Delrapport

AI-regulatorisk sandlåda – en första iteration

ES2024-14

Juni 2024

Innehåll

1. Inledning

Inom EU har det förhandlats fram en förordning för harmonisering av regler gällande användningen av AI[1] (AI-förordningen). Förordningen har ett riskbaserat angreppsätt i syfte att skapa en strukturerad uppdelning mellan olika typer av AI-system och dess användning där vissa är förbjudna, medan andra är tillåtna men med restriktioner och krav i form av bl.a. tillsyn och registrering hos ansvarig myndighet. Förordningen är i delar en produktlagstiftning och i delar en skydds- och rättighetslagstiftning.

Förordningen kommer sannolikt att få betydelse för myndigheters AI-användning och tillföra nya arbetsuppgifter inom offentlig sektor. AI-förordningen innehåller bland annat bestämmelser om så kallade regulatoriska sandlådor för AI (AI-regulatorisk sandlåda). De AI-regulatoriska sandlådorna har ett innovationsfrämjande syfte och varje medlemsstat måste upprätta minst en sandlåda, enskilt eller tillsammans med andra medlemsstater. Ett sätt att förbereda offentlig sektor och Sverige för regleringen är att genomföra en pilot av en AI-regulatorisk sandlåda.

Vidare kan ett genomförande av en pilot av en AI-regulatorisk sandlåda

- möjliggöra för att AI-regulatoriska sandlådor är användbara från AI-förordningens tillämpningsdatum
- främja svenska intressen i EU
- preventivt och gemensamt utvärdera eller utveckla ett AI-system i enlighet med AI-förordningen och
- ge en fördjupad kunskap om AI-förordningen.

Inom eSams ram har Bolagsverket, Skatteverket och Arbetsförmedlingen tillsammans med Integritetsskyddsmyndigheten startat ett initiativ med att genomföra en pilot av en AI-regulatorisk sandlåda. Arbetet med piloten genomförs av en arbetsgrupp bestående av deltagare från nämnda myndigheter och från eSams kansli. Kretsen av deltagare kan komma att utökas.

Utgångspunkten för arbetet med piloten är att i iterationer utvärdera ett eller flera AI-system utifrån de krav som förordningen ställer för 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.

[1] Slutlig version av AI-förordningen har inte publicerats vid tidpunkten för denna rapport.

Förhoppningen är att arbetet kan bidra till kunskap om hur AI-regulatoriska sandlådor bör inrättas och fungera i Sverige och vilka förutsättningar som krävs för att aktörer ska kunna nyttja sådana sandlådor.

I denna rapport beskrivs lärdomar som arbetsgruppen gjort i samband med en första iteration av piloten.

1.1 Avgränsningar

Denna iteration har genomförts innan AI-förordningen börjat att gälla. Reglerna för regulatoriska sandlådor kommer att träda i kraft tjugofyra månader efter att förordningen börjat tillämpas. Arbetet i denna iteration har avgränsats till hypotetiska frågor utifrån AI-förordningen. Bedömningar av beskrivet system ska därmed inte heller uppfattas som bindande eller slutliga. Frågor om testning under verkliga förhållanden har inte tagits upp eller analyserats. Frågor om personuppgiftsbehandling, cybersäkerhet och svensk nationell lagstiftning har inte och kommer inte att tas upp i piloten.

2. Rättsliga förutsättningar för en AI-regulatorisk sandlåda

Slutlig version av AI-förordningen har inte publicerats vid tidpunkten för denna rapport. De rättsliga förutsättningarna nedan är baserade på den korrigerade versionen av AI-förordningen som antagits av Europaparlamentet den 16 april 2024.[2]

2.1 AI-regulatorisk sandlåda

AI är en teknikfamilj i snabb utveckling som kräver tillsyn och ett säkert och kontrollerat område för experiment, med säkerställande av ansvarsfull innovation och integrering av ändamålsenliga skydds- och riskbegränsningsåtgärder. AI-förordningens bestämmelser om regulatoriska sandlådor syftar till att främja AI-innovation genom att tillhandahålla en kontrollerad miljö som erbjuder leverantörer eller potentiella leverantörer av AI-system en möjlighet att utveckla, träna, validera och testa innovativa AI-system under en begränsad tid innan de släpps ut på marknaden eller på annat sätt tas i bruk. Målet är att öka rättssäkerheten för innovatörer och förbättra de behöriga myndigheternas möjligheter att utöva tillsyn samt skapa sig en förståelse för möjligheter, risker och effekter av AI-användning.[3]

AI-regulatoriska sandlådor enligt AI-förordningen får omfatta testning under verkliga förhållanden under tillsyn i sandlådan.[4] Med *testning under verkliga förhållanden* avses tillfällig testning av ett AI-system med avseende på dess avsedda ändamål under verkliga förhållanden utanför ett laboratorium eller en på annat sätt simulerad miljö. Syftet ska vara att samla in tillförlitliga och robusta data och bedöma och kontrollera AI-systemets överensstämmelse med kraven i AI-förordningen.[5] (Testning under verkliga förhållanden får även ske utanför AI-regulatoriska sandlådor under särskilt fastställda villkor).[6]

Genomförandet ska ske enligt en särskild *sandlådeplan* som de potentiella leverantörerna och den behöriga myndigheten kommer överens om. En sandlådeplan beskriver målen, villkoren, tidsplanen, metoden och kraven för den verksamhet som ska bedrivas i sandlådan.[7]

[2] Rättelse till Europaparlamentets ståndpunkt fastställd vid första behandlingen den 13 mars 2024 inför antagandet av Europaparlamentets och rådets förordning (EU) 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 (rättsakt om artificiell intelligens) P9_TA(2024)0138 (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).
[3] Skäl 138, skäl 139, artikel 3 punkt 55, artikel 57 punkt 5.
[4] Artikel 57 punkt 5 och skäl 139.
[5] Artikel 3 punkt 57.
[6] Artikel 60, 61 och 76. Testningen ska bl.a. följa en plan som ska lämnas in till och godkännas av aktuell marknadskontrollmyndighet och testningen får endast utföras under en begränsad period.
[7] Artikel 3 punkt 54, artikel 57 punkt 5.

2.2 Inrättande av AI-regulatorisk sandlåda

Medlemsstaterna ska säkerställa att deras behöriga myndigheter inrättar minst en AI-regulatorisk sandlåda på nationell nivå. Medlemsstaterna kan också fullgöra denna skyldighet genom att delta i redan befintliga regulatoriska sandlådor eller gemensamt inrätta en sandlåda med en eller flera medlemsstaters behöriga myndigheter, i den mån detta deltagande ger de deltagande medlemsstaterna likvärdig nationell täckning.[8] Ytterligare AI-regulatoriska sandlådor får också inrättas på regional eller lokal nivå. Regulatoriska sandlådor bör vara tillgängliga i bred omfattning i hela unionen.[9]

AI-regulatoriska sandlådor kan inrättas i fysisk eller digital form eller i hybridform och kan rymma både fysiska och digitala produkter.[10] Medlemsstaterna ska säkerställa att de behöriga myndigheterna anslår tillräckliga resurser på ett ändamålsenligt sätt och i god tid så att de regulatoriska sandlådorna har tillräckliga resurser för sin funktion, inbegripet ekonomiska och mänskliga resurser.[11]

De behöriga myndigheterna ska, beroende på vad som är lämpligt, tillhandahålla vägledning, tillsyn och stöd inom den AI-regulatoriska sandlådan i syfte att identifiera risker, särskilt när det gäller grundläggande rättigheter, hälsa och säkerhet, testning, riskreducerande åtgärder och deras effektivitet i förhållande till skyldigheterna och kraven i AI-förordningen samt, i förekommande fall, annan unionsrätt och medlemsstatsrätt som är föremål för tillsyn inom sandlådan. De behöriga myndigheterna ska vidare ge leverantörer och potentiella leverantörer som använder den AI-regulatoriska sandlådan, vägledning om rättsliga förväntningar och hur de krav och skyldigheter som fastställs i AI-förordningen ska uppfyllas.[12]

De AI-regulatoriska sandlådorna ska utformas och genomföras på ett sådant sätt att de i relevanta fall underlättar gränsöverskridande samarbete mellan de nationella behöriga myndigheterna.

Kommissionen får tillhandahålla tekniskt stöd, rådgivning och verktyg för inrättande och drift av AI-regulatoriska sandlådor.

2.2.1 Genomförandeakter för sandlådorna

I syfte att minska fragmentering inom EU vad gäller AI-regulatoriska sandlådor ska Europeiska kommissionen anta genomförandeakter som specificerar de närmare

[8] Skäl 138, artikel 57 punkt 1.
[9] Skäl 139.
[10] Skäl 138.
[11] Skäl 138 och artikel 57 punkt 4.
[12] Artikel 57 punkt 6 och 7.

arrangemangen för inrättande, utveckling, genomförande, drift och tillsyn av AI-regulatoriska sandlådor.[13] Dessa genomförandeakter ska innehålla gemensamma principer i följande frågor:

- Behörighets- och urvalskriterier för deltagande i den AI-regulatoriska sandlådan.
- Förfarandet för tillämpning, deltagande, övervakning, utträde ur och avslutande av den AI-regulatoriska sandlådan, inbegripet sandlådeplanen och slutrapporten.
- De villkor som gäller för deltagarna.

2.3 Deltagande i AI-regulatorisk sandlåda

Deltagandet i den AI-regulatoriska sandlådan bör inriktas på problem som skapar rättsosäkerhet[14] för leverantörer och potentiella leverantörer när de ska vara innovativa, experimentera med AI i unionen och bidra till evidensbaserat regulatoriskt lärande.[15]

Medlemsstaterna bör ge små och medelstora företag, inbegripet nystartade företag, som har ett säte eller en filial i unionen prioriterad åtkomst till de AI-regulatoriska sandlådorna. Det är under förutsättningen att de uppfyller behörighetskraven och urvalskriterierna och utan att andra leverantörer och potentiella leverantörer hindras från att få åtkomst till sandlådorna.[16] Tillgången till de AI-regulatoriska sandlådorna ska vara kostnadsfri för dessa företag.[17]

På begäran av leverantören eller den potentiella leverantören av AI-systemet ska den behöriga myndigheten tillhandahålla ett skriftligt bevis på den verksamhet som framgångsrikt utförts i sandlådan. Den behöriga myndigheten ska också tillhandahålla en slutrapport med uppgifter om den verksamhet som bedrivs i sandlådan och tillhörande resultat och läranderesultat.[18]

2.4 Tillsyn i AI-regulatorisk sandlåda

Tillsynen av AI-systemen i den AI-regulatoriska sandlådan bör omfatta utveckling, träning, testning och validering innan systemen släpps ut på marknaden eller tas i bruk.[19] De behöriga myndigheter som är involverade i den AI-regulatoriska sandlådan ska

[13] Artikel 58.
[14] ”Legal uncertainty” i den engelska översättningen.
[15] Skäl 139.
[16] Skäl 139 och 143.
[17] Artikel 58 punkt 2 d. Detta gäller dock utan påverkan på exceptionella kostnader som nationella behöriga myndigheter får återkräva på ett rättvist och proportionellt sätt.
[18] Artikel 57 punkt 7.
[19] Skäl 139.

tillhandahålla vägledning, tillsyn och stöd i syfte att identifiera risker, särskilt när det gäller grundläggande rättigheter, hälsa och säkerhet, testning, riskreducerande åtgärder och deras effektivitet i förhållande till skyldigheterna och kraven i AI-förordningen.[20]

Om ett AI-system som prövas inom ramen för den AI-regulatoriska sandlådan behandlar personuppgifter eller på annat sätt omfattas av andra nationella myndigheters tillsyn, ska dataskyddsmyndigheterna och dessa andra myndigheter involveras i driften och tillsynen av sandlådan. Detta ska ske i enlighet med deras respektive uppgifter och befogenheter.[21] Alla betydande risker för hälsa och säkerhet och grundläggande rättigheter som upptäcks under utvecklingen och testningen av ett AI-system ska leda till lämpliga begränsningsåtgärder. De behöriga myndigheterna ska ha befogenhet att tillfälligt eller permanent avbryta testprocessen eller deltagandet i sandlådan om inga verkningsfulla begränsningsåtgärder är möjliga.[22]

Leverantörer och potentiella leverantörer som deltar i den AI-regulatoriska sandlådan ska förbli ansvariga, enligt tillämplig unionsrätt och nationell rätt om ansvar, för skada som åsamkas tredje part till följd av de experiment som äger rum i sandlådan. Under förutsättning att de potentiella leverantörerna följer den särskilda planen och villkoren för deras deltagande samt i god tro följer de riktlinjer som den nationella behöriga myndigheten ger, ska myndigheterna dock inte ålägga några administrativa sanktionsavgifter för överträdelser av AI-förordningen. Om andra behöriga myndigheter med ansvar för annan unionsrätt och nationell rätt aktivt deltagit i tillsynen av AI-systemet i sandlådan och tillhandahållit vägledning för efterlevnad ska inga administrativa sanktionsavgifter åläggas avseende den rätten.[23]

2.5 Särskilt om vidarebehandling av personuppgifter i en AI-regulatorisk sandlåda

Inom ramen för den AI-regulatoriska sandlådan får personuppgifter som lagligen samlats in för vissa ändamål också behandlas i syfte att utveckla, träna och testa AI-system under vissa förutsättningar.[24] En av dessa förutsättningar är att det aktuella AI-systemet ska utvecklas för ett väsentligt allmänintresse som skyddas av en offentlig myndighet eller annan fysisk eller juridisk person inom något av följande områden: allmän säkerhet och folkhälsa, miljö, energihållbarhet, säkerhet och resiliens och den offentliga förvaltningen.

[20] Artikel 57 punkt 6.
[21] Artikel 57 punkt 10.
[22] Artikel 57 punkt 11.
[23] Artikel 57 punkt 12.
[24] Skäl 140 och artikel 59 punkt 1.

Vidare får personuppgifterna endast behandlas om det är nödvändigt för att uppfylla kraven som ställs på AI-system med hög risk och endast om dessa krav inte kan uppfyllas effektivt genom användning av anonymiserade eller syntetiska data.

Personuppgifterna ska behandlas i en separat och skyddad miljö där endast behöriga personer har tillgång till uppgifterna. Sådana personuppgifter som skapats inom sandlådan får inte delas utanför den.

Personuppgifterna ska skyddas genom lämpliga tekniska och organisatoriska åtgärder och raderas efter att sandlådeprojektet avslutats eller när de inte längre behövs. Loggar över personuppgiftsbehandlingen ska bevaras så länge som sandlådeprojektet bedrivs, om inte annat föreskrivs i unionsrätten eller nationell rätt.

En beskrivning av processen och motiveringen för träning, testning och validering av AI-systemet ska upprättas och bevaras tillsammans med testresultaten. Sådana handlingar ska utgöra en del av den tekniska dokumentation som ska upprättas enligt AI-förordningen.

Vad som framgår i ovan stycken ska dock inte påverka tillämpningen av unionsrätt eller nationell rätt som utesluter behandling av personuppgifter för andra ändamål än dem som uttryckligen anges i den rätten.[25]

[25] Artikel 59 punkt 3.

3. Pilot – den första iterationen

Arbetet med denna pilot av en AI-regulatorisk sandlåda bedrivs iterativt och utforskande. I detta avsnitt beskrivs genomförandet och resultatet av den första iterationen.

3.1 Metod och arbetssätt

Denna första iteration av piloten har genomförts av en arbetsgrupp bestående av tvärfunktionell kompetens, såsom jurister, strateger, AI-utvecklare och avdelningschef. Arbetsgruppen har genomfört sex halvdagsmöten, varav två fysiska möten. Därtill har en styrgrupp bestående av några från arbetsgruppen genomfört möten månadsvis.

Arbetsgruppen har tagit fram en projektplan för: planering, utformning av sandlådan, genomförande, utvärdering och kommunikation. Arbetsgruppen har inledningsvis tagit del av Integritetsskyddsmyndighetens erfarenheter av arbetssätt i deras regulatoriska sandlådor om dataskydd. Efterföljande arbetsgruppsmöten har lagts upp enligt följande: arbetsgruppen har fått en beskrivning av AI-systemet och utifrån beskrivningen diskuterat frågeställningar utifrån en framtagen frågematris.

3.1.1 Frågematris

För att nå syftet med piloten behöver området angripas från flera perspektiv. För detta syfte har arbetsgruppen tagit fram en frågematris utifrån perspektiven: AI-förordningen, regulatorisk sandlåda, verksamhetsfrågor och nationella frågor. Matrisen är en första ansats och avsikten är att utvärdera och utveckla denna i kommande iterationer av arbetet. Matrisen bifogas i bilaga 1. I denna första iteration har arbetsgruppen fokuserat på om beskrivet system utgör ett AI-system, om detta i så fall är ett högrisksystem samt om systemet passar för en regulatorisk sandlåda.

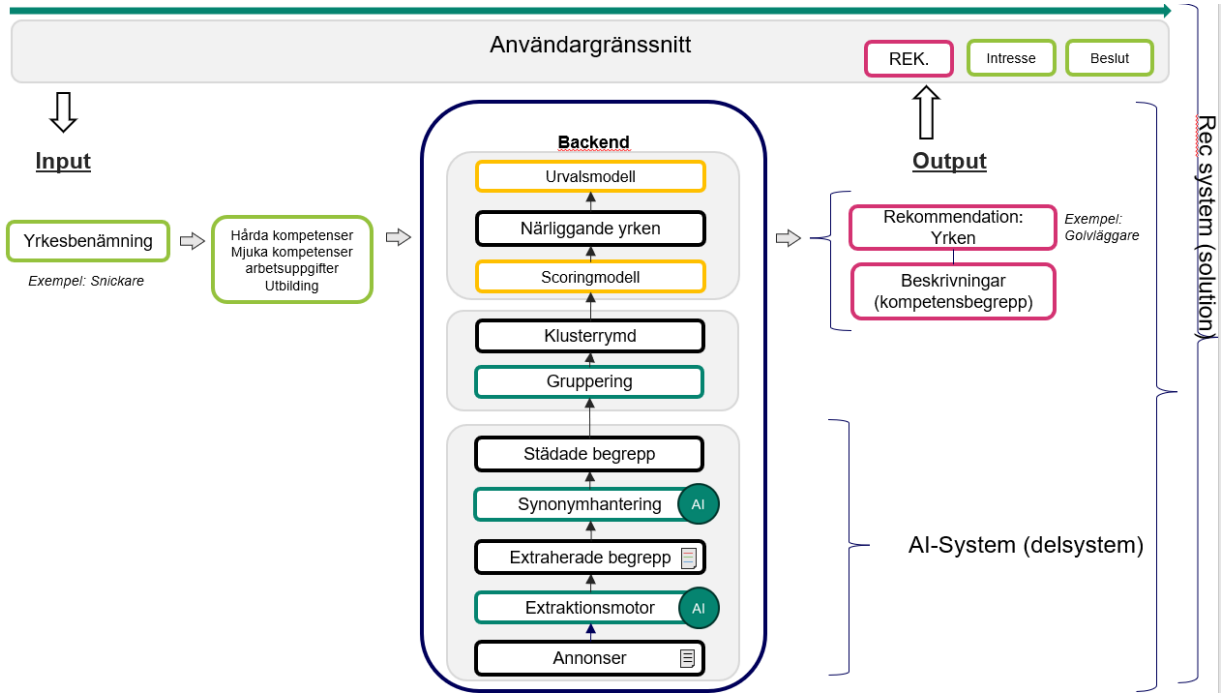
3.2 Beskrivning av AI-systemet

Arbetsgruppen har i denna iteration utgått från ett system som avser att hjälpa arbetssökanden att bredda sin arbetsmarknad genom att hjälpa den arbetssökande att se nya yrken inom ramen för befintlig kompetens.

Lösningen består av flera samverkande komponenter

- Extraktion av kompetensbegrepp från platsannonser
- Berikning
- Granskning

- Synonymhantering
- Rekommendationssystem som baserar sig på jämförelse av vilka begrepp som förknippas med respektive yrke
  - Rankning
  - Urval
- Användargränssnitt



Användarens tidigare yrken används som input (indata).

I lösningen ingår två AI-funktionaliteter:

Extraktionsmodellen extraherar ut entiteter (hårda kompetenser, mjuka kompetenser, arbetsuppgifter och utbildning) och bildar en uppfattning av gemensamma nämnare. Med hjälp av extraktionsmodellen fångas vilka kompetenser som arbetsgivare efterfrågar i platsannonser.

Synonymhanteringen hittar kluster av liknande fraser och slår ihop dessa, dvs. ger en gruppering av begrepp. Syftet med synonymhanteringen är att hitta uttryck som har samma betydelse.

Rekommendationsdelen baserar sig på jämförelse av vilka begrepp som förknippas med respektive yrke, dvs. jämför ursprungsyrke och kopplar entiteter och detta ger nya yrken som kan rekommenderas. En scoringmodell avgör hur de olika entitetsklasserna ska viktas, dvs. räknar ut hur lika två olika yrken är. En breddningsmodell ger vilka

dimensioner (olikheter) som kan användas för breddning. Både scoringmodell och breddningsmodell är idag regelbaserade.

Output (utdata) ges i form av rekommendationerna men också själva beskrivningen (kompetensbegrepp).

3.3 Bedömning

3.3.1 Vad är ett AI-system?

En väsentlig aspekt för ett deltagande i en AI-regulatorisk sandlåda, och för förståelsen av AI-förordningen som helhet, är om den lösning som är tänkt att utvecklas utgör ett AI-system enligt AI-förordningen.

I AI-förordningen definieras ett AI-system som: ett maskinbaserat system som är utformat för att fungera med varierande grad av autonomi, som kan uppvisa anpassningsförmåga efter införande och som, för uttryckliga eller underförstådda mål, på grundval av de indata det tar emot, drar slutsatser om hur utdata såsom förutsägelser, innehåll, rekommendationer eller beslut som kan påverka fysiska eller virtuella miljöer ska genereras.[26]

Begreppet baseras på centrala egenskaper hos AI-system som skiljer det från enklare traditionella programvarusystem eller programmeringsmetoder, och omfattar inte system som bygger på de regler som endast fysiska personer fastställer för att automatiskt utföra operationer. En viktig egenskap hos AI-system är deras förmåga att dra slutsatser. Termen *maskinbaserad* avser det faktum att AI-system körs på maskiner. Definitionen understryker att AI-system kan fungera enligt uttryckliga definierade mål eller underförstådda mål. Med *miljöer* ska förstås de sammanhang där AI-systemen är i drift, medan *utdata* som genereras av AI-systemet återspeglar olika funktioner som utförs av AI-system och inbegriper förutsägelser, innehåll, rekommendationer eller beslut.[27]

AI-system är utformade för att fungera med varierande grad av *autonomi*, vilket innebär att de är oberoende av mänsklig kontroll i viss mån och har förmåga att fungera utan mänskligt ingripande. 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. AI-system kan användas fristående eller som komponent i en produkt, oavsett om systemet är fysiskt integrerat i produkten (inbyggt) eller tjänar produktens funktioner utan att vara integrerat i produkten (ej inbyggt).[28] Europeiska

[26] Artikel 3 punkt 1.
[27] Skäl 12.
[28] Skäl 12.

kommissionen ska utarbeta riktlinjer avseende tillämpningen av definitionen av ett AI-system.[29]

Arbetsgruppen har identifierat att det föreligger flera gränsdragnings- och förståelseutmaningar gällande definitionen av vad ett AI-system är. Flera av begreppen som används i definitionen kan uppfattas på olika sätt utifrån yrkesbakgrund och utbildning och även olika i olika kontext. Ett exempel på detta är "system". Övergripande förstås även begreppet AI olika.

Utifrån definitionen menar arbetsgruppen att det är troligt att de flesta it-system som inte är helt regelbaserade faller in under definitionen av AI-system enligt AI-förordningen.[30] Centralt för bedömningen är att AI-system enligt förordningen skiljer sig från "enklare traditionella programvarusystem eller programmeringsmetoder och inte omfattar system som bygger på de regler som fastställs endast av fysiska personer för att automatiskt utföra operationer".[31]

För att kunna avgöra om ett system är ett AI-system anser arbetsgruppen att en bedömning av systemets *omfattning*, eller gränsen för systemet, behöver göras. Frågan blir alltså vilka delar av lösningen som ska bedömas i sandlådan, om det är den enskilda AI-komponenten eller *hela* systemet.

För gränsdragning om vad som ska anses utgöra AI-systemet kan möjligen viss ledning erhållas från beskrivningen av skillnaden mellan AI-modeller för allmänna ändamål och ett AI-system, dvs. att det för AI-modeller krävs tillägg av ytterligare komponenter, till exempel ett användargränssnitt, för att bli AI-system.[32] Detta talar för att det krävs ett ytterligare sammanhang för att det ska vara fråga om ett AI-system.

För förordningen, som är dels en produktlagstiftning och dels för den enskilde en skyddslagstiftning, är syftet med AI-systemet – det avsedda ändamålet – styrande för vilken riskkategori systemet sorterar under.[33] Avsett ändamål definieras i artikel 3 punkt 12 som: den användning för vilken ett AI-system är avsett av leverantören, inbegripet det specifika användningssammanhanget och de specifika användningsvillkoren, enligt specifikationerna i de uppgifter som tillhandahålls av leverantören i bruksanvisningen, reklam- eller försäljningsmaterial och uttalanden samt i den tekniska dokumentationen.

[29] Artikel 96 punkt 1 f.
[30] OECD har en likartad definition som förklaras i OECD (2024), "Explanatory memorandum on the updated OECD definition of an AI system", OECD Artificial Intelligence Papers, No. 8, OECD Publishing, Paris, https://doi.org/10.1787/623da898-en. som även talar för denna bedömning genom att lägga emfas på automatiserade funktioner, till skillnad från helt regelbaserad it. Enligt skäl 12 i AI-förordningen är definitionen av AI-system avsedd att vara i linje med internationella organisationens definition av AI-system.
[31] Skäl 12.
[32] Skäl 97.
[33] Jämför skäl 65, 84 och 93.

Det avsedda ändamålet bör därmed enligt arbetsgruppen vara centralt för bedömningen av ramen för vad som ska anses utgöra AI-systemet. Om en komponent i en helhet krävs för att systemet ska kunna fungera för det avsedda ändamålet, bör komponenten ses som en del av AI-systemet. Ett AI-system kan därmed innehålla flera komponenter som behöver beaktas i bedömningen av systemets omfattning. I det ovan beskrivna exemplet för denna iteration är det avsedda ändamålet att ge personer som söker eller kan tänkas söka jobb, förslag på tjänster att söka. Om en komponent, som användargränssnitt mot slutanvändare, skulle tas bort från systemet skulle systemet inte kunna nå sitt avsedda ändamål. Därmed måste alla relevanta komponenter för att systemet ska fungera tas med i bedömningen.

Utifrån det avsedda ändamålet behöver det göras en rekvisitgenomgång av definitionen för AI-system i artikel 3 för att besvara om rekvisiten är uppfyllda för det system som ska bedömas. För denna iteration har arbetsgruppen gjort detta på följande sätt.

AI-system:

- ett maskinbaserat system (ja, systemet existerar i maskiner)
- som är utformat för att fungera med varierande grad av autonomi (ja, allt är inte regelbaserat)
- och som kan uppvisa anpassningsförmåga efter införande och som, (ja, systemet har ett visst mått av lärande utifrån indata)
- för uttryckliga eller underförstådda mål, (ja, det finns uttryckliga mål)
- drar slutsatser härledda från den indata det tar emot, (ja, systemet använder sig av data och drar slutsatser som ska leda till de uppsatta målen)
- om hur utdata såsom förutsägelser, innehåll, (ja, skapar innehåll dvs. sammansättningar av utbud av annonser) rekommendationer (ja, ger rekommendationer till slutanvändare) eller beslut (nej, inga beslut)
- som kan påverka fysiska eller virtuella miljöer ska genereras. (ja, människor agerar utifrån utdata)

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

3.3.2 Är det ett högrisksystem?

Beroende på vilken risknivå ett AI-system kategoriseras som omfattas systemet av olika krav. Det är därmed högst relevant att avgöra vilken risknivå som AI-systemet tillhör. Arbetsgruppen har i denna iteration undersökt frågan om beskrivet system kan anses utgöra ett högrisksystem.

Riskerna med AI-system kan ha att göra med hur systemen utformas, men de kan även härröra från hur AI-systemen används.[34]

I AI-förordningen anges förutsättningar för när ett AI-system ska betraktas som ett högrisksystem. 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 är det att anse som ett högrisksystem. Därtill är AI-system som avses i bilaga III i AI-förordningen också att betrakta som högrisksystem.[35] Det gäller AI-system för biometri; kritisk infrastruktur; utbildning och yrkesutbildning; anställning, arbetsledning och tillgång till egenföretagande; tillgång till och åtnjutande av väsentliga privata tjänster och väsentliga offentliga tjänster och förmåner; brottsbekämpning; migration, asyl och gränskontrollförvaltning samt rättskipning och demokratiska processer.

System i bilaga III ska inte betraktas som högrisksystem om det inte utgör en betydande risk för skada på fysiska personers hälsa, säkerhet eller grundläggande rättigheter, inbegripet genom att det inte väsentligt påverkar resultatet av beslutsfattandet. Detta ska vara fallet om AI-systemet är avsett att utföra en snäv processuell uppgift, förbättra resultatet av tidigare fullbordad mänsklig verksamhet, utföra en förberedande uppgift som är relevant för de användningsfall som förtecknas i bilaga III eller upptäcka beslutsmönster eller avvikelser från tidigare beslutsmönster och inte är avsett att ersätta eller påverka tidigare slutförd mänsklig bedömning. AI-system som avses i bilaga III ska dock alltid anses utgöra högrisksystem om AI-systemet utför profilering av fysiska personer.[36]

AI-system som används i utbildning, arbetsledning och tillgång till egenföretagande, i synnerhet när det gäller rekrytering eller urval av personer, för beslutsfattande som påverkar villkoren för arbetsrelaterad befordran eller uppsägning av arbetsrelaterade avtalsförhållanden för fördelning av uppgifter på grundval av individuellt beteende eller personlighetsdrag och egenskaper och för övervakning eller utvärdering av personer i arbetsrelaterade avtalsförhållanden, klassificeras som hög risk, eftersom dessa system märkbart kan påverka framtida karriärutsikter och försörjning för dessa personer samt arbetstagarnas rättigheter. Under hela rekryteringsförfarandet och vid utvärdering, befordran eller bibehållande av personer i arbetsrelaterade avtalsförhållanden, kan sådana system reproducera historiska mönster av diskriminering, exempelvis mot kvinnor, vissa åldersgrupper, personer med funktionsnedsättning eller mot personer på grund av ras, etniskt ursprung eller sexuell läggning. AI-system som används för att övervaka sådana

[34] Skäl 93.
[35] Artikel 6.
[36] Artikel 6 punkt 3.

personers prestation och beteende kan också undergräva deras grundläggande rätt till dataskydd och personlig integritet.[37]

Arbetsgruppen finner att det bör vara syftet med systemet som styr bedömningen, se avsnitt 3.3.1 om avsett ändamål. Syftet med beskrivet AI-system är att hjälpa den arbetssökande att få förslag på nya yrken att söka. AI-systemet har i dess utformning inte funktionalitet så att det kan användas t.ex. av myndighetshandläggare för bedömning av en arbetssökandes aktivitet eller för att ersätta annat myndighetsstöd för en arbetssökande. (Skulle sådan funktionalitet tillföras och syftet med AI-systemet ändras föranleder detta en förnyad bedömning.)

Gruppen finner utifrån artikel 6 och bilaga III att den punkt som det kan finnas anledning att resonera kring är punkt 4 a i bilaga III Anställning, arbetsledning och tillgång till egenföretagande. Enligt denna är det fråga om ett högrisksystem om något av följande är uppfyllt:

”AI-system som är avsedda att användas för rekrytering eller urval av fysiska personer, särskilt för att publicera riktade platsannonser, analysera och filtrera platsansökningar och utvärdera kandidater.”

Arbetsgruppen konstaterar att punkten ser ut att ta sikte på arbetsgivarperspektivet, dvs. när arbetsgivare använder ett AI-system för rekrytering. Beskrivet system är inte tänkt för en arbetsgivare utan som en hjälp för en arbetssökande. Utdata är rekommendationer om yrken, men det finns inget krav på att den arbetssökande söker ett arbete. Det finns inte heller någon koppling till att arbetsgivare använder informationen på något sätt. Gruppen finner att det inte bör vara fråga om ett högrisksystem utifrån denna punkt.

Arbetsgruppen har också resonerat kring punkt 5 a i bilaga III Tillgång till och åtnjutande av väsentliga privata tjänster och väsentliga offentliga tjänster och förmåner. Punkten avser “AI-system som är avsedda att användas av offentliga myndigheter eller för offentliga myndigheters räkning för att utvärdera fysiska personers rätt till väsentliga förmåner och tjänster i form av offentligt stöd, inbegripet hälso- och sjukvårdstjänster, samt för att bevilja, minska, upphäva eller återkalla sådana förmåner och tjänster”. Som angetts ovan har inte beskrivet AI-system sådant syfte. Arbetsgruppens bedömning är därmed att AI-systemet inte anses vara ett högrisksystem enligt denna punkt.

Arbetsgruppen har övervägt även övriga punkter om högrisksystem i bilaga III, men inte funnit att dessa punkter är tillämpliga.

[37] Skäl 57.

3.3.3 Transparenskrav

Arbetsgruppen har tittat på vilka krav som kan föreligga för beskrivet AI-system när det inte bedöms vara ett högrisksystem.

Majoriteten av förordningens krav riktar sig mot högrisksystem, men det finns krav som riktar sig även mot system med lägre risknivå. Mest centralt i sammanhanget är artikel 50 i förordningen, som uppställer transparenskrav riktade mot leverantörer och tillhandahållare av vissa AI-system. Enligt denna artikel ska leverantörer säkerställa att system vars syfte är 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 detta inte är uppenbart för en fysisk person som är normalt informerad och skäligen uppmärksam och medveten, med beaktande av användningens omständigheter och sammanhang.[38] Informationen ska lämnas till de berörda fysiska personerna på ett tydligt och urskiljbart sätt senast vid den första interaktionen eller exponeringen för systemet. Därutöver ska informationen uppfylla tillämpliga tillgänglighetskrav.[39] Bristande efterlevnad av transparenskraven i artikel 50 kan leda till sanktioner.[40]

Arbetsgruppen finner att beskrivet system träffas av det ovan beskrivna transparenskravet, eftersom AI-systemet syftar till att direkt interagera med fysiska personer i ett användargränssnitt där användaren ges rekommendationer om yrken.

3.3.4 Deltagande i en AI-regulatorisk sandlåda

En förutsättning för att delta i en AI-regulatorisk sandlåda är att AI-systemet ännu inte släppts på marknaden. Men vilka andra krav eller riktlinjer finns? Är tanken att sandlådorna främst är till för högrisksystem eller är de även till för lågrisksystem? Kommissionen ska ta fram genomförandeakter för urvalskriterier och det saknas därmed fortfarande tydlig vägledning kring urval för deltagande. Arbetsgruppen har dock fört ett första resonemang kring frågan.

I AI-förordningen uttrycks att AI-regulatoriska sandlådor bör vara tillgängliga i bred omfattning i hela unionen och att små och medelstora företag har prioritet. Det anges vidare att deltagande i den AI-regulatoriska sandlådan bör inriktas på problem som skapar rättsosäkerhet för leverantörer och potentiella leverantörer när de ska vara innovativa, experimentera med AI i unionen och bidra till evidensbaserat regulatoriskt

[38] Artikel 50 punkt 1.
[39] Artikel 50 punkt 5.
[40] Artikel 99 punkt 4 g. Jämför dock punkt 8 som anger att varje medlemsstat ska fastställa regler om i vilken utsträckning administrativa sanktionsavgifter får påföras offentliga myndigheter och organ som är inrättade i medlemsstaten.

lärande.[41] Arbetsgruppens initiala bedömning är att det sannolikt är en ganska låg tröskel för att det ska vara fråga om ”rättsosäkerhet”.

Ett resonemang skulle kunna vara att lågrisksystem inte har så många krav att uppfylla eller i vissa fall endast frivilliga uppförandekrav och att det kan tala för att det i första hand är högrisksystem som är tänkta att delta i sandlådan. Å andra sidan kan det även vid få krav, t.ex. transparenskravet i artikel 50, bli fråga om sanktioner om kraven inte uppfylls, det talar för att det kan finnas en anledning till att även lågrisksystem deltar i sandlådan.

Syftet med de AI-regulatoriska sandlådorna är att det ska vara en slags ventil som underlättar att system snabbt kan komma ut på marknaden. Därför kan det resoneras att tanken inte är att endast högrisksystem deltar utan att sandlådan även är till för att sålla bort de system som inte är högrisk, så de kommer ut på marknaden snabbare.

Det verkar inte finnas någon skyldighet att delta i en AI-regulatorisk sandlåda.[42]

Arbetsgruppen finner att beskrivet AI-system troligen hade kunnat ansöka till en AI-regulatorisk sandlåda.

[41] Skäl 139.
[42] Artikel 60.

4. Lärdomar

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

4.1 Om AI-förordningen

En insikt är att mycket kommer att anses vara ett AI-system. Sannolikt kommer det mesta som inte är helt regelbaserad it att kunna omfattas av AI-förordningens definition av AI-system.

Gränsdragningen av omfattningen av systemet, dvs. vad som ska ingå i systemet och vad som inte är en del av systemet, är komplicerad. Ett exempel kan vara användargränssnitt mot slutanvändare. Arbetsgruppens reflektion är att det avsedda ändamålet bör stå i centrum. Detta sätter ramen för systemet och gör att ett AI-system kan innehålla flera komponenter, modeller och data som behöver beaktas i bedömningen av systemets omfattning. Det gör det också logiskt möjligt att dra gränser för vad ett system är och inte. Om en del, som i sig kan vara en AI-modell, behövs för att systemets avsedda ändamål ska kunna uppnås är det troligen en del av AI-systemet. För att göra den här bedömningen krävs en god helhetsbild över it- och AI-arkitekturen.

Även om det är komplicerat att i sig avgöra om det är ett AI-system är det sannolikt inte denna frågeställning som får mest betydelse. Istället så är det, utifrån AI-förordningens uppbyggnad med det riskbaserade angreppssättet, frågan om vilken risknivå systemet tillhör och särskilt om systemet är ett högrisksystem som är mest central utifrån de konsekvenser som följer med kraven på högrisksystem.

Bedömningen av vad som är ett högrisksystem är inte en subjektiv bedömning utan ska göras utifrån AI-förordningens bestämmelser. På samma sätt som rekvisiten i fråga om vad som är ett AI-system behöver bockas av, så behöver det göras en bedömning av rekvisiten för högrisksystem som är reglerade i AI-förordningen.

För att bedömningarna ska bli så välgrundade som möjligt är det att föredra att den sker i dialog mellan de som kan juridiken och de som kan tekniken. Andra kompetenser kan också komma att behövas inom ramen för en bedömning eftersom det avsedda ändamålet med AI-systemet bör vara centralt för bedömningen.

4.2 Om den AI-regulatoriska sandlådans innehåll

Denna första iteration visar att det är svårt att i förhand bedöma om ett utvecklingsprojekt platsar eller inte i en AI-regulatorisk sandlåda. Som framhålls i avsnitt 4.1 uppstår gränsdragningsfrågor om vad som ska anses vara ett AI-system. En lärdom är att uppfattningen om vad man anser vara ett AI-system skiljer sig både mellan såväl professioner som personer. Det är dock AI-förordningens definition och grundföresatser som ska gälla just för AI-regulatoriska sandlådor. Därför är det viktigt att jurister med kunskap om AI-förordningen deltar från början i analysen. Vid ansökan till en AI-regulatorisk sandlåda kommer den första frågan att vara om det är ett AI-system. Detta är sannolikt en fråga som en leverantör behöver ha en första egen uppfattning om och som behörig myndighet behöver värdera innan någon beviljas att delta i sandlådan. Sannolikt kommer dock tröskeln vara låg för vad som är ett AI-system.

En erfarenhet i arbetet är att beskrivningen av AI-systemet behöver vara tillräckligt definierad för att kunna bedömas. Som konstateras i avsnitt 4.1 kan det vara svårt att initialt veta var gränsen går för vad som utgör AI-systemet. Gränsdragningen för vad som är ett AI-system har samtidigt betydelse för hur lösningen ska beskrivas vid en ansökan till en AI-regulatorisk sandlåda. Det är därmed bättre att börja brett i beskrivningen och snäva in vartefter.

Ytterligare en lärdom är att det i diskussionen om AI-system kan ske en sammanblandning av bedömningarna, dvs. att vad som är ett AI-system skulle ha ett beroende till vilka konsekvenser det innebär utifrån riskbestämmelserna. Det är därmed viktigt att tydliggöra att analysen av om projektet, som är avsett att delta i sandlådan, utgör ett AI-system alltid måste vara första steget. Därefter tas ställning till frågan om vilken risknivå AI-systemet tillhör.

En annan lärdom från denna första iteration är att det är lätt att hamna diskussioner om generella frågor, men det är i fallstudien som det är möjligt att bli konkret.

4.3 Om arbetssättet

Att göra saker i praktiken synliggör ofta andra problem än de som är uppenbara vid planering. Vid uppstart av ett nytt arbetssätt är det nödvändigt att vara öppensinnad och lösningsfokuserad, men ändå hålla fast vid en riktning. Det är inte möjligt att göra allt och lära allt vid ett tillfälle, utan det är genom upprepning och nya försök som lärande skapas. En viktig lärdom för det iterativa arbetssättet är med andra ord att se utmaningarna som finns men inte fastna vid att lösa dem, såvida de inte har en enkel lösning.

Integritetsskyddsmyndighetens arbete med regulatoriska sandlådor om dataskydd har visat att det är viktigt att de som deltar besitter både kompetens, mandat och förmåga till dialog och samma reflektion görs i denna sandlåda. I en sandlåda möts olika kompetenser och myndigheter för att skapa lärande på obruten mark. Det kräver en respektfull dialog där det inte finns några dumma frågor och allas bidrag tilldelas stort värde. Tvärfunktionellt arbete i nära samarbete är helt avgörande för framgång.

Förmågan 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 är avgörande för att arbetet i sandlådan ska kunna gå framåt. Vanligen möts väldigt specialiserade jurister och lika specialiserade tekniker i sandlådan för att gemensamt analysera hur ny teknik och ny lagstiftning samverkar och interagerar med varandra. Det är viktigt med en gemensam begreppsbild då det annars finns risk för olika tolkningar av begreppens betydelse.

Första iterationen har haft en styrgrupp och en arbetsgrupp, som delvis varit överlappande. Det stora lärandet både om arbetssätt och AI-förordningen sker dock i arbetsgruppen, vilken analyserar och arbetar med de materiella frågorna. Om man i uppstart av liknande arbete känner behov av både en styrgrupp och en arbetsgrupp behöver mandaten mellan grupperna vara tydligt.

Alla som deltar i en sandlåda behöver ha en utpekad roll, och man bör undvika att personer sitter med "för att lyssna" eftersom det kan hämma en fri och öppen dialog och skapa onödig prestige. En viktig framgångsfaktor är att alla deltagande är med från början till slut. Tillfälliga inhopp av medarbetare i arbetet behöver begränsas och om de ändå sker ha ett tydligt syfte. Det innebär inte att projektet ska undvika att söka stöd i svåra frågor, till exempel genom att vända sig till myndigheter eller forskare som specialiserar sig på en enskild fråga. Dessa ska dock inte betraktas som sandlådedeltagare.

En reflektion är att, om möjligt, försöka vara tydlig med förväntningarna på deltagarnas bidrag vid arbetsmötena, vilket underlättar och effektiviserar arbetet i sandlådan. Det kan också vara värdefullt med en informerad ledning som har insikt i förväntningarna på deltagandet i en sandlåda.

Det är en framgångsfaktor att dokumentera och producera skriven text under hela arbetet. Många gånger vid dialog tar deltagare till sig och minns olika saker. Genom att dokumentera vad gruppen är överens om uppmärksammas tidigt om det föreligger olika bilder om vad gruppen kommit fram till.

En skillnad från Integritetsskyddsmyndighetens regulatoriska sandlådor om dataskydd är att den AI-regulatoriska sandlådan kommer behöva ta hand om liknande frågor (såsom vad som är AI-system) varje gång, medan i Integritetsskyddsmyndighetens sandlådor byter man frågor. Det bör vara positivt med återkommande frågor, då processen bör kunna struktureras mycket mer.

Bilaga 1 Frågematris

AI-förordningen	Regulatorisk sandlåda	Verksamhetsfrågor	Nationella frågor
Är det ett AI-system?	Hur moget bör AI-systemet vara?	Vilka kompetenser krävs?	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?
Är det hög risk?	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?	Kan system som med hög sannolikhet inte är hög risk vara med i sandlådan? Bör begränsningar göras?		

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, Länsstyrelserna, Migrationsverket, Naturvårdsverket, Patent- och Registreringsverket, Pensionsmyndigheten, Riksarkivet, Rättsmedicinalverket, Sida, Skatteverket, Skolverket, Statens institutionsstyrelse, Statens servicecenter, Statens tjänstepensionsverk, Statens veterinärmedicinska anstalt, Statistiska centralbyrån, Tillväxtverket, Trafikverket, Transportstyrelsen, Tullverket och Universitets- och högskolerådet (februari 2024).