Delrapport

AI-regulatorisk sandlåda – den andra iterationen

ES2024-18

December 2024

Innehåll

1. Sammanfattning

Inom eSams ram drivs en pilot i syfte att bidra med kunskap om AI-regulatoriska sandlådor. I den andra iterationen av piloten för AI-regulatorisk sandlåda har arbetsgruppen tittat på ett system för remissammanställningar.

AI-system
Arbetsgruppen finner att aktuellt system utgör ett AI-system enligt AI-förordningen. Systemet träffas inte av bestämmelserna om oacceptabel risk, hög risk eller transparenskraven utan är ett AI-system med minimal/ingen risk. Därmed finns det inga krav på aktörerna enligt AI-förordningen. (Det kan finnas informationskrav i annan lagstiftning, men rapporten behandlar endast AI-förordningen.) För en myndighet kan det ändå finnas anledning att överväga att informera om AI-systemet, men då på frivillig basis eller utifrån frivilliga uppförandekoder.

Roller enligt AI-förordningen
Arbetsgruppen har övervägt vilken roll involverade aktörer bör anses ha enligt AI-förordningen. Reglerna om rollfördelning är skrivna utifrån en kommersiell marknad där ansvarsfördelning ofta regleras genom avtal snarare än genom samverkan mellan myndigheter. En intressant frågeställning är hur begreppet *kommersiell verksamhet* ska tolkas och vilken betydelse begreppet har för AI-förordningens tillämpning för offentliga myndigheter. Utifrån det samarbete som föreligger mellan involverade aktörer är det inte självklart vilken roll respektive aktör har enligt AI-förordningen, tanken verkar ändå vara att det ska finnas en utpekad leverantör. Det är förmodligen inte är möjligt för flera aktörer att ha ett gemensamt leverantörsansvar enligt AI-förordningen.

AI-modeller för allmänna ändamål
I AI-systemet finns två språkmodeller och arbetsgruppen har därför analyserat bestämmelserna om AI-modeller för allmänna ändamål samt bestämmelserna om öppen källkod, systemrisk och forskning. Arbetsgruppen konstaterar att åtminstone en AI-modell för allmänna ändamål ingår i AI-systemet. När en AI-modell för allmänna ändamål integreras i eller ingår i ett AI-system bör systemet anses vara ett AI-system för allmänna ändamål när systemet, på grund av integreringen, har förmåga att tjäna en rad olika ändamål. Troligen har aktuellt system inte sådan förmåga. Ett AI-system för allmänna ändamål medför dock inte några ytterligare krav på leverantören eller tillhandahållaren. Kraven som gäller baseras istället på AI-systemets risknivå, oavsett om en AI-modell för allmänna ändamål är integrerad i systemet eller inte. Däremot har leverantör i efterföljande led rätt att ställa krav vad gäller information från leverantören av AI-modellen för allmänna ändamål.

Lärdomar

Arbetsgruppens lärdomar från iterationen är bland annat att det kan vara svårt att avgöra gränsdragningen mellan en AI-modell och ett AI-system. Det finns inte någon definition i AI-förordningen om vad som avses med en AI-modell.

Ytterligare lärdomar är att det kan vara svårt att på förhand veta vilka frågeställningar som kan bli aktuella att titta på i en AI-regulatorisk sandlåda. Det kan finnas utmaningar med att få fram all information för en bedömning av systemet i en AI-regulatorisk sandlåda, bland annat för att systemet ska utvecklas under pågående sandlåda och att den dokumentation som kan behövas för en beskrivning ännu inte finns tillgänglig. Det är inte alltid helt enkelt att bedöma vilken roll inblandade aktörer har. Det är endast leverantörer och potentiella leverantörer av AI-system som har möjlighet att söka till en AI-regulatorisk sandlåda. En ansökan kan dock ske i partnerskap med till exempel en tillhandahållare eller en leverantör av en AI-modell för allmänna ändamål.

En reflektion från arbetsgruppen är att det kan ifrågasättas hur harmoniserade de AI-regulatoriska sandlådorna kommer att vara och det kan finnas risk för diskrepans utifrån medlemsländernas kapacitet och fokus. En vanlig fråga som arbetsgruppen stött på är om det kan vara möjligt med samordnade sandlådor. Arbetsgruppen finner att sannolikt kommer det i så fall handla om olika sandlådor som samordnas snarare än en gemensam sandlåda för flera olika rättsområden.

Arbetsgruppen vill understryka behovet av tvärfunktionell kompetens, såsom jurister, AI-utvecklare, teknisk kompetens och projektledningskompetens, i en AI-regulatorisk sandlåda. Det finns uppenbara fördelar med ett kärnteam som arbetar tillsammans över tid. Ett arbetssätt där tolkningssvårigheter identifierats utifrån ett praktiskt fall ger ett större lärande än en teoretisk inläsning av lagstiftningen. En lärdom är att en aktör som ansöker till en sandlåda behöver bidra med visst eget deltagande och kompetens, särskilt vad gäller beskrivning av tilltänkt AI-system. Det behöver finnas ett stöd hos aktörens beslutsfattare och att de på något sätt är involverade. Det är en framgångsfaktor att dokumentera och producera skriven text under hela arbetet, för att tidigt uppmärksamma om det föreligger olika syn på vad arbetsgruppen har kommit fram till.

2. Inledning

Inom eSams ram drivs ett initiativ med att genomföra en pilot av en AI-regulatorisk sandlåda. Utgångspunkten för arbetet med piloten är att i iterationer utvärdera ett eller flera AI-system utifrån de krav som AI-förordningen[1] ställer. Målet är att förstå kraven på AI-system och göra en analys av vad som krävs för att upprätta en AI-regulatorisk sandlåda, inbegripet kompetenser, resurser, dokumentation och eventuellt teknisk infrastruktur.

Förhoppningen är att arbetet kan bidra till kunskap om hur AI-regulatoriska sandlådor bör inrättas och fungera i Sverige samt vilka förutsättningar som krävs för att aktörer ska kunna nyttja sådana sandlådor. Målen med de AI-regulatoriska sandlådorna är att främja AI-innovation och att öka rättssäkerheten.[2]

Den första iterationen finns att ta del av i eSams delrapport *AI-regulatorisk sandlåda – en första iteration*.[3]

Pilotens andra iteration har fokuerat på AI-system som innehåller AI-modeller för allmänna ändamål. Arbetet har genomförts av en arbetsgrupp bestående av deltagare från Bolagsverket, Skatteverket, Arbetsförmedlingen, Ekonomistyrningsverket, Riksarkivet, Integritetsskyddsmyndigheten och eSams kansli.

Rapporten beskriver lärdomar som arbetsgruppen gjort i samband med den andra iterationen av piloten.

2.1 Avgränsningar

AI-förordningens regler om regulatoriska sandlådor ska tillämpas från och med den 2 augusti 2026. Arbetet i piloten har avgränsats till hypotetiska frågor utifrån förordningen. Bedömningar av beskrivet system ska inte uppfattas som bindande eller slutliga.

Piloten tar inte upp frågor om personuppgiftsbehandling, dataskydd, cybersäkerhet och svensk nationell lagstiftning.

[1] Europaparlamentets och Rådets förordning (EU) 2024/1689 av den 13 juni 2024 om harmoniserade regler för artificiell intelligens och om ändring av förordningarna (EG) nr 300/2008, (EU) nr 167/2013, (EU) nr 168/2013, (EU) 2018/858, (EU) 2018/1139 och (EU) 2019/2144 samt direktiven 2014/90/EU, (EU) 2016/797 och (EU) 2020/1828 (förordning om artificiell intelligens).
[2] Skäl 139.
[3] ES2024-14 Delrapport AI-regulatorisk sandlåda – en första iteration.

3. Pilot – den andra iterationen

Arbetet med piloten av en AI-regulatorisk sandlåda bedrivs iterativt och utforskande. I det här avsnittet beskrivs genomförandet och resultatet av den andra iterationen.

3.1 Rättsliga förutsättningar

I delrapporten *AI-regulatorisk sandlåda – en första iteration*[4] redovisas rättsliga förutsättningar för regulatoriska sandlådor för AI (AI-regulatorisk sandlåda) bland annat om inrättande, deltagande, tillsyn och vidarebehandling av personuppgifter i en AI-regulatorisk sandlåda. Läsaren rekommenderas att ta del av den rättsliga redovisningen i den rapporten.

3.2 Metod och arbetssätt

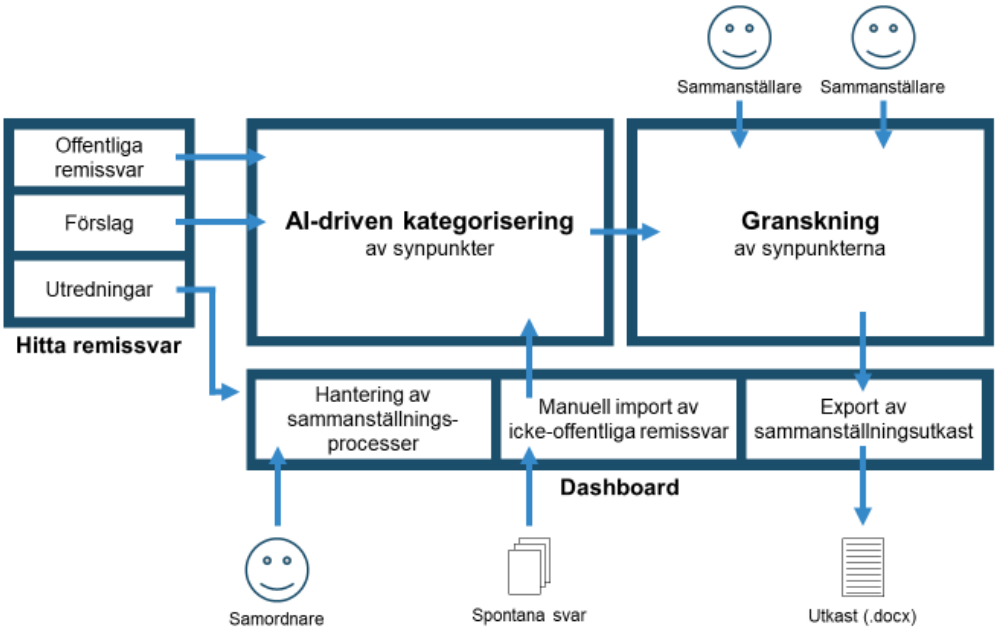
Den andra iterationen av piloten har, på samma sätt som den första iterationen, genomförts av en arbetsgrupp bestående av tvärfunktionell kompetens. Arbetsgruppen har genomfört sex halvdagsmöten där arbetsgruppen fått en beskrivning av AI-systemet och utifrån beskrivningen diskuterat frågeställningar utifrån framtagen frågematris, se bilaga 1.

3.3 Beskrivning av AI-systemet

Arbetsgruppen har i denna iteration utgått från ett system för remissammanställningar.

Utveckling av systemet sker i samarbete mellan Ekonomistyrningsverket och Regeringskansliet. Regeringskansliet genomför över 300 remissprocesser[5] varje år. Förenklat innebär processen att regeringen samlar in synpunkter på olika förslag. Svaren sammanställs sedan av handläggare vid Regeringskansliet i en remissammanställning. Det tar ca 14 dagar för en handläggare att producera en remissammanställning där 90 % av handläggningstiden avser att ”klippa och klistra” synpunkter från remissvar. Målet med projektet är att automatisera klipp-och-klistring med hjälp av AI.

[4] ES2024-14 Delrapport AI-regulatorisk sandlåda – en första iteration.
[5] Processen för remisser beskrivs i SB PM 2021:1 Svara på remiss Om remisser av betänkanden och andra förslag från Regeringskansliet.

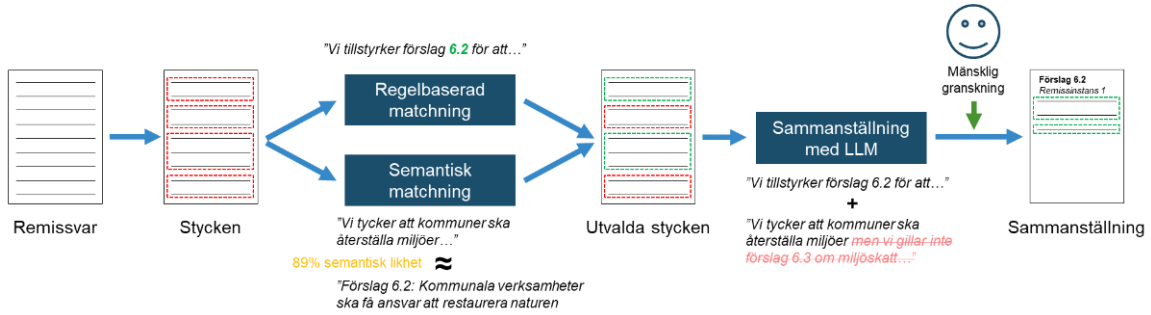


Övergripande skiss av systemet

Den AI-drivna kategoriseringen är tänkt att fungera enligt följande. För varje förslag i utredningen går systemet igenom alla remissvar och letar efter stycken i remissvaret som matchar (dvs. handlar om) förslaget. Matchningen sker dels regelbaserat (t.ex. matchar mot rubrik i förslagen från utredningen), dels semantiskt[6] (dvs. textdelar konverteras till vektorer och stycken som är mest semantiskt lika väljs ut). För den semantiska konverteringen används en språkmodell från Kungliga biblioteket, en s.k. KB-BERT-modell[7].

Styckena sammanställs och språkmodellen GPT-4o[8] bedömer utifrån särskilda instruktioner om dessa är riktiga synpunkter och extraherar, rensar och slår ihop relevanta meningar. Därmed skapas en remissammanställning.

[6] Semantisk likhet är en metod för att jämföra två textdelar som inte överensstämmer ord för ord men som har en liknande betydelse, för jämförelsen behöver de båda textdelarna representeras som vektorer.
[7] Contrastive-Tension/BERT-Base-Swe-CT-STSb · Hugging Face
[8] GPT-4o ("o" för omni) är en flerspråkig, multimodal generativ förtränad transformer, utvecklad av OpenAI och släppt i maj 2024. GPT-4o brukar i vardagligt tal benämnas som en stor språkmodell (eng. *Large Language Model* (LLM)).



Skiss på AI-kategoriseringen

Från den AI-drivna kategoriseringen skickas data till en webbapplikation som visar resultatet, en s.k. granskningsapplikation, där resultatet genomgår en manuell granskning. Det ska också finnas en funktion för att lägga till manuella svar (t.ex. spontansvar), som inte finns publicerade men som ändå bör beaktas. Granskningsapplikationen bygger på Label Studio, som är ett annoteringsverktyg med öppen källkod.

Det kommer att tas fram rutiner kring arbetssätt för att underlätta personalens användning av AI-systemet.

3.4 Bedömning

3.4.1 Är det ett AI-system?

Arbetsgruppen har i den första iterationen av piloten konstaterat att det är troligt att de flesta it-system som inte är helt regelbaserade faller in under definitionen av AI-system enligt AI-förordningen. Arbetsgruppen har därtill konstaterat att det avsedda ändamålet bör vara centralt för bedömningen av ramen för vad som ska anses utgöra AI-systemet.

I den första iterationen konstaterade arbetsgruppen att en central utmaning vad gäller AI-förordningens tillämpning är att dra gränsen för vad som utgör ett AI-system, när ett och samma system innehåller flera AI-komponenter som skulle kunna betraktas som enskilda AI-system. Avgörande för bedömningen bör vara syftet med AI-systemet och vilka komponenter som behövs för att systemet ska kunna uppnå syftet. För gränsdragning av vad som ska anses utgöra AI-systemet konstaterades också i första iterationen att viss ledning möjligen kan erhållas från beskrivningen av skillnaden mellan AI-modeller för allmänna ändamål och AI-system i förordningens skäl 97. Där anges bland annat att begreppet AI-modeller för allmänna ändamål ska särskiljas från begreppet AI-system, eftersom sådana modeller förvisso kan utgöra väsentliga komponenter i AI-system, men inte utgör AI-system i sig. Enligt skäl 97 är AI-modeller vanligtvis integrerade i och utgör en del av ett AI-system. För att det ska röra sig om ett

AI-system krävs ytterligare komponenter (utöver en AI-modell), t.ex. ett användargränssnitt.

Syftet med systemet i denna iteration är att ge en handläggare en sammanställning av remissvar. Det bör innebära att det inte går att särskilja AI-kategoriseringen för sig och betrakta funktionen som ett eget AI-system, eftersom den funktionen inte har ett eget syfte eller användargränssnitt utan tjänar en helhetslösning. Utifrån detta bedömer arbetsgruppen att hela lösningen bör betraktas som ett AI-system enligt AI-förordningen.

För redogörelse av artiklar och skäl hänförliga till definitionens rekvisit hänvisas till genomgången i eSams delrapport *AI-regulatorisk sandlåda – en första iteration*. Arbetsgruppen har utifrån det avsedda ändamålet gjort en rekvisitgenomgång av definitionen för AI-system i artikel 3.1 för att besvara om rekvisiten är uppfyllda för det system som ska bedömas, enligt följande:

AI-system:

- ett maskinbaserat system (Ja, systemet existerar i maskiner.)
- som är utformat för att fungera med varierande grad av autonomi (Ja, här bör ordet *varierande* vara av betydelse. Förekomsten av mänsklig interaktion förtar inte systemets autonomi. Systemet gör automatiska prediktioner, vilket leder till utfall som en människa kan godkänna.)
- och som kan uppvisa anpassningsförmåga efter införande (Ja, här bör ordet *kan* vara av betydelse. I skäl 12 anges att ”den anpassningsförmåga som ett AI-system kan uppvisa när det införts avser förmågan till självlärande, vilket gör det möjligt för systemet att förändras under sin användning.” Båda ingående språkmodeller har förmåga till självlärande. Bedömningen är därmed att AI-systemet kan uppvisa anpassningsförmåga efter införande. Här skulle ett mer generellt resonemang kunna vara att allt som inte är regelbaserat har anpassningsförmåga, dvs. skäl 12 ger en bred tillämpning.)
- och som, för uttryckliga eller underförstådda mål, (Ja, det finns uttryckliga mål för vad systemet ska göra.)
- drar slutsatser härledda från den indata det tar emot, (Ja, i nuläget består indata endast av remissvar i pdf-format, men på sikt kommer indata även kunna hämtas från granskningsprocessen.)
- om hur utdata såsom förutsägelser, innehåll, (Ja, systemet skapar innehåll.) rekommendationer (Nej, systemet ger inga rekommendationer till slutanvändare.) eller beslut (Nej, inga beslut.) som kan påverka fysiska eller

virtuella miljöer ska genereras. (Ja, systemet drar slutsatser om hur utdata, såsom innehåll, som kan påverka fysiska eller virtuella miljöer, ska genereras.)

Utifrån ovan finner arbetsgruppen att det beskrivna systemet kan anses vara ett AI-system utifrån AI-förordningens definition.

3.4.2 Risknivå?

Beroende på vilken risknivå ett AI-system kategoriseras som, omfattas systemet av olika krav. Risknivåerna enligt AI-förordningen brukar delas in i följande:

- Oacceptabel risk (förbjudna)
- Hög risk (hårt reglerade)
- Begränsad risk (transparenskrav)
- Minimal eller ingen risk (AI-förordningen ställer inga krav)

Risknivån oacceptabel risk avser förbjudna AI-användningsområden enligt artikel 5 i förordningen, vilka kortfattat kan beskrivas som förbud mot utsläppande på marknaden, ibruktagande eller användning av AI system som gör något av följande:

- använder subliminala eller vilseledande tekniker med syfte att påverka mänskligt beslutsfattande
- utnyttjar sårbarheter hos vissa särskilt sårbara personer eller samhällsgrupper med målet att väsentligt påverka deras beteende
- utvärderar eller klassificerar fysiska personer eller grupper av personer för social poängsättning
- utför riskbedömning grundat uteslutande på en viss form av profilering, i syfte att förutse om en fysisk person kommer att begå brott
- skapar eller utvidgar databaser för ansiktsigenkänning genom oriktad skrapning av ansiktsbilder
- uttyder fysiska personers känslor på arbetsplatser eller vid utbildningsinstutitioner
- utför biometrisk kategorisering av fysiska personer.

Enligt artikel 5 h är användning av system för biometrisk fjärridentifiering i realtid på allmänt tillgängliga platser för brottsbekämpande ändamål som huvudregel förbjudet, men får användas för vissa utpekade syften och under vissa förutsättningar.

Arbetsgruppen finner att det inte är fråga om ett sådant användande som regleras i artikel 5 och det aktuella systemet utgör därmed inte ett AI-system med oacceptabel risk.

3.4.2.1 Hög risk

Riskerna med AI-system kan ha att göra med hur systemen utformas, men de kan även härröra från hur AI-systemen används.[9]

Ett AI-system som är avsett att användas som en säkerhetskomponent i en produkt, eller om AI-systemet i sig är en produkt, som omfattas av unionens harmoniseringslagstiftning enligt förteckningen i bilaga I och den aktuella produkten behöver genomgå en tredjepartsgranskning för att få släppas ut på marknaden eller tas i bruk är att anse som ett högrisksystem. Därtill är AI-system som avses i bilaga III i AI-förordningen också att betrakta som högrisksystem.[10] I artikel 6.3 finns ytterligare reglering som anger när ett AI-system, trots att det omfattas av bilaga III, inte ska betraktas som ett högrisksystem. Det gäller om systemet inte utgör en betydande risk för skada på fysiska personers hälsa, säkerhet eller grundläggande rättigheter, inbegripet att det inte väsentligt påverkar resultatet av beslutsfattandet. AI-system som avses i bilaga III ska dock alltid anses utgöra högrisksystem om AI-systemet utför profilering av fysiska personer.[11]

För tilltänkt system för remissammanställningar kan regleringen i punkt 8 i bilaga III om rättskipning och demokratiska processer vara relevant att utforska. I punkten 8 a utpekas som högrisksystem: ”AI-system som är avsedda att användas av en rättslig myndighet eller på dess vägnar för att hjälpa en rättslig myndighet att undersöka och tolka fakta och lagstiftning och att tillämpa lagen på konkreta fakta, eller som är avsedda att användas på ett liknande sätt i alternativa tvistlösningar.” I punkten 8 b utpekas som högrisksystem: ”AI-system som är avsedda att användas för att påverka resultatet av ett val eller en folkomröstning eller fysiska personers röstningsbeteenden när de utövar sin rätt att rösta i val eller folkomröstningar.”

Klassificeringen av ett AI-system som ett AI-system med hög risk bör, enligt skäl 61, ”inte omfatta AI-system som är avsedda för rent administrativa stödfunktioner som inte påverkar den faktiska rättskipningen i enskilda fall, exempelvis anonymisering eller pseudonymisering av rättsliga beslut, handlingar eller data, kommunikation mellan anställda, administrativa uppgifter.”

Som arbetsgruppen konstaterat i första iterationen bör det vara syftet med systemet som styr bedömningen. Syftet med systemet i aktuellt fall är att tillhandahålla ett stöd för Regeringskansliet att sammanställa remissvar.

[9] Skäl 93.
[10] Artikel 6.
[11] Artikel 6.3.

Begreppet ”rättslig myndighet” i punkt 8 a i bilaga III finns inte definierat i förordningen och en fråga att ta ställning till är om Regeringskansliet är att anse som en rättslig myndighet. Sett till rubriken verkar bestämmelsen ta sikte på rättskipande myndighet, t.ex. en domstol. Vid en jämförelse mellan olika språkversioner[12], verkar det också vara dömande verksamhet som avses. Rättspraxis från EU-domstolen ger också stöd för en sådan tolkning, där EU-domstolen uttalat att ”med uttrycket rättsliga myndigheter avses traditionellt myndigheter som deltar i rättskipning, till skillnad från bland annat förvaltningsmyndigheter, vilka är en del av den verkställande makten”.[13] Det stöder arbetsgruppens uppfattning att bestämmelsen inte omfattar myndigheter som Regeringskansliet. I skäl 61 talas dessutom om undantag där rättskipningen inte påverkas, dvs. avsikten verkar vara att punkt 8 a tar sikte på rättskipande aktivitet. Inte heller nästa led i punkt 8 a, ”för att hjälpa en rättslig myndighet att undersöka och tolka fakta och lagstiftning och tillämpa lagen på konkreta fakta”, träffar beskrivet AI-system. Dessutom avser undantaget i skäl 61 rent aministrativa stödfunktioner, vilket aktuellt AI-system får anses falla inom.

Punkt 8 b anses inte tillämplig i det här fallet eftersom bestämmelsen tar sikte på val och folkomröstningar, inom vilket remissammanställningsprocessen inte ryms.

Arbetsgruppens bedömning, utifrån rekvisiten i punkten 8 i bilaga III samt undantaget i skäl 61, är att AI-systemet inte faller inom bestämmelsen och systemet utgör därmed inte ett högrisksystem. Arbetsgruppen har även övervägt övriga punkter om högrisksystem i bilaga III, men inte funnit att de punkterna är tillämpliga i det aktuella fallet.

3.4.2.2 Transparenskrav

Arbetsgruppen har vidare tittat på om det föreligger några transparenskrav enligt artikel 50. Transparenskraven kan gälla oaktat om systemet betraktas som ett högrisksystem eller inte.

Enligt artikel 50.1 ska leverantörer ”säkerställa att AI-system som är avsedda att interagera direkt med fysiska personer utformas och utvecklas på ett sådant sätt att de berörda fysiska personerna informeras om att de interagerar med ett AI-system, såvida det inte är uppenbart för en fysisk person som är normalt informerad och skäligen uppmärksam och medveten, med beaktande av användningens omständigheter och sammanhang.”

[12] I det här fallet engelskans ”judicial authority” och tyskans ”Justizbehörde”.
[13] C-16/22 punkt 35 och C-452/16.

En frågeställning är vilken omfattning artikeln har, dvs. vad som innefattas i begreppet ”interagera direkt med fysiska personer” och om det går att säga att det i det här fallet är uppenbart för användarna att det är fråga om en interaktion med ett AI-system.

I skäl 132 anges att vissa AI-system som är avsedda för att interagera med fysiska personer eller generera innehåll, kan utgöra särskilda risker för identitetsmissbruk eller vilseledning oavsett om de kategoriseras som AI-system med hög risk eller inte. Under vissa omständigheter bör därför användningen av sådana system omfattas av särskilda transparensskyldigheter. Det som kännetecknar fysiska personer som tillhör sårbara grupper på grund av ålder eller funktionsnedsättning bör beaktas i den mån AI-systemet även är avsett att interagera med dessa grupper.

I första hand verkar begreppet ”interagera” ta sikte på något mer än att bara använda ett AI-system, såsom kommunikation med en chattbot.[14] När exempel ges kring transparenskravet är det ofta i situationen att personer ska kunna förstå att det är en AI-baserad chattbot på andra sidan och att svaren som ges är AI-genererade. Även expertgruppen på hög nivå för AI-frågor verkar i sin guide Etiska riktlinjer för tillförlitlig AI ta sikte på just dialogformen.[15] Det kan emellertid inte uteslutas att även annan interaktion med ett AI-system skulle kunna omfattas av bestämmelsen.

Aktuellt AI-system för remissammanställningar ska användas av anställda i arbetet. Användargränssnittet är inte baserat på en chattfunktion. Det är tveksamt om denna situation faller inom begreppet ”interagera”. Till skillnad från en AI-baserad chattbot kan det inte framstå som att svaren från det aktuella AI-systemet genereras av någon annan än en maskin. Det får också anses uppenbart för användarna, som är anställda på myndigheten, att det rör sig om ett AI-system. Användningen bygger på att de anställda är involverade i granskningsprocessen av AI-systemets utfall. Arbetsgruppen bedömer att artikel 50.1 inte bör vara tillämplig i aktuellt fall. Det kan dock finnas ett behov av att förtydliga punktens omfattning. Sista meningen i artikel 50.1 tar sikte på brottsbekämpande system och bedöms inte vara aktuell i detta fall.

Enligt artikel 50.2 ska leverantörer av AI-system som genererar syntetiskt ljud-, bild-, video- eller textinnehåll säkerställa att AI-systemets utdata är märkta i ett maskinläsbart format och kan upptäckas som artificiellt genererade eller manipulerade. En frågeställning är vad som menas med syntetisering. I aktuellt fall kan AI-systemet klippa en mening och fylla i en mening, men väsentliga indata ändras inte. Det bör därmed inte

[14] Se exempelvis Kop, Mauritz, *EU Artificial Intelligence Act: The European Approach to AI*, Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 2/2021, sida 5 och Şimşek, Can, *Regulating Artificial Intelligence: Could the EU's "AI Act" lead the way forward?,* SciencePo.
[15] Se t.ex. sid 29 Bedömningslista för tillförlitlig AI, 36 Sociala konsekvenser och 40 Dolda AI-system i Ethics guidelines for trustworthy AI | Shaping Europe's digital future

vara fråga om syntetisering. Systemet ”plockar” snarare än sammanfattar. I andra punkten anges vidare att skyldigheten inte är tillämplig i den mån AI-systemen utför en hjälpfunktion för vanlig redigering eller inte väsentligt ändrar de indata som tillhandahålls av tillhandahållaren eller deras semantik, eller om systemen enligt lag får upptäcka, förebygga, förhindra, utreda eller lagföra brott.[16] Arbetsgruppen bedömer att artikel 50.2 inte är aktuell då det inte är fråga om syntetiska data och att systemet snarare utgör en redigeringsfunktion som inte väsentligt ändrar indata.

Artikel 50.3 avser system för känsloigenkänning eller system för biometrisk kategorisering, vilket det inte är fråga om i aktuellt fall.

Artikel 50.4 första stycket avser AI-system som genererar eller manipulerar bild-, ljud- eller videoinnehåll som utgör en deepfake, vilket det inte är fråga om i aktuellt fall. Andra stycket gäller för tillhandahållare av ett AI-system som genererar eller manipulerar text som offentliggörs i syfte att informera allmänheten om frågor av allmänt intresse, då ska tillhandahållaren upplysa om att texten har genererats artificiellt eller manipulerats. En fråga är vad som ryms i begreppet ”genererar”, det bör även innefatta att sammanställa. Det är arbetsgruppens erfarenhet att remissammanställningarna inte publiceras, men det kan inte uteslutas att så sker. I sista meningen i andra stycket anges att skyldigheten inte är tillämplig om det AI-genererade innehållet har genomgått en process med mänsklig granskning eller redaktionell kontroll och om en fysisk eller juridisk person bär det redaktionella ansvaret för offentliggörandet av innehållet, se även skäl 134. I aktuellt fall kommer en fysisk person att granska sammanställningen innan den blir slutlig. Arbetsgruppen gör bedömningen att artikel 50.4 därför inte är tillämplig.

Arbetsgruppen gör sammanfattningsvis bedömningen att aktuellt AI-system inte omfattas av transparenskraven i artikel 50. Gruppen finner därmed att aktuellt system bör vara ett inget/minimal risk-system. För sådana system finns inga krav på att ge information och inte heller några andra krav på aktörerna. Arbetsgruppen konstaterar samtidigt att det föreligger en otydlighet i hur ”interaktion” ska bedömas. För en myndighet kan det i sådana fall finnas anledning att överväga att ändå informera om AI-systemet, men då på frivillig basis eller utifrån eventuella frivilliga uppförandekoder.[17]

3.4.3 Roller

Utvecklingen av aktuellt system är ett samarbete mellan Ekonomistyrningsverket och Regeringskansliet. En central fråga är vilken roll respektive aktör får enligt AI-

[16] Se även skäl 133.
[17] Artikel 95.

förordningen. Arbetsgruppen har bedömt att de roller som är aktuella att analysera i detta sammanhang är rollerna ”leverantör” och ”tillhandahållare”.

Rollen *leverantör* definieras i artikel 3.3 som ” en fysisk eller juridisk person, en offentlig myndighet, en byrå eller ett annat organ som utvecklar ett AI-system eller en AI-modell för allmänna ändamål eller som har ett AI-system eller en AI-modell för allmänna ändamål och släpper ut det eller den på marknaden eller tar AI-systemet i bruk i eget namn eller under eget varumärke, antingen mot betalning eller kostnadsfritt”.

En *tillhandahållare* är enligt artikel 3.4 ”en fysisk eller juridisk person, offentlig myndighet, en byrå eller annat organ som under eget överinseende använder ett AI-system, utom när AI-systemet används inom ramen för en personlig icke-yrkesmässig verksamhet”.

3.4.3.1 Leverantör

Ekonomistyrningsverket och Regeringskansliet är *offentliga myndigheter*, som är en aktör som räknas upp i leverantörsdefinitionen.

3.4.3.1.1 Utvecklar eller har ett AI-system

Vad gäller kriteriet *som utvecklar ett AI-system* kan både Ekonomistyrningsverket och Regeringskansliet anses vara delaktiga i utvecklingen av AI-systemet. I det här skedet av utvecklingen står Ekonomistyrningsverket för merparten av den tekniska utvecklingen. Regeringskansliet bistår med annoteringsdata och står för kravbilden. I dagsläget är det oklart vilken roll Ekonomistyrningsverket kommer att få inför en implementering och användande av systemet på Regeringskansliet och det är inte klart vem som kommer att förvalta systemet när det väl sätts i drift. Det är dock Regeringskansliet som kommer att använda systemet i sin verksamhet.

Vid en jämförelse av den svenska versionen av AI-förordningen och andra språkversioner konstaterar arbetsgruppen att andra språkversioner[18] uttrycker att en leverantör är någon som ”utvecklar eller låter utveckla” ett AI-system, till skillnad från den svenska versionen som enbart uttrycker att en leverantör är någon ”som utvecklar” ett AI-system. Arbetsgruppen finner att i ljuset av andra språkversioner får innebörden av leverantörsdefinitionen anses innefatta såväl en aktiv utveckling som när en aktör uppdrar åt någon annan att utveckla AI-systemet för aktörens räkning. En fråga att ta ställning till är på vems initiativ AI-systemet utvecklas.

Det saknas ett reglerat ansvarsförhållande mellan Regeringskansliet och Ekonomistyrningsverket, men det finns en överenskommelse mellan myndigheterna att

[18] Till exempel tyskans ”entwickelt oder entwickeln lässt“, franskans ” qui développe ou fait développer”, engelskans ” that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed” och danskans ” der udvikler eller får udviklet”.

lösningen ska tas fram tillsammans. Ekonomistyrningsverket är den part som utför merparten av utvecklingen. Det ingår i Ekonomistyrningssverkets uppdrag att tillhandahålla it-lösningar till Regeringskansliet, men det här projekt har ingen direkt koppling till det uppdraget. Det faktum att Ekonomistyrningsverket utvecklar AI-systemet specifikt för Regeringskansliets behov kan tolkas som att Regeringskansliet är beställare och att Ekonomistyrningsverket då utvecklar systemet på Regeringskansliets uppdrag. Arbetsgruppen finner att det går att argumentera i båda riktningarna, dvs. att såväl Ekonomistyrningsverket som Regeringskansliet skulle kunna anses vara den som utvecklar AI-systemet.

En aktör kan även vara leverantör om den har ett AI-system. Definitionen av leverantör har alltså två alternativa rekvisit i första delen av meningen, dvs. antingen utvecklar ett AI-system eller har ett AI-system. De två alternativa rekvisiten förutsätter dessutom att aktören släpper ut AI-systemet på marknaden eller tar AI-systemet i bruk i eget namn eller under eget varumärke, antingen mot betalning eller kostnadsfritt. Även här är det fråga om två alternativa rekvisit som behöver värderas.

3.4.3.1.2 Utsläppande på marknaden

Utsläppande på marknaden definieras i artikel 3.9 som den första gången ett AI-system tillhandahålls på unionsmarknaden. *Tillhandahållande på marknaden* definieras i artikel 3.10 som leveransen av ett AI-system för distribution eller användning på unionsmarknaden i samband med kommersiell verksamhet, mot betalning eller kostnadsfritt.

En frågeställning är vad som avses med *kommersiell verksamhet* i artikel 3.10 och om en myndighet kan anses syssla med kommersiell verksamhet. Kommersiell verksamhet definieras inte i AI-förordningen, men en jämförelse med andra språkversioner[19] ger att begreppet verkar motsvara näringsverksamhet eller affärsverksamhet. Det finns definitioner i produkt- och marknadskontrollagstiftningen som påminner om artikel 3.10, så viss ledning av betydelsen av begreppet bör kunna hämtas från denna lagstiftning. I EU:s rambeslut om en gemensam ram för saluföring av produkter[20] återfinns i bilaga 1 artikel R1 punkt 1 och 2 liknande definitioner som AI-förordningens definitioner av utsläppande på marknaden och tillhandahållande. Även i EU:s marknadskontrollförordning[21] artikel 3.1 finns liknande skrivningar, där tillhandahållande

[19] Se till exempel danskans ”levering af et AI-system eller en AI-model til almen brug med henblik på distribution eller anvendelse på EU-markedet *som led i erhvervsvirksomhed* mod eller uden vederlag”, tyskans ”die entgeltliche oder unentgeltliche Abgabe eines KI-Systems oder eines KI-Modells mit allgemeinem Verwendungszweck *zum Vertrieb oder zur Verwendung auf dem Unionsmarkt im Rahmen einer Geschäftstätigkeit*" och franskans ”la fourniture d’un système d’IA ou d’un modèle d’IA à usage général destiné à être distribué ou utilisé sur le marché de l’Union *dans le cadre d’une activité commerciale*, à titre onéreux ou gratuit”

[19] Prop. 1989/90:89 s. 60.

[20] Europaparlamentets och rådets beslut nr 768/2008/EG av den 9 juli 2008 om en gemensam ram för saluföring av produkter och upphävande av rådets beslut 93/465/EEG.

[21] Europaparlamentets och rådets förordning (EU) 2019/1020 av den 20 juni 2019 om marknadskontroll och överensstämmelse för produkter och om ändring av direktiv 204/42/EG och förordningarna (EG) nr 765/2008 och (EU) nr 305/2011.

på marknaden definieras som leverans av en produkt för distribution, förbrukning eller användning på unionsmarknaden i samband med kommersiell verksamhet, mot betalning eller gratis.

I EU-kommissionens blåbok om genomförandet av EU:s produktbestämmelser från 2022[22] anges att kommersiell verksamhet innebär tillhandahållande av varor i affärssammanhang och att icke-vinstdrivande organisationer kan anses bedriva kommersiell verksamhet om de är verksamma inom ett sådant sammanhang. Bedömningen kan enligt EU-kommissionen endast göras från fall till fall genom att ta hänsyn till leveransernas frekvens, produktens egenskaper, leverantörens avsikter osv.

Arbetsgruppen noterar en grundläggande skillnad mellan produktlagstiftningarna och AI-förordningen, där produktlagstiftningarna som huvudprincip inte avser att omfatta offentliga myndigheter. Det finns dock vissa avgöranden som ger att även om svenska myndigheter inte är vinstdrivande organisationer, kan de i vissa fall anses tillhandahålla varor och tjänster i affärssammanhang. Det finns till exempel domstolsavgöranden om miljösanktionsavgifter där vissa myndighetsverksamheter ansetts utgöra näringsverksamhet och andra inte.[23]

Enligt definitionen av leverantör i AI-förordningen är det tydligt att offentliga myndigheter är tänkta att omfattas av förordningen. Ett resonemang skulle, utifrån praxis om produktlagstiftningen, kunna vara att när offentliga myndigheter tillgängliggör AI-system så att andra kan använda systemen i sina verksamheter, så verkar myndigheten inom en konkurrerande verksamhet och det skulle därmed kunna anses ske inom ramen för en kommersiell verksamhet. Gruppen ser dock inte att det är en självklar tolkning. Frågan är om det för offentliga myndigheter är tänkt att det ska ställas krav på att utsläppandet ska ske i ett affärssammanhang, i analogi med vad som gäller enligt produktlagstiftningen, för att utsläppandet av myndigheten ska omfattas av AI-förordningen eller om offentliga myndigheter kan anses släppa ett AI-system på marknaden enligt leverantörsdefinitionen trots rekvisitet om kommersiell verksamhet i artikel 3.10.

Sammanfattningsvis anser gruppen att det i dagsläget råder osäkerhet kring hur artikel 3.10 ska tolkas och tillämpas, men gruppen kan samtidigt konstatera att offentliga myndigheter ingår i AI-förordningens definitioner av leverantör respepektive tillhandahållare. Arbetsgruppen gör utifrån detta bedömningen att offentliga myndigheter kan vara en sådan leverantör som släpper ut ett AI-system på marknaden. En annan tolkning skulle innebära att offentliga myndigheter i stor utsträckning skulle

[22] Meddelande från Kommissionen 2022 års blåbok om genomförandet av EU:s produktbestämmelser (2022/C 247/01), sid 19.
[23] Se SOU 2004:37 s. 90.

vara undantagna från AI-förordningens tillämpningsområde när de tillgängliggör AI-system för andra att använda i sina verksamheter. Gruppens uppfattning är att det inte varit avsikten med regleringen i AI-förordningen.

Kravet på samband med kommersiell verksamhet bör innebära att tillhandahållandet ska ske i någon form av konkurrens med andra tillhandahållare. Om Regeringskansliet anses vara den som utvecklar och använder systemet bör det inte vara fråga om tillhandahållande i samband med kommersiell verksamhet och därmed är det inte ett utsläppande på marknaden. Om istället Ekonomistyrningsverket anses vara den som utvecklar systemet för användning uteslutande av Regeringskansliet, är det mer svårbedömt om denna situation är ett tillhandahållande i samband med kommersiell verksamhet och därmed ett utsläppande på marknaden. Arbetsgruppen lutar åt att situationen att Ekonomistyrningsverket inom ett samarbete levererar ett AI-system till Regeringskansliet, sannolikt inte bör ses som ett sådant utsläppande på marknaden som avses i artikel 3.3. Enligt arbetsgruppen är det dock önskvärt med ett förtydligande av hur AI-förordningen ska tillämpas i förhållande till myndigheternas verksamheter.

3.4.3.1.3 Ibruktagande

Ibruktagande definieras i artikel 3.11 som leverans av ett AI-system för första användning direkt till tillhandahållaren eller för eget bruk i unionen för dess avsedda ändamål.

Ekonomistyrningsverket levererar AI-systemet direkt till Regeringskansliet för första användning och det skulle kunna ses som ett ibruktagande. Enligt leverantörsdefinitionen i artikel 3.3 krävs emellertid att systemet tas i bruk ”i eget namn eller under eget varumärke”. Det är oklart om Ekonomistyrningsverkets leverans av systemet kan anses ske i eget namn eller under eget varumärke. I så fall kan Ekonomistyrningsverket anses vara den som tar systemet i bruk, i annat fall uppfyller Ekonomistyrningsverket inte det rekvisitet.

Regeringskansliet kan anses vara den aktör som tar AI-systemet i bruk genom att det är hos Regeringskansliet systemet kommer att användas för dess avsedda ändamål. Mycket talar för att Regeringskansliet är att anse som leverantör, det kan dock inte uteslutas att Ekonomistyrningsverket anses omfattas av leverantörsdefinitionen.

3.4.3.1.4 Gemensamt leverantörsansvar?

En intressant fråga är om AI-förordningen ger utrymme för ett gemensamt leverantörsansvar, likt det gemensamma personuppgiftsansvaret som två eller fler aktörer kan ha enligt artikel 26 i GDPR. AI-förordningens regler om operatörers skadeståndsskyldighet ger inte uttryck för att flera aktörer skulle kunna vara gemensamt

skadeståndsansvariga, och förordningen innehåller inga särskilda regler om regress.[24] Till skillnad från GDPR, enligt vilken flera aktörer kan vara gemensamt personuppgiftsansvariga, saknar AI-förordningen bestämmelser om möjligheten till gemensamt ansvar samt möjligheten att välja vilken aktör som rättigheter kan utkrävas från. Därutöver benämns de olika rollerna i förordningen som huvudregel i singular. Arbetsgruppen har även jämfört med produktlagstiftningens ansvarsreglering där det inte verkar föreligga en reglering om gemensamt ansvar, tvärtom verkar det förutsättas att det tydligt utpekas vilken aktör som bär ansvaret. Mot bakgrund av ovan framförda konstaterar arbetsgruppen att det förmodligen inte är möjligt för flera aktörer att ha ett gemensamt leverantörsansvar enligt AI-förordningen. Ett parallellt ansvar kan dock finnas för andra aktörer såsom importörer eller distributörer.

3.4.3.2 Tillhandahållare

Gruppen finner att Regeringskansliet troligtvis är tillhandahållare, eftersom AI-systemet är anpassat efter och kommer att användas inom myndighetens verksamhet. Intressant i sammanhanget är dock rekvisitet ”under eget överinseende”, eftersom det i dagsläget är oklart i vilken utsträckning Ekonomistyrningsverket kommer att stötta Regeringskansliet gällande drift, förvaltning men även användning av systemet. Om Regeringskansliet över tid kommer använda systemet självständigt utan stöd av Ekonomistyrningsverket, talar det för att Regeringskansliet är tillhandahållare.

Det skulle innebära att Regeringskansliet i praktiken kan bli både leverantör och tillhandahållare av systemet, och därmed omfattas av de krav som kan bli tillämpliga på respektive roll.[25]

3.4.3.3 Summering om roller

Gruppen konstaterar att AI-förordningen, gällande reglerna om rollfördelning, är skriven utifrån en kommersiell marknad där ansvarsfördelning ofta regleras genom avtal snarare än genom samverkan mellan myndigheter.

Utifrån det samarbete som föreligger mellan Ekonomistyrningsverket och Regeringskansliet är det inte självklart vilken roll respektive aktör har enligt AI-förordningen. Vid en jämförelse med annan produktlagstiftning verkar dock tanken vara att det ska finnas en utpekad leverantör. Arbetsgruppens reflektion är att rollerna behöver tydliggöras innan ansökan till en AI-regulatorisk sandlåda, se avsnitt 4.2 om vem som har rätt att ansöka till sådan sandlåda.

[24] Jämför artikel 82.4 – 82.5 i dataskyddsförordningen.
[25] Skäl 83.

3.4.4 Är det en AI-modell för allmänna ändamål?

AI-förordningen innehåller särskilda regler för AI-modeller för allmänna ändamål. Syftet med att definiera sådana AI-modeller och särskilja dem från begreppet AI-system är att skapa rättsäkerhet.[26]

I avsnitt 3.4.1 gör arbetsgruppen bedömningen att hela lösningen anses utgöra ett AI-system, dvs. att de två språkmodellerna KB-BERT-modellen och GPT-4o är en del av AI-systemet. Arbetsgruppen finner därmed anledning att bedöma dessa modeller utifrån rekvisiten i definitionen i artikel 3.63, vilken lyder:

- en AI-modell,
- även när en sådan AI-modell tränas med en stor mängd data med hjälp av självövervakning i stor skala,
- som uppvisar betydande generalitet och
- på ett kompetent sätt kan utföra ett brett spektrum av distinkta uppgifter
- oavsett hur modellen släppts ut på marknaden och
- som kan integreras i en rad system eller tillämpningar i efterföljande led,
- utom AI-modeller som används för forsknings-, utvecklings- eller prototypverksamhet innan de släpps ut på marknaden.

Definitionen baseras på de viktigaste funktionella egenskaperna: generaliteten och förmågan att på ett kompetent sätt utföra ett stort antal olika uppgifter. Dessa modeller tränas vanligtvis med stora mängder data genom olika metoder, såsom självövervakad inlärning. AI-modeller för allmänna ändamål får släppas ut på marknaden på olika sätt, bland annat genom bibliotek, gränssnitt för applikationsprogrammering, som direkt nedladdning eller som fysisk kopia.[27]

Skyldigheterna för leverantörer av AI-modeller för allmänna ändamål gäller när sådana modeller släpps ut på marknaden. När leverantören av en AI-modell för allmänna ändamål integrerar en egen modell i sitt AI-system som tillhandahålls på marknaden eller tas i bruk, bör den modellen anses ha släppts ut på marknaden, och skyldigheterna i AI-förordningen för modeller bör därför fortsätta att gälla utöver skyldigheterna för AI-system. De skyldigheter som föreskrivs för modeller gäller inte när en egen modell används för rent interna processer som inte är väsentliga för att tillhandahålla en produkt eller tjänst till tredje parter och fysiska personers rättigheter inte påverkas.

Det första rekvisitet i artikel 3.63 är att det ska vara fråga om en AI-modell. Arbetsgruppen konstaterar att begreppet ”AI-modell” inte finns ytterligare definierat i

[26] Skäl 97.
[27] Skäl 97.

AI-förordningen och det finns inte heller några bestämmelser som tar sikte på AI-modeller som inte är för allmänna ändamål. Gruppen finner att det bör vara rimligt att utgå från hur en AI-modell i allmänhet beskrivs, dvs. att en algoritm plus träningsdata ger en AI-modell.[28] Utifrån ett sådant synsätt är både KB-BERT-modellen och GPT-4o AI-modeller. Arbetsgruppen finner att båda modellerna också har tränats på stora mängder data med hjälp av självövervakning i stor skala.[29] Båda modellerna kan också integreras i en rad system eller tillämpningar i efterföljande led.

3.4.4.1 Betydande generalitet och ett brett spektrum

Det finns ingen definition av vad som är betydande generalitet eller ett brett spektrum av distinkta uppgifter. Visst stöd för bedömningen finns i skäl 98 och 99.

Skäl 98 anger att en AI-modells generalitet bland annat kan bestämmas av ett antal parametrar, men modeller med minst en miljard parametrar som tränats med en stor mängd data med hjälp av självövervakning i stor skala bör anses uppvisa betydande generalitet och på ett kompetent sätt utföra ett brett spektrum av olika uppgifter. Gruppen finner att skäl 98 nog bör ses som ett exempel och inte utesluta att AI-modeller med mindre antal parametrar ändå kan uppfylla rekvisitet betydande generalitet och ett brett spektrum av uppgifter.

Skäl 99 anger att stora generativa AI-modeller är ett typiskt exempel på en AI-modell för allmänna ändamål, eftersom de möjliggör flexibel generering av innehåll, exempelvis i form av text, ljud, bilder eller video, som lätt kan rymma ett brett spektrum av olika uppgifter.

Gruppen konstaterar att bedömningen är komplex. En modell som bara kan generera bilder kan anses ha ett smalt tillämpningsområde eftersom modellen inte kan generera exempelvis text, men om modellen kan variera ett brett spektrum av bilder kan modellen ändå anses uppfylla kravet på generalitet.[30]

Bedömningen bör utgå från syftet med modellen och vad den kan göra, snarare än vad den faktiskt används till. Stöd för det resonemanget bör kunna hämtas ur skäl 99 som talar om ”möjliggöra” och skäl 100 som talar om ”har förmåga”. De flesta modeller är möjliga att anpassa och skala ned eller specialisera. Om de skulle bedömas utifrån

[28] Se eSams rapport ES2024-1 AI – Utvecklingsprocessen och data.
[29] The KBLab Blog: Introducing a Swedish Sentence Transformer, The KBLab Blog: Swedish Sentence Transformer 2.0, [2303.08774] GPT-4 Technical Report
[30] General-Purpose AI Models in the AI Act – Questions & Answers https://digital-strategy.ec.europa.eu/en/faqs/general-purpose-ai-models-ai-act-questions-answers

faktiskt användning bör syftet med bestämmelserna förfelas, då syftet är att bestämmelserna ska träffa t.ex. de stora generativa modellerna, se skäl 99.

GPT-4o har ett brett användningsområde och över en miljard parametrar. Denna AI-modell bör därmed anses uppfylla kravet på betydande generalitet samt att den på ett kompetent sätt kan utföra ett brett spektrum av olika uppgifter.

KB-BERT har enligt uppgift ca 345 miljoner parametrar och kan användas inom flera olika områden. KB-BERT är en familj av olika modeller. En frågeställning är då om bedömningen ska göras av varje modell för sig eller tillsammans. De kan i dagsläget laddas ner var för sig och därefter integreras. Var för sig har modellerna mer specifika områden snarare än generella. Samtidigt bör en sådan uppdelning innebära en risk för kringgående av regelverket. Liknande argumentation går sannolikt att föra för andra språkmodeller. Det är därtill inte ovanligt med korsbefruktning av olika familjer.

En tanke är att generalitet skulle kunna bedömas utifrån hur modellerna mäts, t.ex. finns det internationella och nordiska rankningar för språkmodeller.[31]

Arbetsgruppens bedömning är att enbart uppgiften om antal parametrar inte bör utesluta att KB-BERT-modellen kan anses vara en AI-modell för allmänna ändamål, jämfört att skäl 98 är uttryckt som exemplifierande. Utifrån den information som just nu finns tillgänglig är betydelsen av rekvisiten mycket oklara och det kan argumenteras i båda riktningarna. Gruppens slutsats är att det kan inte uteslutas att KB-BERT-modellen kan bedömas uppfylla rekvisiten betydande generalitet och brett spektrum av distinkta uppgifter.

3.4.4.2 Forskning

I sista ledet av definitionen för AI-modell för allmänna ändamål anges att den inte omfattar AI-modeller som används för forsknings-, utvecklings- eller prototypverksamhet innan de släpps ut på marknaden. I avsnitt 3.4.3.1.2 redogörs för vad som avses med utsläppande på marknaden.

Utifrån att KB-BERT-modellen tagits fram av Kungliga biblioteket, som är en nationell forskningsinfrastruktur med uppdrag att bidra till den svenska forskningens kvalitet[32] bedömer gruppen att det finns anledning att överväga om rekvisitet om forskning är tillämpligt för KB-BERT-modellen. Arbetsgruppen finner att rekvisitet inte är tillämpligt för GPT-4o, som levereras av en kommersiell aktör i syfte att användas brett.

[31] T.ex. MTEB Leaderboard eller The Scandinavian Embedding Benchmark.
[32] 1 § förordning (2008:1421) med instruktion för Kungl. biblioteket.

AI-förordningen ska respektera forskningens frihet och inte underminera forsknings- och utvecklingsverksamhet.[33] Forskning är därmed generellt undantaget i AI-förordningen. I artikel 2.6 anges att förordningen inte är tillämplig på AI-system eller AI-modeller, inbegripet deras utdata, som specifikt utvecklas och tas i bruk enbart i vetenskapligt forsknings- och utvecklingssyfte.

Bestämmelserna är inte heller tillämpliga på vetenskaplig eller produktorienterad forsknings- och utvecklingsverksamhet avseende AI-system eller AI-modeller innan dessa släpps ut på marknaden eller tas i bruk. Det undantaget påverkar inte skyldigheten att följa AI-förordningen om ett AI-system släpps ut på marknaden eller tas i bruk till följd av sådan forsknings- och utvecklingsverksamhet.[34]

Leverantörer av AI-modeller för allmänna ändamål som är undantagna pga. användning för vetenskapliga forskningsändamål uppmuntras att ändå frivilligt uppfylla kraven.[35]

Artikel 2.6 innehåller kumulativa rekvisit: att AI-modellen specifikt ska utvecklas och tas i bruk enbart i vetenskapligt forsknings- och utvecklingssyfte. Jämför även sista meningen i artikel 3.63 samt skäl 25 som ger uttryck för att undantag ska gälla innan produkten släpps på marknaden eller tas i bruk. Gruppen finner att det bör gå att föra ett resonemang om att KB-BERT-modellen utvecklats specifikt för vetenskapligt forskningssyfte. Av detta följer att så länge AI-modellen inte släpps på marknaden eller tas i bruk så föreligger inga krav. Det skulle eventuellt även gå att resonera kring att KB-BERT-modellen är undantagen AI-förordningen om modellen tas i bruk enbart för vetenskapligt forsknings- och utvecklingssyfte. Det vill säga om modellen enbart hade använts i Kungliga bibliotekets utveckling och forskning hade undantaget kunnat bli tillämpligt.

I aktuellt fall använder Ekonomisstyrningsverket KB-BERT-modellen i sin utveckling. KB-BERT-modellen finns allmänt tillgänglig på Hugging Face[36] och koden är anpassad för att vara lättillgänglig för andra att använda. KB-BERT-modellen har en licens som tillåter användning i vid omfattning och licensen är inte begränsad till användning enbart i forskningssyfte. AI-modellen är därmed tillgängliggjord för andra. Den är således utsläppt på marknaden och möjlig att använda även för andra syften än enbart vetenskapligt forskningssyfte. Gruppen finner att det bör finnas en anledning till att bestämmelserna är utformade med rekvisitet ”enbart”. Om det även finns andra möjliga användningsområden bör AI-förordningen bli tillämplig. Sista ledet i definitionen av AI-modell för allmänna ändamål tar sikte på undantag innan en AI-modell är släppt på

[33] Skäl 25.
[34] Skäl 25.
[35] Skäl 109.
[36] En plattform för maskininlärning för samarbete kring modeller, datauppsättningar och applikationer.

marknaden. Eftersom modellen får anses vara utsläppt på marknaden är undantaget inte tillämpligt.

3.4.4.3 Öppen källkod

KB-BERT-modellen är en AI-modell som tillgängliggörs via öppen källkod, till skillnad från GPT-4o som är en proprietär modell. Om KB-BERT-modellen skulle bedömas som en AI-modell för allmänna ändamål finns det anledning att titta på de bestämmelser som finns kring AI-modeller för allmänna ändamål med öppen källkod.

Av artikel 53 framgår skyldigheterna för leverantörer av AI-modeller för allmänna ändamål. Av artikel 53.1 följer att ”levererantörer av AI-modeller för allmänna ändamål ska

a) utarbeta och uppdatera den tekniska dokumentationen för modellen, inbegripet tränings- och testningsförfarandet samt resultaten av utvärderingen, som åtminstone ska innehålla de uppgifter som anges i bilaga XI, i syfte att på begäran lägga fram den för AI-byrån och de nationella behöriga myndigheterna,
b) utarbeta, uppdatera och göra information och dokumentation tillgänglig för leverantörer av AI-system som avser att integrera AI-modellen för allmänna ändamål i sina AI-system; utan att det påverkar behovet av att respektera och skydda immateriella rättigheter och konfidentiell affärsinformation eller företagshemligheter i enlighet med unionsrätten och nationell rätt ska informationen och dokumentationen
i) göra det möjligt för leverantörer av AI-system att ha en god förståelse av kapaciteten och begränsningarna hos AI-modellen för allmänna ändamål och att fullgöra sina skyldigheter enligt denna förordning, och
ii) minst innehålla de uppgifter som anges i bilaga XII,
c) införa en policy för att följa unionsrätten om upphovsrätt och närstående rättigheter, och i synnerhet för att identifiera och efterleva, inbegripet med hjälp av den senaste tekniken, ett förbehåll för rättigheter som uttryckts enligt artikel 4.3 i direktiv (EU) 2019/790,
d) utarbeta och göra en tillräckligt detaljerad sammanfattning av det innehåll som använts för träning av AI-modellen för allmänna ändamål allmänt tillgänglig, i enlighet med en mall som tillhandahålls av AI-byrån.”

Bestämmelsen tar därmed sikte på att dokumentering ska göras, vad den ska innehålla och vad som ska göras tillgängligt för leverantören av ett AI-system där en AI-modell för allmänna ändamål finns integrerad. Artikelns tredje punkt anger att leverantörerna ska vid behov samverka med EU-kommissionen och nationella behöriga myndigheter.

Enligt artikel 53.2 följer dock att de skyldigheter som anges i 53.1 a och b inte är tillämpliga på leverantörer av AI-modeller som släpps ut med en kostnadsfri licens med öppen källkod som möjliggör åtkomst, användning, ändring och distribution av modellen, och vars parametrar, inbegripet vikter, information om modellarkitekturen och information om modellanvändning, görs allmänt tillgängliga. Undantaget är inte tillämpligt på AI-modeller för allmänna ändamål med systemrisker.

Av skäl 109 framgår en proportionalitetsprincip som innebär att skyldigheterna för leverantörer av AI-modeller för allmänna ändamål ska stå i proportion till typen av modelleverantör.

AI-förordningen har inte någon definition av öppen källkod, men artikel 53.2 ger en ram för bedömningen. I denna punkt talas om kostnadsfri licens, möjlighet att använda och ändra samt om allmänt tillgängliggörande.

AI-förordningen har flera incitament för att öppen källkod har lättnader. Sådana incitament följer bl.a. av skäl 102-108. Det framgår t.ex. av skäl 104 att: ”Leverantörer av AI-modeller för allmänna ändamål som släpps ut med en kostnadsfri licens med öppen källkod och vars parametrar, inbegripet vikter, information om modellarkitekturen och information om modellanvändning, görs allmänt tillgängliga bör omfattas av undantag från de transparensrelaterade krav som gäller för AI-modeller för allmänna ändamål, såvida de inte kan anses utgöra en systemrisk,…” Enligt skäl 89 uppmuntras dock utvecklare av kostnadsfria verktyg, tjänster eller processer med öppen källkod att tillämpa allmänt vedertagen dokumentationspraxis, såsom modellkort och datablad, som ett sätt att påskynda informationsutbytet längs AI-värdekedjan, vilket gör det möjligt att främja tillförlitliga AI-system i unionen (skäl 89 gäller dock inte AI-modeller för allmänna ändamål).

Vilka krav finns då för en AI-modell för allmänna ändamål om det är fråga om öppen källkod? Av artikel 53.2 följer att kraven i artikel 53.1 a och b som avser teknisk dokumentation samt tillgängliggörande av dokumentation till en leverantör som integrerat AI-modellen i sitt AI-system är undantagna. Det innebär att de mer detaljerade kraven i bilaga XI (teknisk dokumentation hänförlig till punkt a) och bilaga XII (transparensinformation enligt punkt b) inte blir tillämpliga. De dokumentationskrav som kvarstår är då kraven i punkterna c och d, som innebär att det ska finnas en policy för upphovsrätt samt krav på leverantören att utarbeta och göra en tillräckligt detaljerad sammanfattning av det innehåll som använts för träning av AI-modellen för allmänna ändamål allmänt tillgänglig.

3.4.4.4 Systemrisk

Vad gäller AI-modeller för allmänna ändamål är undantaget i artikel 53.2 för öppen källkod inte tillämpligt om det är fråga om en AI-modell för allmänna ändamål med systemrisk. Systemrisk definieras i artikel 3.65 som ”en risk som är specifikt kopplad till den kapacitet för hög påverkansgrad som finns hos AI-modeller för allmänna ändamål och som påverkar unionsmarknaden i betydande grad på grund av sin räckvidd eller på grund av faktiska eller rimligen förutsebara negativa effekter på folkhälsa, säkerhet, allmän säkerhet, grundläggande rättigheter eller samhället som helhet, och som kan spridas i stor skala i hela värdekedjan”.

Särskilda regler för AI-modeller för allmänna ändamål och för AI-modeller för allmänna ändamål som medför systemrisker, gäller även när modellerna är integrerade eller ingår i ett AI-system.[37] De särskilda kraven gäller för leverantören av modellen, även om systemet tillhandahålls av någon annan.

Av skäl 110 framgår att AI-modeller för allmänna ändamål kan medföra systemrisker som omfattar faktiska eller rimligen förutsebara negativa effekter i samband med allvarliga olyckor, störningar i kritiska sektorer och allvarliga konsekvenser för folkhälsan och säkerheten, alla faktiska eller rimligen förutsebara negativa effekter på demokratiska processer, allmän och ekonomisk säkerhet, och spridning av olagligt, falskt eller diskriminerande innehåll. Det anges också att det bör antas att systemrisker ökar med modellkapacitet och modellräckvidd, att sådana risker kan uppstå under modellens hela livscykel och påverkas av förhållanden med felaktig användning, modellens tillförlitlighet, rättvisa och säkerhet, dess grad av autonomi, tillgång till verktyg, nya eller kombinerade metoder, strategier för utsläppande och distribution, potentialen att avlägsna skyddsmekanismer och andra faktorer.

En AI-modell för allmänna ändamål enligt artikel 51.1 ska klassificeras som en AI-modell för allmänna ändamål med systemrisk om den: a) har kapacitet med hög påverkansgrad som utvärderats på grundval av lämpliga tekniska verktyg och metoder, inbegripet indikatorer och riktmärken eller b) har, baserat på ett beslut av kommissionen, på eget initiativ eller efter en kvalificerad varning från den vetenskapliga panelen, kapacitet eller inverkan som motsvarar den som avses i led a med beaktande av kriterierna i bilaga XIII vilka är:

a) Antalet parametrar i modellen.
b) Datasetets kvalitet eller storlek, till exempel mätt genom token.

[37] Skäl 97.

c) Den beräkningsmängd som används för att träna modellen, mätt i flyttalsberäkningar eller angiven med en kombination av andra variabler, såsom beräknad träningskostnad, uppskattad tid som krävs för träningen eller uppskattad energiförbrukning för träningen.
d) Modellens in- och utmatningsmetoder, såsom text till text (stora språkmodeller), text till bild, multimodalitet och tröskelvärden som motsvarar den senaste utvecklingen för att fastställa kapacitet med hög påverkansgrad för varje metod, och den specifika typen av in- och utdata (t.ex. biologiska sekvenser).
e) Riktmärken för och utvärderingar av modellens kapacitet, inbegripet med beaktande av antalet uppgifter utan ytterligare träning, anpassningsförmåga att lära sig nya, distinkta uppgifter, dess grad av autonomi och skalbarhet samt de verktyg som den har tillgång till.
f) Huruvida modellen har stor inverkan på den inre marknaden på grund av sin räckvidd, vilket ska förutsättas om den har gjorts tillgänglig för minst 10 000 registrerade företagsanvändare som är etablerade i unionen.
g) Antalet registrerade slutanvändare.

Kapacitet med hög påverkansgrad definieras i artikel 3.64 som ”kapacitet som motsvarar eller överstiger den kapacitet som registrerats i de mest avancerade AI-modellerna för allmänna ändamål”. Med *flyttalsberäkning* avses enligt artikel 3.67 ”varje matematisk operation eller tilldelning som inbegriper flyttal, som är en delmängd av de reella talen som typiskt representeras på datorer genom ett heltal med fast precision multiplicerad med en heltalsexponent med en fast talbas”.

I den svenska språkversionen anges i artikel 51.2 att en AI-modell för allmänna ändamål ska förutsättas ha kapacitet med hög påverkansgrad enligt artikel 51.1 a om den sammanlagda beräkningsmängd som används för dess träning mätt i flyttalsberäkningar är större än 1025.[38] Vid en jämförelse med övriga språkversioner ser det ut att vara ett översättningsfel och att den korrekta siffran ska vara 10 upphöjt till 25 (dvs. 10^{25}).

Arbetsgruppen finner att regleringen om systemrisk inte är helt enkel att tillämpa, särskilt som det anges många tekniska parametrar både i artikel 51 och i bilaga XIII. Syftet med regleringen verkar dock vara att försöka rama in de riktigt stora modellerna såsom Googles Gemini och Open AI:s GPT-4o samt att ge ett objektivt mått på storleken eftersom AI-förordningen verkar utgå från att modellens storlek har stor betydelse för risken. Måttet 10 upphöjt till 25 (enligt andra språkversioner) är ett gigantiskt tal. Arbetsgruppen finner att det sannolikt finns en anledning till just detta tal då det är där måtten just nu ligger för de största modellerna. Arbetsgruppen konstaterar att sannolikt är Mistrals senaste modeller inte långt ifrån måttet och på samma sätt kan det vara för Llamas senaste modell. Av intresse är att det i skäl 111 anges att tröskelvärdet bör

[38] Se även skäl 111 och skäl 112.

justeras med tiden för att återspegla tekniska och industriella förändringar, såsom algoritmiska förbättringar eller ökad hårdvaruеffektivitet, och bör kompletteras med riktmärken och indikatorer för modellkapacitet.

Gruppens bedömning är att GPT-4o bör anses vara en AI-modell för allmänna ändamål med systemrisk. Om KB-BERT-modellen bedöms vara en AI-modell för allmänna ändamål bör den inte anses vara en AI-modell med systemrisk.

3.4.4.5 Summering av AI-modell för allmänna ändamål

Arbetsgruppen bedömer att GPT-4o är en AI-modell för allmänna ändamål utifrån definitionen och det som anges i skälen samt att det rör sig om en AI-modell med systemrisk. Vad gäller KB-BERT-modellen är bedömningen mer osäker, men det kan inte uteslutas att modellen kan anses vara en AI-modell för allmänna ändamål. Arbetsgruppen anser att KB-BERT-modellen i vart fall inte är en AI-modell med systemrisk och att modellen därför omfattas av begränsningen för leverantörens skyldigheter utifrån bestämmelserna om öppen källkod. Kraven för AI-modeller för allmänna ändamål ska uppfyllas av leverantören av AI-modellen och inte av den leverantör som senare använder modellen i ett AI-system.

Gruppen konstaterar att aktuellt AI-system innehåller åtminstone en AI-modell för allmänna ändamål. När en sådan modell integreras i eller ingår i ett AI-system bör systemet anses vara ett *AI-system för allmänna ändamål* när systemet, på grund av integreringen, har förmåga att tjäna en rad olika ändamål.[39] Enligt artikel 3.66 definieras ett AI-system för allmänna ändamål som ”ett AI-system som bygger på en AI-modell för allmänna ändamål och som har kapacitet att tjäna en rad olika ändamål, både för direkt användning och för integrering i andra AI-system”. Det är enligt arbetsgruppen inte helt tydligt hur artikeln ska tolkas, men mycket talar för att uttrycket ”som har kapacitet att tjäna en rad olika ändamål” syftar till systemets generella förmåga. Det AI-system som nu prövas kommer att utvecklas för ett specifikt ändamål och kan troligtvis inte bedömas ha en sådan generell förmåga.

Regleringen är emellertid också oklar med hänsyn till att definitionen inte verkar ha någon särskild betydelse enligt AI-förordningen. Arbetsgruppen har inte kunnat finna att omständigheten att ett visst system utgör ett AI-system för allmänna ändamål leder till att leverantören eller tillhandahållaren av AI-systemet tillförs några ytterligare krav. De krav som föreligger utgår istället från AI-systemets risknivå. Däremot har leverantör i efterföljande led rätt att ställa krav vad gäller information från leverantören av AI-

[39] Skäl 100.

modellen för allmänna ändamål, t.ex. vad gäller tillgång till teknisk dokumentation, se artikel 53.

3.4.5 Övergångsbestämmelser

Det är viktigt att skilja på tidpunkterna för när vissa bestämmelser i AI-förordningen ska börja tillämpas och vad som gäller för AI-system och AI-modeller som redan släppts på marknaden dessförinnan.

Bestämmelserna om AI-modeller för allmänna ändamål ska börja tillämpas från och med den 2 augusti 2025, dvs. 12 månader från förordningens ikraftträdande.

I artikel 111.3 regleras vad som gäller för AI-modeller för allmänna ändamål som släppts ut på marknaden före den 2 augusti 2025. Leverantörer av sådana AI-modeller som har släppts ut på marknaden före den 2 augusti 2025 ska vidta nödvändiga åtgärder för att uppfylla de skyldigheter som fastställs i AI-förordningen senast den 2 augusti 2027.

Det innebär att för en befintlig AI-modell för allmänna ändamål, såsom GPT-4o, gäller att åtgärder ska vara vidtagna senast den 2 augusti 2027. För en motsvarande AI-modell som släpps ut på marknaden den 3 augusti 2025 eller senare behöver däremot kraven vara uppfyllda genast, dvs. vid samma tidpunkt som när den släpps.

4. Lärdomar

I avsnittet redogörs för några av de lärdomar arbetsgruppen gjort i samband med den andra iteration av piloten.

4.1 Om AI-förordningen

Arbetsgruppen finner att insikten från den första iterationen om att mycket kommer att anses vara ett AI-system står sig och har förstärkts i den andra iterationen. Likaså har insikten om att bedömningen av vad som är omfattningen av systemet kräver en god helhetsbild över it- och AI-arkitekturen förstärkts.

En lärdom är att en leverantör eller tillhandahållare av ett AI-system inte tillförs några ytterligare krav med anledning av att det i AI-systemet integrerats en AI-modell för allmänna ändamål, utan de krav som föreligger är utifrån AI-systemets risknivå.

Det är syftet med systemet som avgör risknivå, dvs. risknivån är oberoende av om AI-systemet innefattar en AI-modell för allmänna ändamål. I aktuellt fall betyder det, utifrån bedömningen att det är fråga om ett AI-system med minimal eller ingen risk, att det inte föreligger några krav på leverantören eller tillhandahållaren. Detta trots det faktum att det i AI-systemet ingår en stor språkmodell som utgör en AI-modell för allmänna ändamål med systemrisk.

Transparenskravet enligt artikel 50 är tillämpligt oaktat om det är ett högrisksystem eller inte. De olika risknivåerna ska därmed inte tolkas gälla i fallande ordning, dvs. att transparenskrav först prövas för det fall det inte är fråga om ett högrisksystem, tvärtom ska artikeln tillämpas även vid ett högrisksystem.

En insikt är att AI-förordningen inte har någon definition på vad som avses med en AI-modell, utan enbart en definition av en AI-modell för allmänna ändamål. Det finns inte heller några tillämpliga bestämmelser om det inte är fråga om en AI-modell för allmänna ändamål. I värdekedjan saknas alltså krav på den som tagit fram AI-modellen, dvs. leverantören för AI-systemet har inte några bestämmelser att använda för att ställa krav gentemot den som tar fram AI-modellen om det inte är en AI-modell för allmänna ändamål. En reflektion från arbetsgruppen är att även om det inte finns krav, är det inte osannolikt att leverantören ändå väljer att ha transparens för modellen i försäljningssyfte, dvs. beställarkrav kommer troligtvis leda till att det finns dokumentation. Sannolikt är det ett medvetet glapp med anledning av att AI-förordningen inte reglerar modeller utan AI-system. Fokus är istället på den som tar in en AI-modell i sitt system och den påverkan det då kan medföra.

En ytterligare reflektion är att det kommer vara svårt att bedöma gränsdragningen mellan AI-modeller och AI-system, särskilt när det uppstår hybrider. Sådan bedömning blir svår om det inte finns transparens.

Det finns en begreppsförvirring utifrån de svenska översättningarna av ”provider” (leverantör) och ”deployer” (tillhandahållare). Tanken går lätt fel att det är tillhandahållaren som tillhandahåller ett AI-system. Så är dock inte fallet utan en tillhandahållare är den som använder systemet i sin verksamhet, medan den som tillhandahåller ofta är en distributör eller återförsäljare.

Det kan vara vanskligt att i själva förordningen reglera mått för bedömning, som t.ex. antalet parametrar i skäl 98, eftersom måtten kan ta tid att ändra. Gruppens reflektion är att dessa mått riskerar att snabbt tappa aktualitet. Det sker nu en snabb uppskalning av många språkmodeller. Parametrar är visserligen ett enkelt sätt att mäta, men utvecklingen går också emot att få effektivare modeller redan vid ett färre antal parametrar och därmed kan måttet bli missvisande i förhållande till vad som är betydande generalitet. Förordningen definierar begrepp utifrån en kunskap om vad som är ”state of the art” just nu och har eventuellt inte tagit höjd för en framtida utveckling.

4.2 Om den AI-regulatoriska sandlådans innehåll

En lärdom från den första iteration i piloten är att det kan vara svårt att på förhand bedöma om ett utvecklingsprojekt platsar eller inte i en AI-regulatorisk sandlåda.

En insikt i den andra iterationen är att det är svårt att på förhand veta vilka frågeställningar som kan bli aktuella att titta på. Arbetsgruppens initala bedömning var att det skulle vara fråga om ett case av mindre komplexitet, men det har visat sig att vissa frågeställningar varit svåra att bedöma och arbetsgruppen har sett ett behov av ytterligare förtydligande t.ex. i kommande kompletterande lagstiftning från EU eller nationellt. Till exempel är bedömningen av AI-modeller för allmänna ändamål en sådan fråga.

En reflektion är att det kan finnas utmaningar med att få fram all information för en bedömning av systemet i en AI-regulatorisk sandlåda, bland annat för att systemet ska utvecklas under pågående sandlåda och att den dokumentation som kan behövas för en beskrivning inte ännu finns tillgänglig.

Det har i denna iteration visat sig att det inte alltid är enkelt att bedöma vilken roll inblandade aktörer har. Bedömningen av vilken roll en aktör har, får betydelse för vem som kan ansöka till en AI-regulatorisk sandlåda. Det är endast leverantörer och potentiella leverantörer av AI-system som har möjlighet att söka till en AI-regulatorisk

sandlåda.[40] En aktör som tar fram en AI-modell för allmänna ändamål kan därmed inte söka på egen hand, inte heller en tillhandahållare. En ansökan kan dock ske i partnerskap, så dessa aktörer kan tillsammans med en leverantör eller potentiell leverantör delta i en AI-regulatorisk sandlåda.[41]

Inför den andra iterationen tog arbetsgruppen ställning till ett annat case som avsåg att börja använda ett redan färdigt AI-system. För det systemet konstaterade arbetsgruppen att det inte borde vara ett case för en AI-regulatorisk sandlåda, då sandlådan är till för system som ännu inte är släppta på marknaden eller tagits i bruk.

En reflektion från arbetsgruppen avser hur harmoniserade de AI-regulatoriska sandlådorna kommer att vara. Det kommer troligtvis finnas risk för diskrepans utifrån medlemsländernas kapacitet och vilket fokus respektive medlemsstat tar. I AI-förordningen finns det flera olika möjligheter för utformningen av sandlådor, t.ex. att en teknisk infrastruktur för träning och validering av ett AI-system kan tillhandahållas.[42] Det är inte helt tydligt var kravnivån kommer att ligga och det kommer troligtvis att förtydligas genom en eller flera genomförandeförordningar. Däremot är det klart att slutrapporten från en AI-regulatorisk sandlåda kan användas av leverantören för att visa på överenstämmelse med förordningen.[43] Om leverantörerna följer sandlådeplanen och villkoren för deras deltagande samt i god tro följer de riktlinjer som den nationella behöriga myndigheten ger, ska de inte heller åläggas några administrativa sanktionsavgifter för överträdelser.[44] Det är sannolikt att ett deltagande i sandlådorna kommer att vara attraktivt av de här anledningarna. Om sandlådorna får tillräcklig kapacitet för att kunna hantera många ärenden bör de kunna bidra till att AI-system snabbt kommer ut på marknaden vilket i sin tur kan stärka EU:s innovationskraft.

En vanlig fråga som arbetsgruppen stött på är om det kan vara möjligt med samordnade sandlådor. Till viss del finns sådana incitament, t.ex. verkar dataskyddsmyndigheterna komma att bli involverade i stor utsträckning.[45] Men det bör uppstå en utmaning att hantera nationella bestämmelser, då tanken är att man ska kunna söka vilken sandlåda som helst inom EU. Arbetsgruppen finner att sannolikt kommer det i så fall handla om olika sandlådor som samordnas snarare än en gemensam sandlåda för flera rättsområden. Även med samordnade sandlådor (t.ex. AI och dataskydd) kommer antalet frågor som kan utredas inom ramen för sandlådan vara begränsade. Aktörerna kommer säkert vara

[40] Artikel 57.7.
[41] Artikel 58.2 a.
[42] Artikel 57.
[43] Artikel 57.7.
[44] Artikel 57.12.
[45] Artikel 57.10.

hjälpta av harmoniserade sandlådor men fortfarande ha flera frågor kvar att besvara efter en avslutad sandlåda.

Det är en relevant fråga vad det är värt att som leverantör delta i en AI-regulatorisk sandlåda om man bara får svar på en fråga. Arbetsgruppens erfarenhet hittills är att även om alla frågor inte kan besvaras får leverantören ändå en stor hjälp genom att delta i en sandlåda och ett deltagande kan hjälpa leverantören att komma vidare med sitt projekt. Samtidigt är det tydligt att frågor gällande andra regler än de i AI-förordningen ofta kommer upp. Det finns sannolikt goda skäl att i någon mån samordna sandlådor och andra stöd så att leverantörer får ett helhetsperspektiv.

4.3 Om arbetssättet

I rapporten för den första iterationen konstaterade arbetsgruppen att i detta utforskande arbetssätt är det nödvändigt att vara öppensinnad och lösningsfokuserad, men ändå hålla fast vid en riktning. Det är viktigt att de som deltar besitter kompetens, mandat och förmåga till dialog. Det är också viktigt att det finns en förmåga att förklara sin specialistkompetens på ett enkelt och pedagogiskt sätt så att de som inte har samma sakkunskap förstår och kan använda sig av den kunskapen för att analysera frågor som är avgörande för att arbetet i sandlådan ska kunna gå framåt.

Denna lärdom har ytterligare förstärkts i den andra iterationen. Arbetsgruppen vill understryka behovet av tvärfunktionell kompetens, såsom jurister, AI-utvecklare, teknisk kompetens och projektledningskompetens. De som har dessa roller behöver också i sig ha eller vilja utveckla en tvärfunktionell förståelse. En AI-regulatorisk sandlåda kommer alltid att behöva bemannas av jurister, men det har blivit än mer tydligt i denna iteration att också AI-utvecklingskompetens och teknisk kompetens behövs för att kunna ge stöd till den aktör som ansökt till den AI-regulatoriska sandlådan. Det är inte tillräckligt att den ansökande aktören bidrar med sådan kompetens utan bör även tillhandahållas av den som tillhandahåller sandlådan. Av förklarliga skäl kommer de som ansöker till den AI-regulatoriska sandlådan ha ett intresse av att produkten får ett godkännande, det intresset kan behöva balanseras med en objektivitet för de behöriga myndigheter som tillhandahåller sandlådan.

Det finns också uppenbara fördelar med ett kärnteam som arbetar tillsammans över tid. Det skapar en trygghet och effektivitet i arbetet. Arbetsgruppens reflektion är att arbetsgruppsmötena i piloten ger berikande möten med mycket lärande. Ett arbetssätt där tolkningssvårigheter identifierats utifrån ett praktiskt fall ger ett större lärande än en teorisk inläsning av lagstiftningen. Arbetsmetoden upplevs fungera väl för uppdraget. Arbetsgruppen hade en effektivare uppstart i den andra iterationen, vilket tyder på att

modellen att utgå från en frågematris fungerar, innefattat att matrisen behöver uppdateras vartefter nya frågor identifieras. Sannolikt bör metoden även kunna tillämpas på andra sandlådor eller annan innovativ utveckling.

En lärdom är att en aktör som ansöker till en sandlåda behöver komma med visst eget deltagande och kompetens, särskilt vad gäller beskrivning av tilltänkt system. Det behöver finnas ett stöd hos aktörens beslutsfattare och att de på något sätt är involverade. I den första iterationen konstaterades också att det är en framgångsfaktor att dokumentera och producera skriven text under hela arbetet. Genom att dokumentera vad gruppen är överens om uppmärksammas tidigt om det föreligger olika bilder om vad gruppen kommit fram till. Även denna lärdom har förstärkts i den andra iterationen. Särskilt när det varit diskussion om de mer tekniska bestämmelserna, som arbetsgruppen i denna iteration har upplevt mer komplexa att tolka, har gruppen kunnat uppmärksamma där vi har olika uppfattningar eller har missat att diskutera något rekvisit.

Bilaga 1 Frågematris

AI-förordningen	Regulatorisk sandlåda	Verksamhetsfrågor	Nationella frågor
Är det ett AI-system?	Hur moget bör AI-systemet vara?	Vilka kompetenser krävs i en sandlåda?	Vilken/vilka myndigheter är eller bör vara myndigheter att upprätta sandlådor?
Var går ”gränsen” för systemet?	Vilka bestämmelser i förordningen kan prövas? (är det bara bestämmelserna för leverantörer och potentiella leverantörer)?	Behövs en teknisk infrastruktur?	Hur ska privata företag komma med?
Vilken risknivå har AI-systemet?	Vad ska ingå i att utarbeta planen för sandlådan?	Hur mycket kostar det att ha en sandlåda?	Kan sandlådorna fungera som en motor för att få centraliserade juridiska bedömningar av AI-system?
Hanteras data korrekt enligt art. 10?			

eSam är ett medlemsdrivet program för samverkan mellan myndigheter för att underlätta och påskynda digitaliseringen inom det offentliga. eSam bildades 2015 som en frivillig fortsättning på E-delegationen. En viktig uppgift för eSam är att ta fram stöd och vägledningar som ger förutsättningar för att öka den digitala samverkan inom offentlig förvaltning.

Alla stöddokument finns på esamverka.se

I eSam ingår Arbetsförmedlingen, Arbetsmiljöverket, Bolagsverket, Boverket, Centrala Studiestödsnämnden, Domstolsverket, E-hälsomyndigheten, Ekonomistyrningsverket, Finansinspektionen, Folkhälsomyndigheten, Försäkringskassan, Havs- och vattenmyndigheten, Inspektionen för vård och omsorg, Jordbruksverket, Kemikalieinspektionen, Kriminalvården, Kronofogdemyndigheten, Kustbevakningen, Lantmäteriet, Livsmedelsverket, Länsstyrelserna, Migrationsverket, Naturvårdsverket, Pensionsmyndigheten, Riksantikvarieämbetet, Riksarkivet, Rättsmedicinalverket, Sida, Skatteverket, Skolverket, Statens institutionsstyrelse, Statens servicecenter, Statens tjänstepensionsverk, Statens veterinärmedicinska anstalt, Statistiska centralbyrån, Tillväxtverket, Trafikverket, Transportstyrelsen, Tullverket, Universitets- och högskolerådet samt Utbetalningsmyndigheten. (Juni 2024.)