade till användning av språkmodeller inom myndighetskommunikation, särskilt risken att felaktig information presenteras med hög säkerhet eller att användare lämnar känsliga personuppgifter i system som uppfattas som säkra.

Stora internationellt utvecklade språkmodeller innehåller dock redan omfattande språklig och semantisk kunskap som skulle vara mycket kostsam och tidskrävande att reproducera nationellt. Som påpekats i föregående avsnitt kräver träning av avancerade språkmodeller

³¹ Kungl. biblioteket, *Fördjupat samarbete med Språkbanken* (KB 2026-293).

³² Statistiska centralbyrån, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

från grunden betydande mängder data och omfattande beräkningskapacitet, vilket i praktiken kan vara svårt att upprätthålla för mindre språkområden. Genom att i stället utgå från redan tränade modeller och därefter anpassa eller finjustera dessa för specifika språk och användningsområden kan nationella utvecklingsinsatser i stället fokusera på nationella och verksamhetsnära behov. Inom KB:s verksamhet med utveckling av språkmodeller för forskningsändamål har denna typ av metodik tillämpats genom att stora språkmodeller använts som teknisk grund och därefter kompletterats med nationella datamängder från KB:s samlingar. Syftet har varit att bättre spegla svenska språkliga och samhällseliga förhållanden samt stärka modellernas relevans för svenska användningskontexter. Denna utvecklingsmodell skapar bättre förutsättningar att kombinera generella språkliga förmågor med högre lokal precision och terminologisk träffsäkerhet. I en sådan utvecklingsmodell fungerar den stora språkmodellen som teknisk bas, medan nationell anpassning bidrar till ökad transparens, spårbarhet och verksamhetsrelevans. Detta möjliggör för en mer kostnadseffektiv och ändamålsenlig utveckling där internationella tekniska framsteg kombineras med nationella språkresurser.

Som påtalats skiljer sig därmed KB:s språkmodellsverksamhet i flera avseenden från den utvecklingslogik som präglar stora delar av den internationella AI-utvecklingen. Medan många kommersiella språkmodeller utvecklas i sammanhang där lanseringstakt, marknadspositionering och teknologiskt försprång är centrala drivkrafter, utgår KB:s arbete från myndighetens långsiktiga samhällsuppdrag och forskningsstödande funktion. Utvecklingen har i huvudsak inriktats på funktionalitet, stabilitet och användbarhet i forskningsmiljöer. Prioriteringarna har därmed i större utsträckning utgått från identifierade forskningsbehov, praktiska användningsområden och långsiktig förvaltning än från kortsiktiga innovationscykler eller tekniska demonstrationseffekter. Den inriktningen har även skapat förutsättningar för en bred användning inom offentlig sektor.

KB:s cirka 80 utvecklade modeller för forskningsändamål inom samtliga modaliteter har fått en omfattande spridning och används i dag inom akademi, offentlig sektor och andra delar av samhället. Det är tydligt att modellerna i många fall integrerats som komponenter i system och tjänster för löpande användning, vilket talar för att de i praktiken används i ett stort antal digitala tillämpningar. Det breda genomslaget, trots avsaknad av egentliga marknadsföringsinsatser, kan tolkas som en indikation på att modellerna uppfattas som tekniskt användbara och språkligt relevanta i praktiska sammanhang.

De språkmodeller med öppna vikter som KB utvecklar för forskningsändamål tillgängliggörs öppet i enlighet med principerna för öppen vetenskap.³³ En grundläggande utgångspunkt för öppen vetenskap är att FoU som finansieras med offentliga medel, i den mån det är möjligt och förenligt med gällande rätt, bör komma samhället till del genom öppna och tillgängliga resultat. Det innebär att språkteknologiska resurser och forskningsresultat ska kunna återanvändas, vidareutvecklas och anpassas inom forskning, offentlig sektor så väl som inom andra delar av samhället. Samtidigt aktualiserar öppen vetenskap, i relation till utvecklingen av språkmodeller för forskningsändamål, frågor om relationen mellan offentliga resurser,

³³ Kungl. biblioteket, *Nationella riktlinjer för öppen vetenskap* (KB 2024-42).

forskningsresultat och privat kontroll över teknologiska tillämpningar. KB:s verksamhet illustrerar med andra ord hur nationellt anpassade språkmodeller för forskningsändamål utvecklas genom att kombinera internationella tekniska framsteg med nationella språkresurser och verksamhetsnära forskningsbehov. Sammanfattningsvis visar KB:s verksamhet för forskningsändamål att öppna modeller med öppna vikter är det mest ändamålsenliga arbetssättet, eftersom det möjliggör vidareutveckling, anpassning och kumulativ kapacitetsuppbyggnad utan att rådata behöver lämna förvaltande institution.

Digitalt kulturarv som del av nationell AI-infrastruktur för offentlig sektor

Språkmodeller är i flera avseenden en form av digital infrastruktur, där till exempel Södertörns högskola ser utvecklingen av svenska språkmodeller som en strategiskt viktig forskningsinfrastruktur och Göteborgs universitet ser språkmodeller som en grundläggande del av digital forskningsinfrastrukturen.³⁴ De är inte enbart avgränsade tekniska produkter, utan fungerar som grundläggande komponenter i ett brett spektrum av system och verksamhetsprocesser. På motsvarande sätt som annan digital infrastruktur kan de möjliggöra utveckling av nya tillämpningar, dels som Sveriges lantbruksuniversitet framför inom forskning och utbildning där de stödjer ny forskningsmetodik och förändrade former för kunskapsproduktion,³⁵ dels inom offentlig sektor där de stödjer nya sätt att hantera information och kunskapsstöd. Språkmodellens generella användbarhets värde ökar när de integreras med andra system, datakällor och digitala verktyg. På motsvarande sätt som för annan samhällsviktig digital infrastruktur kan det därför uppstå behov av tydliga krav på robusthet, transparens och kontinuitet i hur modeller utvecklas, förvaltas och används. Förstås KB:s verksamhet med språkmodellutveckling för forskningsändamål i relation till andra nationella satsningar på infrastrukturella resurser inom forskning förändras också utgångspunkterna för hur utvecklingen av nationella modeller betraktas, där det finns, som exempelvis Blekinge tekniska högskola poängterar, en nödvändighet att utveckling av både generella och områdesspecifika modeller ges stöd.³⁶

KB är enligt instruktionen en nationell forskningsinfrastruktur och kan, i egenskap av nationalbibliotek, närmare beskrivas som en samlingsbaserad forskningsinfrastruktur. Med detta avses en infrastruktur där forskning och kunskapsutveckling i stor utsträckning utgår från tillgången till omfattande kulturarvssamlingar. Kärnan i denna infrastruktur utgörs av KB:s fysiska och digitala samlingar, såsom böcker, handskrifter, kartor, audiovisuellt material (såsom radio, TV, film med mera) och digitalt födda resurser. Samlingarna utgör en central grund för forskning inom ett brett spektrum av vetenskapliga områden, inte minst inom humaniora och samhällsvetenskap. För förståelsen av KB:s infrastrukturella roll är det

³⁴ Göteborgs universitet, skrivelse via mejl, inkommen 2025-05-20 (KB 2025-930); Södertörns högskola, skrivelse via mejl, inkommen 2025-05-26 (KB 2025-930).

³⁵ Sveriges lantbruksuniversitet, skrivelse via mejl, inkommen 2025-05-20 (KB 2025-930).

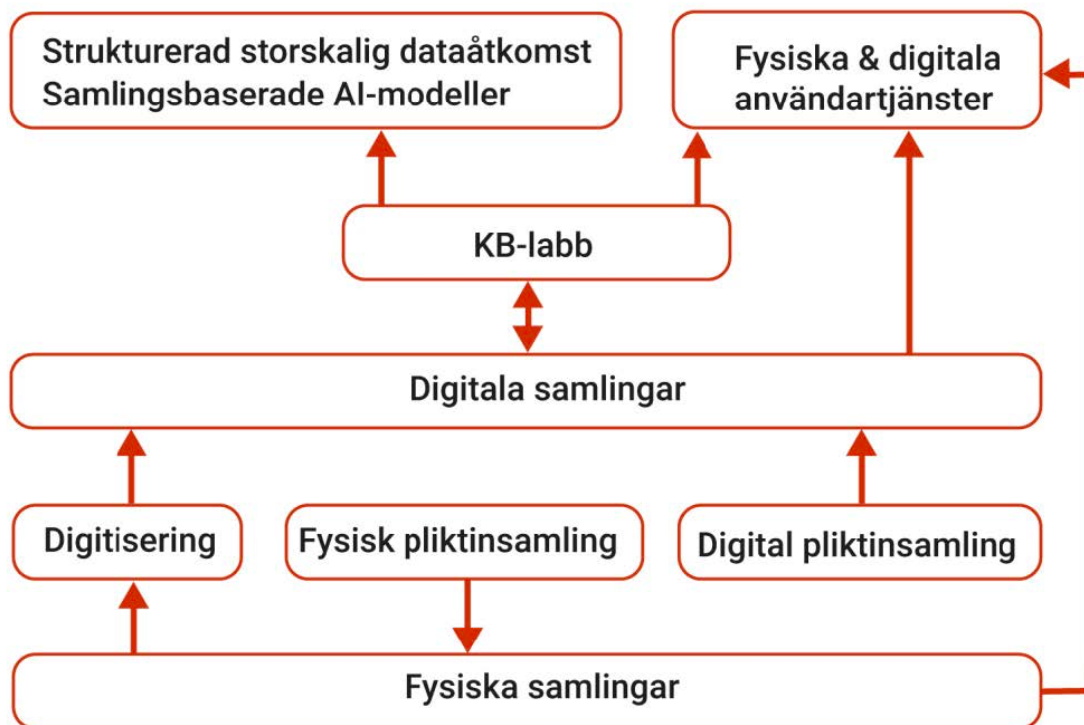
³⁶ Blekinge tekniska högskola, skrivelse via mejl, inkommen 2025-05-25 (KB 2025-930).

samtidigt viktigt att notera att samlingarna, i enlighet med bestämmelserna i Tryckfrihetsförordning (1949:105), inte utgör allmänna handlingar.³⁷

KB:s infrastrukturella funktion är att både bevara samlingarna över tid och att möjliggöra ett så ändamålsenligt tillhandahållande som möjligt. Det inkluderar utveckling och förvaltning av digitala plattformar, databaser och tekniska gränssnitt som stödjer olika forskningsmetoder, inklusive digitala angreppssätt. Kombinerat med ett nationalbiblioteks grunduppdrag av att säkerställa långsiktig lagring och dokumentation av materialets proveniens och integritet skapar detta förutsättningar för att forskare ska kunna identifiera, analysera, kombinera och återanvända kulturarvsmaterial i såväl traditionella som digitala forskningsmiljöer.

Språkmodeller och språkteknologiska resurser är avgörande komponenter för en samlingsbaserad forskningsinfrastruktur i samtiden på ett nationalbibliotek. Digital analyskapacitet är idag en integrerad del av hur samlingar tillgängliggörs, används och vidareutvecklas som forskningsresurs, vilket innebär att infrastrukturen inte enbart stödjer tillgång till enskilda objekt i samlingarna, utan också möjliggör nya former av analys av stora text- och datamängder för att svara mot forskarsamhällets behov och metodologiska utveckling. Ett nationalbibliotek av idag kan från detta perspektiv förstås som en nod där samlingar, teknisk infrastruktur och specialistkompetens samverkar i en dynamisk forskningsmiljö där historiskt och samtida material kan undersökas genom såväl etablerade vetenskapliga metoder som nya digitala och beräkningsintensiva angreppssätt. Förenklat kan KB:s flöde som en samlingsbaserad forskningsinfrastruktur beskrivas enligt följande bild:

³⁷ 2 kap. 14 § första stycket 3 Tryckfrihetsförordning (1949:105).



Bilden illustrerar att utgångspunkten för KB:s forskningsinfrastrukturverksamhet är samlingarna, vilka utgör det avgörande fundamentet oavsett vilken forskningsmetodik enskilda forskare nyttjar. Ett lika grundläggande beroende till samlingarna finns vid utveckling av språkmodeller, där tillgång till stora och kvalitativa datamängder är en förutsättning för att utveckla modeller med hög språklig relevans och precision. I den offentliga debatten har det mot denna bakgrund framförts att digitisering av arkiv- och kulturarvssamlingar bör förstås som del av uppbyggnaden av nationell AI-infrastruktur.³⁸ KB, likväl som andra kulturarvsmyndigheter som RA, har därför en central funktion att fylla som bara delvis har tagits i beaktande. Isöf har inom ramen för utredningen särskilt framhållit att ”digitalisering inte är en separat fråga vid sidan av AI-utvecklingen, utan en nödvändig förutsättning för den”.³⁹ I handlingsplanen för Sveriges AI-strategi framhålls tillgång till data och datainfrastruktur som grundläggande förutsättningar för utveckling och användning av AI.⁴⁰ Samtidigt behandlar handlingsplanen i begränsad utsträckning det mer specifika behovet av kulturarvsdata och den betydelse som språkliga resurser från kulturarvsinstitutioner har för utvecklingen av svenska språkmodeller och andra språkteknologiska tillämpningar. För att den ”inlåsta” potentialen i kulturarvet ska kunna realiseras finns det därför behov av en närmare koppling mellan KB:s övergripande

³⁸ Amanda Vesterlund, ”AI-boomens bränsle är inte digitaliserat”, *Dagens industri*, <https://www.di.se/debatt/ai-boomens-bransle-ar-inte-Digitiserat/>, kontrollerat 2026-04-28.

³⁹ Institutet för språk och folkminnen, skrivelse via mejl, inkommen 2026-05-07 (KB 2025-930).

⁴⁰ Regeringskansliet, *Handlingsplan för Sveriges AI-strategi*, 2026, sidorna 3-5.

digitiseringsbehov för bevarande och de nationella satsningar på data- och AI-infrastruktur som föreslås inom ramen för handlingsplanen.

Betydelsen av digitaliserade kulturarvssamlingar

Trots den växande betydelsen av kulturarvssamlingar för utveckling av språkmodeller för språklig precision är stora delar av det svenska kulturarvet fortfarande inte digitaliserat. Som visades i delredovisningen är i dag endast omkring 3-4 procent av KB:s samlade bestånd digitaliserat. För RA gäller cirka 7 %, ⁴¹ medan Isof uppskattar att 36,2 procent av samlingarna är digitaliserade, varav 9,8 procent är digitalt tillgängliga. ⁴² Detta kan till exempel jämföras med Norge, som har hela sitt kulturarv digitaliserat (se sidan 42). Samtidigt förvaltar arkiv, bibliotek och andra kulturarvsinstitutioner omfattande samlingar av texter och dokument som fortsatt endast finns i analog form. I praktiken är betydande delar av det svenska kulturarvet ännu inte tillgängliga som dataresurs för utveckling av språkmodeller och andra språkteknologiska tillämpningar. En begränsad tillgång till nationellt relevanta textresurser kan påverka möjligheterna att utveckla språkmodeller som speglar svenska språkliga, juridiska, historiska och samhällseliga förhållanden. För mindre språkområden, där tillgången till kvalitativa språkresurser redan är mer begränsad, får detta särskild betydelse för möjligheten att utveckla modeller med hög relevans och precision i nationella kontexter. I ett längre perspektiv kan omfattningen av digitalisering inom kulturarvssektorn påverka Sveriges förutsättningar att utveckla AI-lösningar anpassade till nationella behov. Om kulturarvsmaterial i större utsträckning digitaliseras, struktureras och görs maskinläsbart skapas däremot bättre förutsättningar för utveckling av mer representativa och språkligt anpassade språkmodeller för svenska förhållanden.

Digitaliserade samlingar är med andra ord i förlängningen en förutsättning för utveckling och användning av språkmodeller inom offentlig sektor. Data utgör i dag ett strategiskt råmaterial i den digitala ekonomin, ⁴³ vilket innebär att frågor om datatillgång, datakvalitet och långsiktig dataförvaltning har fått ökad betydelse såväl internationellt som i Sverige. Samtidigt har kulturarvsdata historiskt i Sverige inte haft samma framträdande roll i diskussionen om nationell digital infrastruktur och AI-utveckling. ⁴⁴ Digitalisering av kulturarvssamlingar bör däremot betraktas som strategiska dataresurser, där digitalisering av kulturarv inte längre enbart betraktas som en kulturpolitisk eller forskningsrelaterad fråga, utan även som del av en bredare nationell digital infrastruktur. Om endast en begränsad del av kulturarvet är digitalt tillgängligt riskerar forskning och teknikutveckling att i praktiken

⁴¹ Riksarkivet, *Årsredovisning 2025 för Riksarkivet*, 2026, sidan 20.

⁴² Institutet för språk och folkminnen, skrivelse via mejl, inkommen 2026-05-07 (KB 2025-930). I slutet av 2025 omfattade Isof:s samlingar 246 468 accessions. Av dessa är 89 165 digitala, vilket motsvarar 36,2 procent av samlingarna. Andelen av samlingarna som är digitalt tillgänglig uppgår till 9,8 procent.

⁴³ Europeiska kommissionen, *En EU-strategi för data*, COM(2020) 66; OECD, *Going Digital to Advance Data Governance for Growth and Well-being*, 2022.

⁴⁴ Regeringen, Finansdepartementet, *Data – en underutnyttjad resurs för Sverige: En strategi för ökad tillgång av data för bl.a. artificiell intelligens och digital innovation*, bilaga till beslut II 5 vid regeringssammanträde den 20 oktober 2021, I2021/02739.

styras mot det material som redan råkar vara digitiserat, snarare än det material som är mest relevant utifrån en specifik forskningsfråga eller det övergripande samhällsbehovet.

Även om endast omkring 3-4 procent av KB:s samlingar i dag är digitiserade utgör detta redan en mycket omfattande mängd material i ett svenskt perspektiv. Samlingarnas omfattning innebär att KB, möjligen med undantag för RA, förvaltar en av de mest betydande samlade resurserna av digitiserat textmaterial i Sverige. Därutöver erbjuder KB:s samlingar multimodala resurser, där text kompletteras av bild-, ljud- och audiovisuellt material. Det skapar särskilda förutsättningar för utveckling av språkmodeller och andra AI-tillämpningar som i ökande grad bygger på kombinationer av olika datatyper och modaliteter. Situationen är i flera avseenden likartad hos Isof, vars samlingar omfattar dialektinspelningar, folkminnesmaterial, ortnamn, personnamn, frågelistsvar, ordboksresurser och andra språkliga och kulturhistoriska källor.⁴⁵ Dessa resurser innehåller såväl historiska som samtida data om språk och kultur i Sverige och kan därmed skapa förutsättningar för utveckling av språkmodeller som bättre speglar språkliga variationer.

Digitisering av kulturarv som del av nationell AI-infrastruktur behöver också förstås i relation till bredare frågor om samverkan inom offentlig sektor och mer specifikt inom ABM-sektorn (arkiv, bibliotek och museer). Utvecklingen av språkmodeller och andra AI-tillämpningar är beroende av tillgång till relevanta dataresurser, men det är inte enbart mängden data som är avgörande. Även datas struktur, kvalitet, dokumentation och långsiktiga förvaltning har stor betydelse för möjligheterna att utveckla tillförlitliga och användbara språkmodeller. Digitisering av kulturarvssamlingar innebär betydligt mer än att enbart skapa digitala kopior av analogt material. För att kulturarvsdata ska kunna användas effektivt vid utveckling av språkmodeller krävs ofta omfattande förberedande och kvalitetssäkrande insatser för att kunna användas i maskinella processer.

AI-kommissionens betänkande *Färdplan för Sverige* (SOU 2025:12) belyser att betydande utmaningar kvarstår när det gäller nationell samordning och koordinering av övergripande initiativ inom AI-området.⁴⁶ Betänkandet identifierar ett behov av organisatoriska och politiska reformer för att stärka Sveriges kapacitet inom AI, men innehåller i mer begränsad utsträckning analyser av hur långsiktig tillgång till träningsdata och andra strategiska dataresurser bör säkerställas. Frågan om dataförsörjning är ett område där ytterligare förtydliganden och strategiska ställningstaganden är nödvändiga. Denna aspekt uppmärksammas även av Finansinspektionen i myndighetens remissvar på SOU 2025:12. I remissyttrandet framhålls behovet av ytterligare förslag för att stärka förutsättningarna för dataförsörjning och särskilt de roller som KB och RA kan spela i arbetet med att tillgängliggöra data och utveckla språkmodeller.⁴⁷ Sveriges digitaliseringsstrategi pekar i en liknande riktning och på att användningen av AI inom offentlig sektor behöver förstås i relation till den bredare digitala omställningen av statsförvaltningen.⁴⁸ Strategin framhåller

⁴⁵ Institutet för språk och folkminnen, skrivelse via mejl, inkommen 2026-05-07 (KB 2025-930).

⁴⁶ SOU 2025:12, *AI-kommissionens Färdplan för Sverige*, sidorna 11, 46 och 149-150.

⁴⁷ Finansinspektionen, *Remissvar: AI-kommissionens Färdplan för Sverige* (SOU 2025:12), FI dnr 25-7411, sidan 2.

⁴⁸ Regeringskansliet, Finansdepartementet, *Sveriges Digitaliseringsstrategi 2025–2030*, 2025, sidan 26.

särskilt betydelsen av datadriven utveckling, interoperabilitet och automatisering som centrala komponenter i framtidens offentliga tjänster och verksamhetsprocesser. Nyttan är samtidigt nära kopplad till hur väl dataresurserna är strukturerade, dokumenterade och långsiktigt förvaltade för att kunna stödja interoperabilitet och användning inom offentlig sektor.

En bredare förståelse behövs för betydelsen av hur digitaliserade kulturarvssamlingar kan skapa bättre förutsättningar för sammanhängande strategier kring hur nationella dataresurser utvecklas, förvaltas och nyttiggörs i utvecklingen av svenska språkmodeller. Frågan handlar därmed inte enbart om digitalisering i sig, utan om hur kulturarvsdata kan integreras i ett större ekosystem för nationell AI-utveckling och digital infrastruktur. I vissa europeiska länder, däribland Nederländerna, uppmärksammades tidigt kopplingarna mellan kulturarvsresurser, språkdata och nationell AI-kapacitet som en del av utvecklingen av nationella AI-strategier.⁴⁹ Den svenska AI-strategin framhåller visserligen betydelsen av datadelning, digital infrastruktur och förbättrad tillgång till data som centrala förutsättningar för AI-utveckling. Samtidigt har mer riktade satsningar på digitalisering av kulturarvssamlingar hittills haft en relativt begränsad plats i genomförandet av dessa ambitioner. Ett undantag utgörs av den anslagsförstärkning som tillfördes RA i den senaste budgetpropositionen,⁵⁰ vilket illustrerar att frågor om kulturarvsdata och digitalisering i viss utsträckning har börjat uppmärksammas även i relation till bredare digitaliserings- och AI-politiska mål.

Betydelsen av reviderad plikttagstiftning

En viktig och hittills relativt underutnyttjad resurs för utvecklingen av svenska språkmodeller utgörs av det material som samlas in genom e-plikttagstiftningen.⁵¹ Genom e-plikt har KB tillgång till omfattande och delvis unika digitala resurser som speglar samtida svenska språk-, medie- och samhällsförhållanden. Materialet omfattar bland annat artiklar, ljudinnehåll och video från svenska medieföretags och programbolags webbplatser samt innehåll som publiceras i sociala medier inom ramen för professionell medieverksamhet. Därutöver ingår filmer, ljudfiler och textdokument som publiceras exklusivt på internet, under förutsättning att de har gjorts tillgängliga av svenska yrkesmässiga medieaktörer och riktar sig till en svensk publik. Detta innefattar även vissa typer av poddar och andra digitalt födda medieformat. E-pliktinsamlingen omfattar även myndigheters elektroniskt framställda material, såsom rapporter, broschyrer, årsredovisningar och andra typer av publikationer. För myndigheter gäller att elektroniskt framställt material ska levereras även när en fysisk motsvarighet redan har inkommit genom annan pliktinsamling. Sammantaget innebär detta att e-pliktssystemet successivt bygger upp omfattande samlingar av samtida digitalt material som kan bidra till att stärka förutsättningarna för utveckling av språkmodeller med hög relevans för svenska språkliga, institutionella och samhälleliga förhållanden.

⁴⁹ Ministerie van Economische Zaken en Klimaat, *Strategic Action Plan for Artificial Intelligence*, 2019, sidorna 34-36.

⁵⁰ Regeringskansliet, Finansdepartementet, *Budgetpropositionen för 2026*, Utgiftsområde 17, Kultur, medier, trossamfund och fritid (Prop. 2025/26:1), sidan 77.

⁵¹ Lag (2012:492) om pliktexemplar av elektroniskt material.

Likt i Sverige har många europeiska länder gett sina nationalbibliotek ett utökat uppdrag att samla in digitalt publicerat material och därmed utvecklat system för e-plikt.⁵² De samlingar som har byggts upp och alltjämt byggs upp som en del av e-plikten utgör i praktiken en av de mest kvalitetssäkrade källorna till språkliga data på nationell nivå. KB som Sveriges nationalbibliotek förvaltar via e-pliktlagstiftningen en av de mest omfattande, och kontinuerligt växande, samlingarna av svenskt digitalt publicerat material.

Idag saknas däremot lagstöd för insamling av tryckförlagor om det samtidigt trycks ett fysiskt exemplar. E-pliktlagstiftningen är subsidiär och underordnad lagen om pliktexemplar av dokument.⁵³ Detta innebär i praktiken att insamling av fysiska förlagor ges företräde framför insamling av digitala förlagor och digitala format. Om ett verk ges ut både i digital och i fysisk form, omfattas inte den digitala förlagan av den nuvarande lagstiftningen. Detta leder exempelvis till att KB behöver digitisera samtida dagstidningar för att göra dessa tillgängliga för forskning trots att de finns i en digital tryckförlag. Skulle ett sådant lagstöd finnas minskar behovet av framtida digitisering av fysiska pliktexemplar av böcker, tidningar med mera som löpande inkommer till KB. I stället kan digitiseringssatsarna koncentreras till det historiska materialet.⁵⁴ Myndigheten för tillgängliga medier (MTM) har i ett remissvar till utredningen om ny pliktmateriallagstiftning (SOU 2021:32), uttryckt att en exkludering av tillgängliggjorda medier i det material som pliktas innebär att en viktig pusselbit saknas för framtida forskning.⁵⁵

Trots det insamlade materialets storlek, är med andra ord dess omfång fortfarande begränsat jämfört med länder som har nått betydligt längre i digitiseringen och lagstiftningen. Antologirapporten *Digital Cultural Heritage or: E-Legal Deposit in European National Libraries* lyfter fram exempel på flera europeiska länder som har moderniserat sin pliktlagstiftning för att stödja nationalbibliotekens digitisering och forskningens tillgång till samlingarna.⁵⁶ Pliktlagstiftningen i länder som Storbritannien, Frankrike, Spanien, Tyskland, Norge och Estland har övergått till att, utöver de fysiska exemplaren, i stor utsträckning även omfatta insamling av digitala förlagor och alternativa digitala utgivningsformat.

Kulturarvets låga digitiseringsgrad hämmar i förlängningen utvecklingen av svenska språkmodeller. Ett lagstöd för insamling av digitala förlagor hade inneburit bättre långsiktiga förutsättningar för svensk språkteknologisk forskning i och med att behovet för digitisering av framtida fysiska pliktexemplar av böcker, tidningar, tidskrifter, och annat analogt material som löpande inkommer till KB minskar. I stället kan digitiseringssatsarna koncentreras till det historiska materialet. Effekten blir sammantaget att mer material blir tillgängligt i digital form för utveckling av svenska språkmodeller för forskningsändamål och därmed i förlängningen skulle detta stödja offentlig sektors effektiviseringsbehov. Förslag att ge sådant

⁵² CENL, *Digital Cultural Heritage or: E-Legal Deposit in European National Libraries*, 2025.

⁵³ Lag (1993:1392) om pliktexemplar av dokument; Lag (2012:492) om pliktexemplar av elektroniskt material.

⁵⁴ SOU 2021:32, *Papper, poddar och ... Pliktmateriallagstiftning för ett tryggt källmaterial*, 2021, sidorna 244-245.

⁵⁵ Myndigheten för tillgängliga medier, *Remiss av SOU 2021:32 Papper, poddar och... Pliktmateriallagstiftning för ett tryggt källmaterial*, MTM 2021/371, sidan 3.

⁵⁶ CENL, *Digital Cultural Heritage or: E-Legal Deposit in European National Libraries*, 2025.

lagstöd återfinns i pliktutredning SOU 2021:32 som för närvarande bereds inom Regeringskansliet.⁵⁷

I direktivet till SOU 2021:32 framhålls att pliktregelverket tar sin utgångspunkt i såväl forskningens behov av tillgång till bevarat material för forskningsändamål som det bredare samhällsintresset av att säkerställa tillgång till material som publicerats i Sverige eller rör svenska förhållanden.⁵⁸ Utifrån ett perspektiv där digitiserat kulturarv förstås som del av nationell AI-infrastruktur aktualiserar detta även frågan om vilken roll pliktinsamling och digitisering har för utvecklingen av svenska språkmodeller och andra språkteknologiska resurser. Från ett bredare perspektiv kan digitisering av kulturarvssamlingar och en vidareutveckling av pliktregelverket förstås som långsiktiga investeringar i nationell datakapacitet. Om KB:s digitiseringsinsatser i ökad utsträckning planeras långsiktigt i relation till språkteknologiska behov kan detta skapa bättre förutsättningar för att utveckla datamängder som är särskilt användbara för utveckling av svenska språkmodeller för forskningsändamål. Det kan exempelvis handla om prioritering av material som är språkligt representativt, speglar olika samhällsområden eller har särskild relevans för offentlig verksamhet och myndighetskommunikation.

Ett konkret exempel på den samhällsnytta som kan uppstå till följd av långsiktiga investeringar inom digitisering och datatillgång kan åskådliggöras genom KB:s utveckling av tal-till-text-modellen KB-Whisper.⁵⁹ Modellen tränades på över 50 000 timmar svenska med tillhörande transkriptioner. De tillhörande transkriptionerna kom från undertexter, protokolltexter, samt ordagranna annotationer från tidigare forskningsprojekt. Andelen fel på ordnivå vid transkribering på svenska halverades i den största modellvariant som KB tränade (kb-whisper-large), jämfört med en stor motsvarande internationell grundmodell. Datamängden på över 50 000 timmar gjorde det möjligt att, utöver den största varianten, även träna välpresterande modeller av mindre storlek. Dessa mindre modeller kan köras lokalt på vanliga laptops, med svensk transkribering som är av högre eller likvärdig kvalitet jämfört med de bästa stora presterande internationella modellerna. Detta trots en minskning i antalet modellparametrar med faktor tio eller mer.

Utvecklingen av KB-Whisper har möjliggjorts av en kedja av beroenden där långsiktiga investeringar inom arkivhållning, digitisering, infrastruktur, och forskning har resulterat i att värdefulla datamängder blivit tillgängliga att beforskas. En betydande del (cirka 40 procent) av träningsdatamängden för KB-Whisper utgjordes av inspelningar av debatter i riksdagen, med tillhörande protokoll över de anföranden som har hållits. Den kedja av beroenden som gjorde materialet tillgängligt för modellutveckling inbegrep investeringar i:

- arkivföring, bevarande och lagring av riksdagstexter, ljud- och videoupptagningar från riksdagen.

⁵⁷ SOU 2021:32, *Papper, poddar och ... Pliktmateriallagstiftning för ett tryggt källmaterial*, 2021, sidorna 244-245.

⁵⁸ Regeringskansliet, Utbildningsdepartementet, *Översyn av regelverket för pliktleverans och hantering av annat material vid Kungl. biblioteket*, dir. 2019:84, sidan 2.

⁵⁹ Kungl. biblioteket, *Speech recognition for Swedish using wav2vec 2.0, Whisper and Conformers* (KB 2023-717).

- digitisering av äldre riksdagstryck, ljud- och videoupptagningar från analoga format.
- infrastruktur för tillgängliggörande av det digitiserade materialet genom riksdagens öppna data, och delvis genom samarbete med KB.⁶⁰
- bearbetning, vidareförädling och berikning av det digitiserade materialet i forskningsprojekt.⁶¹

I den alternativa verklighet där en eller flera länkar i ovanstående kedja saknas, förblir materialet otillgängligt för storskalig kvantitativ forskning, och hämmar därmed språkmodellsutvecklingen. De konsekvenser som uppstår, och den samhällsnytta som går förlorad, till följd av nuvarande låga digitiseringsgrad av kulturarvet är svåra att konkretisera och överblicka. De utgår, på liknande vis som ovanstående exempel, från en hypotetisk verklighet där effekterna av ett ännu ej digitiserat kulturarv tydligt går att åskådliggöra.

Nuvarande pliktlagstiftning innebär att KB, i många fall, saknar möjlighet att samla in och bevara kulturarvet i de format som är mest användbara och bäst lämpade för att stödja framtida forskning. För närvarande saknar, till exempel, KB lagstöd att begära in undertexter som en del av pliktleveransen av tv-sändningar. KB saknar även stöd att samla in ljudböcker och e-textböcker, i det fall boken parallellpubliceras i tryckt format. För att återknyta till det tidigare exemplet med ljud och tillhörande transkriptioner: KB bedömer att hundratusentals, sannolikt över en miljon, timmar ljud och video saknar tillhörande texter och förblir därmed otillgängliga för den typ av forskning som har möjliggjort utvecklingen av KB-Whisper.

Sammantaget bör digitalt kulturarv betraktas som en central del av Sveriges nationella AI-infrastruktur. För ett mindre språkområde som det svenska är tillgången till representativa, kvalitetssäkrade och långsiktigt förvaltade språkresurser avgörande för möjligheten att utveckla språkmodeller med hög relevans och precision. Karlstads universitet lyfter exempelvis fram att ”kvalitetssäkrade svenska språkresurser är en grundförutsättning för forskning och utveckling av språkmodeller på svenska”.⁶² KB:s samlingar, e-pliktmaterial och språkteknologiska verksamhet utgör i detta sammanhang en strategisk resurs för både forskning och offentlig sektor. En mer sammanhållen politik för digitisering, pliktinsamling, kulturarvsdata och AI-infrastruktur skulle därför kunna stärka Sveriges förutsättningar att utveckla öppna, transparenta och verksamhetsanpassade språkmodeller på svenska villkor.

⁶⁰ Riksdagstryck från 1521–1970 har digitiserats av riksdagsförvaltningen och tillgängliggjorts av KB på riksdagstryck.kb.se.

⁶¹ Det RJ-finansierade projektet SWERIK (Swedish Riksdag 1867–2022: An Ecosystem of Linked Open Data) har bearbetat, berikat och strukturerat de inscannade digitiserade protokollen med information som har varit viktig som utgångspunkt för att kunna skapa ett dataset utifrån materialet.

⁶² Karlstads universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

Andra nationalbiblioteks utveckling av språkmodeller

Eftersom uppdragsdirektivet utgår från KB:s verksamhet med utveckling av språkmodeller för forskningsändamål har den internationella jämförelsen avgränsats till nationalbibliotek. Mot bakgrund av att denna verksamhet på KB bedrivs inom ramen för myndighetens FoU-verksamhet har nationalbibliotek med mer begränsad forsknings- och utvecklingskapacitet inte heller inkluderats i analysen. Urvalsprocessen har därmed inriktats mot aktörer som bedömts ha särskilt relevanta erfarenheter av digitisering, forskningsinfrastruktur och språkteknologisk utveckling. Urvalet har vidare avgränsats till EU-länder samt EES-landet Norge, mot bakgrund av att dessa länder har implementerat, eller i Norges fall är i färd med att implementera, Europaparlamentets och rådets direktiv om upphovsrätt på den digitala inre marknaden (DSM-direktivet, 2019/790).⁶³ Tre nationalbibliotek har därmed identifierats som särskilt relevanta för en fördjupad analys: Bibliothèque nationale de France (BnF) i Frankrike, Det Kongelige Bibliotek i Danmark (KB-DK) samt Nasjonalbiblioteket i Norge (NB).

Urvalet har även motiverats av frågor om språklig representation och utveckling av språkmodeller för olika språkområden. Medan franska utgör ett internationellt större språk representerar Danmark och Norge mindre språkområden vars förutsättningar i flera avseenden är mer jämförbara med svenska förhållanden. Särskilt relevant är även att NB arbetar med utveckling av språkresurser och språkmodeller för samiska språk, vilket har bäring på frågor om nationella minoritetsspråk och därmed anknyter till KB:s uppdrag.

Informationsinhämtningen har genomförts genom en kombination av genomgångar av offentligt tillgängligt material och fördjupade intervjuer med representanter från BnF och NB. Syftet har varit att skapa en fördjupad förståelse för hur olika nationalbibliotek arbetar med frågor som rör digitisering, dataförvaltning, språkmodeller och rättsliga förutsättningar för AI-relaterad verksamhet. Av de undersökta nationalbiblioteken är det endast NB som har ingått ett särskilt avtal avseende användning av vissa materialkategorier (tidningsmaterial) för utveckling av språkmodeller för forskningsändamål i och med att artikel 3 i DSM-direktivet inte tillämpas.

En gemensam utgångspunkt för de tre nationalbibliotekens arbete med språkmodeller är strävan efter att utveckla språkteknologiska resurser som kombinerar hög språklig precision med representativitet, tillgång till kvalitativa data, transparens och långsiktig kontroll över nationella språkresurser. I de studerade exemplen framträder återkommande frågor om datakvalitet, dokumentation, spårbarhet och möjligheten att utveckla modeller som speglar nationella språkliga och kulturella förhållanden. Frågor om digital suveränitet, förstått som kapaciteten att nationellt kunna utveckla, förvalta och påverka centrala digitala resurser, utgör i varierande grad också en del av dessa resonemang. Av de nationalbibliotek som

⁶³ Europaparlamentets och rådets direktiv (EU) 2019/790 av den 17 april 2019 om upphovsrätt och närstående rättigheter på den digitala inre marknaden och om ändring av direktiven 96/9/EG och 2001/29/EG.

behandlas i denna redovisning har särskilt BnF och NB mer uttalade uppdrag eller etablerade strukturer för att tillsammans med andra myndigheter, forskningsaktörer, organisationer och i vissa fall näringsliv bidra till utveckling och vidareutveckling av nationella språkmodeller. Den samlade bedömningen är att den utvecklingsverksamhet som bedrivs av KB för forskningsändamål ryms inom ett växande fält där kulturarvsinstitutioner använder sina samlingar och forskningsinfrastrukturer för språkteknologisk utveckling inom ramen för forskning.

Bibliothèque nationale de France, Frankrike

BnF har sedan 2021 bedrivit en särskild AI-relaterad verksamhet genom DataLab,⁶⁴ som fungerar som en miljö för experiment, utveckling och analys av digitala metoder och AI-tillämpningar kopplade till bibliotekets samlingar. Samtidigt visar genomgången att biblioteket arbetat med frågor relaterade till AI och kulturarvsdata sedan åtminstone 2018. Tillsammans med NB var BnF dessutom en av de sex grundande institutionerna bakom AI4LAM,⁶⁵ ett internationellt initiativ med fokus på AI inom ABM-sektorn. BnF deltar även i arbetsgruppen AI in Libraries, där även NB deltar.⁶⁶ BnF bedriver i dag utveckling och träning av ett flertal modeller, inklusive mer experimentella och forskningsnära tillämpningar i samarbete med akademiska institutioner. Utöver forskningsrelaterad modellutveckling bedrivs även interna projekt med fokus på verksamhetsstöd, samlingsförvaltning och större utvecklingsprojekt tillsammans med externa aktörer inom privat sektor.

En betydande del av BnF:s nuvarande kapacitet att bedriva datadrivet forskningsstöd och AI-relaterad utveckling härrör från de omfattande franska digitiseringsinsatserna som genomfördes under 2010-talet.⁶⁷ Bibliotekets långsiktiga arbete med att digitisera, strukturera och tillgängliggöra samlingarna digitalt har också skapat de förutsättningar för mer avancerade former av digital analys och forskningsstöd som idag utvecklas. I dialoger har det framhållits att AI-relaterade verksamheter i stor utsträckning kunnat byggas vidare på tidigare investeringar i digital infrastruktur och dataförvaltning. Ett exempel är det digitala biblioteket Gallica,⁶⁸ som tillgängliggör miljontals dokument ur BnF:s samlingar och nu utgör grund för bibliotekets digitala forskningsmiljö, vilket inkluderar forskningsprojekt för utveckling av språkmodeller. Erfarenheterna från BnF illustrerar därmed hur långsiktiga investeringar, lik desamma som låg till grund för utvecklingen av KB-Whisper, kan skapa förutsättningar för senare utveckling av AI-relaterade tillämpningar och språkteknologiska resurser.

I och med sin framträdande roll har BnF även getts en aktiv roll i det franska arbetet med frågor kopplade till upphovsrätt, dataåtkomst och ersättningsmodeller vid utveckling av språkmodeller. BnF deltog exempelvis som konsultativ part i en rapport framtagen för

⁶⁴ BnF DataLab, <https://www.bnf.fr/fr/bnf-datalab>, kontrollerad 2026-05-26.

⁶⁵ AI4LAM, <https://ai4lam.org/>, kontrollerad 2026-05-26.

⁶⁶ CENL, AI in Libraries Network Group, <https://www.cenl.org/networkgroups/ai-in-libraries-network-group/>, kontrollerad 2026-04-20.

⁶⁷ Marie Carlin och Arnaud Laborderie, "Le BnF DataLab, un service aux chercheurs en humanités numériques", *Humanités numériques*, 4, 2021: <https://doi.org/10.4000/revuehn.2684>

⁶⁸ Gallica, <https://gallica.bnf.fr/accueil/fr/html/accueil-fr>, kontrollerad 2026-03-30.

Conseil supérieur de la propriété littéraire et artistique, ett rådgivande organ under det franska kulturministeriet med ansvar för immaterialrättsliga frågor.⁶⁹ Rapporten behandlade bland annat rättsliga och ekonomiska förutsättningar för användning av kulturarvsdata i AI-relaterade sammanhang och identifierade BnF samt Institut national de l'audiovisuel (INA) som centrala aktörer för tillgång till dataresurser vid utveckling av språkmodeller.

BnF deltar även i konsortiet ArGiMi inom ramen för det nationella investeringsprogrammet France 2030 som är ett samarbete mellan offentliga och privata aktörer.⁷⁰ Konsortiet är inriktat på utveckling av storskaliga franskspråkiga språkmodeller med särskilt fokus på regelefterlevnad, transparens och etiskt ansvarsfull användning av generativ AI. Inom ramen för ArGiMi har BnF tillsammans med INA tagit fram en juridisk vägledning för användning av kulturarvsdata i AI-relaterade sammanhang och som är speciellt intressant för denna utredning. Vägledningen, publicerad den 26 april 2026, behandlar rättsliga förutsättningar för användning av kulturarvsdata vid träning av generativa modeller samt tillämpningen av undantagen för TDM för forskningsändamål, inklusive användning av pliktinsamlat material hos biblioteks- och kulturarvsinstitutioner. Rapporten konstaterar att:

”L’application de l’exception de fouilles de textes et de données « recherche » aux données patrimoniales exige un accès licite à ces données. Cet accès licite peut avoir été autorisé par les titulaires de droits, dans les emprises de l’entité bénéficiaire de l’exception ou plus largement. La condition de l’accès licite peut également être remplie pour les données patrimoniales conformément à deux exceptions aux droits de propriété intellectuelle : les exceptions dites « dépôt légal » et « bibliothèques ». En effet, bien que ne faisant l’objet d’aucun texte légal ou réglementaire, le principe d’une combinaison de deux exceptions a déjà été admis par la Cour de justice de l’Union européenne. Ce cumul d’exceptions suppose toutefois de respecter les conditions qui y sont attachées ; ainsi, l’utilisation et la conservation des données patrimoniales dans le cadre de l’exception de fouille ne sera possible que dans les emprises de l’organisme bénéficiaire des exceptions, dès lors que les exceptions « bibliothèque » et « dépôt légal » n’autorisent pas la consultation des œuvres en-dehors de ces emprises.”⁷¹

”Tillämpningen av undantaget för text- och datautvinning för forskningsändamål på kulturarvsdata förutsätter laglig tillgång till dessa data. Denna lagliga tillgång kan ha beviljats av rättighetshavarna, antingen inom den enhet som omfattas av undantaget eller i ett vidare sammanhang. Villkoret om laglig tillgång kan även vara uppfyllt för kulturarvsdata genom två undantag från immaterialrätten: de så kallade undantagen för pliktleverans och bibliotek. Även om detta inte uttryckligen regleras i lag eller annan författning har principen om att kombinera två undantag redan godtagits av Europeiska unionens domstol. En sådan kumulation av undantag förutsätter dock att de villkor som är knutna till respektive undantag iakttas. Det innebär att användning och bevarande av kulturarvsdata inom ramen för undantaget för text- och datautvinning endast är möjligt inom den organisation som omfattas

⁶⁹ CSPLA, *Rapport de mission sur la rémunération des contenus culturels utilisés par les systèmes d’IA* (2025).

⁷⁰ ArGiMi, <https://huggingface.co/argimi>, kontrollerad 2026-05-25.

⁷¹ INA och BnF, *Guide juridique sur l’utilisation des données patrimoniales dans le cadre de l’intelligence artificielle generative* (2026), sidorna 27-28.

av undantagen, eftersom undantagen för bibliotek och pliktleverans inte medger att verk görs tillgängliga för konsultation utanför dessa organisatoriska ramar.”

Vägledningen kommer med andra ord fram till samma slutsatser som de externa experterna gör i denna utredning.

Det Kongelige Bibliotek, Danmark

KB-DK verkar under organisatoriska förutsättningar som i flera avseenden skiljer sig från dem som gäller för KB. Genom en organisationsförändring år 2017 integrerades nationalbiblioteket med flera universitetsbibliotek inom ramen för en gemensam organisation, vilket skapar särskilda förutsättningar för samverkan. KB-DK omfattar därmed, utöver nationalbibliotekets verksamhet, även universitetsbiblioteken vid Köpenhamns universitet, Roskilde universitet, Aarhus universitet och IT-Universitetet i Köpenhamn.

KB-DK har under en längre tid genomfört omfattande digitiseringsinsatser som successivt kompletterats med infrastrukturer för tillgängliggörande och återanvändning av data i forskningssammanhang. I KB-DK:s strategi för perioden 2024–2027 betonas vikten av att stärka den integrerade servicen mellan nationalbiblioteket och universitetsbiblioteken samt att förbättra stödet till forskares användning av kulturarvssamlingar, bland annat genom utvecklade möjligheter för TDM.⁷² KB-DK:s AI-relaterade verksamhet kan övergripande ses utifrån tre huvudsakliga områden: utveckling och tillgängliggörande av data för intern och extern användning, AI-relaterade tillämpningar kopplade till samlingarna samt en experimentell verksamhet inom det så kallade KB Labs etablerat 2016 med målet att utforska möjligheterna i dess digitala samlingar.⁷³

Idag erbjuder KB-DK tjänster för utlämning av dataset samt dokumentation och tekniskt stöd som underlättar användning av materialet i digitala forskningsmetoder. I vissa fall tillhandahålls även vägledningar via GitHub för att stödja forskare i visualisering, textutvinning och annan databaserad analys.⁷⁴ Dataset där upphovsrättsliga begränsningar inte längre föreligger tillgängliggörs genom repositoriet Library Open Access Repository, där forskningsdata och andra datamängder från KB-DK samlas tillsammans med material från bland annat Aarhus universitet, Aalborg universitet och Nationalmuseet i Danmark.⁷⁵ En del av bibliotekets dataset av historiska och offentliga data har dessutom använts i utvecklingen av de så kallade Danish Foundation Models, ett samverkansinitiativ för utveckling av danska språkmodeller mellan flera lärosäten och forskningsnära aktörer inom näringslivet.⁷⁶

⁷² Det Kgl. Bibliotek, *Vi viser vej til viden*, <https://www.kb.dk/om-os/opgaver-og-maal/strategi>, kontrollerad 2026-04-20.

⁷³ KB Labs, <https://labs.kb.dk/>, kontrollerad 2026-05-05.

⁷⁴ Det Kgl. Bibliotek, *Cultural Heritage Data Guides – Python*, <https://kb-dk.github.io/cultural-heritage-data-guides-Python/intro.html>; Det Kgl. Bibliotek, *Cultural Heritage Data Guides – R*, <https://github.com/kb-dk/cultural-heritage-data-guides-R>, kontrollerad 2026-05-05.

⁷⁵ Det Kgl. Bibliotek, *Home LOAR*, <https://loar.kb.dk/home>, kontrollerat 2026-05-05.

⁷⁶ Danish Foundation Models, *Building the Foundation of Danish AI*, <https://www.foundationmodels.dk/>, kontrollerad 2026-05-05.

Upphovsrättsliga frågor har också haft en framträdande roll i KB-DK:s arbete. KB-DK har exempelvis haft med en representant i den expertgrupp om upphovsrätt och AI som tillsattes av det danska kulturministeriet. Slutrapporten överlämnades den 15 september 2025 och visar att expertgruppen gör en tydlig avgränsning mellan kommersiell AI-utveckling och modellträning för vetenskapliga forskningsändamål. Rapporten drar samma slutsats som BnF:s vägledning och det framhålls att forskning har berättigad tillgång till material för TDM.⁷⁷

Expertgruppen presenterar dessutom tio rekommendationer med fokus på reglering och hantering av användning av upphovsrättsligt skyddat material i kommersiella AI-system.⁷⁸ Rekommendationerna omfattar bland annat ökad transparens kring träningsdata, mekanismer för samtycke och förbehåll, utveckling av kollektiv avtalslicensiering samt modeller för hantering av rättighetskonflikter och rättsliga prövningar. Vidare behandlas frågor om skydd mot digital imitation av personliga attribut, säkerhetsmekanismer, stärkt användning av mänskligt skapat innehåll samt förbättrad vägledning kring AI och upphovsrätt.

Det är tydligt hur den danska expertgruppens slutsatser skiljer på kommersiell AI-utveckling och forskningsnära modellutveckling och att dessa behandlas utifrån skilda utgångspunkter. Expertgruppen förordar uttryckligen att särskilda undantagsbestämmelser fortsatt ska gälla för forskningsverksamhet.⁷⁹ Även denna expertgrupp kommer med andra ord fram till samma slutsatser som de externa experterna gör i denna utredning.

Nasjonalbiblioteket, Norge

Norge har via en nationell satsning i princip digitiserat merparten av sitt publicerade kulturarv på NB och 2020 utökat satsningen till att inkludera flera materialkategorier.⁸⁰ 2018 etablerades också en forskningsnära och laborativ miljö för AI-relaterad utveckling genom inrättandet av AI Lab. Verksamheten bedrivs inom ramen för nationalbibliotekets FoU-verksamhet och har till uppgift att utforska och utveckla användningen av AI inom ABM-sektorn.⁸¹ AI Lab har därmed fungerat som en miljö för experiment, metodutveckling och praktisk tillämpning av AI i relation till kulturarvsdata och NB:s omfattande digitala samlingar. NB:s satsningar på språkmodeller och AI-relaterad utveckling uppvisar i flera avseenden likheter med de utgångspunkter som återfinns i KB:s arbete. NB har ett brett strategiskt perspektiv på språkmodellutveckling, där frågor om digital suveränitet, nationell handlingsfrihet och minskat beroende av externa aktörer ges en framträdande roll. I detta

⁷⁷ Kulturministeriet, Danmark, Ekspertgruppe om ophavsret og kunstig intelligens, *Anbefalinger fra Ekspertgruppe om ophavsret og kunstig intelligens*, 2025, sidan 12.

⁷⁸ Kulturministeriet, Danmark, Ekspertgruppe om ophavsret og kunstig intelligens, *Anbefalinger fra Ekspertgruppe om ophavsret og kunstig intelligens*, 2025, sidor 57-66.

⁷⁹ Kulturministeriet, Danmark, Ekspertgruppe om ophavsret og kunstig intelligens, *Anbefalinger fra Ekspertgruppe om ophavsret og kunstig intelligens*, 2025, sidorna 65-66.

⁸⁰ Nasjonalbiblioteket, *Digitaliseringsplan for den norske dokumentbaserte kulturarven*, 2020. https://abmdig.no/content/uploads/sites/14/2020/08/Digitaliseringsplan-for-ABM-omr%C3%A5det_.pdf, kontrollrad 2026-05-25.

⁸¹ Nasjonalbiblioteket har även ett labb för digitalhumaniora, DH-LAB, som grundades 2017 och arbetar i synnerhet med att tillhandahålla REST-API:er, Python-paket och webbapplikationer för forskning på material ur bibliotekets samlingar.

sammanhang betonas särskilt betydelsen av öppna, transparenta och väldokumenterade språkmodeller, inte minst i relation till offentlig sektor och forskning.

NB:s FoU-verksamhet bedrivs, i likhet med KB:s FoU-verksamhet, i nära samverkan med externa forskningsmiljöer och akademiska aktörer. Samarbeten har bland annat etablerats med Universitetet i Oslo och Norges teknisk-naturvitenskapelige universitet (NTNU), där nationalbibliotekets samlingar och dataresurser används i FoU-projekt kopplade till AI och språkmodeller. Verksamheten omfattar även utveckling av samiska språkresurser och språkmodeller.⁸² Under 2024 fick NB dessutom i uppdrag att genomföra ett utvecklingsprojekt för att undersöka kvaliteten på upphovsrättsligt skyddat material vid träning av norska generativa språkmodeller.⁸³ Uppdraget resulterade i etableringen av Mimir-projektet, som bedrevs i samarbete med Universitetet i Oslo och NTNU genom forskningsmiljön NorwAI.⁸⁴

Även om Mimir-projektet ursprungligen inte hade som uttalat syfte att analysera ersättningsmodeller för användning av upphovsrättsligt skyddat material vid träning av generativa språkmodeller kom projektet att få betydande policykonsekvenser. En följd av arbetet blev att NB fick i uppdrag att vidareutveckla norska språkmodeller samt mandat att förhandla med rättighetsinnehavare om tillgång till relevanta datamängder för modellträning.⁸⁵ Av särskilt intresse i det norska sammanhanget är därför det avtal som NB ingick med Kopinor, den kollektiva förvaltningsorganisation som företräder rättighetsinnehavare, på uppdrag av Mediebedriftenes Landsforening. Avtalet avser användning av norskt tidningsmaterial för träning av språkmodeller.

Den överenskomna ersättningen för användning av materialet uppgår till 45 miljoner norska kronor per år,⁸⁶ en kostnad som hanteras genom motsvarande anslagsförstärkning till NB.⁸⁷ Avtalet ger tillgång till tidningsmaterial äldre än ett år för träning av språkmodeller och omfattar upphovsrättsliga förfoganden vid vidareanvändning av modellerna av såväl offentliga som privata aktörer. Avtalet har även inneburit tillgång till digitala förlagor av tidningsmaterial. Samtidigt innehåller avtalet vissa kontrollmekanismer och avgränsningar. Uppsägningstiden uppgår till ett år och enskilda tidningsaktörer har möjlighet att avstå från att omfattas av överenskommelsen.

⁸² NorLLM, *Prosjekt «MIMIR» evaluerer verdien av opphavsrett*, <https://www.ntnu.no/norllm/prosjekt-mimir-evaluerer-verdien-av-opphavsrett>, kontrollerad 2026-05-06; Norsk rikskringkasting (NRK), *Samisk språk løftes i nye KI-modeller*, <https://www.nrk.no/sapmi/samisk-sprak-loftes-i-nye-ki-modeller-1.17701717>, kontrollerad 2026-05-06.

⁸³ Regjeringen, Kultur- og likestillingsdepartementet, *Statsbudsjettet 2024 - tildelingsbrev til Nasjonalbiblioteket* (23/392-7), sidan 4.

⁸⁴ Nasjonalbiblioteket, *Mimir-prosjektet. Evaluering av virkningen av opphovsrettsbeskyttet materiale på generative store språkmodeller for norske språk* (2024), https://www.nb.no/content/uploads/2024/08/Mimirprosjektet_teknisk-rapport.pdf, kontrollerad 2026-05-13.

⁸⁵ Regjeringen, Kultur- og likestillingsdepartementet, *Statsbudsjettet 2025 - tildelingsbrev til Nasjonalbiblioteket* (23/5753), sidan 5.

⁸⁶ Nasjonalbiblioteket, *Nasjonalbiblioteket inngår historisk avtale med landets aviser om innhold til kunstig intelligens*, <https://www.nb.no/pressemeldinger/nasjonalbiblioteket-inngar-historisk-avtale-med-landets-aviser-om-innhold-til-kunstig-intelligens/>, kontrollerad 2026-04-20.

⁸⁷ Regjeringen, Kultur- og likestillingsdepartementet, *Statsbudsjettet 2026 - tildelingsbrev til Nasjonalbiblioteket* (25/6156), sidan 5.

Ramarna för det norska avtalet har samtidigt aktualiserat fortsatta diskussioner om hur andra typer av upphovsrättsligt skyddat material kan inkluderas i språkmodellutveckling. Dialoger och förhandlingar har bland annat inletts med bokförlag, men vid tidpunkten för denna redovisning hade dessa ännu inte resulterat i något motsvarande avtal för användning av litterära verk eller annat förlagsutgivet material.

Avtalet om tidningsmaterial har träffats genom Mediebedriftenes Landsforening och medieföretagen, vilka företräder utgivarsidan snarare än de professionella grupper som skapat innehållet, såsom journalister, översättare och andra skribenter. Detta innebär att frågor om hur ekonomisk ersättning och rättigheter ska fördelas mellan olika aktörer i viss utsträckning kvarstår även efter att den övergripande avtalslösningen nåtts. Om utvecklingen på sikt skulle leda till separata avtal för olika materialkategorier kan avtalsstrukturen samtidigt bli mer komplex och administrativt omfattande. Behovet av flera parallella avtalsrelationer med olika rättighetsföreträdare kan påverka kostnader, förutsägbarhet och möjligheten att långsiktigt planera språkmodellutvecklingen, vilket är speciellt problematiskt inom ett område där teknikutvecklingen sker snabbt. Detta har inte minst visat sig i det pågående arbetet i Norge med att implementera DSM-direktivet.

Till skillnad från Sverige är Norge inte medlem i EU men omfattas, via EES-avtalet, av delar av EU:s regelutveckling på det digitala området. Som en del av detta pågår ett arbete med att implementera DSM-direktivet i norsk rätt. Den 27 mars 2026 överlämnades ett lagförslag till Stortinget avseende ändringar i den norska upphovsrättslagstiftningen,⁸⁸ där införlivandet av bland annat DSM-direktivet färgat av nämnda avtalslösning.

⁸⁸ Regjeringen, Kultur- og likestillingsdepartementet, *Endringer i åndsverkloven mv. (gjennomføring av nett- og videresendingsdirektivet og digitalmarkedsdirektivet) og samtykke til godkjenning av EØS-komiteens beslutning nr. 332/2023 og 333/2023 om innlemmelse i EØS.-avtalen av direktiv (EU) 2019/789 og (EU) 2019/790, og samtykke til tiltredelse til WIPO-traktatene av 20. Desember 1996 om opphavsrett (WCT) og om fremføringer og fonogrammer (WPPT)*, Prop. 41 LS (2025 – 2026).

Ekonomiska aspekter vid KB:s utveckling av språkmodeller

Kostnader för utveckling och förvaltning av språkmodeller

En eventuell utvidgning av KB:s verksamhet med språkmodeller för forskningsändamål i syfte att även bidra till att möta behov inom andra relevanta delar av offentlig sektor innebär att KB i ökad utsträckning skulle behöva balansera sitt uppdrag som nationell forskningsinfrastruktur med förväntningar på att samtidigt bidra till uppbyggnaden av en bredare nationell infrastruktur för uppdragsforskning från offentlig sektor för utveckling av språkmodeller och språkteknologiska resurser.

Kostnader uppstår enligt KB:s bedömning inom ett flertal områden inte minst kopplat till löpande förvaltning. Resursbehoven ökar dessutom successivt i takt med att modeller förväntas kunna användas inom fler verksamhetsområden och möta högre krav på säkerhet, transparens, robusthet och kontinuerlig tillgänglighet. Erfarenheterna från den verksamhet för forskningsändamål KB har bedrivit ger att utveckling av språkmodeller inte enbart innebär initiala investeringar i modellträning, utan i hög grad handlar om långsiktig förvaltning av data, kompetens och tekniska miljöer. Detta innebär att kostnadsbilden behöver förstås i ett livscykelperspektiv, där betydande resurser krävs även efter att en modell har utvecklats och tagits i bruk. I slutbetänkandet *Fler möjligheter till ökat välbefinnande* (SOU 2025:96) bedömer Produktivitetskommissionen att KB:s arbete med språkmodeller, om det ska kunna skalas upp för att möta bredare behov inom offentlig sektor, kan behöva förstärkas med mellan 50 och 100 miljoner kronor per år.⁸⁹ Bedömningen är en grov uppskattning utifrån kommissionens erfarenhet av vad som krävs för viss slagkraft vid denna typ av satsningar och verksamhet, även om den förstås bör ses som en övergripande uppskattning snarare än en detaljerad kostnadsberäkning.⁹⁰

KB bedömer att Produktivitetskommissionens uppskattning i huvudsak framstår som rimlig i relation till de resursbehov som utveckling och förvaltning av språkmodeller innebär. KB:s arbete med språkmodeller för forskning innebär redan idag betydande ekonomiska åtaganden särskilt vad gäller investeringar i infrastruktur (beräkningsresurser och lagringskapacitet), datatillgång (digitisering av samlingarna) och högspecialiserad kompetens (kompetensförsörjning). Som nationalbibliotek har KB förvisso redan ansvar för att samla in, bevara och tillgängliggöra material, vilket redan har gjort KB till en naturlig aktör för att bygga högkvalitativa träningsdatamängder för forskningsändamål som kan hantera språkliga och kulturella aspekter. Med en breddning av KB:s verksamhet till att möta behov från offentlig sektor, följer däremot ett ökat krav på skalbarhet, robusthet och kontinuerlig tillgänglighet i de tekniska lösningarna där ytterligare investeringar i beräkningsresurser, lagringsmiljöer samt system för drift, säkerhet, dokumentation och användarstöd. KB

⁸⁹ SOU 2025:96, *Fler möjligheter till ökat välbefinnande*, Slutbetänkande av Produktivitetskommissionen, 2025, sidan 197.

⁹⁰ Mejlkontakt med Produktivitetskommissionens huvudsekreterare Kristian Seth, 13 april 2026 (dnr. 2025-930).

beräknar att med nuvarande anslagsnivå kan KB upprätthålla nuvarande kapacitet för att utveckla språkmodeller för forskningsändamål.

I den senaste budgetpropositionen tilldelades KB en anslagsförstärkning på 30 miljoner kronor för att möta ökade kostnader för digital lagring och stärkt beräkningskapacitet.⁹¹ Förstärkningen svarar mot delar av de behov som identifierats i relation till AI-relaterad forskning och digital infrastruktur, men att den samtidigt inte framstår som tillräcklig för att möta mer långtgående ambitioner för storskalig digitisering och utveckling av språkmodeller för bredare offentlig användning. Regeringen framhåller i propositionens motivbild att den snabba utvecklingen inom beräkningsteknik och AI har ökat relevansen av KB:s digitala samlingar för forskning samt pekar på behovet av fortsatt digitisering, vidareutveckling av språkmodeller för forskningsändamål och anpassning till offentlig verksamhet.⁹²

För en uppskalning uppstår främst tre typer av kostnader: utökad teknisk driftorganisation för lagring och beräkning inklusive hantering av data med visst skyddsvärde (till en beräknad kostnad av ytterligare 15 miljoner kronor per år utöver nuvarande anslag); utökad laborativ kapacitet för långsiktigt konkurrenskraftig utveckling och breddning av den datavetenskapliga förmågan för modellutveckling och modellinferens (till en beräknad kostnad av ytterligare 15 miljoner kronor per år utöver nuvarande anslag); samt utökad AI-understödjande digitisering och tillhörande strukturering och kvalitetssäkring av datan (till en beräknad kostnad av ytterligare 10 miljoner kronor per år utöver nuvarande anslag).

Utökad teknisk driftorganisation för lagring och beräkning

Ökad beräknings- och lagringskapacitet medför samtidigt ökade kostnader för lokaler, energiförsörjning och teknisk drift. Uppbyggnaden av avancerad språkteknologisk infrastruktur är en successiv och långsiktig process som förutsätter kontinuerliga investeringar i teknik, data, kompetens och organisatorisk kapacitet. Om kontinuiteten i verksamheten bryts finns risk att upparbetad kompetens och teknisk kapacitet successivt förloras. Att återuppbygga specialistkompetens inom språkteknologi, AI och dataförvaltning kan vara både tidskrävande och kostsamt, särskilt i ett område där den internationella utvecklingstakten är hög och den nationella konkurrensen om kompetens är betydande. Även kortare avbrott i investeringar eller verksamhetsutveckling kan därmed få långsiktiga konsekvenser för möjligheten att upprätthålla nationell kapacitet. Det arbete som KB inledde i projektform med att successivt bygga upp en forskningsinfrastruktur för digitala humaniora har bidragit till att etablera och upprätthålla kompetens, metoder och teknisk kapacitet. Samtidigt framstår verksamhetens långsiktiga relevans som beroende av att utvecklingen kan ske i takt med förändringar i omvärlden. Om investeringar och organisatoriska förutsättningar inte utvecklas i relation till den tekniska utvecklingens tempo finns en risk att nyttan och användbarheten av KB:s verksamhet successivt minskar. För ett land av Sveriges storlek,

⁹¹ Regeringskansliet, Finansdepartementet, *Budgetpropositionen för 2026*, Utgiftsområde 16, Utbildning och universitetsforskning (Prop. 2025/26:1), sidan 246.

⁹² Regeringskansliet, Finansdepartementet, *Budgetpropositionen för 2026*, Utgiftsområde 16, Utbildning och universitetsforskning (Prop. 2025/26:1), sidan 235.

med mer begränsade resurser än större internationella aktörer, framstår frågan om långsiktig finansiering och institutionell stabilitet därför som särskilt strategisk.

Utökad laborativ kapacitet för långsiktigt konkurrenskraftig utveckling

Själva utvecklingen av språkmodeller utgör däremot endast ett första steg i en långsiktig verksamhet där kontinuerlig förvaltning och vidareutveckling är avgörande för modellernas användbarhet och relevans över tid. För att språkmodeller ska kunna användas på ett tillförlitligt sätt krävs löpande insatser för uppdatering, dokumentation, säkerhetsgranskning, kvalitetssäkring och anpassning till förändrade behov, regelverk och tekniska förutsättningar. Drift- och förvaltningskostnader utgör därmed en betydande och ofta underskattad del av den samlade kostnadsbilden. Om tillräcklig finansiering inte säkerställs långsiktigt finns en risk att modeller successivt tappar i relevans, kvalitet och funktionalitet i takt med att språkbruk, tekniska standarder och verksamhetskrav förändras. Detta framstår som särskilt betydelsefullt vid en breddning mot offentlig sektor, där krav på robusthet, rättssäkerhet och informationssäkerhet är höga.

Utökad AI-understödjande digitisering

Enbart omkring 3-4 procent av KB:s samlade bestånd är digitiserat, dvs överförda till digital form. I Sverige har inte någon större nationell satsning på digitisering av kulturarv genomförts. Återstartsutredningen *Från kris till kraft* (SOU 2021:77) föreslog att KB, RA och Riksantikvarieämbetet (RAÄ) gemensamt skulle tilldelas 800 miljoner kronor över en treårsperiod för en nationell satsning på digitisering av kulturarvet.⁹³ Jämförelsen illustrerar storleksordningen på de investeringar som tidigare har bedömts nödvändiga för att åstadkomma en mer genomgripande uppbyggnad av digital kulturarvsinfrastruktur. Erfarenheter från Norge pekar i samma riktning. Den långsiktiga nationella digitiseringssatsning som genomförts vid NB (se avsnittet om NB) illustrerar hur storskalig digitisering kan fungera som en grundläggande infrastruktur för både forskning och AI-relaterad utveckling, inklusive språkmodeller baserade på nationellt kulturarvsmaterial. KB:s bedömning är att motsvarande ambitionsnivå i Sverige sannolikt skulle kräva betydande investeringar över längre tidshorisonter, särskilt mot bakgrund av omfattningen och heterogeniteten i KB:s samlingar.

Kostnaderna för digitisering av analogt material utgör samtidigt endast en del av de investeringar som krävs för att kulturarvssamlingar ska kunna användas som underlag för träning av språkmodeller. För att digitala datamängder ska vara användbara i AI-relaterade sammanhang krävs omfattande bearbetning, strukturering och kvalitetssäkring. Det handlar därmed inte enbart om att omvandla analogt material till digital form, utan även om att säkerställa att materialet kan återfinnas, tolkas och bearbetas maskinellt på ett tillförlitligt sätt. Som Isof framhöll i dialogen med KB präglas deras samlingar, liksom KB:s, ofta av betydande heterogenitet.⁹⁴ Variationen innebär särskilda utmaningar i relation till

⁹³ SOU 2021:77, *Från kris till kraft: Återstart för kulturen*, 2021, sidorna 28, 250-251.

⁹⁴ Institutet för språk och folkminnen, skrivelse via mejl, inkommen 2026-05-07 (KB 2025-930).

språkmodellutveckling, eftersom datamängderna i många fall kräver omfattande bearbetning innan de kan användas för TDM och modellträning. Frågor om datakvalitet och interoperabilitet är viktiga förutsättningar för att kulturarvsdata ska kunna användas effektivt vid utvecklingen av språkmodeller. Isof ser exempelvis ett tydligt behov av gemensam nationell infrastruktur och stärkt samordning mellan berörda myndigheter och aktörer.

En ytterligare kostnadsdrivande faktor utgör kompetensförsörjningen. Tillgången till specialistkompetens är redan i dag en kritisk förutsättning för KB:s arbete med språkmodeller för forskningsändamål, och behovet bedöms öka ytterligare vid en eventuell breddning av verksamheten. Utveckling, träning och förvaltning av språkmodeller förutsätter kompetens inom språkteknologi, AI, maskininlärning, dataförvaltning och digital forskningsinfrastruktur, men även inom områden såsom metadatahantering, storskalig digitisering, juridik, informationssäkerhet och långtidslagring. Samtidigt är tillgången till sådan kompetens begränsad och föremål för betydande nationell konkurrens.

Effektiviseringspotential i offentlig sektor

KB konstaterar att utveckling och förvaltning av språkmodeller är förenade med betydande kostnader, men att det samtidigt finns en betydande potential för effektivitetsvinster inom offentlig sektor genom ett bredare nyttiggörande av KB:s arbete på området. Från ett samhällsekonomiskt perspektiv kan investeringar i språkmodeller och språkteknologiska resurser skapa förutsättningar för effektiviseringar i flera delar av den offentliga förvaltningen, särskilt i verksamheter där stora mängder textbaserad information hanteras och där administrativa, analytiska eller kunskapsintensiva processer i hög grad är beroende av informationsbearbetning. KB:s samlade bedömning är att språkmodeller därför i första hand bör förstås som en möjliggörande teknik för förbättrad informationshantering och kunskapsbearbetning snarare än som ett verktyg för generell automatisering av offentliga verksamheter.

Inhämtade kunskapsunderlag från myndigheter visar samstämmigt att de mest efterfrågade användningsområdena i offentlig sektor rör förbättrad åtkomst till intern information, mer ändamålsenlig hantering av omfattande dokumentmängder samt stöd för analys- och utredningsarbete. Flera myndigheter beskriver betydande praktiska svårigheter att lokalisera relevant information i stora och fragmenterade informationsmiljöer där dokumentation återfinns i olika format, äldre system och separata lagringslösningar. Detta gäller exempelvis Skatteverket och Finansinspektionen, som i skrivelser till utredningen framhåller att relevanta kunskapsunderlag ofta är spridda över flera tekniska miljöer, vilket innebär att information som redan finns i digital form i praktiken kan vara svår att återfinna och nyttiggöra i det löpande arbetet.⁹⁵ Även E-hälsomyndigheten beskriver betydande utmaningar kopplade till att identifiera relevant intern information, särskilt i verksamheter där stora mängder styrande dokument, kunskapsunderlag och analysmaterial behöver användas samlat.⁹⁶ Språkmodeller i

⁹⁵ Finansinspektionen, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930); Skatteverket, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

⁹⁶ E-hälsomyndigheten, skrivelse via mejl, inkommen 2026-04-22 (KB 2025-930).

sådana miljöer kan bidra till väsentliga effektivitetsvinster genom semantisk sökning, strukturering av information, dokumentjämförelser och sammanfattningar som gör befintlig kunskap mer sökbar, begriplig och användbar.

Den inhämtade informationen visar samtidigt att effektiviseringspotentialen i hög grad återfinns inom analysintensiva verksamheter där stora mängder kvalitativ information behöver bearbetas, kategoriseras och förstås. SCB illustrerar detta exempel särskilt tydligt.⁹⁷ Myndigheten bedömer att språkmodeller kan bidra till effektivisering i flera delprocesser nära statistikproduktionen, bland annat genom extraktion av strukturerade data ur textmaterial, klassificering av fritextsvar och dokument samt identifiering av teman och mönster i kvalitativa datamängder. SCB framhåller även potentialen i automatiserade sammanställningar av forsknings- och metodlitteratur som stöd för kunskapsuppbyggnad och analysarbete. Det här illustrerar hur språkmodeller kan minska tidsåtgången för kunskapsbearbetning och samtidigt stärka kvaliteten i analys- och utredningsarbetet genom att stora informationsmängder blir mer hanterbara och överblickbara.

Liknande behov framträder i dialogen med Statskontoret och i skrivelsen från Finansinspektionen. Statskontoret beskriver hur myndighetens verksamhet är beroende av tillgång till omfattande samhälls-, policy- och rättsinformation från många olika källor samt hur dagens fragmenterade informationsstruktur innebär betydande administrativa kostnader genom behovet av att utveckla egna lösningar för insamling och strukturering av offentlig information.⁹⁸ Finansinspektionen aktualiserar på motsvarande sätt behovet av språkmodeller som kan stödja analysarbete i regulatoriska miljöer där stora dokumentmängder behöver bearbetas och sättas i relation till rättsliga och institutionella sammanhang.⁹⁹ Språkmodeller med stor förståelse för svenska förvaltningsmiljöer och juridiska kontexter kan därför bidra till mer effektiv informationsbearbetning inom sådana kunskapsintensiva verksamheter, särskilt när modellerna används som stöd för analys, informationssyntes och identifiering av relevanta samband i omfattande dokumentmaterial.

En återkommande iakttagelse är vidare att språkmodellers effektiviseringspotential i betydande utsträckning är kopplad till förbättrad intern kunskapshantering. E-hälsomyndigheten beskriver ett omfattande behov av språkmodeller för smart och semantisk sökning i myndighetens samlade informationsmängder, kompletterat med funktioner för dokumentjämförelser, sammanfattningar och strukturering av information.¹⁰⁰ Ekobrottsmyndigheten framhåller på motsvarande sätt behovet av bättre möjligheter att navigera i omfattande och komplex dokumentation som stöd i det dagliga utredningsarbetet, där semantisk sökning och strukturering av stora informationsmängder bedöms kunna minska tidsåtgången för informationsinhämtning och förbättra överblicken i komplexa ärenden.¹⁰¹ Skatteverket lyfter samtidigt fram att ostrukturerad information, skannade handlingar och dokument spridda över flera lagringsmiljöer i dag utgör ett praktiskt hinder för att hitta rätt

⁹⁷ Statistiska centralbyrån, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

⁹⁸ Dialog KB-Statskontoret, online, 2026-04-29.

⁹⁹ Finansinspektionen, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

¹⁰⁰ E-hälsomyndigheten, skrivelse via mejl, inkommen 2026-04-22 (KB 2025-930).

¹⁰¹ Ekobrottsmyndigheten, skrivelse via mejl, inkommen 2026-05-05 (KB 2025-930).

information i rätt tid.¹⁰² Dessa behov sammantaget pekar mot att anpassade språkmodeller inom offentlig sektor sannolikt kommer innebära stora besparingar.

Underlagen visar samtidigt att offentlig sektors effektiviseringspotential är kopplade till språkmodeller som är tydligt källförankrade och integrerade med verksamhetsspecifika kunskapsmiljöer snarare än fristående generativa system. Skatteverket framhåller särskilt värdet av språkmodeller som kombineras med relevanta kunskapsbaser, rättsliga vägledningar och styrande dokument, där svar kan härledas till verifierbara informationskällor. Finansinspektionen aktualiserar liknande behov genom att efterfråga lösningar där språkmodeller kombineras med myndighetsspecifika dokumentmiljöer för att stärka precisionen i analys- och tillsynsarbetet. Även E-hälsomyndigheten betonar behovet av språkmodeller som kan användas tillsammans med myndighetens samlade kunskapsunderlag för att förbättra träffsäkerheten i informationssökning och analysarbete. Språkmodeller ger sannolikt störst nytta när de används som stöd för informationsbearbetning, sökning och analys i tydligt avgränsade arbetsprocesser där resultaten kan granskas, verifieras och sättas i relation till etablerade kunskapskällor

Samtidigt som den inhämtade informationen visar på betydande effektiviseringsmöjligheter aktualiserar de också en mer grundläggande avvägning mellan effektivitet, rättssäkerhet och tillit i offentlig sektor. De potentiella vinsterna i form av minskad administration, förbättrad informationsåtkomst och stärkt analyskapacitet behöver vägas mot krav på rättssäkerhet, informationssäkerhet, metodologisk kvalitet och demokratisk legitimitet. Flera myndigheter framhåller att användning av språkmodeller inom offentlig verksamhet ställer särskilt höga krav på transparens och kontroll eftersom felaktiga slutsatser eller otillräckligt förankrade svar kan få rättsliga, ekonomiska eller förtroendemässiga konsekvenser.

Ekobrottsmyndigheten illustrerar särskilt tydligt de krav som uppstår i verksamheter där känslig information hanteras och där analyser eller bedömningar kan få betydande rättsliga och ekonomiska konsekvenser. Myndigheten framhåller behovet av språkmodeller som kan användas i kontrollerade och säkerhetsanpassade miljöer där åtkomst till data styrs utifrån verksamhetsbehov, säkerhetsklassning och rättsliga förutsättningar. Samtidigt aktualiserar myndigheten risker kopplade till automatiserad kommunikation där språkmodeller används för att generera svar till användare eller externa aktörer. Särskilt framhålls risken för att svar kan bli felaktiga, sakna tillräcklig relevans eller inte i tillräcklig grad beakta den specifika kontext som krävs i enskilda ärenden. Liknande perspektiv framkommer i Skatteverkets redogörelse, och pekar särskilt på riskerna med övertillit till språkmodeller inom myndighetskommunikation och serviceverksamhet. Myndigheten framhåller att en betydande andel inkommande frågor är återkommande och därmed potentiellt lämpar sig för språkbaserat stöd i kommunikationen med medborgare och företag. Samtidigt betonas betydande risker kopplade till hallucinationer, alltför vaga svar eller felaktiga formuleringar som kan påverka myndighetens trovärdighet och leda till missvisande vägledningar.

¹⁰² Skatteverket, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

Upphovsrättsliga aspekter vid KB:s utveckling av språkmodeller

KB har i delredovisningen redogjort för sin bedömning av de upphovsrättsliga förutsättningarna för den utveckling av språkmodeller som myndigheten bedrivit. Av delredovisningen framgår att KB ser att myndigheten kan tillämpa 15 b § URL vid TDM för forskningsändamål.

Flera upphovsrättsorganisationer framför att rättsläget är osäkert och de har en oro för vilka konsekvenser utvecklingen av språkmodeller som grundar sig i deras medlemmars data får på sikt.¹⁰³ Flera av organisationerna har även skickat in så kallade förbehåll till KB, vilket är möjligt att göra enligt 15 a § URL. Givet att KB uteslutande tillämpar 15 b § URL ser myndigheten att dessa förbehåll inte har någon verkan, eftersom förbehåll inte kan lämnas om TDM sker för forskningsändamål inom ramen för 15 b § URL.

AI-utvecklingen i ett större perspektiv innebär ett omfattande och omdanande teknikskifte för inte minst kreatörer och utgivare. Ett teknikskifte som också innebär att processer för utgivning med mera kan förändras och effektiviseras. I detta sammanhang är det viktigt att betona att KB:s bedömning helt och hållet avser KB:s roll som forskningsstödjande myndighet med uppdrag att bidra till svensk forsknings kvalitet och behandlar inte vad den bredare AI-utvecklingen kan innebära för berörda sektorer eller aktörer.

Rättsliga utlåtanden av sakkunniga

KB har sedan delredovisningen inhämtat utlåtanden från externa jurister med expertis inom upphovsrätt och EU-rätt. KB har ställt följande tre frågor till de sakkunniga:

1. Såvitt gäller den utveckling av språkmodeller som vilar på vetenskaplig grund och som beskrivs delredovisning, gör ni en annan tolkning av 15 b § URL än den vilken KB redovisat för regeringen (se ovan nämnda delredovisning)?
2. Har det någon betydelse för tillämpningen av 15 b § URL vilken slags språkmodell som forskningen avser?
3. Förutsatt att det sker inom ramen för forskningsprojekt, i KB:s forskningsinfrastruktur, leds av disputerad forskare och har en forskningsfråga, är ni av uppfattningen att KB kan tillämpa 15 b § URL, såsom redogjorts för i delredovisningen, även för uppdragsforskning (oaktat om det rör grundforskning/tillämpad forskning)?

De rättsliga utlåtandena har inhämtats från sakkunniga med särskilt hög akademisk och praktisk kompetens inom upphovsrätt och EU-rätt.

En av de sakkunniga är Jan Rosén, professor emeritus i civilrätt vid Stockholms universitet, som under flera decennier haft en central och normbildande roll inom svensk upphovsrätt.

¹⁰³ Kungl. biblioteket, justerade mötesanteckningar mellan KB och Läromedelsförfattarna, 2025-11-26 (KB 2025-930); Tidningsutgivarna, skrivelse inkommen via mejl, 2025-12-16 (KB 2025-930).

Jan Rosén har bland annat varit regeringens särskilde utredare i upphovsrättsliga frågor och är författare och redaktör till flera ledande verk och lagkommentarer på området.

De två övriga sakkunniga är Henrik Bengtsson, advokat och delägare vid advokatfirman Delphi, med omfattande erfarenhet av upphovsrätt och annan immaterialrätt i kvalificerad juridisk rådgivning och tvistlösning, samt Helena Andersson, jur. dr och delägare vid samma byrå, med särskild expertis inom EU-rätt.

Svar på de ställda frågorna har skriftligen inkommit till KB och bifogas denna utredning, se Bilaga 3 och 4.

De båda utlåtandena ger ett tydligt och samstämmigt stöd för KB:s rättsliga bedömning i delredovisningen avseende myndighetens tillämpning av 15 b § URL vid utveckling av språkmodeller för forskningsändamål. Av utlåtandena framgår vidare att tillämpningen av bestämmelsen inte är beroende av vilken typ av språkmodell som forskningen avser – lagstiftningen är teknikneutral. Vidare framhålls att tillämpad forskning och uppdragsforskning kan bedrivas med stöd av inskränkningen, bland annat under förutsättning att forskningsresultaten sedermera kommer allmänheten till del. När det gäller användning av forskningsresultaten, dvs. när man lämnat tillämpning av 15 b § URL, så understryks att om framtagna språkmodell inte innehåller eller kan generera exemplar av skyddade verk så finns inget behov av rättighetsklarering inför ett tillgängliggörande.

Sammantaget visar sakkunnigutlåtandena att de upphovsrättsliga frågeställningar som aktualiseras av KB:s utveckling av språkmodeller för forskningsändamål har analyserats ingående och vilar på ett väl underbyggt rättsligt underlag.

Rättsläget

KB konstaterar att det inom EU alltså finns begränsad tillgång till rättspraxis på aktuellt område. I Sverige har, vad KB känner till, tillämpningen av 15 b eller 15 c §§ URL inte varit föremål för rättslig prövning.

I slutet av februari 2026 publicerade EU-parlamentet en rapport¹⁰⁴ varigenom det lyfts fram ett antal viktiga aspekter man anser att EU-kommissionen bör beakta framåt. Rapporten är inte bindande men kan anses utgöra en politisk riktning, som skulle kunna fungera som underlag till framtida lagförslag. Rapporten behandlar generativ AI, och berör bland annat behov kring ökad transparens och stärkta rättigheter för skapare.

Av relevans för KB som forskningsinfrastruktur är att det i punkt fyra i rapporten lyfts fram en uppmaning till EU-kommissionen att säkerställa att, i linje med principerna i artikel 13 i EU:s stadga om de grundläggande rättigheterna (bland annat akademisk frihet), verksamheter som bedrivs för vetenskapliga forsknings- eller utbildningsändamål inte begränsas, särskilt

¹⁰⁴ https://www.europarl.europa.eu/doceo/document/A-10-2026-0019_EN.html

vad gäller arbete av forskningsorganisationer och kulturarvsinstitutioner, eller inom ramen för icke-kommersiell innovation.

Samtidigt som rapporten i stort uppmanar till tydligare reglering och stärkta rättigheter för upphovspersoner, så understryker EU-parlamentet således betydelsen av att gällande rätt i fråga om inskränkningen för TDM i forskningssyfte består. Detta med hänsyn till respekten för akademisk frihet med dess fundamentala betydelse för innovation och dess relevans för samhället i stort. Det får här erinras om att det bakomliggande skälet till att det finns inskränkningar i upphovsrätten för forskningsändamål är akademisk frihet som en fundamental rättighet och att denna särställning har kommit till uttryck i flera andra lagstiftningar.

Rapporten utgör, enligt KB:s uppfattning, ett tydligt EU-politiskt stöd för att utveckling av språkmodeller på vetenskaplig grund och av kulturarvsinstitutioner ska kunna fortsätta med stöd av art. 3 DSM-direktivet och såsom det genomförts i 15 b och 15 c §§ URL.

Vägen framåt för KB

KB:s bedömning av innebörden av gällande rätt vinner stöd av EU-parlamentets rapport och de rättsutlåtanden KB har inhämtat. Innebörden av detta är att KB dels kan fortsätta att bedriva forskning om AI såsom hittills skett, dels utveckla andra språkmodeller och göra detta i form av tillämpad forskning och uppdragsforskning, allt till gagn för en bredare användning av myndighetens språkmodeller för AI i offentlig sektor såsom uttryckt av regeringen enligt regeringsbeslut U2025/01837. Härvid bör noteras att den inskränkning som KB tillämpar för TDM inte är förenad med en ersättningsrätt.

Slutligen vill KB återigen påpeka att utvecklingen av språkmodeller för AI för forskningsändamål aktualiserar flertalet andra rättsområden utöver upphovsrätt, vilka KB har att beakta och följa i verksamheten. Eftersom aktuellt uppdrag omfattar analys av de upphovsrättsliga aspekterna redogörs inte för övriga rättsområden i denna rapport.

Redogöra för KB:s roll i utvecklingen av språkmodeller inom offentlig sektor i förhållande till andra aktörers arbete med att utveckla språkmodeller inom offentlig sektor

Redogörelse för KB:s roll i relation till forskningsinfrastrukturer och excellenta miljöer

I Vetenskapsrådets rapport *Forskning i samverkan* framhålls att samverkan ofta präglas av strukturella och institutionella utmaningar, där skillnader i organisatoriska logiker, regelverk och incitament kan försvåra effektiv användning av gemensamma resurser.¹⁰⁵ Även om rapporten huvudsakligen behandlar samverkan mellan akademi och näringsliv bedömer KB att flera av slutsatserna är direkt relevanta även för utvecklingen av språkmodeller och AI-relaterad infrastruktur inom offentlig sektor. Utveckling av språkmodeller förutsätter typiskt sett tillgång till omfattande datamängder, språkteknologisk kompetens, högpresterande beräkningskapacitet och tekniska miljöer för träning, utvärdering och förvaltning, det vill säga resurser som sällan finns samlade hos en enskild aktör. Vetenskapsrådets analyser av e-infrastruktur för forskning pekar samtidigt på att fragmentering, projektbaserad finansiering och otydlig ansvarsfördelning riskerar att hämma långsiktig utveckling och effektivt nyttiggörande av gemensamma resurser.¹⁰⁶ Liknande utmaningar kan identifieras även inom utvecklingen av språkmodeller mellan olika nationella forskningsinfrastrukturer, där bristande samordning riskerar att leda till parallella investeringar, begränsad återanvändning och ökat beroende av internationella infrastrukturer.

KB:s roll bör i första hand ses som infrastrukturell och möjliggörare snarare än som utförare av AI-relaterad spetsforskning. KB:s uppdrag enligt instruktionen är huvudsakligen att skapa åtkomst till samlingar och dataresurser under reglerade former, även om graden av faktisk tillgänglighet påverkas av upphovsrättsliga, integritetsrelaterade och tekniska förutsättningar. KB ska skapa rättsligt, tekniskt och organisatoriskt hållbara former för forskning och modellutveckling genom reglerad åtkomst till dataresurser. I de fall rättsliga eller tekniska förutsättningar begränsar möjligheterna att dela träningsdata aktualiseras i stället behovet av infrastrukturer där analys, träning och utvärdering kan ske i kontrollerade miljöer nära data så som sker på KB-labb.

Utredningens dialog med Vetenskapsrådet visar också på vikten av strategiska frågor om långsiktig förvaltning och nyttiggörande av data och AI-relaterade forskningsresurser.¹⁰⁷ Vetenskapsrådet framhöll särskilt behovet av hållbara former för dokumentation,

¹⁰⁵ Vetenskapsrådet, *Forskning i samverkan: Drivkrafter, utmaningar och möjligheter för samverkan inom grundforskning mellan akademi och näringsliv*, 2025, sidorna 14 och 16.

¹⁰⁶ Vetenskapsrådet, *Ökad samordning och organisatorisk förändring av e-infrastruktur för forskning Svar på regeringsuppdrag* (dnr. 1.1.2-2024-06310), 2025.

¹⁰⁷ Dialog KB-Vetenskapsrådet, online, 2026-04-23.

versionshantering och återanvändning av data över tid samt utmaningar kopplade till kompetensförsörjning och rättsliga osäkerheter. I dialogen aktualiserades även betydelsen av infrastrukturer där forskning kan bedrivas genom kontrollerad åtkomst till känsliga eller rättsligt begränsade datamängder, snarare än genom fri överföring av underliggande material. Myndigheten betonade även behovet av infrastrukturer som kan svara mot olika forskningsändamål och möjliggöra användning av språkmodeller i skilda forskningskontexter. Dessa frågor är även relevanta utanför forskningssektorn där användningen av språkmodeller inom offentlig sektor aktualiserar motsvarande krav på rättssäkerhet, spårbarhet och möjlighet till uppföljning av resultat över tid.

Chalmers tekniska högskola betonar i sin skrivelse till utredningen liknande aspekter som de Vetenskapsrådet förmedlar,¹⁰⁸ det vill säga behovet av tydlig dokumentation kring träningsdata, licensvillkor, filtreringsmetoder, finjustering, begränsningar och utvärderingsmetoder för att möjliggöra vetenskaplig granskning och ansvarsfull användning. Vikten av metadata, versionshantering och interoperabilitet i linje med etablerade FAIR-principer framhålls också. Särskilt reproducerbarhet anges som en central princip för forskningsrelaterad AI-infrastruktur, där benchmarktester, väldokumenterade gränssnitt och tillgång till modeller och dataset för verifiering lyfts fram som viktiga komponenter. Avseende träningsdata och forskningsinfrastruktur framhåller dessutom Chalmers och Linnéuniversitet behovet av långsiktig tillgång till kvalitetssäkrade svenskspråkiga dataset, korpusar och metadata i strukturerad form.¹⁰⁹

Ett tydligt behov som uttrycks av bland annat Chalmers och Högskolan Dalarna är kontrollerbar, transparent och granskningsbar AI-infrastruktur som kan integreras i akademiska forskningsmiljöer.¹¹⁰ Samarbeten med andra forskningsinfrastrukturer har varit, och är, viktiga för uppbyggnaden av KB:s nuvarande kapacitet, och dess vidareutveckling, inom språkmodellområdet. Ett pågående samarbete med forskningsinfrastrukturen Språkbanken Text vid Göteborgs universitet, som omtalades tidigare (se sidan 27) stärker exempelvis förutsättningarna för bearbetning, dokumentation och kvalitetssäkring av träningsdata i linje med principer om reproducerbarhet, vetenskaplig transparens och metodologisk spårbarhet. Samarbetet omfattar även utveckling av metoder för dataförädling, versionshantering och modellutvärdering, vilket är betydelsefullt för kvalitet och uppföljning över tid.

På motsvarande sätt kan NAISS utgöra en grundläggande förutsättning för träning och utvärdering av större språkmodeller genom tillgång till högpresterande beräkningsresurser om än KB specifikt i stället har utnyttjat de möjligheter som givits via EuroHPC JU. För denna utredning kontaktades därför NAISS i egenskap av en central nationell aktör för utveckling av AI och språkmodeller. Särskilt fokus låg på frågor kopplade till EU:s datastrategi för perioden 2025–2030, där etableringen av datalabb inom ramen för de AI-

¹⁰⁸ Chalmers tekniska högskola, skrivelse via mejl, inkommen 2026-05-26 (KB 2025-930).

¹⁰⁹ Chalmers tekniska högskola, skrivelse via mejl, inkommen 2026-05-26 (KB 2025-930); Linnéuniversitetet, skrivelse via mejl, inkommen 2026-05-27 (KB 2025-930).

¹¹⁰ Chalmers tekniska högskola, skrivelse via mejl, inkommen 2026-05-26 (KB 2025-930); Högskolan Dalarna, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

fabriker som utvecklas utgör en viktig komponent, samt på hur AI-fabriken Mimer arbetar för att ge forskare och företag tillgång till dataset för AI-utveckling. Via Mimer erbjuder NAISS stöd till forskare och organisationer som bedriver forskning och utveckling. Stödet omfattar bland annat tillgång till avancerade beräkningsresurser samt stöd i hantering, bearbetning och kurering av data för modellutveckling. För att tydliggöra KB:s möjliga roll i relation till NAISS efterfrågades information om hur KB:s resurser och kompetenser skulle kunna bidra till att möta de behov som forskare och andra användare av NAISS beräkningsinfrastruktur har vid utveckling av språkmodeller. Mot bakgrund av KB:s uppdrag att analysera hur verksamheten kan bidra till att stödja offentlig sektor ställdes även frågor om i vilken utsträckning de språkmodeller som utvecklas inom det pågående samarbetet mellan NAISS och WASP skulle kunna användas, vidareutvecklas eller anpassas för olika verksamhetsbehov inom offentlig sektor.¹¹¹ Efter att NAISS bekräftat mottagandet av KB:s frågor inkom dock inga ytterligare svar till utredningen.

I relation till Vinnovas uppdrag att stödja samverkansbaserade excellensmiljöer inom strategiska teknikområden framstår tillgång till högkvalitativa data som en grundläggande förutsättning för utveckling av konkurrenskraftiga språkmodeller.¹¹² Även om KB:s samlingar utgör en unik nationell resurs genom sin omfattning, språkliga bredd och dokumenterade proveniens bör KB:s roll återigen främst beskrivas som stödjande och möjliggörande snarare än som utförare av AI-relaterad spetsforskning. KB:s främsta bidrag bedöms ligga i att skapa infrastrukturella förutsättningar genom språkresurser, dataförvaltning, metadata, forskningsstöd och samverkan med forskningsmiljöer och tekniska infrastrukturer.

Redogörelse för KB:s roll i relation till andra kulturarvsmyndigheter

Relationen mellan kulturarvsmyndigheter och utveckling av språkmodeller i offentlig sektor aktualiserar i stor utsträckning frågor om digitisering, dataförvaltning och institutionell samverkan. I budgetpropositionen kopplas uppdragen till KB och RA samman genom betoningen av digitisering som en strategisk förutsättning för forskning och utveckling av AI-relaterade tillämpningar. Den snabba utvecklingen inom beräkningsteknik och AI bedöms samtidigt ha ökat betydelsen av digitiserade kulturarvsresurser som grund för språkteknologisk utveckling. Tillgång till stora, representativa och väldokumenterade datamängder framstår därmed som avgörande för möjligheten att utveckla språkmodeller med relevans för svenska språkliga, kulturella och institutionella förhållanden. Begränsad digitisering av kulturarvssamlingar riskerar samtidigt att påverka såväl forskningens möjligheter som förutsättningarna att utveckla nationellt anpassade AI-lösningar.

Dialogen med Isof visar att myndigheten redan bedriver ett omfattande arbete med att digitisera, bearbeta och tillgängliggöra språk- och kulturarvsmaterial med relevans för språkteknologisk utveckling. Ett konkret exempel är samarbetet kring tal-till-text-modellen

¹¹¹ <https://www.naiss.se/news/2026/02/naiss-and-wasp-in-joint-project-to-train-quality-swedish-llm/>

¹¹² Regeringen, Klimat- och näringslivsdepartementet, *Uppdrag till Verket för innovationssystem att genomföra insatser för samverkansbaserad forskning och innovation inom excellenskluster för banbrytande och strategisk teknik* (KN2025/01161).

KB-Whisper (se sidan 36), där Isof bidrog med dialektinspelningar för att förbättra möjligheterna till automatisk transkribering och ökad sökbarhet i ljudsamlingar. Samtidigt framhåller Isof att digitisering i AI-relaterade sammanhang innebär betydligt mer än att konvertera material till digital form.¹¹³ För att kulturarvsdata ska kunna användas vid språkmodellutveckling krävs även strukturering, metadatahantering, rättighetsbedömningar, kvalitetssäkring och kontextualisering.

Dialogen med Isof pekar på betydande potential i AI-baserade metoder för automatisk transkribering, handskriftsigenkänning, metadataextraktion, ämnesklassificering, semantisk sökning och analys av stora materialmängder.¹¹⁴ Samtidigt framhåller Isof att denna potential är beroende av gemensamma infrastrukturer för språk- och kulturarvsdata som stödjer standardiserade format, metadatahantering, API-baserad åtkomst, versionshantering och olika nivåer av dataåtkomst. För att kulturarvsdata långsiktigt ska kunna nyttiggöras inom forskning och språkmodellutveckling bedöms behovet av semantisk interoperabilitet vara särskilt betydelsefullt.

Dialogerna med Isof och RA aktualiserade även frågor som går utöver aspekter kopplade till upphovsrätt och teknisk dataåtkomst.¹¹⁵ Samlingsförvaltande myndigheter hanterar omfattande material där känsligt material kan förekomma i digitiserade texter, metadata och andra dokumentationsformer. När data används i gemensamma utvecklingsmiljöer uppstår därför behov av tydliga principer för riskbedömningar, åtkomstkontroll, anonymisering och informationssäkerhet. Gemensamma institutionella ramverk för datastyrning och ansvarsfrågor kommer sannolikt att få ökad betydelse i takt med att kulturarvsdata används för språkmodellutveckling. För kulturarvsinstitutioner med omfattande datamängder innebär detta behov av tydliga principer för åtkomst och användning, särskilt när material delas mellan olika aktörer eller används för nya ändamål.

Den föreslagna lagen om interoperabilitetskrav för datadelning inom offentlig förvaltning¹¹⁶ bör därför inte enbart förstås som en fråga om teknisk kompatibilitet mellan olika system. Minst lika viktigt är hur data definieras, dokumenteras, förstås och tolkas på ett enhetligt sätt mellan olika verksamheter, myndigheter och informationsmiljöer. Interoperabilitet handlar därmed inte enbart om att möjliggöra informationsutbyte, utan även om att säkerställa att information kan användas och tolkas konsekvent över organisatoriska och tekniska gränser. I detta sammanhang blir förståelsen för variationer i metadata, terminologi, klassificeringsprinciper och dokumentationsstrukturer särskilt betydelsefull. Olika myndigheter och verksamheter förvaltar data utifrån skilda uppdrag, regelverk och administrativa logiker, vilket kan påverka hur information beskrivs och struktureras.

¹¹³ Institutet för språk och folkminnen, skrivelse via mejl, inkommen 2026-05-07 (KB 2025-930).

¹¹⁴ Dialog KB-Isof, online, 2026-05-06.

¹¹⁵ Dialog KB-RA, fysiskt, 2026-05-04.

¹¹⁶ Regeringen, Finansdepartementet, *Nya krav på interoperabilitet vid datadelning inom den offentliga förvaltningen* (Prop. 2025/26:244).

Redogörelse för KB:s roll i relation till rätts- och säkerhetsvårdande myndigheter

Utvecklingen och användningen av språkmodeller inom rätts- och säkerhetsrelaterade verksamheter behöver förstås i relation till frågor om datadelning, informationssäkerhet, nationell säkerhet och långsiktig digital motståndskraft. Det föreslagna regelverket om interoperabilitetskrav för datadelning inom offentlig förvaltning innehåller exempelvis bestämmelser som begränsar skyldigheten att tillämpa sådana krav när det är olämpligt med hänsyn till Sveriges säkerhet eller när ett tillräckligt skydd för cybersäkerhet och personuppgifter inte kan säkerställas.¹¹⁷ För samlingsförvaltande myndigheter såsom KB aktualiseras därmed behovet av former för datadelning som möjliggör nyttiggörande av kulturarvs- och språkdata samtidigt som krav på informationssäkerhet och kontrollerad åtkomst kan upprätthållas. KB:s roll i detta sammanhang bör ses som stödjande och infrastrukturell snarare än som en aktör med operativa uppgifter inom säkerhets- eller underrättelseverksamhet.

Genom Vinnovas uppdrag att etablera och stödja excellenskluster inom strategisk och banbrytande teknik tydliggörs också sambandet mellan AI-utveckling, nationell säkerhet, teknologisk konkurrenskraft och strategisk autonomi. Inom EU betonas på motsvarande sätt behovet av egen AI-kapacitet för forskning och vetenskapliga ändamål, bland annat för att minska beroendet av ett begränsat antal globala teknikaktörer och stärka kontrollen över kritisk digital infrastruktur.¹¹⁸ Myndigheten för totalförsvarsanalys (MTFA) poängterar i sitt inspel till utredningen att språkmodeller bör förstås som en central komponent i samhällets informationshantering och därmed som en fråga med relevans för både cybersäkerhet och nationell säkerhet.¹¹⁹ MTFA pekar särskilt på risker kopplade till begränsad insyn i träningsdata och modellarkitektur samt osäkerhet kring långsiktig tillgång till internationella AI-tjänster, vilket enligt myndigheten kan skapa särskilda sårbarheter i verksamheter med höga krav på robusthet, spårbarhet och kontroll. Liknande perspektiv framförs av Myndigheten för psykologiskt försvar (MPF), som lyfter att tillgång till språkmodeller tränade på svensk data bör betraktas som en del av den nationella digitala infrastrukturen, särskilt mot bakgrund av myndigheters behov av tillförlitliga och transparenta språkteknologiska resurser.¹²⁰ KB:s bedömning är att dessa perspektiv sammantaget stärker slutsatsen att investeringar i svenska språkresurser och nationellt förvaltade språkmodeller har betydelse inte enbart för forskning och offentlig förvaltning, utan även för Sveriges och Europas samlade digitala motståndskraft.

De inhämtade informationsunderlagen från rätts- och säkerhetsvårdande myndigheter aktualiserar samtidigt en strategisk avvägning mellan nationell autonomi, resurseffektivitet och behovet av teknologisk konkurrenskraft. Polismyndigheten betonar att Sverige, mot

¹¹⁷ Regeringen, Finansdepartementet, *Nya krav på interoperabilitet vid datadelning inom den offentliga förvaltningen* (Prop. 2025/26:244).

¹¹⁸ Europeiska kommissionen, *En europeisk strategi för artificiell intelligens inom vetenskapen Vägen mot en resurs för AI-vetenskap i Europa (Raise)*, COM(2025) 724.

¹¹⁹ Myndigheten för totalförsvarsanalys, skrivelse via mejl, inkommen 2026-05-07 (KB 2025-930).

¹²⁰ Myndigheten för psykologiskt försvar, skrivelse via mejl, inkommen 2026-05-08 (KB 2025-930).

bakgrund av de betydande kostnaderna och den snabba internationella utvecklingstakten, i första hand bör bidra till utvecklingen av konkurrenskraftiga europeiska språkmodeller snarare än att självständigt utveckla svenska grundmodeller från grunden.¹²¹ Myndigheten framhåller samtidigt att beroenden till internationella AI-leverantörer kan innebära strategiska sårbarheter, särskilt om externa aktörer ensidigt förändrar tillgången till modeller eller villkoren för användning. I dialogen med MPF aktualiserades ett delvis likartat men samtidigt mer pragmatiskt perspektiv på nationell språkmodellutveckling. MPF framhåller att det inte framstår som rimligt eller resurseffektivt att offentlig sektor utvecklar helt egenutvecklade språkmodeller från grunden som ska konkurrera med internationella kommersiella modeller i generell kapacitet. I stället lyfter myndigheten möjligheten att vidareutveckla existerande internationella språkmodeller som bas och därefter vidareträna och finjustera dessa med svensk data i säkra miljöer (se tidigare resonemang om vidareträning av internationella modeller sidan 25).

Ett sådant förhållningssätt ligger i flera avseenden nära KB:s nuvarande verksamhet, där språkmodeller utvecklas genom vidareträning och anpassning av befintliga modeller med svenska språkdata. Tack vare KB:s samlingar och kompetens inom språkteknologi, metadatahantering och långsiktig informationsförvaltning finns förutsättningar att bidra med resurser som är särskilt relevanta för utvecklingen av svenska språkmodeller, utan att myndigheten behöver ansvara för operativ utveckling av generella grundmodeller. KB:s bedömning är att denna metod bidrar till en mer kostnadseffektiv och långsiktigt hållbar utveckling av svenska språkmodeller, samtidigt som krav på transparens, spårbarhet och svensk språklig relevans bättre kan tillgodoses. Samtidigt framhåller MTFA vikten av att offentlig sektors utveckling av språkmodeller och språkteknologiska resurser sker i linje med etablerade säkerhetsramverk och nationella strategier. Myndigheten betonar särskilt betydelsen av kontroll över data och modeller, robusthet mot manipulation, skydd mot informationspåverkan samt säker hantering av känslig information.

En gemensam ståndpunkt som lyfts av de rätts- och säkerhetsvårdande myndigheter är att språkmodeller i första hand efterfrågas som stöd för analys, informationsbearbetning och kunskapsunderlag snarare än som ersättning för mänskliga bedömningar.

Ekobrottsmyndigheten lyfter exempelvis ett behov av språkbaserat AI-stöd för att effektivisera utredningsverksamhet och stärka förmågan att bekämpa systemhotande ekonomisk brottslighet. Myndigheten betonar särskilt potentialen i språkmodeller för analys av omfattande dokumentmängder, identifiering av mönster, översättning samt bearbetning av komplexa ekonomiska informationsflöden, inklusive analys av pengaströmmar och omfattande transaktionsunderlag. KB:s bedömning är att detta tydliggör hur språkmodeller i rättsvårdande verksamheter främst efterfrågas som verktyg för att stödja analytiskt arbete i informationsintensiva miljöer där stora mängder strukturerad och ostrukturerad information behöver sammanställas och bearbetas under höga krav på precision och rättssäkerhet. Ekobrottsmyndighetens perspektiv visar också att språkmodeller sannolikt får störst betydelse inom arbetsmoment som dokumentanalys, informationssyntes och kunskapsstöd,

¹²¹ Polismyndigheten, skrivelse via mejl, inkommen 2026-05-13 (KB 2025-930).

snarare än inom verksamheter där automatiserade bedömningar direkt påverkar rättstillämpning eller myndighetsutövning. Liknande behov aktualiseras i dialogen med MTFA, som framhåller betydelsen av språkmodeller för att analysera stora informationsmängder. Myndigheten pekar särskilt på behovet av att kunna bearbeta omfattande text- och ljudbaserade informationsflöden inom säkerhets- och beredskapsområdet, där stora mängder information behöver analyseras och sammanställas under tidspress och i komplexa informationsmiljöer.

Polismyndigheten tydliggör samtidigt att offentlig sektors behov av språkmodeller inte enbart handlar om generell språklig kapacitet, utan i stor utsträckning om domänspecifik förståelse och specialisering. Myndigheten framhåller ett tydligt behov av vidaretränade och specialiserade modeller anpassade till svenska förhållanden, där förståelse för svenska språkvariationer såsom dialekter, sociolekter, slang och minoritetsspråk kombineras med förmåga att hantera myndighetsspråk och generera text i enlighet med klarspråksprinciper. Polismyndigheten lyfter därutöver fram behovet av mindre och mer specialiserade modeller för specifika funktioner kopplade till utredningsverksamheten. Med andra ord kommer offentlig sektors behov sannolikt inte att kunna mötas genom en enskild generell språkmodell, utan snarare genom ett ekosystem av modeller med varierande funktionalitet, storlek och användningsområden. Detta var något som också kom upp i dialogen med BnF, där de såg en framtid med fler, mer specialiserade och tillämpade modeller för specifika ändamål.

Dialogen med Myndigheten för civilt försvar (MCF) aktualiserar samtidigt språkmodellers betydelse från ett bredare säkerhets-, beredskaps- och totalförsvarsperspektiv.¹²² Myndigheten framhåller att AI och språkmodeller har betydande potential att stödja effektivisering inom riskprediktion, resursoptimering, beslutsstöd och efteranalys vid kriser och samhällsstörningar. Särskilt inom räddningstjänst och katastrofhantering bedöms språkmodeller på sikt kunna bidra till stärkt operativ förmåga genom förbättrad informationsbearbetning och analys av omfattande textbaserade informationsflöden. Samtidigt betonar MCF att språkmodeller inte bör förstås som tekniskt neutrala verktyg, utan som resurser vars funktion påverkas av träningsdata, utvecklingsmiljö och institutionell kontext. MCF understryker att språkmodeller är bärare av språkliga, kulturella och normativa perspektiv, vilket aktualiserar frågor om nationell kontroll över data och språkliga resurser, särskilt i verksamheter där korrekt användning av svenska språket, svensk rättstradition och nationella institutionella förhållanden är centrala.

De inhämtade underlagen visar vidare att språkmodeller också får betydelse för myndigheters möjligheter att identifiera, analysera och förstå informationspåverkan och andra säkerhetsrelaterade fenomen i det digitala informationslandskapet. Såväl MTFA som MPF pekar på behovet av att kunna analysera omfattande informationsmiljöer där stora mängder text- och ljudbaserat material behöver bearbetas för att upptäcka mönster, avvikande informationsflöden och potentiellt koordinerade påverkansaktiviteter. Samtidigt bör språkmodeller i detta sammanhang förstås som analysstöd snarare än som verktyg för

¹²² Myndigheten för civilt försvar, skrivelse via mejl, inkommen 2026-05-08 (KB 2025-930).

självständiga slutsatser eller operativa säkerhetsbedömningar. Det är viktigt att påtala att KB inte ska ikläda sig en säkerhetsvårdande roll med ett operativt uppdrag att bedriva analys eller motverka säkerhetshot via sin verksamhet att utveckla språkmodeller. KB:s bidrag i detta avseende bör bestå i att skapa förutsättningar för andra aktörer att utveckla, pröva och använda sådana metoder genom tillgång till språkresurser, dokumenterade modeller och metodstöd, samtidigt som ansvarsfördelningen mellan myndigheterna upprätthålls.

Behovet av språkteknologisk kapacitet är samtidigt i ökande grad multimodalt och förutsätter mer komplexa analysförmågor än traditionell textbearbetning. MPF framhåller att analys av informationspåverkan och digitala informationsmiljöer i allt större utsträckning förutsätter förmåga att bearbeta och analysera kombinationer av text, ljud, bild och video från digitala plattformar och sociala medier. Liknande behov aktualiseras av Polismyndigheten, som framhåller ett växande behov av modeller som kan hantera flera modaliteter inom ramen för utrednings- och analysverksamhet, särskilt i relation till omfattande och heterogena informationsmiljöer. MCF lyfter dessutom fram vikten av teknisk redundans och mobilitet, där modeller i olika storlekar och med varierande tekniska krav behöver kunna användas lokalt på olika typer av tekniska plattformar i beredskapssyfte.

De framförda synpunkterna till utredningen visar samtidigt att utveckling och användning av språkmodeller inom rätts-, säkerhets- och beredskapsområdet behöver integrera säkerhetsmässiga och rättsliga perspektiv redan i utformningen av modeller, träningsmiljöer och infrastrukturer. MTFA betonar vikten av att offentlig sektors utveckling av språkmodeller och språkteknologiska resurser sker i linje med etablerade säkerhetsramverk och nationella strategier. MPF framhåller på motsvarande sätt att särskilt avancerade modeller med koppling till analys av informationspåverkan och säkerhetsområdet kan kräva särskilda former för åtkomst och användning, där vissa modeller kan behöva utvecklas och användas inom skyddade miljöer. Detta aktualiserar behovet av differentierade former för distribution och användning av språkmodeller inom offentlig sektor, där vissa resurser kan göras öppet tillgängliga medan andra, beroende på träningsdata, kapacitet eller användningsområde, kan behöva omfattas av kontrollerad åtkomst för särskilt behöriga myndigheter.

Behovet av dokumentation och reproducerbarhet är också centrala för att språkmodeller ska kunna användas med höga krav på rättssäkerhet och tillit. MCF framhåller vikten av att språkmodeller som används inom beredskapsrelaterade sammanhang är dokumenterade och möjliga att granska, dels för att säkerställa rättslig grund och efterlevnad av svenska regelverk, dels för att möjliggöra uppföljning och reproducerbarhet över tid. MCF och andra medverkande i utredningen, exempelvis Högskolan i Halmstad, betonar samtidigt behovet av systematiska analyser av bias i såväl träningsdata som modellbeteende för att motverka risker kopplade till stereotypisering, osaklig behandling eller underrepresentation av grupper och minoriteter.¹²³

¹²³ Högskolan i Halmstad, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930); Myndigheten för civilt försvar, skrivelse via mejl, inkommen 2026-05-08 (KB 2025-930).

Digitisering av kulturarvssamlingar och utveckling av svenska språkmodeller behöver dessutom förstås i relation till samhällets beredskap, motståndskraft och totalförsvaret. I rapporten om behovet av beredskapsåtgärder för skydd av kulturarv framhåller Riksantikvarieämbetet (RAÄ) att kulturarvets resurser inte enbart har ett kulturellt eller forskningsrelaterat värde,¹²⁴ utan även betydelse för samhällets motståndskraft och förmåga att hantera kriser och ytterst krig. Tillgång till nationellt förvaltade språkresurser, dokumenterade datamängder och modeller tränade på svenska förhållanden kan bidra till att stärka samhällets informationsberedskap. Ett långtgående beroende av externa och kommersiella system utan tillräcklig insyn eller nationell förankring kan från ett beredskapsperspektiv innebära sårbarheter. KB:s bedömning är därför att investeringar i digitisering av kulturarv också bör förstås som en del av en bredare samhällslig resiliens. MCF:s underlag till utredningen stärker samtidigt bilden av att beroenden till internationella AI-leverantörer kan innebära särskilda risker från ett beredskapsperspektiv. Myndigheten framhåller begränsad insyn i hur data används, osäkerhet kring vilka aktörer som kan få tillgång till information samt begränsade möjligheter att anpassa modeller till verksamhetsspecifika och nationella behov som särskilt betydelsefulla riskområden.

MCF framhåller samtidigt att multimodala AI-system, där text, ljud, bild och video kombineras, sannolikt kommer att få ökad betydelse i beredskaps- och krissammanhang, exempelvis för simuleringar, analys av operativa situationer eller stöd i preventivt civilt försvar. Textbaserade språkmodeller är dock en grundläggande komponent även i mer avancerade multimodala system, vilket understryker betydelsen av långsiktiga investeringar i svenska språkresurser. I takt med att kulturarv och offentliga informationsresurser digitiseras förändras däremot samtidigt riskbilden.

KB:s samverkansroll

Dialogen med Totalförsvarets forskningsinstitut (FOI) visar att det finns ett värde i samordning kring grundläggande metoder och tekniska förutsättningar för säker användning av AI-relaterade teknologier.¹²⁵ I relation till språkmodeller aktualiseras möjligheter till samverkan mellan KB och myndigheter med ansvar för cybersäkerhet, informationssäkerhet, beredskap och analys av informationsmiljöer. Sådan samverkan kan exempelvis avse metodutveckling, tekniska pilotprojekt, kunskapsutbyte eller gemensamma analyser av hur språkmodeller kan användas på ett säkert och ändamålsenligt sätt. Samtidigt behöver sådan samverkan utformas med tydliga ansvarsförhållanden där KB:s roll även i detta sammanhang bör ses som möjliggörande och infrastrukturell snarare än operativ.

Utredningsdirektivet till utredningen *Stärkt skydd för uppgifter som rör forskning och innovation* (Dir. 2026:15) aktualiserar därutöver frågor om informationssäkerhet i strategiska teknikområden där AI, kvantteknik och bioteknik i ökande grad präglas av geopolitisk konkurrens.¹²⁶ Information om forskningsinriktningar, metodutveckling, dataresurser och

¹²⁴ Riksantikvarieämbetet, *Behov av beredskapsåtgärder för skydd av kulturarv, samt en första prioritering - Delredovisning 2 från regeringsuppdraget Rådet för skydd av kulturarv*, 2026, sidan 6.

¹²⁵ Dialogmöte KB-FOI, 2026-03-24.

¹²⁶ Regeringen, kommittédirektiv, *Stärkt skydd för uppgifter som rör forskning och innovation*, Dir. 2026:15.

tekniska framsteg kan därmed få ett säkerhetsstrategiskt värde. Samtidigt vilar svensk offentlig förvaltning på offentlighetsprincipen, vilket innebär att insyn och granskning utgör grundläggande demokratiska värden och centrala förutsättningar för transparens, rättssäkerhet och förtroende för offentliga institutioner. Bedömningen är att AI-baserade analysmetoder förändrar balansen mellan öppenhet och skydd, eftersom även öppna och till synes harmlösa uppgifter kan få ny betydelse när de sammanställs, bearbetas och analyseras i stor skala. Den tekniska utvecklingen möjliggör systematisk kartläggning av forskningsinriktningar, identifiering av kunskapsmiljöer, analys av metodutveckling och upptäckt av potentiella sårbarheter i en omfattning och hastighet som tidigare inte var möjlig. Detta innebär att information som var för sig framstår som begränsat känslig, genom storskalig analys i vissa sammanhang kan få säkerhetsmässig eller strategisk betydelse.

Denna förändrade riskbild aktualiserar behovet av att utveckla samverkansformer för riskbedömning, informationshantering och skydd av forsknings- och innovationsrelaterade uppgifter inom strategiska teknikområden, samtidigt som offentlighetsprincipens grundläggande värden upprätthålls. I relation till språkmodeller aktualiserar detta särskilt frågor om kontroll över träningsdata, dokumentation, åtkomstformer och säker användning av forskningsnära resurser. Mot denna bakgrund framstår språkmodeller utvecklade inom en offentlig och rättsligt reglerad kontext, med dokumenterad proveniens och tydliga styrningsformer, som särskilt betydelsefulla för att möjliggöra transparens, spårbarhet och institutionell kontroll.

Dialogerna med rätts-, säkerhets- och beredskapsmyndigheter accentuerar ett tydligt behov av särskilda miljöer för utveckling, testning och vidareträning av språkmodeller. MCF framhåller att AI-system som används i sammanhang med koppling till nationell säkerhet och civil beredskap i vissa fall kan kräva särskilda tekniska miljöer och former för åtkomstkontroll. Myndigheten uttrycker samtidigt ett tydligt intresse för gemensamma miljöer där offentliga aktörer under säkra former kan vidareträna eller anpassa språkmodeller utifrån verksamhetsspecifika behov. Liknande perspektiv framförs av MPF, som betonar värdet av gemensamma infrastrukturer där myndigheter kan testa, vidareutveckla och anpassa språkmodeller i skyddade miljöer. MPF förordar särskilt etableringen av en sandlådemiljö där relevanta myndigheter tillsammans kan utveckla och vidareträna språkmodeller under former som möjliggör kontroll över data, metodutveckling och säker användning.

Polismyndigheten och MTFA stärker samtidigt denna bild av ett behov av gemensamma och säkra miljöer för vidareträning och anpassning av språkmodeller. Polismyndigheten framhåller att möjligheten att dela vidaretränade modeller samt tränings- och testdata skulle kunna bidra till mer effektiv kapacitetsuppbyggnad och snabbare utveckling av robusta modeller anpassade till svenska behov. MTFA bedömer på motsvarande sätt att gemensamma miljöer kan vara relevanta under förutsättning att höga krav på informationssäkerhet, åtkomstkontroll och hantering av känslig information kan upprätthållas. I relation till KB:s roll aktualiserar detta möjligheten att myndigheten kan bidra med grundläggande språkresurser, dokumenterade träningsdata och metodkompetens, medan vidare anpassning sker inom särskilt skyddade miljöer i samverkan med berörda myndigheter.

Dessa perspektiv stärker sammantaget argumenten för en differentierad modellstruktur inom offentlig sektor, där vissa språkmodeller kan göras öppet tillgängliga medan andra, beroende på träningsdata, användningsområde eller säkerhetsrelevans, tillgängliggörs under mer kontrollerade former för särskilt behöriga myndigheter. En sådan ordning ligger i linje med proportionalitetsprincipen mellan öppenhet och skyddsbehov och aktualiserar behovet av tydliga styrningsprinciper för distribution, användning och vidareträning beroende på modellernas karaktär och potentiella riskprofil. Samtidigt framhåller Polismyndigheten att anonymisering och pseudonymisering av personuppgifter i vissa sammanhang kan innebära betydande kvalitetsförluster i dataunderlaget, vilket tydliggör behovet av fortsatt rättslig analys av förutsättningarna för att använda känsliga data i kontrollerade utvecklingsmiljöer, särskilt inom säkerhets- och brottsbekämpande verksamheter.

Redogörelse för KB:s roll till andra relevanta myndigheter och KB:s medverkan i en svensk AI-verkstad för offentlig sektor

KB:s medverkan i AI-verkstaden behöver förstås mot bakgrund av myndighetens särskilda kompetens inom språkdata, metadatahantering, digitala samlingar och utveckling av språkmodeller för forskningsändamål.¹²⁷ Språkmodeller utgör i ökande grad en grundläggande komponent i många AI-baserade tillämpningar inom offentlig verksamhet, exempelvis inom textanalys, informationssökning, sammanfattning, klassificering och kunskapsstöd. KB kan genom sin verksamhet bidra till att svenska språkresurser och språkteknologiska komponenter integreras i AI-verkstadens utvecklingsmiljöer.

Finansinspektionens underlag till utredningen illustrerar samtidigt att behoven av språkmodeller inom offentlig sektor varierar beroende på verksamhetsområde, datamiljö och regulatoriska krav.¹²⁸ Finansinspektionen framhåller att myndighetens analysarbete i stor utsträckning redan baseras på digital information, där dokument och handlingar huvudsakligen förekommer i digital form såsom textdokument, pdf-filer och databaser, men där även andra datatyper, exempelvis transaktionsdata och ljudinspelningar från telefonsamtal, används i undersöknings- och analysverksamhet. Utmaningen för AI-verkstaden är därmed kopplad till åtkomst och interoperabilitet snarare än till avsaknad av data. Finansinspektionen framhåller exempelvis att delar av relevant information är spridd mellan olika lagringsytor eller i praktiken svåråtkomlig genom äldre systemmiljöer och otillräcklig metadata.

MPF framhåller samtidigt att offentlig sektor i begränsad utsträckning separat bör utveckla språkmodeller inom varje myndighet eller beredskapssektor. Myndigheten bedömer att resurser, kompetens och teknisk kapacitet i stället behöver samlas hos ett mindre antal aktörer med särskilda uppdrag och infrastrukturella förutsättningar. KB:s anser att detta stärker argumenten för en mer samordnad nationell modell där vissa myndigheter fungerar som noder för språkresurser, data och modellutveckling, samtidigt som verksamhetsnära

¹²⁷ Regeringskansliet, Finansdepartementet, *Budgetpropositionen för 2026*, Utgiftsområde 16, Utbildning och universitetsforskning (Prop. 2025/26:1).

¹²⁸ Finansinspektionen, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

tillämpningar utvecklas i samverkan med andra aktörer. I detta sammanhang framstår KB:s roll som särskilt relevant genom myndighetens ansvar för språkdata, digital kulturarvsförvaltning och utveckling av språkmodeller för forskningsändamål. Samtidigt ligger ansvar för operativa tillämpningar fortsatt hos respektive sektorsmyndighet. En sådan ansvarsfördelning kan bidra till att minska dubbelarbete, stärka interoperabilitet och skapa bättre förutsättningar för långsiktigt hållbar AI-utveckling inom offentlig sektor.

I dialogen med Statskontoret aktualiserades samtidigt ett närliggande perspektiv som rör tillgången till och långsiktig användbarhet av offentlig information som underlag för AI-relaterad utveckling.¹²⁹ Statskontoret framhåller att det i dag saknas ett sammanhållet system för beständig och långsiktig digital åtkomst till centrala publikationer från staten, såsom regleringsbrev, instruktioner, regeringsuppdrag, budgetunderlag, årsredovisningar, rapporter och rättsinformation. Fragmenteringen i hur sådan information publiceras och tillgängliggörs innebär enligt myndigheten att betydande resurser behöver läggas på att lokalisera, samla in och strukturera data från ett stort antal olika källor. Denna typ av strukturella hinder kan även få betydelse för utvecklingen av språkmodeller och andra AI-relaterade tillämpningar inom offentlig sektor. När offentlig information är spridd, varierande dokumenterad eller saknar beständig åtkomst försvåras möjligheterna att utveckla språkmodeller med hög kvalitet och god förståelse för svenska offentliga verksamheter.

Statskontoret framhåller vidare betydelsen av språkmodeller som har en god förståelse för svensk förvaltning, svensk juridik och offentliga beslutsprocesser. Enligt myndigheten stärks sådana förmågor dels genom träning på relevanta svenska dataresurser, såsom offentliga dokument, rättspraxis och förarbeten, dels genom att sådana material finns tillgängliga som kontextuell kunskapsbas vid användning av modeller. Statskontoret efterfrågar även starkare svenska språkmodeller för semantisk sökning, särskilt i situationer där modeller används för att identifiera relevant kontextuell information i mer avancerade och agentiska arbetsflöden. Myndigheten framhåller ett stort behov av mindre och mer specialiserade språkmodeller med stark domänkunskap om svenska samhällsförhållanden och offentlig verksamhet. I dialogen aktualiserades även att internationella språkmodeller, såväl öppna som proprietära, i vissa fall tenderar att utgå från anglosaxiska rättstraditioner även när frågor ställs på svenska.

Dialogen med Myndigheten för tillgängliga medier (MTM) illustrerar samtidigt de rättsliga begränsningar som kan påverka samverkan kring data för språkmodeller.¹³⁰ MTM har bedömt att de i sin verksamhet vanligen inte kan tillämpa 15 b § URL för TDM i samband med språkmodellträning då myndighetens digitala bestånd huvudsakligen består av produktioner som skett med stöd av 17, 17 a och 17 e §§ URL. Detta begränsar användningsområdet för beståndet till de ändamål som följer av 17, 17 a och 17 e §§ URL. Skillnader i rättsliga förutsättningar mellan myndigheter påverkar med andra ord möjligheterna till myndighetssamverkan vid utveckling av svenska språkmodeller.

¹²⁹ Dialog KB-Statskontoret, online, 2026-04-29.

¹³⁰ Dialog KB-MTM, 2026-05-06.

Den rapport som Försäkringskassan och Skatteverket presenterade i januari 2026 visar, i likhet med tidigare analyser från SCB,¹³¹ att behovet av gemensam infrastruktur, kompetensförsörjning och samordning alltjämt är betydande inom offentlig sektor. Trots ett växande intresse för AI och ökade investeringar i digital utveckling kvarstår utmaningar kopplade till organisatorisk mognad, teknisk kapacitet och tillgång till gemensamma resurser. Samtidigt bör KB:s roll avgränsas tydligt i förhållande till AI-verkstanen och andra myndigheters uppdrag. KB har inte mandat eller organisatoriska förutsättningar att utveckla, drifta och tillhandahålla fullständiga AI-tjänster för offentlig sektor i bred mening. Ansvar för verksamhetsnära implementering, gemensamma digitala tjänster och förvaltningsövergripande samordning ligger närmare de strukturer som utvecklas genom uppdragen till DIGG, Post- och telestyrelsen (PTS),¹³² Försäkringskassan och Skatteverket. KB kan bidra med språkresurser, träningsdata, modeller, dokumentation och metodstöd, medan utveckling av operativa tjänster och användning i myndigheters processer bör ske inom respektive ansvarig myndighets uppdrag.

E-hälsomyndigheten framhåller samtidigt att de största hindren för ökad användning av språkmodeller inte primärt är tekniska, utan snarare juridiska, säkerhetsmässiga och organisatoriska. Osäkerhet kring rättsliga förutsättningar, särskilt i relation till allmänt tillgängliga molnbaserade tjänster utanför Sverige och EU, tillsammans med höga krav på informationssäkerhet, påverkar myndighetens möjligheter att nyttja språkmodeller fullt ut. Därtill aktualiseras frågor om kompetensförsörjning och organisatorisk omställningsförmåga. Dessa perspektiv ligger nära de utmaningar som identifierats även av andra myndigheter och stärker argumenten för gemensamma, rättsligt hållbara och säkerhetsmässigt kontrollerade infrastrukturer för offentlig sektors AI-användning.

Inom ramen för det utökade anslaget på 10 miljoner kronor som KB tilldelats för medverkan i AI-verkstanen bedömer KB att myndigheten kan bidra med ett antal avgränsade men strategiskt viktiga leveranser. Ett första område är tillgängliggörande och vidareutveckling av specialiserade svenska språkmodeller och språkresurser via uppdragsforskning. Det kan exempelvis avse modeller eller komponenter för svensk språkförståelse, tal-till-text, textklassificering, informationsutvinning och semantisk sökning. Sådana resurser kan användas i AI-verkstansens utvecklingsmiljöer och minska behovet av att varje myndighet bygger upp egna grundläggande språkmodeller från början.

Ett andra område gäller hantering av träningsdata. KB kan bidra genom kvalitetssäkrade och dokumenterade datamängder, metadata och språkteknologiska komponenter som stödjer utvecklingen av modeller anpassade till svenska förhållanden. Inom ramen för ett begränsat anslag bör detta dock främst avse pilotinsatser, kontrollerade åtkomstformer och vidareutveckling av befintliga resurser, inte etablering av en fullskalig nationell

¹³¹ Försäkringskassan och Skatteverket, *Svar på regeringsuppdraget att utreda förutsättningarna för en AI-verkstad för den offentliga förvaltningen* (Fi 2025/01523), FK 2025/016953 / SKV 8-8583-2026, sidan 12.

¹³² Regeringen, Finansdepartementet, *Uppdrag till Myndigheten för digital förvaltning att föreslå en samverkansstruktur för utveckling av tjänster för data och artificiell intelligens* (Fi2026/00137); Regeringen, Finansdepartementet, *Uppdrag till Myndigheten för digital förvaltning och Post- och telestyrelsen att stödja genomförandet av Sveriges AI-strategi* (Fi2026/00398).

språkmodellinfrastruktur. KB:s bedömning är att satsningen på 10 miljoner kronor bör ses som ett steg mot ökad samordning och nyttiggörande av redan etablerade resurser.

Ett tredje område gäller metodstöd och expertis. KB har byggt upp en avsevärd kompetens inom språkteknologi, maskininlärning, digitala humaniora, metadatahantering och dataförvaltning. Denna kompetens kan bidra till AI-verkstadens arbete med gemensamma arbetssätt för datakvalitet, dokumentation, spårbarhet, modellutvärdering och språkrepresentation. En särskild styrka är att KB kan bidra till att svenska språkliga variationer, kulturarvsdimensioner och krav på transparens beaktas i utvecklingen av AI-relaterade lösningar. KB vill framhålla att detta är särskilt viktigt för att offentlig sektors AI-användning ska bli långsiktigt hållbar, rättssäker och anpassad till svenska förvaltningsförhållanden.

AI-verkstaden bör samtidigt inte förstås som en ersättning för befintliga infrastrukturer eller myndighetsinitiativ, utan som en samordnande och förstärkande struktur. I Försäkringskassans och Skatteverkets redovisning framhålls betydelsen av samverkan med nationella och europeiska initiativ.¹³³ KB anser att medverkan bör utformas i linje med detta, genom att myndigheten bidrar med de språkresurser och den kompetens som redan byggts upp inom KB:s forskningsinfrastruktur. På så sätt kan KB stödja AI-verkstaden att bidra till att överbrygga avståndet mellan forskning, dataförvaltning och praktisk användning i offentlig sektor.

Sammantaget bedömer KB att sin roll i förhållande till andra relevanta myndigheter bör vara att fungera som en nationell nod för språkresurser, språkdata och språkteknologisk kompetens. En sådan roll kan bidra till att minska fragmentering, stärka interoperabilitet och säkerställa att svenska språkmodeller utvecklas med tillräcklig kvalitet, dokumentation och långsiktig förvaltning. Genom medverkan i AI-verkstaden kan KB bidra till att offentlig sektor får bättre tillgång till öppna, transparenta och nationellt anpassade språkresurser, samtidigt som utveckling av verksamhetsnära AI-tjänster fortsatt sker inom andra myndigheters ansvar.

¹³³ Försäkringskassan och Skatteverket, *Svar på regeringsuppdraget att utreda förutsättningarna för en AI-verkstad för den offentliga förvaltningen* (Fi 2025/01523), FK 2025/016953 / SKV 8-8583-2026, sidan 58.

Bilaga 1: Genomförda dialoger och inhämtad information

KB har under arbetets gång inhämtat information från följande organisationer och personer:

Organisationer

Advokatfirman Delphi
Bibliothèque nationale de France, Frankrike
Blekinge tekniska högskola
Chalmers tekniska högskola
Cybercampus/KTH
E-hälsomyndigheten
Ekobrottsmyndigheten
Finansinspektionen
Försäkringskassan
Göteborgs universitet
Högskolan Dalarna
Högskolan i Borås
Högskolan i Halmstad
Högskolan Kristianstad
Institutet för språk och folkminnen
Karlstads universitet
Kungl. Tekniska högskolan
Linnéuniversitetet
Lunds universitet
Läromedelsförfattarna
Mittuniversitetet
Myndigheten för civilt försvar
Myndigheten för digital förvaltning
Myndigheten för tillgängliga medier
Myndigheten för psykologiskt försvar
Myndigheten för totalförsvarsanalys
Mälardalens universitet
Nasjonalbiblioteket, Norge

Polismyndigheten
Post- och telestyrelsen
Riksarkivet
Skatteverket
Statistiska centralbyrån
Statskontoret
Stockholms universitet
Sveriges Författarförbund
Sveriges Tidskrifter
Svenska Förläggareföreningen
Sveriges lantbruksuniversitet
Säkerhetspolisen
Södertörns högskola
Tidningsutgivarna
Totalförsvarets forskningsinstitut
Uppsala universitet
Vetenskapsrådet

Personer

Professor emeritus Jan Rosén
Produktivitetskommissionens huvudsekreterare Kristian Seth

Inom utredningen har dessa externa dialoger genomförts

AI-verkstad (Försäkringskassan/Skatteverket), online, 2025-03-15, 2026-01-26, 2026-03-06
Bibliothèque nationale de France, Frankrike, online, 2026-05-19
Institutet för språk och folkminnen, online, 2026-05-06
Kungliga bibliotekets forskningsråd, hybrid, 2026-03-11
Kungliga bibliotekets insynsråd, fysiskt, 2025-11-20, 2026-03-12, 2026-05-20
Läromedelsförfattarna, fysiskt, 2025-11-26
Myndigheten för digital förvaltning, fysiskt, 2026-03-11, online 2026-04-07
Myndigheten för tillgängliga medier, online, 2026-05-06
Nasjonalbiblioteket, Norge, online, 2026-04-09, 2026-04-13
Riksarkivet, fysiskt, 2026-05-04
Statskontoret, online, 2026-04-29

Svenska Förläggareföreningen, fysiskt, 2025-11-24
Sveriges Författarförbund, fysiskt, 2025-12-08
Sveriges Tidskrifter, fysiskt, 2026-01-23
Tidningsutgivarna, fysiskt, 2025-12-01
Totalförsvarets forskningsinstitut, fysiskt, 2026-03-24
Vetenskapsrådet, online, 2026-04-23

Bilaga 2: Källor

AI4LAM, <https://ai4lam.org/>, kontrollerad 2026-05-26.

ArGiMi, <https://huggingface.co/argimi>, kontrollerad 2026-05-25.

Blekinge tekniska högskola, skrivelse via mejl, inkommen 2025-05-25 (KB 2025-930).

BnF DataLab, <https://www.bnf.fr/fr/bnf-datalab>, kontrollerad 2026-05-26.

Carlin, Marie och Arnaud Laborderie, "Le BnF DataLab, un service aux chercheurs en humanités numériques", *Humanités numériques*, 4, 2021:

<https://doi.org/10.4000/revuehn.2684>.

CENL, *AI in Libraries Network Group*, <https://www.cenl.org/networkgroups/ai-in-libraries-network-group/>, kontrollerad 2026-04-20.

CENL, *Digital Cultural Heritage or: E-Legal Deposit in European National Libraries*, 2025.

Finansinspektionen, *Remissvar: AI-kommissionens Färdplan för Sverige (SOU 2025:12)*, FI dnr 25-7411.

Chalmers tekniska högskola, skrivelse via mejl, inkommen 2026-05-26 (KB 2025-930).

Chang et al., *When Is Multilinguality a Curse? Language Modeling for 250 High- and Low-Resource Languages*. EMNLP 2024.

CSPLA, Rapport de mission sur la rémunération des contenus culturels utilisés par les systèmes d'IA (2025).

Danish Foundation Models, *Building the Foundation of Danish AI*, <https://www.foundationmodels.dk/>, kontrollerad 2026-05-05.

Det Kgl. Bibliotek, *Cultural Heritage Data Guides - Python*, <https://kb-dk.github.io/cultural-heritage-data-guides-Python/intro.html>, kontrollerad 2026-05-05.

Det Kgl. Bibliotek, *Cultural Heritage Data Guides - R*, <https://github.com/kb-dk/cultural-heritage-data-guides-R>, kontrollerad 2026-05-05.

Det Kgl. Bibliotek, *Home LOAR*, <https://loar.kb.dk/home>, kontrollerat 2026-05-05.

Det Kgl. Bibliotek, *Vi viser vej til viden*, <https://www.kb.dk/om-os/opgaver-og-maal/strategi>, kontrollerad 2026-04-20.

Diskrimineringsombudsmannen, *Transparens, träning och data: Myndigheters användning av AI och automatiserat beslutsfattande samt kunskap om risker för diskriminering*, Rapport 2022:1.

E-hälsomyndigheten, skrivelse via mejl, inkommen 2026-04-22 (KB 2025-930).

Ekobrottsmyndigheten, skrivelse via mejl, inkommen 2026-05-05 (KB 2025-930).

Europaparlamentets och rådets direktiv (EU) 2019/790 av den 17 april 2019 om upphovsrätt och närstående rättigheter på den digitala inre marknaden och om ändring av direktiven 96/9/EG och 2001/29/EG.

Europeiska kommissionen, *En EU-strategi för data*, COM(2020) 66.

Europeiska kommissionen, *En europeisk strategi för artificiell intelligens inom vetenskapen Vägen mot en resurs för AI-vetenskap i Europa (Raise)*, COM(2025) 724.

Finansinspektionen, *Remissvar: AI-kommissionens Färdplan för Sverige* (SOU 2025:12), FI dnr 25-7411.

Finansinspektionen, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

Försäkringskassan och Skatteverket, *Svar på regeringsuppdraget att utreda förutsättningarna för en AI-verkstad för den offentliga förvaltningen* (Fi 2025/01523), FK 2025/016953 / SKV 8-8583-2026.

Gallica, <https://gallica.bnf.fr/accueil/fr/html/accueil-fr>, kontrollerad 2026-03-30.

Göteborgs universitet, skrivelse via mejl, inkommen 2025-05-20 (KB 2025-930).

Högskolan Dalarna, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

Högskolan i Borås, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

Högskolan i Halmstad, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

Högskolan Kristianstad, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

INA och BnF, *Guide juridique sur l'utilisation des données patrimoniales dans le cadre de l'intelligence artificielle generative* (2026).

Institutet för språk och folkminnen, skrivelse via mejl, inkommen 2026-05-07 (KB 2025-930).

Karlstads universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

KB Labs, <https://labs.kb.dk/>, kontrollerad 2026-05-05.

Kulturministeriet, Danmark, Ekspertgruppe om ophavsret og kunstig intelligens, *Anbefalinger fra Ekspertgruppe om ophavsret og kunstig intelligens*, 2025.

Kungl. biblioteket, *Analys för en bredare användning av Kungl. bibliotekets språkmodeller i svensk offentlig sektor: Delredovisning av regeringsuppdrag enligt regeringsbeslut U2025/01837* (KB 2025-930), 2026. <https://www.kb.se/om-oss/nyheter/nyhetsarkiv/2026-01-16-delredovisning-av-uppdrag-om-sprakmodeller-i-offentlig-sektor.html>, kontrollerad 2026-05-26.

Kungl. biblioteket, *Fördjupat samarbete med Språkbanken* (KB 2026-293).

Kungl. biblioteket, justerade mötesanteckningar mellan KB och Läromedelsförfattarna, 2025-11-26 (KB 2025-930).

Kungl. biblioteket, *Nationella riktlinjer för öppen vetenskap* (KB 2024-42).

Kungl. biblioteket, *Speech recognition for Swedish using wav2vec 2.0, Whisper and Conformers* (KB 2023-717).

Lag (1960:729) om upphovsrätt till litterära och konstnärliga verk.

Lag (1993:1392) om pliktexemplar av dokument.

Lag (2012:492) om pliktexemplar av elektroniskt material.

Linnéuniversitetet, skrivelse via mejl, inkommen 2026-05-27 (KB 2025-930).

Longpre et al., *ATLAS: Adaptive Transfer Scaling Laws for Multilingual Pretraining, Finetuning, and Decoding the Curse of Multilinguality*. ArXiv preprint, 2025.

Lunds universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

Mejlkontakt med Produktivitetskommissionens huvudsekreterare Kristian Seth, 13 april 2026 (dnr. 2025-930).

Ministerie van Economische Zaken en Klimaat, *Strategic Action Plan for Artificial Intelligence*, 2019.

Mittuniversitetet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

Myndigheten för civilt försvar, skrivelse via mejl, inkommen 2026-05-08 (KB 2025-930).

Myndigheten för psykologiskt försvar, skrivelse via mejl, inkommen 2026-05-08 (KB 2025-930).

Myndigheten för tillgängliga medier, *Remiss av SOU 2021:32 Papper, poddar och... Pliktmateriallagstiftning för ett tryggt källmaterial*, MTM 2021/371.

Myndigheten för totalförsvarsanalys, skrivelse via mejl, inkommen 2026-05-07 (KB 2025-930).

Mälardalens universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

Nasjonalbiblioteket, *Digitaliseringsplan for den norske dokumentbaserte kulturarven*, 2020. https://abmdig.no/content/uploads/sites/14/2020/08/Digitaliseringsplan-for-ABM-omr%C3%A5det_.pdf, kontrollerad 2026-05-25.

Nasjonalbiblioteket, *Mimir-prosjektet. Evaluering av virkningen av opphovsrettsbeskyttet materiale på generative store språkmodeller for norske språk* (2024), https://www.nb.no/content/uploads/2024/08/Mimirprosjektet_teknisk-rapport.pdf, kontrollerad 2026-05-13.

Nasjonalbiblioteket, *Nasjonalbiblioteket inngår historisk avtale med landets aviser om innhold til kunstig intelligens*, <https://www.nb.no/pressemeldinger/nasjonalbiblioteket-inngar-historisk-avtale-med-landets-aviser-om-innhold-til-kunstig-intelligens/>, kontrollerad 2026-04-20.

NorLLM, *Prosjekt «MIMIR» evaluerer verdien av opphavsrett*, <https://www.ntnu.no/norllm/prosjekt-mimir-evaluerer-verdien-av-opphavsrett>, kontrollerad 2026-05-06.

Norsk rikskringkasting (NRK), *Samisk språk løftes i nye KI-modeller*, <https://www.nrk.no/sapmi/samisk-sprak-loftes-i-nye-ki-modeller-1.17701717>, kontrollerad 2026-05-06.

OECD, *Competition in Artificial Intelligence Infrastructure*, OECD Roundtables on Competition Policy Papers, No. 330, 2025.

OECD, *Going Digital to Advance Data Governance for Growth and Well-being*, 2022.

Regeringen, Finansdepartementet, *Data – en underutnyttjad resurs för Sverige: En strategi för ökad tillgång av data för bl.a. artificiell intelligens och digital innovation*, bilaga till beslut II 5 vid regeringssammanträde den 20 oktober 2021 (I2021/02739).

Regeringen, Finansdepartementet, *Uppdrag till Försäkringskassan och Skatteverket att etablera en AI-verkstad för den offentliga förvaltningen* (Fi2026/00018).

Regeringen, Finansdepartementet, *Uppdrag till Myndigheten för digital förvaltning att föreslå en samverkansstruktur för utveckling av tjänster för data och artificiell intelligens* (Fi2026/00137).

Regeringen, Finansdepartementet, *Uppdrag till Myndigheten för digital förvaltning och Post- och telestyrelsen att stödja genomförandet av Sveriges AI-strategi* (Fi2026/00398).

Regeringen, Infrastrukturdepartementet, *Uppdrag att kartlägga användningen av artificiell intelligens respektive analys av stora datamängder i Sverige* (I2019/01964/D).

Regeringen, Klimat- och näringslivsdepartementet, *Uppdrag till Verket för innovationssystem att genomföra insatser för samverkansbaserad forskning och innovation inom excellenskluster för banbrytande och strategisk teknik* (KN2025/01161).

Regeringen, kommittédirektiv, *Stärkt skydd för uppgifter som rör forskning och innovation*, Dir. 2026:15.

Regeringen, Finansdepartementet, *Nya krav på interoperabilitet vid datadelning inom den offentliga förvaltningen* (Prop. 2025/26:244).

Regeringen, Utbildningsdepartementet, *Uppdrag till Kungl. biblioteket att analysera och förbereda för en bredare användning av myndighetens språkmodeller i svensk offentlig sektor* (U2025/01837).

Regeringskansliet, Finansdepartementet, *Budgetpropositionen för 2026*, Utgiftsområde 16, Utbildning och universitetsforskning (Prop. 2025/26:1).

Regeringskansliet, Finansdepartementet, *Budgetpropositionen för 2026*, Utgiftsområde 17, Kultur, medier, trossamfund och fritid (Prop. 2025/26:1).

Regeringskansliet, Finansdepartementet, *Sveriges Digitiseringsstrategi 2025–2030*, 2025.

Regeringskansliet, *Forskning och innovation för framtid, nyfikenhet och nytta* (Prop. 2024/25:60).

Regeringskansliet, *Handlingsplan för Sveriges AI-strategi*, 2026.

Regeringskansliet, *Sveriges AI-strategi*, 2026.

Regeringskansliet, Utbildningsdepartementet, *Översyn av regelverket för pliktleverans och hantering av annat material vid Kungl. biblioteket* (2019:84).

Regeringen, Kultur- og likestillingsdepartementet, *Endringer i åndsverkloven mv. (gjennomføring av nett- og videresendingsdirektivet og digitalmarkedsdirektivet) og samtykke til godkjenning av EØS-komiteens beslutning nr. 332/2023 og 333/2023 om innlemmelse i EØS.-avtalen av direktiv (EU) 2019/789 og (EU) 2019/790, og samtykke til tiltredelse til WIPO-traktatene av 20. Desember 1996 om opphavsrett (WCT) og om fremføringer og fonogrammer (WPPT)*, Prop. 41 LS (2025 – 2026).

Regjeringen, Kultur- og likestillingsdepartementet, *Statsbudsjettet 2024 - tildelingsbrev til Nasjonalbiblioteket* (23/392-7).

Regjeringen, Kultur- og likestillingsdepartementet, *Statsbudsjettet 2025 - tildelingsbrev til Nasjonalbiblioteket* (23/5753).

Regjeringen, Kultur- og likestillingsdepartementet, *Statsbudsjettet 2026 - tildelingsbrev til Nasjonalbiblioteket* (25/6156).

Riksantikvarieämbetet, *Behov av beredskapsåtgärder för skydd av kulturarv, samt en första prioritering - Delredovisning 2 från regeringsuppdraget Rådet för skydd av kulturarv*, 2026.

Riksarkivet, *Årsredovisning 2025 för Riksarkivet*, 2026.

Skatteverket, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

SOU 2021:32, *Papper, poddar och ... Pliktmateriallagstiftning för ett tryggt källmaterial*, 2021.

SOU 2021:77, *Från kris till kraft: Återstart för kulturen*, 2021.

SOU 2025:12, *AI-kommissionens Färdplan för Sverige*, 2025.

SOU 2025:96, *Fler möjligheter till ökat välbefinnande*, Slutbetänkande av Produktivitetskommissionen, 2025.

Statistiska centralbyrån, *AI-användning i företag och offentlig sektor*, 2023.

Statistiska centralbyrån, *Förhöjd innovation – En mikrodataanalys om AI och innovation*, 2024.

Statistiska centralbyrån, skrivelse via mejl, inkommen 2026-05-11 (KB 2025-930).

Statskontoret, *Myndigheterna och AI: En studie om möjligheter och risker med att använda AI i statsförvaltningen*, skriftserien Om offentlig sektor, 2024.

Stockholms universitet, skrivelse via mejl, inkommen 2025-05-21 (KB 2025-930).

Sveriges lantbruksuniversitet, skrivelse via mejl, inkommen 2025-05-20 (KB 2025-930).

Södertörns högskola, skrivelse via mejl, inkommen 2025-05-26 (KB 2025-930).

Tidningsutgivarna, skrivelse inkommen via mejl, 2025-12-16 (KB 2025-930).

Tryckfrihetsförordning (1949:105).

Uppsala universitet, skrivelse via mejl, inkommen 2025-05-19 (KB 2025-930).

Vesterlund, Amanda, "AI-boomens bränsle är inte Digitiserat", *Dagens industri*, <https://www.di.se/debatt/ai-boomens-bransle-ar-inte-Digitiserat/>, kontrollerat 2026-04-28.

Vetenskapsrådet, *Forskning i samverkan: Drivkrafter, utmaningar och möjligheter för samverkan inom grundforskning mellan akademi och näringsliv*, 2025.

Vetenskapsrådet, *Forskningsbarometern 2025: Svensk forskning i internationell jämförelse* (dnr. 3.5-2025-07677), 2025.

Vetenskapsrådet, *Ökad samordning och organisatorisk förändring av e-infrastruktur för forskning Svar på regeringsuppdrag* (dnr. 1.1.2-2024-06310), 2025.

Vinge, Simon and Maja Fjaestad, "Introduction: From Algocracy to Democracy", *Algorithmic rule: AI and the Future of Democracy in Sweden and Beyond* (2025).

Bilaga 3: Advokatfirman Delphis utlåtande

Delphi

PM

Till:	Kungliga biblioteket
Från:	Advokatfirman Delphi KB, advokat, jur. dr. Helene Andersson och advokat Henrik Bengtsson
Datum:	8 maj 2026
Angående:	Upphovsrättsliga aspekter av Kungliga bibliotekets träning av språkmodeller

1. Inledning och bakgrund

- 1.1 Kungliga biblioteket ("KB") genomför för närvarande en utredning inom ramen för ett regeringsuppdrag. KB har uppdragits att analysera och förbereda för en bredare användning av myndighetens AI-språkmodeller i offentlig sektor och att redogöra för de upphovsrättsliga aspekterna av att utveckla sådana modeller. En delredovisning lämnades i januari 2026, där myndigheten redogjorde för sin tillämpning av upphovsrättslagen i förhållande till den verksamhet som i dag bedrivs för forskningsändamål.
- 1.2 Slutredovisningen ska lämnas senast den 1 juni 2026. Den har ett framåtblickande perspektiv och kommer bland annat att behandla de upphovsrättsliga frågorna kring en möjlig breddning av KB:s AI-utveckling för att möta behov inom offentlig sektor.
- 1.3 Inför slutredovisningen har KB identifierat ett behov av att komplettera sin analys med externa rättsliga utlåtanden. Dessa utlåtanden avser de upphovsrättsliga aspekterna av en eventuell bredare användning av myndighetens AI-språkmodeller i offentlig sektor. Advokatfirman Delphi ("Advokatfirman Delphi" eller "vi") har fått i uppdrag att ta fram ett sådant utlåtande.

2. Uppdrag

- 2.1 Uppdraget omfattar följande frågor:
 - (a) Skiljer sig Delphis tolkning av 15 b § upphovsrättslagen från den som KB tidigare redovisat för regeringen, avseende utveckling av språkmodeller på vetenskaplig grund?
 - (b) Har valet av språkmodell betydelse för tillämpningen av 15 b § upphovsrättslagen?
 - (c) Kan KB tillämpa 15 b § upphovsrättslagen även för uppdragsforskning förutsatt att det sker inom ramen för KB:s forskningsinfrastruktur och leds av en disputerad forskare?

3. Sammanfattning

- 3.1 Vår övergripande bedömning är att KB ingår i den avnämmerkrets som avses i 15 b § upphovsrättslagen¹ och artikel 2.3 i Europaparlamentets och rådets direktiv (EU) 2019/790

¹ Lag (1960:729) om upphovsrätt till litterära och konstnärliga verk.

(DSM-direktivet), varför KB är berättigat att utföra text- och datautvinning (TDM) för forskningsändamål på det material det har lovlig tillgång till.

- 3.2 Delphi gör inte en annan tolkning av 15 b § upphovsrättslagen än den KB redovisat för regeringen. KB:s TDM-verksamhet sker för forskningsändamål i den mening som avses i 15 b § upphovsrättslagen och samtliga förutsättningar för inskränkningens tillämplighet är härigenom uppfyllda. Denna rätt är, i enlighet med 15 b § tredje stycket upphovsrättslagen, tvingande; avtalsvillkor som inskränker KB:s rätt till text- och datautvinning saknar därmed verkan.
- 3.3 Begreppet "forskningsändamål" i 15 b § upphovsrättslagen är teknikneutralt och tar sikte på verksamhetens vetenskapliga ändamål snarare än den specifika modelltyp som används. Den typ av språkmodell som forskningen avser är därmed som utgångspunkt utan avgörande självständig betydelse för tillämpligheten av bestämmelsen, förutsatt att utvinningen sker för vetenskapliga forskningsändamål.
- 3.4 Avgörande för tillämpligheten av 15 b § är att den text- och datautvinning som KB utför sker för forskningsändamål, det vill säga att utvinningen inte vidtas i annat syfte än att möjliggöra vetenskaplig forskning. Förutsatt att forskningen uppfyller kraven på vetenskaplig grund, däribland att dess integritet upprätthålls och att resultaten inte exklusivt reserveras för enskilda intressen, är det vår bedömning att KB kan tillämpa 15 b § upphovsrättslagen även för uppdragsforskning.

4. Undantaget för text- och datautvinning i 15 b § upphovsrättslagen

- 4.1 Bestämmelsen i 15 b § upphovsrättslagen genomför artikel 3 i DSM-direktivet och ger vissa kategorier av institutioner rätt att utan rättighetshavarens samtycke utföra text- och datautvinning på verk de har lovlig tillgång till, förutsatt att detta sker för forskningsändamål.
- 4.2 För att undantaget ska vara tillämpligt måste tre kumulativa förutsättningar vara uppfyllda: (i) den aktör som utför utvinningen måste tillhöra den avnämmarkrets som anges i bestämmelsen, (ii) utvinningen måste ske för forskningsändamål i bestämmelsens mening, och (iii) institutionen måste ha lovlig tillgång till det material som bearbetas.
- 4.3 Vad gäller det tredje kriteriet, lovlig tillgång, utgår detta utlåtande, i enlighet med uppdragsbeskrivningen, från att det material som bearbetas antingen har inhämtats med stöd av lagen (2012:492) om pliktexemplar av elektroniskt material eller lagen (1993:1392) om pliktexemplar av dokument (pliktlagstiftningen), eller att KB på annat sätt har lovlig tillgång till verken. Någon närmare analys av detta kriterium görs därför inte i det följande.
- 4.4 Det kan dock noteras att KB:s pliktexemplarssamlingar utgör en central del av det material som är aktuellt i detta sammanhang och att KB:s rätt att inneha dessa verk har en direkt och tydlig rättslig grund i pliktlagstiftningen. Mot denna bakgrund är det vår bedömning att förutsättningen om lovlig tillgång är uppfylld i förhållande till det material som avses med uppdraget.
- 4.5 I det följande analyseras de två återstående kriterierna, avnämmarkrets och forskningsändamål, i tur och ordning i förhållande till KB:s verksamhet.
- 4.6 Av underlaget framgår att de modeller som KB utvecklar och tillämpar fortlöpande kontrolleras så att verk som kan omfattas av upphovsrätt inte ingår i modellerna som sådana och att modellerna inte genererar resultat som gör intrång i upphovsrättsligt skyddade verk. De upphovsrättsliga förfoganden som sker inom ramen för modellerna sker för text-

och datautvinningsändamål och några andra upphovsrättsliga förfoganden som behöver bedömas i detta utlåtande sker således inte.

5. Forskningsorganisation och/eller kulturarvsinstitution?

- 5.1 Det första kriteriet som ska analyseras avser den avnämmarkrets som omfattas av undantaget i 15 b § upphovsrättslagen. I det följande redogörs inledningsvis för hur avnämmarkretsen definieras i DSM-direktivet och i 15 b § upphovsrättslagen, varefter en bedömning görs av huruvida KB träffas av någon av de angivna definitionerna.
- 5.2 **DSM-direktivet**
- 5.3 Artikel 3 i DSM-direktivet ålägger medlemsstaterna att införa ett obligatoriskt undantag från ensamrätten till mångfaldigande och utvinning för forskningsorganisationer och kulturarvsinstitutioner.
- 5.4 Med forskningsorganisationer avses, enligt artikel 2.1, universitet (inklusive deras bibliotek), forskningsinstitut eller andra enheter vars främsta mål är att bedriva vetenskaplig forskning eller utbildning som innefattar sådan forskning. Verksamheten ska bedrivas på icke vinstdrivande grund, genom återinvestering av vinst i forskningen eller inom ramen för ett uppdrag i det allmännas intresse. Vidare ska organisationen vara strukturerad så att kommersiella aktörer med ett avgörande inflytande inte ges privilegierad tillgång till forskningsresultaten.
- 5.5 Enligt artikel 2.3 definieras kulturarvsinstitutioner som offentligt tillgängliga bibliotek och museer, arkiv samt institutioner för film- eller ljudarv.
- 5.6 **Upphovsrättslagen**
- 5.7 I Sverige har dessa bestämmelser implementerats genom 15 b § upphovsrättslagen. Bestämmelsen ger forskningsorganisationer, arkiv, institutioner för film- eller ljudarvet samt offentligt tillgängliga bibliotek och museer rätt att utföra text- och datautvinning för forskningsändamål på verk de har lovlig tillgång till.² En viktig aspekt av 15 b § är dess tvingande natur; enligt bestämmelsens tredje stycke är avtalsvillkor som inskränker denna rätt ogiltiga. För att omfattas av bestämmelsens avnämmarkrets och åtnjuta undantaget är det alltså tillräckligt att KB träffas av en (1) av de angivna definitionerna.
- 5.8 Som framgår av instruktionsförordningen för KB,³ är biblioteket Sveriges nationalbibliotek. KB är enligt bibliotekslagen (2013:801) del av det allmänna biblioteksväsendet och allmänheten ska ges tillgång till dess samlingar.⁴ Biblioteket är således tillgängligt för allmänheten och ingår därmed i en av de i 15 b § angivna avnämmarkretsarna.
- 5.9 Mot denna bakgrund saknas det i princip skäl att pröva om KB även omfattas av andra delar av definitionen. Det kan dock noteras att KB:s verksamhet, som innefattar att samla in, bevara och tillgängliggöra det svenska kulturarvet i form av bl.a. handskrifter, böcker,

² Begreppet "lovlig tillgång" innefattar tillgång via licens, avtal, upphovsrättsliga inskränkningar eller annan rättslig grund, men utesluter material som erhållits i strid med tekniska skyddsåtgärder som rättighetshavaren tillämpat. KB får anses ha lovlig tillgång till materialet i sina samlingar, bland annat genom det pliktexemplarssystem som regleras i lag (2012:492) om pliktexemplar av dokument samt genom de licenser och avtal som biblioteket ingår för att komplettera sina samlingar.

³ Förordning (2008:1421) med instruktion för Kungl. Biblioteket.

⁴ Bibliotekslagen, 1 och 2 §§.

tidningar, musik, tv-program och bilder,⁵ sannolikt utgör ett arkiv i bestämmelsens mening.⁶ Denna bedömning stärks av att KB är utsedd mottagare av pliktexemplar av i princip allt som publiceras i Sverige enligt pliktlagstiftningen; en utpräglad arkivfunktion som direkt svarar mot de uppgifter som kännetecknar en arkivinstitution i direktivets mening. KB omfattas vidare av arkivlagen (1990:782). Av proposition 2013/14:206, s. 43 framgår uttryckligen att KB utgör en institution för film- och ljudarv.⁷

- 5.10 Det är inte uteslutet att KB även skulle kunna anses utgöra en forskningsorganisation i bestämmelsens bemärkelse. Vi har i detta utlåtande dock inte tagit ställning till huruvida vetenskaplig forskning utgör KB:s främsta mål, vilket är ett nödvändigt kriterium för att klassificeras som en forskningsorganisation enligt 15 b § upphovsrättslagen. Frågan saknar dock avgörande betydelse i sammanhanget, eftersom KB redan på annan grund utgör en del av bestämmelsens avnämnskrets.

5.11 **Slutsatser kring avnämnskretsen**

- 5.12 Sammantaget är det vår uppfattning att KB i vart fall är att anse (i) som ett *offentligt tillgängligt bibliotek* respektive (ii) *en institution för film- eller ljudarvet* i den mening som avses i 15 b § upphovsrättslagen och artikel 2.3 i DSM-direktivet, varför KB är berättigat att utföra text- och datautvinning för forskningsändamål på det material det har lovlig tillgång till. Därtill kommer att KB sannolikt också är att betrakta som *ett arkiv* enligt 15 b § upphovsrättslagen. Denna rätt är, i enlighet med 15 b § tredje stycket, tvingande; avtalsvillkor som inskränker KB:s rätt till text- och datautvinning saknar därmed verkan.

6. Vetenskaplig forskning

- 6.1 Det obligatoriska undantaget i 15 b § upphovsrättslagen avser text- och datautvinning som vidtas för *forskningsändamål*. Någon närmare definition av begreppet ges dock inte i bestämmelsen.
- 6.2 Bestämmelsen genomför artikel 3 i DSM-direktivet, som i den svenska språkversionen använder motsvarande formulering. En jämförelse med andra officiella språkversioner av direktivet visar dock att bestämmelsen uttryckligen avser vetenskaplig forskning.⁸ Mot bakgrund av att samtliga språkversioner har samma värde inom EU-rätten och ska tolkas enhetligt, får den svenska formuleringen inte ges en vidare innebörd än dessa. Begreppet "forskningsändamål" i artikel 3 ska således förstås som hänförligt till vetenskaplig forskning. Detta stöds även av direktivets systematik, särskilt definitionen av forskningsorganisation i artikel 2.1. Detsamma bör gälla vid tolkningen av 15 b § upphovsrättslagen.⁹

⁵ <https://www.kb.se/om-oss/kort-om-kb.html>.

⁶ Begreppet "arkiv" ges i skäl 11 i Europaparlamentets och rådets direktiv 2012/28/EU av den 25 oktober 2012 om viss tillåten användning av anonyma verk en bred definition och omfattar bland annat "radio- och televisionsföretag i allmänhetens tjänst och som producerats av dessa" vilket talar för att begreppet "arkiv" ska ges en vid tolkning.

⁷ "Förutom bibliotek, utbildningsanstalter och museer omfattar paragrafen även institutioner för film- eller ljudarvet. Beträffande dessa finns inget krav på att de ska vara tillgängliga för allmänheten. Exempelvis Stiftelsen Svenska Filminstitutet och Kungliga biblioteket utgör sådana institutioner" (vår understrykning).

⁸ Se, till exempel "scientific research" (engelska), "recherche scientifique" (franska) och "wissenschaftliche Forschung" (tyska).

⁹ Detta stöds också av lagens förarbeten som anger att "med forskning avses forskning på vetenskaplig grund och inte allmänt sökande efter kunskap eller information", se prop. 2021/22:278, sid. 222.

- 6.3 Begreppet *vetenskaplig forskning* saknar en enhetlig och uttömmande definition i svensk rätt. Istället får dess innebörd bestämmas genom en samlad tolkning av relevant lagstiftning, internationella riktlinjer och vedertagen akademisk praxis.
- 6.4 I det följande redogörs därför för centrala rättskällor och normer som tillsammans ger vägledning för hur begreppet vetenskaplig forskning bör förstås, inklusive svensk lagstiftning, internationella standarder för forskning och utveckling (FoU) samt riktlinjer på EU-nivå.
- 6.5 **Begreppet i svensk lagstiftning och akademisk praxis**
- 6.6 Högskolelagen (1992:1434) reglerar hur universitet och högskolor ska bedriva sin verksamhet. Lagen anger inte någon definition av vetenskaplig forskning men anger i 1 kap. 6 § att den akademiska friheten ska främjas och värnas och att för forskningen ska som allmänna principer gälla att (i) forskningsproblem fritt får väljas, (ii) forskningsmetoder fritt får utvecklas och (iii) forskningsresultat fritt får publiceras.
- 6.7 Etikprövningslagen (2003:460) som fastställer etiska regler för forskning avseende människor definierar begreppet forskning enligt följande:
- ”Forskning: vetenskapligt experimentellt eller teoretiskt arbete eller vetenskapliga studier genom observation, om arbetet eller studierna görs för att hämta in ny kunskap, och utvecklingsarbete på vetenskaplig grund, dock inte sådant arbete eller sådana studier som utförs endast inom ramen för högskoleutbildning på grundnivå eller på avancerad nivå.”¹⁰*
- 6.8 Forskning och utveckling (FoU) definieras av Vetenskapsrådet som ett kreativt och systematiskt arbete som syftar till att öka kunskapsmängden och hitta nya tillämpningar av befintlig kunskap inom vetenskapens alla fält.¹¹ FoU innefattar därmed grundforskning, tillämpad forskning och utvecklingsverksamhet. För att en aktivitet ska kvalificeras som FoU krävs att den uppfyller vissa kriterier, såsom att vara nyskapande, baserad på originella koncept eller hypoteser, bedrivs systematiskt samt i princip vara reproducerbar.
- 6.9 **OECD:s Frascati Manual**
- 6.10 Vetenskapsrådets beskrivning återspeglar den som anges i OECD:s Frascati Manual. Manualen togs fram första gången 1963 och har sedan uppdaterats regelbundet, senast 2015. Begreppet forskning och utveckling (FoU) definieras där enligt följande: “Research and experimental development (R&D) comprise creative and systematic work undertaken in order to increase the stock of knowledge – including *knowledge of humankind, culture and society* – and to devise new applications of available knowledge.”¹²
- 6.11 Enligt OECD kännetecknas FoU-verksamhet av ett antal gemensamma drag. Fritt översatt ska sådan verksamhet syfta till att uppnå givna mål, vilka kan vara antingen specifika eller mer generella. Verksamheten ska vidare syfta till att generera nya rön, baserade på ursprungliga idéer (och tolkningar av dessa) eller hypoteser. Den ska i hög grad präglas av osäkerhet beträffande det slutliga resultatet, eller åtminstone beträffande den tid och de resurser som krävs för att nå detta. Därtill ska verksamheten vara planerad och

¹⁰ Lag (2003:460) om etikprövning av forskning som avser människor, 2§.

¹¹ Se Vetenskapsrådets webbplats under adressen <https://www.vr.se/analys/svensk-forskning-i-siffror.html> (maj 2026).

¹² Frascati Manual 2015, sid. 44, https://www.oecd.org/content/dam/oecd/en/publications/reports/2015/10/frascati-manual-2015_g1g57dcb/9789264239012-en.pdf.

budgeterad (även när den bedrivs av enskilda individer) samt syfta till att frambringa resultat som antingen kan spridas eller omsättas i praktisk tillämpning, inklusive kommersiell sådan.

6.12 För att en aktivitet ska utgöra FoU-verksamhet måste den enligt OECD uppfylla fem kärnkriterier. Den ska vara:

- (i) nyhetsskapande
- (ii) kreativ
- (iii) osäker
- (iv) systematisk
- (v) överförbar (transferable) och/eller reproducerbar

6.13 Samtliga fem kriterier ska, åtminstone i princip, vara uppfyllda varje gång en FoU-aktivitet genomförs, oavsett om verksamheten bedrivs kontinuerligt eller endast vid enstaka tillfällen.

6.14 Vetenskaplig forskning får anses utgöra ett snävare och mer kvalificerat begrepp än FoU i allmänhet. Medan FoU även kan omfatta tillämpat utvecklingsarbete, förutsätter vetenskaplig forskning typiskt sett att verksamheten inte bara uppfyller ovanstående kriterier utan också bedrivs enligt vetenskapliga metoder, med krav på transparens, prövbarhet och möjlighet till kritisk granskning. OECD:s kriterier kan därför ses som nödvändiga, men inte i sig tillräckliga, för att en verksamhet ska kvalificeras som vetenskaplig forskning i rättslig mening.

6.15 **Begreppets definition i EU:s dataskyddsförordning**

6.16 Vägledning kan vidare sökas i tillämpningen av EU:s dataskyddsförordning, GDPR,¹³ där särskilda regler gäller för vetenskaplig forskning. Europeiska dataskyddsstyrelsen (eng. *European Data Protection Board, EDPB*), ett oberoende EU-organ som ansvarar för att säkerställa en enhetlig tillämpning av GDPR i hela EU, har nyligen enats om riktlinjer för begreppet vetenskaplig forskning.

6.17 Enligt EDPB kännetecknas vetenskaplig forskning av ett metodiskt och systematiskt arbetssätt, efterlevnad av etiska normer, verifierbarhet och transparens, självständighet och oberoende, ett syfte att bidra till allmän kunskap samt potential att tillföra ny kunskap eller nya tillämpningar. Riktlinjerna uppställer sex kriterier som typiskt sett ska vara uppfyllda och som ingår i en samlad bedömning:

- (i) Metodiskt och systematiskt arbetssätt

Forskningen ska bedrivas strukturerat enligt vedertagen metod inom det aktuella forskningsområdet, exempelvis genom forskningsplan, hypotesprövning eller tydligt formulerade mål.

¹³ Europaparlamentets och rådets förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter och om upphävande av direktiv 95/46/EG (allmän dataskyddsförordning).

(ii) Förenlighet med etiska standarder

Verksamheten ska uppfylla relevanta forskningsetiska krav, såsom skydd för deltagare, informerat samtycke där det krävs, ansvarstagande, transparens och tillsyn.

(iii) Verifierbarhet och öppenhet

Resultat, metoder, data och slutsatser ska kunna granskas kritiskt och i princip verifieras. Resultaten bör normalt delas genom publicering eller annan spridning, med hänsyn till sekretess, immaterialrätt och liknande legitima begränsningar.

(iv) Självständighet och oberoende

Forskningen ska bedrivas utan otilldbörig påverkan från externa intressen eller forskarnas egna förutfattade meningar. Forskarna ska ha frihet att formulera frågor, välja metod och publicera resultat. Relevanta kvalifikationer hos forskarna betonas också.

(v) Syfte att bidra till allmän kunskap och samhällsnytta

Forskningen ska syfta till att öka samhällets samlade kunskap och välbefinnande. Kommersiella intressen utesluts inte, men får inte vara det enda syftet.

(vi) Potential att bidra med ny kunskap eller nya tillämpningar

Verksamheten ska ha vetenskapligt värde genom att kunna tillföra ny kunskap, utveckla befintlig kunskap eller använda känd kunskap på nya sätt. Detta kan ofta provas genom extern granskning, finansieringsprövning eller etikprövning.

6.18 EDPB:s riktlinjer är inte rättsligt bindande i sig, men har ett betydande tolkningsvärde vid tillämpningen av GDPR. De uttrycker en gemensam syn hos EU:s tillsynsmyndigheter och beaktas ofta av nationella tillsynsmyndigheter och kan även tillmätas betydelse av domstolar. Riktlinjerna kan därmed fungera som ett viktigt stöd vid tolkning av begreppet vetenskaplig forskning, även om den slutliga rättsliga bedömningen ytterst ankommer på domstol.

6.19 **DSM-direktivet och upphovsrättslagen**

6.20 Vid tolkningen av begreppet vetenskaplig forskning finns det även skäl att beakta den EU-rättsliga regleringen i anslutning till upphovsrätten, särskilt bestämmelserna om text- och datautvinning i artikel 3 i DSM-direktivet, vilka har genomförts i svensk rätt genom 15 b § upphovsrättslagen.

6.21 I detta sammanhang definieras inte endast forskningsverksamheten, utan även vilka aktörer som ska anses utgöra forskningsorganisationer. Av direktivet framgår att sådana organisationer ska ha som främsta mål att bedriva vetenskaplig forskning eller utbildning som innefattar sådan forskning. En central förutsättning är vidare att forskningen bedrivs utan vinstsyfte, eller att eventuell vinst återinvesteras i forskningen, alternativt att verksamheten bedrivs inom ramen för ett uppdrag i det allmännas intresse. Därutöver uppställs krav på att kommersiella aktörer inte får ha ett sådant inflytande över organisationen att de ges privilegierad tillgång till forskningsresultaten.¹⁴

¹⁴ Se bl.a. artikel 2.1 samt direktivets skäl 11 och 12. Skäl 12 är av särskilt intresse eftersom begreppet "vetenskaplig forskning" ges en vid definition: "Termen vetenskaplig forskning bör med avseende på tillämpningen av detta direktiv anses omfatta såväl naturvetenskaplig som humanistisk forskning. Eftersom enheterna är av så

- 6.22 Även om dessa kriterier formellt tar sikte på begreppet forskningsorganisation, ger de uttryck för underliggande principer som även är relevanta vid tolkningen av begreppet forskningsändamål som sådant. Särskilt framträder att vetenskaplig forskning i detta sammanhang förutsätter ett självständigt kunskapssökande syfte, som inte primärt är kommersiellt och där resultaten inte är förbehållna enskilda aktörer.
- 6.23 **Slutsatser kring begreppet vetenskaplig forskning**
- 6.24 Sammanfattningsvis bör vetenskaplig forskning förstås som ett kvalificerat, systematiskt och metodbundet kunskapssökande arbete som bedrivs i enlighet med etablerade vetenskapliga normer, med krav på transparens, prövbarhet och möjlighet till kritisk granskning. Verksamheten ska syfta till att generera ny, generell och verifierbar kunskap och typiskt sett bedrivs under former som säkerställer oberoende samt att resultaten inte exklusivt tillkommer enskilda intressen.
- 6.25 I detta sammanhang bör distinktionen mellan forskning som bedrivs på forskarens eget initiativ och uppdragsforskning uppmärksammas. Vid forskning på eget initiativ väljer forskaren själv frågeställning, metod och inriktning, vilket typiskt sett säkerställer ett högt mått av vetenskapligt oberoende. Vid uppdragsforskning formuleras forskningsuppgiften helt eller delvis av en uppdragsgivare, vilket kan ge upphov till frågor om huruvida forskningen är tillräckligt oberoende och huruvida resultaten exklusivt tillkommer uppdragsgivaren. Dessa omständigheter kan i det enskilda fallet ha betydelse för bedömningen av om verksamheten uppfyller kraven på vetenskaplig forskning i 15 b § upphovsrättslagen.
- 6.26 Så länge forskningen, oavsett om den sker på forskarens eget initiativ eller inom ramen för ett uppdrag, bedrivs enligt de principer som angetts ovan är det i vår mening dock ovidkommande i vilken form den organiseras. Avgörande är att forskningens vetenskapliga integritet upprätthålls och att resultaten inte är föremål för avgörande kontroll av kommersiella företag eller exklusivt reserveras för enskilda kommersiella intressen. Denna förståelse bör läggas till grund vid tolkningen av 15 b § upphovsrättslagen.¹⁵
- 7. KB:s text- och datautvinning för forskningsändamål**
- 7.1 Med den ovan fastställda innebörden av begreppet *vetenskaplig forskning* som utgångspunkt ska det prövas huruvida den TDM som KB utför vid träning av språkmodeller uppfyller kravet på *forskningsändamål* i 15 b § upphovsrättslagen. Forskningen ska utföras på vetenskaplig grund och inte bara utgöras av allmänt sökande efter kunskap eller information.¹⁶

olika slag är det viktigt att ha en gemensam syn på forskningsorganisationer. De bör till exempel, förutom universitet och andra institutioner för högre utbildning och deras bibliotek, även omfatta enheter som forskningsinstitut och sjukhus som bedriver forskning. Trots olika juridiska former och strukturer har forskningsorganisationer i medlemsstaterna rent allmänt det gemensamt att de agerar antingen på icke vinstdrivande grund eller inom ramen för ett uppdrag i det allmännas intresse som erkänns av staten. Ett sådant uppdrag i det allmännas intresse skulle exempelvis kunna återspeglas genom offentlig finansiering eller genom bestämmelser i nationell lagstiftning eller offentliga upphandlingskontrakt. I motsats till detta bör organisationer över vilka kommersiella företag har ett avgörande inflytande som tillåter dessa företag att utöva kontroll på grund av strukturella faktorer, till exempel deras egenskap av aktieägare eller medlemmar, vilket skulle kunna leda till förmånstillträde till forskningsresultaten, inte betraktas som forskningsorganisationer i den mening som avses i detta direktiv.”

Se även prop. 2021/22:278, sid 43.

¹⁵ Jfr skäl 12 sista meningen i direktivet (citrat i fotnot 17).

¹⁶ Prop. 2021/22:278, sid. 222.

- 7.2 Av skäl 12 i DSM-direktivet följer att artikel 3 i samma direktiv ska tolkas brett i disciplinär mening (alla vetenskapsgrenar), men avgränsat i funktionell mening (ändamålet måste vara att generera ny kunskap genom vetenskaplig metodik, inte att utveckla kommersiella produkter). Gränsen för begreppet "vetenskaplig forskning" (som tolkningen av begreppet "forskningsändamål" av logiska skäl måste utgå från) går inte mellan ren och tillämpad forskning, utan mellan å ena sidan forskning som syftar till kunskapsgenerering och å andra sidan verksamhet som syftar till direkt kommersiellt utnyttjande. Trots detta omfattas forskning som inte fullt ut sker i allmänhetens intresse, exempelvis i offentlig-privata partnerskap och offentliga institutioners samarbeten med privata parter när det gäller att utföra text- och datautvinning (se skäl 11 i DSM-direktivet) av begreppet vetenskaplig forskning. Eftersom artikel 3 på detta vis har ett bredare tillämpningsområde än enbart offentligfinansierade forskningsinstitutioner talar det med styrka för att begreppet "forskningsändamål" inte ska ges en snäv tolkning, i synnerhet inte när forskningen bedrivs av en allmännyttig offentligfinansierad institution som KB.
- 7.3 KB bedriver sin TDM-verksamhet i två delvis överlappande roller: dels som utövare av egna interna forskningsprojekt, dels som tillhandahållare av infrastruktur för externa forskare och forskningsprojekt. Det är inte alltid möjligt att från tillgängligt underlag exakt avgränsa vilken roll som är förhärskande i ett enskilt projekt. Denna distinktion saknar emellertid avgörande betydelse för bedömningen. Det avgörande för tillämpligheten av 15 b § är att den TDM som KB utför sker för forskningsändamål, det vill säga att utvinningen inte vidtas i annat syfte än att möjliggöra vetenskaplig forskning. Detsamma gäller oavsett om forskningen bedrivs internt av KB eller av externa forskare som nyttjar KB:s infrastruktur.
- 7.4 Det framgår av tillgängligt underlag att KB endast samarbetar i forskningsprojekt där slutmålet är förenligt med KB:s uppdrag att bidra till den svenska forskningens kvalitet, att projekten alltid leds av en disputerad forskare och att de framtagna modellerna publiceras öppet i enlighet med principerna om öppen vetenskap. Härtill ska beaktas att KB tillämpar en försiktighetsprincip vid distribuering av sina modeller och att de modeller som i dag släpps kan användas på ett meningsfullt sätt enbart för deskription och prediktion, inte för avancerad generering. KB:s TDM-verksamhet syftar således inte till att framställa breda generativa modeller för allmänna ändamål, utan är specifikt anpassad för att möta forskarsamhällets behov och sker därmed för forskningsändamål.
- 7.5 Av underlaget framgår att KB:s forskning och uppdragsforskning inte kontrolleras av kommersiella företag och att kommersiella intressen inte heller har förhandstillgång till forskningsresultaten.
- 7.6 Sammantaget framstår det som välgrundat att KB:s TDM-verksamhet sker för forskningsändamål i den mening som avses i 15 b § upphovsrättslagen och att samtliga förutsättningar för inskränkningens tillämplighet härigenom är uppfyllda. Vi delar därmed KB:s uppfattning om hur bestämmelsen ska tolkas.
- 8. Valet av språkmodell**
- 8.1 Frågan är om valet av språkmodell, exempelvis huruvida forskningen avser en generativ modell, en maskininlärningsmodell eller annan typ av modell, har självständig betydelse för tillämpningen av 15 b § upphovsrättslagen.
- 8.2 Varken 15 b § upphovsrättslagen eller artikel 3 i DSM-direktivet innehåller någon hänvisning till specifika tekniker, metoder eller modelltyper. Bestämmelsen är teknikneutral och tar sikte på ändamålet med utvinningen, det vill säga att den sker för vetenskapliga

forskningsändamål, snarare än på de verktyg eller metoder som används för att uppnå detta ändamål. Det avgörande är enligt vår mening således inte vilken typ av modell som tränas, utan att träningen som sådan sker inom ramen för och i syfte att möjliggöra vetenskaplig forskning.

- 8.3 Det sagda innebär att valet av språkmodell som utgångspunkt saknar självständig betydelse för huruvida undantaget i 15 b § upphovsrättslagen är tillämpligt. Avgörande är i stället att den text- och datautvinning som föregår träningen uppfyller de krav på forskningsändamål som tidigare redovisats. Förutsatt att dessa krav är uppfyllda påverkar inte modellvalet bedömningen.

9. Slutsatser

- 9.1 Sammanfattningsvis är det vår bedömning att KB uppfyller samtliga förutsättningar för att tillämpa undantaget i 15 b § upphovsrättslagen. KB omfattas av bestämmelsens avnämningsskrets och den TDM-verksamhet som bedrivs sker för forskningsändamål i bestämmelsens mening. Rätten att utföra TDM är tvingande och avtalsvillkor som inskränker denna rätt saknar verkan.
- 9.2 Valet av språkmodell saknar som utgångspunkt självständig betydelse för tillämpligheten av undantaget, eftersom bestämmelsen är teknikneutral. Det avgörande är ändamålet med utvinningen, inte de tekniska metoder eller modeller som används.
- 9.3 Distinktionen mellan forskning på forskarens eget initiativ och uppdragsforskning föranleder ingen annan bedömning. Avgörande är forskningens vetenskapliga ändamål och integritet, inte dess organisatoriska form. Förutsatt att forskningen bedrivs enligt etablerade vetenskapliga normer och att kommersiella företag inte har avgörande kontroll över forskningen och att resultaten inte exklusivt reserveras för enskilda kommersiella intressen, är 15 b § upphovsrättslagen tillämplig även på sådan uppdragsforskning.
-

Bilaga 4: Professor emeritus Jan Roséns utlåtande

Jan Rosén*

jur.dr, professor

Stockholm 2026.05.15

Kungliga biblioteket (KB)
Humlegårdsgatan 26
102 41 Stockholm

att. verksjuristen Jerker Rydén

Rättsligt utlåtande – utveckling av språkmodeller för forskning och upphovsrätt, KB dnr 2025-930.

Kungl. biblioteket (KB) genomför en utredning med anledning av ett regeringsuppdrag enligt regeringsbeslut U2025/01837 att analysera och förbereda för en bredare användning av myndighetens språkmodeller för AI i offentlig sektor. I uppdraget har ingått att redogöra för de upphovsrättsliga aspekterna av utvecklingen av språkmodeller. Regeringsuppdraget har *delredovisats* i januari 2026, vari myndighetens tillämpning av upphovsrättslagen, i relation till den utveckling som idag bedrivs för forskningsändamål, redogjorts för.¹

I anledning härav har jag 2026-03-20 ombetts av KB att avge ett rättsligt utlåtande för besvarande av tre i KB:s till mig riktade anbudsunderlag specificerade spörsmål. Frågorna är de följande:

1. *Såvitt gäller den utveckling av språkmodeller som vilar på vetenskaplig grund och som beskrivs i ovan nämnda delredovisning, gör ni en annan tolkning av 15 b § Upphovsrättslagen (URL) än den vilken KB redovisat för regeringen (se ovan nämnda delredovisning)?*
2. *Har det någon betydelse för tillämpningen av 15 b § URL vilken slags språkmodell som forskningen avser?*
3. *Förutsatt att det sker inom ramen för forskningsprojekt, i KB:s forskningsinfrastruktur, leds av disputerad forskare och har en forskningsfråga, är ni av uppfattningen att KB kan tillämpa*

*Jan Rosén, är emeriterad professor i civilrätt vid Stockholms universitet och är där knuten till Stockholm Centre for Commercial Law, SCCL, där han förestår den medierättsliga forskningsgruppen. Han har under 25 år varit ordförande för Svenska Föreningen för Upphovsrätt, SFU, 1998 – 2023, och är sedan många år Vice President i Association Littéraire et Artistique International, ALAI, samt var bl. a. regeringens ensamutredare vid översyn av URL, vilket resulterade i de två betänkandena SOU 2010:24, Avtalad upphovsrätt, och SOU 2011:32, En ny upphovsrättslag. Bland hans bokutgåvor kan nämnas Upphovsrättslagen. En kommentar (initialt jte Henry Olsson † 2021), senaste upplaga nr 5, 2023, Norstedts Gula Biblioteket (även online i Juno); Immaterialrätt och otillbörlig konkurrens, jte Bernitz († 2022) m.fl., senaste upplaga nr 16/2023; Upphovsrättens avtal, uppl. 3, 2006; Medie- och upphovsrätt, Skrifter utgivna av juridiska fakulteten vid Stockholms universitet nr 78, 2012.

¹ KB:s dokument Analys för en bredare användning av Kungl. bibliotekets språkmodeller i svensk offentlig sektor, Diarienummer: KB 2025-930.

15 b § URL, såsom redogjorts för i delredovisningen, även för uppdragsforskning (oaktat om det rör grundforskning/tillämpad forskning)?

Jag får mot den bakgrunden anföra det följande.

Fråga 1

1. Såvitt gäller den utveckling av språkmodeller som vilar på vetenskaplig grund och som beskrivs i ovan nämnda delredovisning, gör ni en annan tolkning av 15 b § Upphovsrättslagen (URL) än den vilken KB redovisat för regeringen (se ovan nämnda delredovisning)?

Frågan rör KB:s ”tolkning” av § 15 b URL och fokuserar närmare besett på en rad av de grundläggande förutsättningarna för att den paragrafen ska kunna tillämpas på korrekt sätt. KB drar härvid för sitt vidkommande vissa slutsatser *dels* rörande frågan om vilka subjekt som kan vara berättigade till TDM enligt nämnd paragraf, *dels* om utvecklandet av språkmodeller kan rymmas inom dylik TDM, *dels* bedöms begreppet lovlig tillgång vid användning av verk för TDM *och dels* huruvida de forskningsändamål som KB kan engagera sig i ryms inom ramen för TDM-undantaget. KB hänför sig här, helt adekvat, främst till de tre kumulativa förutsättningar som krävs för att undantaget ska kunna tillämpas på korrekt vis, alltså *att* den aktör som utför utvinningen måste tillhöra den avnämmarkrets som anges i bestämmelsen, *att* utvinningen måste ske för forskningsändamål i bestämmelsens mening *och att* institutionen måste ha lovlig tillgång till det material som bearbetas.

I huvudsak blir mitt svar att KB:s slutsatser härvidlag är korrekta, dock med följande noteringar och preciseringar.

(i) Subjekt.

KB framhåller att TDM i enlighet med undantaget i § 15 b URL kan ske hos KB eftersom den paragrafen berättigar bland andra ”bibliotek - - som är tillgängliga för allmänheten”, något som vid en bokstavsläsning väl träffar KB. Som framgår av instruktionsförfordningen för KB är biblioteket Sveriges nationalbibliotek.² KB är enligt bibliotekslagen del av det allmänna biblioteksväsendet och allmänheten ska ges tillgång till dess samlingar.³ KB är således tillgängligt för allmänheten och ingår redan därmed i en av de i 15 b § URL angivna avnämmarkretsarna. Så mycket mera som KB med säkerhet också är en ”kulturarvsinstitution”, vilken enligt artikel 3.1 DSM-direktivet jämte forskningsorganisationer ska gynnas av TDM-undantaget när man genomför TDM för forskningsändamål.⁴ Därmed är inte sagt att KBs mångfacetterade uppdrag låter det vara en sådan forskningsorganisation som definieras i § 15 c URL.

Det bör dock understrykas att denna inskränkande regel tillkommit till förmån för forskningsorganisationer och, än viktigare, vid övriga privilegierade institutioner när de faktiskt

² Förfordningen (2008:1421) med instruktion för Kungl. Biblioteket.

³ Se §§ 1 och 2 i Bibliotekslagen (2013:801).

⁴ Jfr p 12 till DSM-direktivet

presterar vetenskaplig forskning. KB är förvisso tillgängligt för allmänheten, men det är när man företar text- och datautvinning för forskningsändamål som man kan åberopa sig på § 15 b URL.

Frågan kan i o f sig ställas om KB kan anses utgöra en *forskningsorganisation* i den mening som faller inom ramen för definitionen i § 15 c URL. Detta skulle förutsätta att vetenskaplig forskning utgör KB:s *främsta mål*, vilket är ett nödvändigt kriterium för att klassificeras just som en forskningsorganisation enligt definitionen i § 15 c § URL, vilken paragraf bl a genomför artikel 2.1 DSM-direktivet. Men även om det så vore att KB inte kan omtalas som en forskningsorganisation så utgör ju KB en del av avnämarskretsens hos § 15 b URL, varför spørsmålet i stället är i vad mån KB utför sådan vetenskapligt baserad forskning som berättigar till korrekt tillämpning av den bestämmelsen.

Artikel 3.1 DSM-direktivet medger, i svensk språkdräkt, exemplarframställning och utdrag som forskningsorganisationer och kulturarvsinstitutioner genomför för *forskningsändamål* (i syfte att utföra text- och datautvinning). Samma uttryck används i § 15 b och för övrigt också i § 16 andra stycket p 1 URL. Den senare bestämmelsen medger, även den som en undantagsbestämmelse, bibliotek och arkiv som är tillgängliga för allmänheten att framställa exemplar för *kompletterings- och forskningsändamål*.

Det kan här konstateras att redan i § 16 § andra stycket p 1 URL används uttrycket forskningsändamål på ett *kvalificerande* vis, även om det aldrig funnits specifika bestämmelser som närmare anger kriterier för sådan kvalifikation.⁵ Det kvalificerande elementet har dock funnits hos denna bestämmelse ända sedan den i viss form förekom redan i § 12 § URL i lagens ursprungliga lydelse liksom när den kom till konkret uttryck i början av 1990-talet, i samband med att praktiskt taget alla bestämmelser i 2 kap. URL arbetades om genom SFS 1994:190, och uttrycket forskningsändamål hos arkiv och bibliotek kom att framträda i § 16 §.⁶ Nedan under (iv) behandlas mer koncentrerat begreppet forskningsändamål såvitt angår tillämpningen av § 15 b § URL.

(ii) Språkmodeller

Att ta fram AI-språkverktyg eller språkmodeller på basen av § 15 b § URL kan väl förekomma under förutsättning att den sker för forskningsändamål.⁷ Villkoren för bestämmelsens tillämpning, t.ex. beträffande lämplig säkerhetsnivå, lovlig tillgång, inflytande över forskningsinsatsen, tillgänglighet för allmänheten, utpekandet av berättigade organisationer etc., rör inte språkmodellen som sådan. Bestämmelsen är såtillvida teknikneutral som den inte tar sikte på den specifika modelltyp som används eller åsyftas.⁸ Kopieringen i fråga kan i princip ske med vilken teknik som helst, låt vara att TDM typiskt sett förutsätter digitalt format för

⁵ Se *Olsson/Rosén*, Upphovsrättslagstiftningen. En kommentar, femte upplagan 2023, s 181 ff.

⁶ Se *Regner*, Upphovsrätt till litterära och konstnärliga verk, 1961, 123 ff; *Olsson*, Upphovsrättslagstiftningen. En kommentar, första upplagan 1996, s 128 f.

⁷ Att lagstiftaren i EU också ser att TDM är ett tillvägagångssätt för att ta fram just språkmodeller för AI framgår bl a i skäl 105 till AI-förordningen, som hänvisar direkt till inskränkningarna för TDM i DSM-direktivet.

⁸ Se artikel 2.2 och artikel 3.1 DSM-direktivet; jfr *Rosati*, Copyright in the Digital Single Market. Article-by-Article Commentary to the Provisions of Directive 2019/790, 2021, s 41 f.

analysens genomförande och som definitionen i § 15 c URL talar om TDM som ”automatiserad teknik för att analysera text och data i digital form”. Vidare kan TDM ha avseende på vilka upphovsrättsligt skyddade verkstyper som helst, utom datorprogram, såsom text, ljud, bild, video eller annan data.⁹ Den typ av språkmodell som forskningen avser är därmed som utgångspunkt utan avgörande självständig betydelse för tillämpligheten av bestämmelsen, förutsatt, som sagt, att utvinningen sker för vetenskapliga forskningsändamål och även uppfyller övriga uppställda villkor. En mer fördjupad kommentar ges dock i svaret till fråga 2.

Här bör dock inflikas följande. Språkmodeller av olika typer kan förvisso tas fram hos KB inom ramen för TDM-undantaget och detta med exemplarframställning av verk som KB har lovlig tillgång till. Men undantaget medger helt enkelt inte, kort sagt, att forskningsresultatet – språkmodellen – förmår framställa exemplar av eller tillgängliggöra för allmänheten skyddade verk, det senare typiskt sett genom överföring eller framföranden till allmänheten, d v s sådana verksexemplar som språkmodellen tränats på, såvida inte detta medgivits av rättsinnehavarna.

Närmare besett torde inte heller en språkmodell av typen generativ AI, se nedan under fråga 3, kunna framtas av KB ens för användning på plats inom KB även om själva modellen, vid en strikt läsning av § 15 b URL, helt korrekt skulle kunna *tränas* inom ramen för TDM-undantaget och under dess givna förutsättningar. Men exemplar av upphovsrättsligt skyddade verk som just en sådan typ av generativ språkmodell kan producera torde alltså kräva rättsinnehavarnas medgivande – framställda exemplar för TDM-användningen får ju inte behållas längre än nödvändigt för forskningsändamålet eller för andra nyttjanden eller tillgängliggöranden inför en allmänhet. KB:s breda och i o f sig välmotiverade och oavvisliga uppdrag förmår således inte tänja gränserna för undantagsregeln i § 15 b URL. Det må åter understrykas att om framtagna språkmodell inte innehåller eller kan generera exemplar av skyddade verk så finns givetvis inget behov av rättighetsklarering inför ett tillgängliggörande.

Det bör dock noteras att vissa andra undantagsregler i URL i o f sig underlättar för KB att framställa exemplar och att åstadkomma tillgängliggöranden. Enligt 16 § URL får exemplar framställas för *kompletterings-* eller *forskningsändamål* och för att tillgodose lånesökandes önskemål om *enskilda artiklar* eller *korta avsnitt* eller om material av säkerhetsskäl inte bör lämnas ut eller för användning i ”läsapparater”. Enligt 21 § andra stycket URL kan KB tillgängliggöra för ”enstaka besökare” i myndighetens egna lokaler och genom ”tekniskt hjälpmedel” verk som ingår i de egna samlingarna.

Undantaget i 16 § andra stycket p. 1 URL för kompletterings- eller forskningsändamål gäller för anställda och uppdragstagare, inte externa användare, och måste ske för institutionens räkning eller inom ramen för dess verksamhet. Den bestämmelsen kan i vart fall inte användas för AI-träning av språkmodeller och vidgar således inte TDM-undantaget i § 15 b URL. En helt annan sak är att ”forskningsändamål” är ett begrepp som i o f sig torde sammanfalla med samma uttryck i såväl § 15 b URL som 16 § andra stycket p. 1 URL, se nedan under (iv).

⁹ Jfr prop. 2021/21:278 s 45 f.

(iii) Lovlig tillgång

Exemplarframställning inför eller under TDM-användning, vid träning av språkmodeller, förutsätter att användaren har lovlig tillgång till de skyddade verk som används. Med lovlig tillgång avses att samtycke föreligger från rättsinnehavaren eller med stöd av lag. Lovlig tillgång kan avse *alla slags verk*, utom datorprogram, och på *alla sätt* lovligen åtkomna verk, såsom inköpta på öppna marknaden eller rekviderade från öppna och lovliga källor. Dessutom behöver sålunda tillgängliga verk varken ha getts ut eller offentliggjorts före TDM-användningen. Därmed kan manuskript eller brevsamlingar som deponerats hos KB komma till TDM-användning i enlighet med 15 b § URL. Kort sagt kan lovlig tillgång således avse inte bara på marknaden inköpta verksexemplar, licensierad data, inhämtade exemplar från nätmiljön etc. utan även material av mer privat eller svåråtkomlig natur där lovligheten likväl är manifest. Det material som bearbetas genom TDM-tekniken kan således ha inhämtats med stöd av plikttagstiftningen, lagen (2012:492) om pliktexemplar av elektroniskt material eller lagen (1993:1392) om pliktexemplar av dokument, eller att KB på annat sätt har lovlig tillgång till verken. Detta material omfattar såväl text som bild, ljud och video och som samlats in enligt hos KB etablerade rutiner.

Även den som har/får tillgång till ett verk med stöd av en *inskränkning* i upphovsrätten eller med stöd av en avtalslicens har lovlig tillgång till ett verk. En inskränkning kan dock innehålla villkor för hur verket får användas. I så fall gäller dylika villkor.¹⁰

Detsamma bör gälla så kallad kompletterande data, utöver pliktmaterial, som KB kan samla in eller ta emot på andra sätt, såsom automatiserad insamling, exempelvis genom robotsvep av webben, nyttjande av öppna datakällor som är fritt tillgängliga och sådant som donerats från externa aktörer. Genomgående förutsätter detta, som sagt, att sålunda insamlad data är lovlig, något som inte är fallet beträffande material som olovligt publicerats på internet.

Till lovliga underlag för TDM-förfoganden torde normalt även kunna räknas sådan data som KB inhämtat för forskningsändamål. Det kan röra sig om kompletterande dataleveranser från andra bibliotek eller myndigheter, insamlad data för ett specifikt forskningsprojekt eller hos KB genererad data, som skett antingen manuellt, med AI eller i kombination. Slutligen torde även ”extern” data kunna tillämpas, om den är lovlig, exempelvis inom forskningssamarbeten där data delas via gemensamma lagrings- och beräkningsmiljöer. Sådan datas lovlighet kontrolleras möjligen inte så lätt av KB, men dess lovlighet kan ju väl existera. Det ankommer dock på KB att kontrollera lovligheten hos det material man använder med åberopande av § 15 b URL.

(iv) Forskningsändamål

KB understryker att man har uppdraget att skapa och tillhandahålla en forskningsinfrastruktur, främst genom att erbjuda ”resurser, verktyg och system” som behövs för att forskare ska kunna genomföra forskning. Detta anges bl a innebära och kräva datainsamling, analys och lagring samt att främja forskningssamarbeten över organisatoriska gränser, allt för att stödja

¹⁰ Se Olsson/Rosén, Upphovsrättslagstiftningen. En kommentar, femte upplagan 2023, s 175 ff.

kunskapsutveckling i samhället. Det finns ingen anledning att här ifrågasätta detta uppdrags bredd, djup, omfattning eller dess välmotiverade syften. Men frågan är om § 15 b URL matchar eller ryms inom detta breda uppdrag?

Det har redan konstaterats att § 15 b URL avser TDM för forskningsändamål och att detta *kvalificerande* begrepp låter förstå att det finns anledning att göra en distinktion mellan vetenskaplig forskning och annan forskning och att TDM-undantaget härför sig till den förra. Bestämmelsen genomför ju artikel 3 i DSM-direktivet som i sin tur återgår på artikel 2.1 DSM-direktivet där de privilegierade institutionerna ska prestera ”vetenskaplig forskning”. I den svenska språkversionen talas visserligen i artikel 3 blott om forskningsändamål, men i de engelska, franska och tyska versionerna, exempelvis, används uttrycken *scientific research*, *recherche scientifique* respektive *wissenschaftliche Forschung*. Begreppet ”forskningsändamål” i artikel 3 ska således förstås som hänförligt till vetenskaplig forskning. F ö ska alla språkversioner inom EU-rätten tolkas enhetligt. Detta stöds även av direktivets systematik, särskilt definitionen av forskningsorganisation i artikel 2.1. DSM-direktivet, där även den svenska språkversionen av direktivet anger enahanda uttryck. Detsamma bör gälla vid tolkningen av 15 b § upphovsrättslagen, något som också ligger i linje med definitionen i § 15 c URL.

Därmed är utgångspunkten att TDM för forskningsändamål måste vila på vetenskaplig grund. Detta stöds också av motivuttalanden till § 15 b URL som anger att ”med forskning avses forskning på vetenskaplig grund och inte allmänt sökande efter kunskap eller information”.¹¹ Forskningen ska alltså utföras på vetenskaplig grund och inte bara utgöras av allmänt sökande efter kunskap eller information, ens om dylikt sökande i o f sig utförs av professionella forskare. Man kan åter tillägga att uttrycket ”forskningsändamål” av gammalt använts i URL för att ange kvalificerad vetenskaplig forskning.

Någon mer auktoritärt utformad definition av begreppet vetenskaplig forskning förekommer dock inte i URL eller eljest i svensk rätt. Inte heller DSM-direktivet anger vad som i tillämpningen ska avses med ”vetenskaplig forskning”. Ytterst är det förstås EU-domstolen som har att avgöra såväl vilka forskningsområden som omfattas liksom vad vetenskaplig forskning ska avse i TDM:s sammanhang.

Skäl 11 och 12 till DSM-direktivet anger exemplifierande ett stort antal forskningsdiscipliner som kan komma ifråga och som öppnar brett för såväl så kallade formalvetenskaper som tillämpad forskning och grundforskning och detta inom all slags naturvetenskap och humaniora. Sätillvida anges ett mycket brett verksamhetsfält för vetenskapligt genomförd forskning. Detta torde väl omfatta vad som brukar omtalas som *tillämpad forskning* såväl som *grundforskning*.¹² Att forskningen, såsom § 15 b URL måste förstås, ska vila på vetenskaplig grund ger, enligt

¹¹ se prop. 2021/22:278, s 222 f.

¹² *Tillämpad forskning* har målet, populärt uttryckt, att ta fram ny kunskap som i första hand syftar till en bestämd praktisk tillämpning, såsom att söka efter tillämpningar av de resultat som man kommit fram till inom grundforskningen eller att hitta på nya metoder för att lösa ett problem. *Grundforskning* avser teoretisk kunskap, kort uttryckt, att försöka förstå hur saker fungerar, under det att tillämpad forskning omvandlar teoretisk kunskap till praktiska, effektiva lösningar.

min mening, inget skäl till att göra en distinktion mellan så kallad grundforskning och tillämpad forskning. Vetenskapligt välgrundad metodik och genomförande på sätt sistnämnd paragraf efterfrågar kan förekomma hos båda blocken. Eftersom allmänt sökande efter kunskap och information däremot inte är något vetenskapligt förankrat så är det förmodligen så att exempelvis predikativa marknadsanalyser för kommersiella ändamål, såsom studier av generella beteendemönster på marknaden eller riskanalys, knappast kan omtalas som vetenskapliga projekt enligt DSM-direktivet.¹³ Dessutom uppfyller ju dylikt typiskt sett inte villkorskraven enligt § 15 c URL.

Det finns dock åtskilliga källor som ger tämligen entydigt stöd för vad vetenskaplig forskning i *tillämpningen* rimligen ska innebära sådant begreppet förekommer i DSM-direktivet och därmed hur det bör uppfattas sådant de genomförts i §§ 15 b och 15 c URL. Detta således vid sidan av de tre specifika och i sammanhanget mycket viktiga villkor för forskningsverksamheten som sådan som uppställs i § 15 c URL och som återgår på artikel 2 DSM-direktivet, nämligen **1)** att forskningen ska bedrivas på icke vinstdrivande grund, **2)** på sådant sätt att all vinst återinvesteras i forskningen eller **3)** sker i enlighet med ett uppdrag i det allmännas intresse som har erkänts i en medlemsstat.

Beträffande den vetenskapliga forskningens genomförande kan enligt min uppfattning vägledning exempelvis vinnas ur EU:s dataskyddsförordning, GDPR,¹⁴ som erbjuder särskilda regler för just vetenskaplig forskning åtminstone på så vis att Europeiska dataskyddsstyrelsen, EDPB - ett oberoende EU-organ med ansvar för att säkerställa en enhetlig tillämpning av GDPR i EU – nyligen har antagit riktlinjer för begreppet vetenskaplig forskning i sex punkter. Enligt EDPB kännetecknas sålunda vetenskaplig forskning av **1)** ett metodiskt och systematiskt arbetssätt, **2)** efterlevnad av etiska normer, **3)** verifierbarhet och transparens, **4)** självständighet och oberoende, **5)** ett syfte att bidra till allmän kunskap samt **6)** potential att tillföra ny kunskap eller nya tillämpningar.

Detta anknyter ganska väl till den beskrivning som anges i OECD:s så kallade Frascati Manual, med fokus på begreppet forskning och utveckling, FOU.¹⁵ Kortfattat sägs detta bygga på att forskningsverksamhet ska vara nyhetsskapande, kreativ, osäker, systematisk, överförbar och/eller reproducerbar, något som dock torde öppna för ett något vidare fält än det som just vetenskaplig forskning påbjuder.¹⁶

En del element ur det nyss angivna framträder således direkt i de villkor som såväl §§ 15 b och 15 c URL uppställer. Sammantaget bör kunna konstateras att TDM på basen av nämnda paragrafer förutsätter ett kvalificerat, systematiskt och metodbundet kunskapssökande arbete

¹³ Jfr t.ex. Kempas, *Artificiell intelligens och immaterialrätt*, 2023, s 196 ff.

¹⁴ Europaparlamentets och rådets förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter och om upphävande av direktiv 95/46/EG (allmän dataskyddsförordning).

¹⁵ Finns under <https://www.oecd.org>; <https://doi.org/10.1787/9789264239012-en>.

¹⁶ Jfr exempelvis den svenska Etikprövningslagen (2003:460) som definierar begreppet forskning sålunda: "Vetenskapligt experimentellt eller teoretiskt arbete eller vetenskapliga studier genom observation, om arbetet eller studierna görs för att hämta in ny kunskap, och utvecklingsarbete på vetenskaplig grund, dock inte sådant arbete eller sådana studier som utförs endast inom ramen för högskoleutbildning på grundnivå eller på avancerad nivå."

på basen av etablerade vetenskapliga normer med krav på transparens, prövbarhet och möjlighet till kritisk granskning. Arbetet ska syfta till att generera ny, generell och verifierbar kunskap och ska typiskt sett bedrivas under former som säkerställer oberoende samt att resultaten inte exklusivt tillkommer enskilda intressen. Viktigt är också att forskningen sker, om inte nödvändigtvis enbart på KB:s initiativ så i vart fall under KB:s kontroll och med säkerställande av dess vetenskapliga oberoende vad gäller metod och genomförande.

Vad KB uppgett i sin delrapport om sitt arbete med att framta språkmodeller med TDM-teknik tyder på att man väl uppfyller de krav på vetenskaplig forskning som § 15 b URL påbjuder och förutsätter. Sammantaget kan kommentarerna (i) – (iv) ovan, såsom svar på fråga 1, sammanfattas med att KB under alla omständigheter är att anse som ett offentligt tillgängligt bibliotek eller en institution för film- eller ljudarvet eller annan kulturarvsinstitution eller arkiv i den mening som avses i 15 b § URL och artikel 2.3 i DSM-direktivet. KB är därmed berättigat att utföra text- och datautvinning för forskningsändamål på sådant material som KB har lovlig tillgång till. Åtminstone de språkmodeller som KB hittills tagit fram med TDM-teknik framstår som väl placerade inom bestämmelsens tillämpningsområde. Det kan tilläggas att denna rätt som KB utrustats med är, enligt 15 b § tredje stycket URL, tvingande varför avtalsvillkor som inskränker KB:s rätt till text- och datautvinning saknar verkan.

Fråga 2.

2. Har det någon betydelse för tillämpningen av 15 b § URL vilken slags språkmodell som forskningen avser?

KB säger sig idag arbeta med två typer av språkmodeller: *deskriptiva* och *preskriptiva*. Däremot presterar man inte *generativa* modeller. Det torde dock vara så att det knappast går att dra knivskarpa linjer mellan dessa typer av språkmodeller. Såvitt här intresserar är det främst angeläget att observera huruvida en framtagna språkmodell innehåller generativa moment i form av kapacitet att återge eller framställa exemplar av skyddade verk som den tränats på.

Deskriptiva modeller, framhåller KB, analyserar och beskriver innehåll. De kan rätta OCR-fel, identifiera namn på personer, platser och organisationer, skapa metadata och kategorisera texter. Ett felaktigt läst ord i en källa kan korrigeras genom att modellen tolkar sammanhanget. På liknande sätt kan taligenkänning upptäcka namn även i brusiga ljudinspelningar, och bildanalys kan identifiera motiv utan att en bildtext finns tillgänglig.

Preskriptiva modeller går enligt KB ett steg längre genom att ge rekommendationer. De kan inte bara känna igen ett företagsnamn, utan också förstå att det handlar om en viss typ av företag, såsom ett fintechföretag, och kan därmed föreslå ämnesklassificering, länkning till relaterade texter eller prioritering av digitalisering inom ett visst område. De används också för att kvalitetssäkra OCR, föreslå arbetsflöden och bidra till bevarandestrategier genom att identifiera material som riskerar att förstöras.

Dylika språkmodeller som således tas fram hos KB anges inte generera data som de tränats på vid användning. Modellerna kan, enligt KB, inte generera alster som liknar något upphovsrättsskyddat material. Såvida KB:s färdiga modeller inte innehåller exemplar av skyddade verk som de tränats på så skulle de följaktligen inte åstadkomma eller medverka till upphovsrättsintrång vid användning varken inom KB eller vid tillämpningar av en allmänhet.

Slutsatsen är att KB:s publicering av dylika färdiga språkmodellerna inte innebär något upphovsrättsligt förfogande, under förutsättning att språkmodellen i fråga inte genererar upphovsrättsskyddat material som inte rättighetsklarerats med rättsinnehavarna. I den mån språkmodellen i sig själv skulle kunna omtalas som ett skyddat verk (av beskrivande art, således litterärt i upphovsrättslig mening), närmast dess algoritm, så lär ju rättsskyddet tillkomma KB som uppdragsgivare (till upphovspersonen) för språkmodellens framtagande. Det är dock möjligt att de skulle kunna utgöra en bearbetning eller kopia av ett verk (algoritm), vars rättsskydd kan behöva utrönas.

Såvida KB inte publicerar, för allmänna ändamål, *generativa* modeller eller andra typer av språkmodeller utan generativa inslag, konkurrerar således modellernas användning inte med kreatörers skapande av verk eller avsättningen på marknaden av skyddade verk.

I framtiden säger sig KB i sin delrapport kunna ta fram *instruktionstränade* modeller för uppgifter som rör sammanfattning, förklaring och modernisering av språk. För att inte komma i konflikt med upphovsrättens intressenter måste dylika modeller antingen utformas så att de inte kan reproducera exempelvis skyddade texter, utan endast genererar ett nytt innehåll baserat på övergripande mönster, eller så har rättsinnehavarnas medgivanden inhämtats till generativa element som modellerna må innehålla. En instruktionstränad modell skulle kunna ge forskare en snabb överblick över hur ett tema eller en aktör behandlats över tid, utan att användaren får tillgång till exempelvis en upphovsrättsligt skyddad originaltext. Såvida det sistnämnda likväl sker föreligger följaktligen ett behov av att ha rättsinnehavarnas medgivande till det upphovsrättsliga förfogandet.

Korrekt tillämpning av § 15 b URL förutsätter knappast att träningen sker blott på någon viss typ av språkmodell. Som framgick av svaret (ii) under fråga nr 1 ovan så kan vid framtagande av själva modellen, med en strikt läsning av § 15 b URL, densamma helt korrekt *tränas* inom ramen för detta TDM-undantag och under dess givna förutsättningar, i princip också såvitt gäller exempelvis en språkmodell av typen generativ AI. Men en sådan språkmodell kan knappast framtas av KB ens för användning på plats inom KB utan vederbörliga tillstånd från rättsinnehavarna. Exemplar av upphovsrättsligt skyddade verk får i och för sig framställas vid träningsprocessen, inkluderande förberedelsematerial för TDM-processen, alltså även så kallade *datasets* eller strukturerade samlingar av data (datamängd), också av en generativ språkmodell. Men framställda exemplar för TDM-användningen får ju inte behållas längre än nödvändigt för det specifika forskningsändamålet – att framställa språkmodellen. Den generativa språkmodellens framtagande kräver därför i praktiken rättighetsklarerings av berörd upphovsrätt redan vid strikt tillämpning inom KB, liksom givetvis vid nyttjanden av eller tillgängliggöranden inför en allmänhet.

Såvitt jag erfarit finns inte någon motsvarighet till en generativ språkmodell tillgänglig hos KB för sökningar i KB:s samlingar och studier av dokument i fulltext etc på datorskärm, något som forskare kunde vara betjänta av för att bättre kunna söka i KB:s enorma samlingar. Denna begränsning följer också av att mängden digitalt material inte är så omfattande hos KB, någonstans mellan 3-4% totalt. Ännu mindre lär vara OCR:at så att man kan söka i annat än metadata. I grunden lär detta bero på upphovsrättsliga orsaker. Upphovsrättsligt skyddat material hos KB är typiskt sett inte alls tillgängligt digitalt på distans.

Hos KB finns störst tillgång till digitiserat material beträffande dagstidningar och av-media. För dagstidningarna gäller att ca 25% finns digitiserade och ca 80 % av det audiovisuella materialet i SMDDB finns tillgängligt digitalt. Det är också digitiserat tidningsmaterial som lär vara träningsdata för den TDM som KB utnyttjar vid utveckling av språkmodeller.

Svaret på fråga nr 2 kan alltså sammanfattas med ett lakoniskt nej; det är inte den eftersökta språkmodellens natur som avgör tillämpligheten av TDM-undantaget i § 15 b URL. Men tillämpning och tillgängliggörande av resultatet av arbetet kan förutsätta medgivande av innehavare av upphovsrätt till verk som i förekommande fall genereras vid en språkmodells användning.

Fråga 3.

3. Förutsatt att det sker inom ramen för forskningsprojekt, i KB:s forskningsinfrastruktur, leds av disputerad forskare och har en forskningsfråga, är ni av uppfattningen att KB kan tillämpa 15 b § URL, såsom redogjorts för i delredovisningen, även för uppdragsforskning (oaktat om det rör grundforskning/tillämpad forskning)?

Som konstaterades vid fråga nr 1 ovan är KB berättigat att utföra text- och datautvinning för vetenskapligt baserade forskningsändamål på sådant material som KB har lovlig tillgång till och att åtminstone de språkmodeller som KB hittills tagit fram med TDM-teknik framstår som väl placerade inom tillämpningsområdet för § 15 b URL. Fråga 3 anknyter främst till de tre specifika villkor för forskningsverksamheten som sådan som uppställs i § 15 c URL och som återgår på artikel 2 DSM-direktivet, nämligen **1)** att forskningen ska bedrivas på icke vinstdrivande grund, **2)** på sådant sätt att all vinst återinvesteras i forskningen eller **3)** sker i enlighet med ett uppdrag i det allmännas intresse som har erkänts i en medlemsstat.

§ 15 c URL genomför alltså artikel 2 TDM-direktivet, vilket har en förklarande bas i skäl 11 och 12 till detta direktiv. Sålunda kan forskningsverksamheten ske inte bara vid universitet och rena forskningsorganisationer utan även inom en avgränsad verksamhet hos en juridisk person eller en myndighet med annan verksamhet. I lagmotiven nämns exempelvis en forskningsverksamhet inom ett sjukhus som omfattat av definitionen.¹⁷ Inget hindrar således att en institution eller ett forskningsinstitut omfattas av definitionen som organisatoriskt utgör en del av en större verksamhet som huvudsakligen bedriver en annan verksamhet.

¹⁷ Se prop. 2021/22:278 s 223. Jfr skäl 12 till DSM-direktivet.

Huvudsaklighetskravet för just definitionen av en ”forskningsorganisation” är, som redan framhållits under (i) vid fråga **nr 1** ovan, inte ett oavvisligt krav för den kategori av verksamheter som kan vara berättigade att åberopa sig på undantaget i § 15 b URL. Såväl offentliga som privata aktörer kan omfattas av definitionen. Men forskningen måste drivas utan vinstsyfte eller så att all vinst återinvesteras i vetenskaplig forskning alternativt att den sker i det allmännas intresse, § 15 c andra stycket URL. Saknas dylika syften eller intressen så faller forskningen utanför en sådan som privilegieras genom möjligheten att tillämpa TDM-undantaget i § 15 b URL.

För *uppdragsforskning* gäller – det vill säga när en extern intressent initierar och söker (finansierar) ett forskningsresultat hos en forskningsinstitution – att dylikt väl kan vara omfattat av villkoren i § 15 c URL. Ett sådant uppdrag kan återspeglas genom offentlig finansiering eller genom bestämmelser i nationell lagstiftning eller offentliga upphandlingskontrakt. Men också helt privata intressen kan framträda här och tillämpligheten hos § 15 b URL förringas exempelvis inte genom samarbeten som sker genom offentlig-privata partnerskap såsom uttryckts i DSM-direktivet.¹⁸ Således helt oavsett om forskningen är offentligt finansierad eller sker helt eller delvis med privata medelstillskott.

Det avgörande, såvitt här intresserar, är ju att forskningen som sådan måste bedrivas under full kontroll och vetenskapligt inflytande av den formellt TDM-berättigade organisationen, något som även medför att detta och forskningsprojektets allmänna angelägenhet måste övervägas och garanteras av KB. Dessutom kan en finansiär och intressent av forskningsresultatet, exempelvis aktieägare i en bidragsgivande organisation, inte få tillgång till resultaten av forskningen framför andra.¹⁹ Det ligger i linje med att forskningen måste drivas utan vinstsyfte eller så att all vinst återinvesteras i vetenskaplig forskning alternativt att den sker i det allmännas intresse. KB synes uppfylla dessa villkor i tillräcklig mån, något som understryks av att forskningsresultaten, språkmodellerna, tillgängliggörs i öppen access.

Slutsatsen är att KB väl skulle kunna uppfylla dessa villkor för att bedriva forskning på vetenskaplig grund också inom ramen för vad som kan omtalas som *uppdragsforskning*. Det behöver inte heller göras en distinktion mellan grundforskning och tillämpad forskning, såsom konstaterades under fråga nr 1 ovan.



Jan Rosén

¹⁸ Se skäl 11 till DSM-direktivet.

¹⁹ Jfr artikel 2.1. a) - b) DSM-direktivet.

Bilaga 5: Uppdragsdirektiv



Regeringen

Utbildningsdepartementet

Regeringsbeslut

2025-09-18
U2025/01837

II:5

Kungl. biblioteket

2025 -09- 2 5

Dnr: KB 2025-930

Kungl. biblioteket
Box 5039
102 41 Stockholm

Uppdrag till Kungl. biblioteket att analysera och förbereda för en bredare användning av myndighetens språkmodeller i svensk offentlig sektor

Regeringens beslut

Regeringen uppdrar åt Kungl. biblioteket att analysera och förbereda för en bredare användning av myndighetens språkmodeller i svensk offentlig sektor.

Uppdraget ska senast den 15 januari 2026 delredovisas och senast den 1 juni 2026 slutredovisas till Regeringskansliet (Utbildningsdepartementet med kopia till Finansdepartementet).

Närmare om uppdraget

I uppdraget ingår att genomföra en fördjupad kartläggning av möjligheter och utmaningar med Kungl. bibliotekets arbete med språkmodeller för forskning. Vidare ingår att belysa ekonomiska och upphovsrättsliga aspekter samt att ge förslag på hur verksamheten kan bidra till att möta behoven av att använda Kungl. bibliotekets språkmodeller även från andra relevanta verksamheter i offentlig sektor. I uppdraget ingår också att redogöra för Kungl. bibliotekets roll i utvecklingen av språkmodeller i förhållande till andra aktörers arbete med att utveckla språkmodeller inom offentlig sektor.

Inom ramen för uppdraget ska Kungl. biblioteket föra nära dialog med och inhämta synpunkter från Vetenskapsrådet och lärosäten samt med andra relevanta myndigheter och aktörer, i synnerhet Myndigheten för digital förvaltning, Riksarkivet, Institutet för språk och folkminnen, Försäkringskassan och Skatteverket.

Telefonväxel: 08-405 10 00
Webb: www.regeringen.se

Postadress: 103 33 Stockholm
Besöksadress: Drottninggatan 16
E-post: u.registrator@regeringskansliet.se

För uppdraget får Kungl. biblioteket under 2025 använda högst 700 000 kronor som ska redovisas mot det inom utgiftsområde 22 Kommunikationer för budgetåret 2025 uppförda anslaget 2:4 Informationsteknik och telekommunikation, anslagsposten 4 Informationsteknik – del till Kammarkollegiets disposition. Medlen betalas ut engångsvis efter rekvisition till Kammarkollegiet senast den 1 december 2025.

Medel som inte har använts för avsett ändamål ska återbetalas senast den 1 juni 2026 till Kammarkollegiet. Vid samma tidpunkt ska en ekonomisk redovisning av använda medel lämnas till Kammarkollegiet.

Skälen för regeringens beslut

Under de senaste åren har det blivit allt tydligare att artificiell intelligens (AI) och maskininlärning kommer att revolutionera forskningen och avsevärt påverka samhället i stort. Sverige har goda förutsättningar att vara ledande inom detta område genom tillgång till olika typer av data, beräkningskapacitet och modeller som kan användas för att utveckla och använda AI som vetenskapligt verktyg. För att säkerställa en transparent och ansvarsfull AI-utveckling och för att underlätta och dra nytta av användningen av AI inom forskningen och offentlig sektor bedömer regeringen att det finns behov av nationell samordning och vägledning.

Kungl. biblioteket utvecklar AI-modeller för forskningens behov med utgångspunkt i de omfattande samlingarna av historiskt och aktuellt material, som böcker, tidningar och radioprogram. Dessa modeller kan spegla variationer av det svenska språket och minoritetsspråk. AI-modellerna görs öppna tillgängliga och används även utanför forskningsområdet, vilket ger bred samhällsnytta. Det finns nu anledning att förbereda för hur Kungl. biblioteket i än större utsträckning skulle kunna möta offentlig sektors behov med sina språkmodeller och därigenom stärka och bredda sin roll som resurs för utvecklingen och användningen av AI i Sverige.

På regeringens vägnar



Lotta Edholm



Beate Eellend

Kopia till

Justitiedepartementet/L3
Finansdepartementet/BA, ESA, OFA
Socialdepartementet, SAM
Klimat- och näringslivsdepartementet/BK, CKS, E, NIM
Kulturdepartementet/SNA

3 (3)



KB

Besöksadress: Humlegården
Box 5039, 102 41 Stockholm
Telefon: 010-709 30 00
E-post: info@kb.se
www.kb.se