

Vidarebehandling av person- uppgifter i vårdnadsärenden för att träna en AI-modell

Slutrapport från Integritetsskyddsmyndighetens
regulatoriska sandlåda om dataskydd

Diarienummer
IMY-2025-23536

Datum
2025-12-09



Diarienummer:
IMY-2025-23536

Datum:
2025-12-09

Vidarebehandling av personuppgifter i vårdnadsärenden för att träna en AI-modell

Innehållsförteckning

Sammanfattning.....	2
1. Inledning.....	4
2. Projektet "Vidarebehandling av personuppgifter i vårdnadsärenden för att träna en AI-modell"	5
2.1. Projektets deltagare	5
2.2. Projektets mål.....	5
2.3. Rättsliga frågeställningar	5
2.4. Avgränsningar i projektet och annan upplysning.....	6
3. Beskrivning av projektet.....	7
3.1 Syntetisering av personuppgifter.....	7
3.1.1 För- och nackdelar med syntetisering	7
3.1.2 Anonymisering.....	8
3.2 Teknisk översikt av syntetiseringsprocessen i projektet	8
4. Är ändamålet med vidarebehandlingen av personuppgifter som samlats in inom ramen för ärendehandläggning förenligt med det ändamål som uppgifterna ursprungligen samlades in för?.....	11
4.1 Principen om ändamålsbegränsning enligt dataskyddsförordningen.....	11
4.2 Kan behandlingen av personuppgifter anses ske för statistiska ändamål?	11
4.2.1 Statistiska ändamål enligt dataskyddsförordningen	11
4.2.2 IMY:s kommentarer	12
4.3 Är behandlingen förenlig med de ursprungliga ändamålen enligt artikel 6.4 i dataskyddsförordningen?.....	13
4.3.1 Bedömningen av förenliga ändamål enligt dataskyddsförordningen	13
4.3.2 IMY:s kommentarer	14
<i>Träning av AI-modellen med personuppgifter</i>	14
<i>Behandling av personuppgifter för syntetisering</i>	15
5. Avslutande kommentar	16

Postadress:
Box 8114
104 20 Stockholm

Webbplats:
www.imy.se

E-post:
imy@imy.se

Telefon:
08-657 61 00

Sammanfattning

- Projektet "Vidarebehandling av personuppgifter i vårdnadsärenden för att träna en AI-modell" är Integritetsskyddsmyndighetens (IMY:s) femte projekt i myndighetens regulatoriska sandlåda om dataskydd. Familjens Jurist har varit deltagare i projektet. Målet har varit att undersöka några av de juridiska frågor som uppkommer när man utvecklar en AI-lösning som ett digitalt verktyg för att effektivisera och förbättra byråns hantering av vårdnadsärenden.
- Vårdnadsärenden präglas ofta av konflikter mellan parterna och har stor påverkan på de enskilda, särskilt de berörda barnen. Familjens Jurist har en omfattande kunddatabas med stora mängder ärendedata, och de vill med hjälp av AI sammanställa och analysera sina avslutade vårdnadsärenden för att hitta mönster och riskfaktorer i ärendena. På så sätt skulle de kunna få evidensbaserad kunskap om vårdnadsärenden, och använda den för att tidigt identifiera ärenden med hög risk för att bli konfliktdrabbade, hamna i domstol eller generera långvariga processer, och sätta in lämpliga resurser i ärendena. Därigenom skulle klientverksamheten kunna effektiviseras och förbättras. Det skulle samtidigt förkorta processerna och minska den påverkan som vårdnadsprocesser ofta innebär för de enskilda.
- Projektet omfattar två scenarier. Familjens Jurist vill träna en AI-modell som ska kunna analysera och förbättra processerna i vårdnadsärenden. För att det ska vara möjligt enligt dataskyddsförordningen vill de göra detta genom att skapa syntetiska uppgifter, det vill säga artificiellt framställda data, och sedan träna AI-modellen på dessa uppgifter. Målet är att skapa anonyma uppgifter, som dataskyddsförordningen inte är tillämplig på. Att behandla personuppgifter för att framställa syntetiska data är dock en personuppgiftsbehandling som behöver leva upp till kraven i dataskyddsförordningen. Familjens Jurist har ställt frågor till IMY om bedömningen av principen om ändamålsbegränsning vid vidarebehandling av personuppgifter för att skapa syntetiska uppgifter.
- Möjligheten att vidarebehandla personuppgifter är relevant för många verksamheter som vill utveckla och använda AI. Vid träning på syntetiska uppgifter blir dock AI-modellens potential till viss del begränsad i jämförelse med om träningen istället hade utförts på de ursprungliga uppgifterna, eftersom precisionen och representativiteten riskerar att försämrats. Det är därför inte alltid aktuellt för verksamheter som vill utveckla AI att träna en AI-modell på syntetiska uppgifter. Av denna anledning har IMY valt att även bedöma de rättsliga frågeställningarna utifrån scenariot att AI-modellen tränas på faktiska personuppgifter.

• Vägledningen i projektet har fokuserat på rättsliga frågeställningar kopplade till principen om ändamålsbegränsning. För att en vidarebehandling av personuppgifterna ska få ske för att träna AI-modellen med personuppgifter, eller för att skapa syntetiska uppgifter, får behandlingen inte vara oförenlig med de ursprungliga ändamålen.

• **Fråga 1: Kan behandlingen av personuppgifter anses ske för statistiska ändamål?** IMY har som utgångspunkt bedömt att mycket talar för att varken träningen av AI-modellen med personuppgifter eller syntetiseringen av uppgifterna kan anses ske för statistiska ändamål.

• **Fråga 2: Är behandlingen förenlig med de ursprungliga ändamålen enligt artikel 6.4 i dataskyddsförordningen?** IMY har även gjort en förenlighetsbedömning enligt artikel 6.4 i dataskyddsförordningen, och bedömt att mycket talar för att varken träningen av AI-modellen med personuppgifter eller syntetiseringen av uppgifterna kan anses vara en vidarebehandling som är förenlig med de ursprungliga ändamålen. IMY har dock bedömt att en tillåten vidarebehandling skulle kunna ske om ytterligare skyddsåtgärder i form av en opt-out lösning ges i samband med att information lämnas till de registrerade i enlighet med artiklarna 13.3 och 14.4.

• IMY vill framhålla att syntetiseringen av personuppgifter för träning av AI-modeller är ett integritetsvänligt sätt för att minska riskerna för de registrerades rättigheter och är en viktig åtgärd i bedömningen av om vidarebehandlingen kan anses vara tillåten eller inte.

IMY:s regulatoriska sandlåda om dataskydd

Med regulatorisk sandlåda avser IMY **fördjupad vägledning om hur dataskyddsregelverket bör tolkas och tillämpas**. Inget undantag ges från reglerna i dataskyddsförordningen eller dess kompletterande lagstiftning utan vägledningen baseras på gällande rätt.

Kännetecknande för arbetssättet är att:

- IMY och innovationsaktörerna tillsammans identifierar de rättsliga frågor som vägledningen ska fokusera på,
- vägledning ges muntligt vid flera tillfällen under några månaders tid i form av workshops eller andra dialogbaserade former, och
- resonemang och bedömningar från vägledningen sammanfattas och publiceras i en publik rapport för att möjliggöra lärande för många.

1. Inledning

Frågor om barn och vem som ska ha vårdnaden om ett barn väcker ofta starka känslor för de inblandade och kan ha stor påverkan på enskilda och ge upphov till konflikter. Även om varje fall är unikt, går det i vårdnadsärenden ofta att identifiera faktorer som påverkar konfliktgraden. I vårdnadsärenden kan de juridiska ombuden spela stor roll för hur konflikterna utvecklar sig och hur parterna samarbetar. Att hitta sätt att identifiera riskfaktorer och därmed tidigt i processen kunna vidta lämpliga åtgärder utifrån situationen, skulle kunna bidra till att skapa bättre stöd för alla inblandade parter, särskilt för de barn som ofta påverkas negativt av processen.

Användningen av artificiell intelligens (AI) ökar i samhället, och allt fler verksamheter upptäcker nyttan med AI. Familjens Jurist har identifierat områden i sin verksamhet där AI skulle kunna vara till stor nytta. Detta projekt i IMY:s regulatoriska sandlåda har haft som syfte att utforska om personuppgifter i vårdnadsärenden kan behandlas för att skapa syntetiska data i syfte att använda dem för att träna en AI-modell som ska kunna analysera och förbättra processerna i vårdnadsärenden.

Utvecklingen och användningen av AI kräver ett noggrant förarbete där bland annat dataskyddsregelverket behöver beaktas. Dataskyddsregelverket och särskilt dataskyddsförordningen (GDPR)¹ ställer krav på hur personuppgifter får behandlas och under vilka förutsättningar personuppgifter exempelvis får lämnas ut. Detta innebär att utveckling och användning av AI behöver genomföras med omsorg för att säkerställa att integritetsskyddet upprätthålls. Många verksamheter har samlat in personuppgifter för ett specifikt ändamål och vill sedan vidarebehandla dessa uppgifter för andra ändamål, till exempel för att utveckla AI. IMY har sedan tidigare tagit fram en allmän vägledning om GDPR och AI.² Denna vägledning kan användas av verksamheter som ett stöd i utvecklingen och användningen av AI.

För att leva upp till kraven i dataskyddsförordningen behöver verksamheter göra noggranna bedömningar och avvägningar mellan integritetsintrånget och nyttan av personuppgiftsbehandlingen. Familjens Jurist önskar träna en AI-modell som ska kunna analysera och förbättra processerna i vårdnadsärenden. För att göra det vill de skapa syntetiska uppgifter som inte längre möjliggör identifiering av enskilda personer, det vill säga uppgifter som är anonyma i dataskyddsförordningens mening. Den aktuella AI-modellen ska sedan tränas på de syntetiska uppgifterna. Anonyma uppgifter är inte personuppgifter, och dataskyddsförordningen är inte tillämplig på behandling av sådana uppgifter. Att behandla personuppgifter för att framställa syntetiska data behöver dock leva upp till kraven i dataskyddsförordningen. Familjens Jurist har ställt frågor till IMY om bedömningen av principen om ändamålsbegränsning i dataskyddsförordningen vid en vidarebehandling av personuppgifter för att skapa syntetiska uppgifter.

Bedömningen av principen om ändamålsbegränsning är relevant för många verksamheter som vill utveckla och använda AI. Vid träning på syntetiska uppgifter finns det dock risk för att möjligheterna med AI-modellen till viss del begränsas jämfört med om träningen istället hade utförts på de ursprungliga uppgifterna eftersom precisionen och representativiteten kan bli sämre. Det kan i sin tur påverka AI-modellens förmåga att producera korrekta resultat. Det är därför inte alltid aktuellt för verksamheter som vill utveckla AI att träna modellen på syntetiska uppgifter. Av denna anledning har IMY valt att även bedöma de rättsliga frågeställningarna utifrån scenariot att AI-modellen tränas med faktiska personuppgifter.

¹ Europaparlamentets och rådets förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter och om upphävande av direktiv 95/46/EG (allmän dataskyddsförordning).

² Vägledning om GDPR och AI (<https://www.imy.se/verksamhet/dataskydd/innovationsportalen/vagledning-om-gdpr-och-ai/>).

2. Projektet "Vidarebehandling av personuppgifter i vårdnadsärenden för att träna en AI-modell"

2.1. Projektets deltagare

Familjens Jurist är en av Sveriges största juristbyråer med ca 200 kontor i hela Sverige. Byrån specialiserar sig på juridik inom privat- och familjeliv, företagande och boende.

Observatörer under projektet

RISE Research Institutes of Sweden och SynData AB har under projektets gång deltagit som observatörer. RISE:s roll har framför allt varit att lyssna in under projektets gång och skaffa sig en bild över hur arbetet med IMY:s regulatoriska sandlåda om dataskydd bedrivs i praktiken. RISE har även varit observatör i ett tidigare projekt i IMY:s regulatoriska sandlåda.³ SynData AB, som utvecklar teknik för syntetisering av data, deltog som observatör med fokus på de tekniska och metodmässiga aspekterna av syntetisering och anonymisering. Bolaget bidrog med teknisk expertis.

2.2. Projektets mål

Familjens Jurists syfte med det aktuella projektet är att utveckla en AI-modell som ska kunna sammanställa och analysera data i ärenden, däribland vårdnadsärenden, som till sin natur är särskilt känsliga. Vårdnadsärenden präglas ofta av konflikter mellan parterna och har stor påverkan på enskilda, särskilt de berörda barnen. Familjens Jurist har en omfattande kunddatabas med stora mängder ärendedata, och de vill med hjälp av AI sammanställa och analysera sina avslutade vårdnadsärenden för att hitta mönster och riskfaktorer i ärendena. På så sätt skulle de kunna få evidensbaserad kunskap om vårdnadsärenden, och använda den för att tidigt identifiera ärenden med hög risk för att bli konfliktdrabbade, hamna i domstol eller generera långvariga processer, och sätta in lämpliga resurser i ärendena. Det kan till exempel handla om att tidigt i ett vårdnadsärende använda en modell för att avgöra om det är lämpligt att föreslå medlingssamtal redan innan den förväntade konflikten uppstår. Därigenom skulle Familjens Jurist, och i förlängningen den enskilda juristens arbete i vårdnadsärenden, kunna utvecklas. Det skulle också kunna förkorta processerna och minska den påverkan som vårdnadsprocesser ofta innebär för de enskilda. Resultatet kommer att behandlas av Familjens Jurist internt för att förbättra byråns egna processer. Resultatet skulle även kunna användas för forskningsändamål inom ramen för samarbete mellan Familjens Jurist och universitet och högskolor.

2.3. Rättsliga frågeställningar

Familjens Jurist vill som utgångspunkt träna AI-modellen på syntetiska uppgifter som är anonyma i dataskyddsförordningens mening. Anonyma uppgifter är inte personuppgifter, men att behandla personuppgifter för att framställa syntetiska uppgifter utgör en personuppgiftsbehandling som måste leva upp till kraven i dataskyddsförordningen. Många verksamheter har dock önskemål om att träna AI-

³ RISE deltog som observatör i ett projekt som mynnade ut i rapporten "Utlämnande av allmänna handlingar med hjälp av AI" (<https://www.imy.se/publikationer/utlamnande-av-allman-na-handlingar-med-hjalp-av-ai/>).

modeller på personuppgifter. IMY har därför valt att undersöka de rättsliga frågeställningarna i projektet utifrån två scenarier: träning av en AI-modell med personuppgifter samt behandling av personuppgifter för att skapa syntetiska uppgifter, vilka utgör anonyma data, och som sedan används som träningsdata.

Följande frågeställningar har behandlats:

Är ändamålet med vidarebehandlingen av personuppgifter som samlats in inom ramen för ärendehandläggning förenligt med det ändamål som uppgifterna ursprungligen samlades in för?

- Kan behandlingen av personuppgifter anses ske för statistiska ändamål?
- Är behandlingen förenlig med de ursprungliga ändamålen enligt artikel 6.4 i dataskyddsförordningen?

2.4. Avgränsningar i projektet och annan upplysning

I denna rapport tar IMY endast ställning till om behandlingen av personuppgifter för att skapa syntetiska uppgifter är förenlig med dataskyddsförordningens krav. IMY tar inte ställning till om de syntetiska uppgifterna utgör anonyma uppgifter i dataskyddsförordningens mening. IMY:s bedömningar utgår dock från att syntetiseringen medför att personuppgifterna blir anonyma. Det bör också understrykas att den personuppgiftsbehandling som är aktuell i detta projekt, det vill säga behandlingen av personuppgifter för att ta fram syntetiska uppgifter, inte pågår i nuläget.

Dataskyddsförordningen ställer vissa krav för att uppgifter ska anses vara anonyma, vilket Europeiska dataskyddsstyrelsen (EDPB) ger vägledning om i riktlinjerna om avidentifieringstekniker.⁴ Det bör noteras att det för närvarande pågår ett arbete med att ta fram nya riktlinjer om anonymisering av uppgifter i EDPB.

Rapporten kommer inte att beröra AI-förordningen⁵ som trädde i kraft den 1 augusti 2024.

Denna rapport innehåller en redogörelse för IMY:s uppfattning i rättsliga frågor där det kan saknas vägledande domstolspraxis eller vägledning från EDPB. Bedömningarna görs mot bakgrund av nuvarande rättsläge och kan komma att ändras om det skulle komma ny lagstiftning, domstolspraxis eller vägledning från EDPB.

⁴ WP29 Opinion 05/2014.

⁵ 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).

3. Beskrivning av projektet

Projektet har undersökts utifrån två olika scenarier. Familjens Jurist har för avsikt att behandla personuppgifter som ursprungligen samlats in i vårdnadsärenden för att utveckla ett AI-verktyg. Som utgångspunkt vill Familjens Jurist undersöka om det är möjligt att skapa syntetiska uppgifter baserade på sådana personuppgifter som verksamheten har tillgång till, och därefter träna AI-modellen med dessa syntetiska uppgifter. Syftet med syntetiseringen är att uppgifterna ska bli anonyma, och att AI-modellen således inte tränas på personuppgifter. Att skapa syntetiska data innebär en behandling av personuppgifter som måste vara förenlig med kraven i dataskyddsförordningen. I projektet har IMY även valt att undersöka om det finns rättsligt stöd för att träna AI-modellen med personuppgifter. Familjens Jurist har dock inte för avsikt att träna den aktuella AI-modellen med faktiska personuppgifter.

3.1. Syntetisering av personuppgifter

Syntetisering av personuppgifter innebär att man skapar artificiella data som efterliknar egenskaperna hos en verklig datamängd, men utan att direkt återge de ursprungliga värdena. Syntetiska datamängder genereras ofta med hjälp av avancerade statistiska metoder eller maskininlärningsmodeller som exempelvis generativa neurala nätverk, på engelska *Generative Adversarial Networks (GAN)*, där syftet är att bevara de statistiska mönster och korrelationer som finns i det ursprungliga datamaterialet utan att exponera underliggande identifierbara personuppgifter.

Tekniken används bland annat i syfte att möjliggöra analys, testning och utveckling av modeller i situationer där det av olika skäl inte är möjligt eller önskvärt att använda verkliga data.

3.1.1. För- och nackdelar med syntetisering

Syntetisering är ett exempel på en integritetshöjande teknik, på engelska *Privacy Enhancing Technology (PET)*. Syntetisering av datamängder har betydande fördelar. Den främsta fördelen i samband med utveckling av AI är att tekniken kan reducera behovet av att behandla personuppgifter vid modellträningen och därmed minska risken för bland annat dataintrång, obehörig åtkomst och andra integritetskränkningar.

En utmaning med syntetisering är att säkerställa att de syntetiska uppgifterna inte kan kopplas tillbaka till verkliga individer, samtidigt som uppgifterna bibehåller tillräcklig datakvalitet för att vara användbara vid till exempel träning av en AI-modell. Om syntetiseringen inte utförs korrekt, finns risk för återidentifiering, särskilt om syntetiska data kombineras med andra datakällor.

Det är viktigt att notera att hur syntetiseringen genomförs kan påverka hur väl den syntetiska datamängden speglar precisionen och representativiteten i den ursprungliga datamängden. Brister i processen kan leda till felaktiga slutsatser eller bias (partiskhet) i analys och modellutveckling. Kvaliteten på syntetiska uppgifter mäts i hur väl den bevarar korrelationer, distributioner och mönster från aktuella ursprungsdata.

Det kan vara svårt att uppnå en perfekt balans mellan säkerställandet av att datamängden inte längre innehåller personuppgifter och bibehållen nytta med dessa data, vilket gör att syntetisering inte alltid kan ersätta verkliga data.

Även om syntetiseringen inte alltid innebär att uppgifterna blir helt anonyma i dataskyddsförordningens mening, kan syntetiseringen utgöra en användbar PET-

lösning som minskar riskerna för de registrerade och som är ett bra komplement till det övriga datasäkerhetsarbetet.

3.1.2. Anonymisering

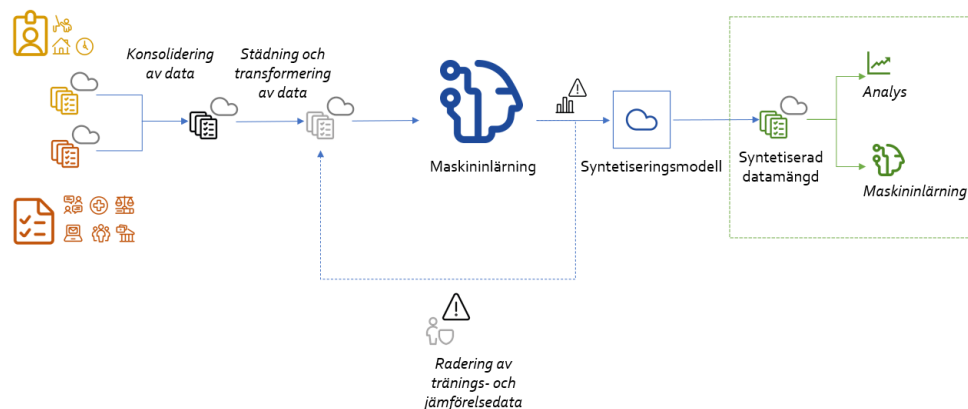
Som nämnts ovan pågår det för närvarande ett arbete inom EDPB med att ta fram riktlinjer om anonymisering av uppgifter och om vilka krav som ställs för att uppgifter ska betraktas som anonyma enligt dataskyddsförordningen. Viss vägledning om vad som krävs finns dock i skäl 26 i dataskyddsförordningen, där det framgår att för att avgöra om en fysisk person är identifierbar bör man beakta alla hjälpmedel, som till exempel utgallring, som, antingen av den personuppgiftsansvarige eller av en annan person, rimligen kan komma att användas för att direkt eller indirekt identifiera den fysiska personen. För att fastställa om hjälpmedel med rimlig sannolikhet kan komma att användas för att identifiera den fysiska personen bör man beakta samtliga objektiva faktorer, såsom kostnader och tidsåtgång för identifiering, med beaktande av såväl tillgänglig teknik vid tidpunkten för behandlingen som den tekniska utvecklingen.

Det kan vara en utmaning för verksamheter som vill anonymisera uppgifter att uppfylla kraven i dataskyddsförordningen. Även i situationer där full anonymitet inte kan uppnås, och uppgifterna alltså utgör personuppgifter, är integritetshöjande åtgärder positiva ur ett integritetssperspektiv. Särskilt vid behandling av personuppgifter som medför höga risker för enskilda kan sådana åtgärder minska riskerna för de registrerade och vara en förutsättning för att behandlingen ska uppfylla kraven i dataskyddsförordningen och kunna genomföras.

I det aktuella projektet har IMY inte bedömt om de syntetiska uppgifterna utgör anonyma uppgifter i dataskyddsförordningens mening, utan endast prövat om det, utifrån omständigheterna i projektet, finns juridiska förutsättningar att genomföra en anonymisering av personuppgifter i syfte att använda dessa uppgifter för träning av en AI-modell.

3.2. Teknisk översikt av syntetiseringsprocessen i projektet

Processen att vidarebehandla uppgifterna i syfte att syntetisera dem sker i flera steg.



När ett vårdnadsärende inleds hos Familjens Jurist samlas all relevant information i verksamhetens ärende- och dokumenthanteringssystem. Där registreras bland annat uppgifter om namn, adress, ålder, ärendetyp, handläggningstid och uppgifter om

fakturering. Även känsliga personuppgifter, såsom psykisk ohälsa och missbruksproblematik, och andra integritetskänsliga uppgifter, som till exempel uppgifter om våld inom relationen, kan framgå i systemen. Utifrån ärendets karaktär och utveckling kan även ytterligare uppgifter dokumenteras, exempelvis:

- interna rättsutredningar
- korrespondens med klienter och motparter
- kommunikation med myndigheter och försäkringsbolag
- beslut från myndigheter och domstolar
- uppgifter om involverade barn och andra familjemedlemmar.

Konsolidering

Uppgifter från avslutade ärenden som är relevanta för ändamålet med träningen är fördelade över flera system. För att möjliggöra behandlingen behöver relevant information från varje enskilt ärende sammanställas (konsolideras) i ett separat system. I detta steg sker ett visst urval genom att vissa ärendetyper väljs ut. Familjens Jurist har uppgett att det behövs cirka 15–20 000 ärenden för att uppnå en statistiskt korrekt modell.

Städning och transformering

De uppgifter som har samlats in från olika system består av både strukturerad och ostrukturerad information av varierande kvalitet, vilket gör det nödvändigt att utföra en städning av uppgifterna. Endast den information som är relevant för syftet med träningen väljs ut ur datamängden utifrån förvalda variabler. Syftet med städningen är att säkerställa att all träningsdata håller tillräckligt hög kvalitet och innehåller så få felaktiga, svårtolkade och ostrukturerade uppgifter som möjligt.

Uppgifterna korrigeras och struktureras till ett enhetligt och standardiserat format. Uppgifter som är irrelevanta, bristfälliga, har för låg kvalitet eller riskerar att leda till missvisande slutsatser sorteras bort.

Modellträning

Efter att uppgifterna har extraherats, städats och transformerats kommer en syntetiseringsmodell att utvecklas utifrån den bearbetade datamängden. Modellträningen kan delvis baseras på generativa neurala nätverk som tränas för att generera en ny syntetisk datamängd. Denna syntetiska datamängd ska bevara de statistiska egenskaperna från originaldata, men utan att kunna kopplas till enskilda individer.

För att ytterligare stärka integritetsskyddet och försvåra möjligheten för återidentifiering kan kompletterande tekniker tillämpas, såsom tillförandet av brus⁶ i de personuppgifter som förekommer i aktuella träningsdata.

⁶ Denna metod brukar kallas för *differential privacy*.

Efter avslutad träning säkerställs det att syntetiseringsmodellen kan generera användbara syntetiska data som speglar den statistiska informationen i de ursprungliga uppgifterna, och att det samtidigt är svårt att härleda de statistiska egenskaperna till de personuppgifter som de ursprungligen härrör från. Utvärdering sker genom jämförelser mellan de syntetiska och de ursprungliga uppgifterna, där statistiska likheter och skillnader analyseras med hjälp av olika beräkningsmetoder. Dessutom genomförs riskbedömningar för att upptäcka möjligt informationsläckage i den syntetiska datamängden, även i kombination med andra datamängder som skulle kunna möjliggöra identifiering av enskilda individer. Denna riskbedömning görs för att hantera och minska risken för återidentifiering. När syntetiseringsmodellen uppfyller uppsatta krav kommer verksamhetens referensdata, det vill säga datamängden som använts vid träning och utvärdering, att raderas.

Syntetiseringsprocessen är tänkt att återupprepas regelbundet för att syntetiseringsmodellen ska kunna tränas på uppdaterade uppgifter. Nya uppgifter kommer då att extraheras från ärendesystemen, konsolideras, städas och transformeras, medan äldre data kan återanvändas.

Användning

Syntetiseringsmodellen kan användas för att skapa syntetiska data som har likvärdiga fördelningar och statistiska egenskaper som personuppgifterna i vårdnadsärendena. Den syntetiska datamängden ska kunna användas för att träna den AI-modell som Familjens Jurist önskar utveckla.

4. Är ändamålet med vidarebehandlingen av personuppgifter som samlats in inom ramen för ärendehandläggning förenligt med det ändamål som uppgifterna ursprungligen samlades in för?

4.1. Principen om ändamålsbegränsning enligt dataskyddsförordningen

Vid all typ av personuppgiftsbehandling krävs att behandlingen är förenlig med de grundläggande principerna i artikel 5 i dataskyddsförordningen. En av de grundläggande principerna är principen om ändamålsbegränsning i artikel 5.1 b. Den innebär att personuppgifter ska samlas in för särskilda, uttryckligt angivna och berättigade ändamål och inte senare behandlas på ett sätt som är oförenligt med dessa ändamål.

Bestämmelsen innehåller alltså två krav, där det ena avser ändamålet med den ursprungliga insamlingen av personuppgifter, och det andra avser den senare behandling av uppgifterna.⁷ Bestämmelsen innebär att insamlingsändamålen sätter gränserna för hur personuppgifterna får användas för nya ändamål.

Av skäl 39 i dataskyddsförordningen följer att de specifika ändamål som personuppgifterna behandlas för bör vara tydliga och legitima och ha bestämts vid den tidpunkt då personuppgifterna samlades in. Att ändamålet ska vara uttryckligt angivet innebär att det måste framgå tydligt. Ett tydligt angivet ändamål är en förutsättning för att kunna bedöma om en viss behandling är laglig, det vill säga om det finns en rättslig grund för behandlingen enligt artikel 6 i dataskyddsförordningen.⁸ Kravet på att ändamålet ska vara berättigat innebär att det måste garantera en laglig behandling av personuppgifter.⁹

4.2. Kan behandlingen av personuppgifter anses ske för statistiska ändamål?

4.2.1. Statistiska ändamål enligt dataskyddsförordningen

Familjens Jurist har ställt frågan om vidarebehandlingen av personuppgifterna för att träna AI-modellen, eller syntetiseringen, kan anses ske för statistiska ändamål.

Av artikel 5.1 b i dataskyddsförordningen följer att ytterligare behandling för arkivändamål av allmänt intresse, vetenskapliga eller historiska forskningsändamål eller statistiska ändamål inte ska anses som oförenlig med de ursprungliga, förutsatt att behandlingen omfattas av lämpliga skyddsåtgärder enligt artikel 89.1.

Av skäl 162 i dataskyddsförordningen framgår att med statistiska ändamål avses varje åtgärd som vidtas för den insamling och behandling av personuppgifter som är nödvändig för "statistiska undersökningar" eller för "framställning av statistiska resultat". Vidare framgår att ett statistiskt ändamål innebär att resultatet av behandlingen för statistiska ändamål inte består av personuppgifter, utan av

⁷ EU-domstolen, dom av den 20 oktober 2022, mål C-77/21, Digi, EU:C:2022:805, p. 26.

⁸ Prop. 2017/18:105, s. 47 f.

⁹ EU-domstolen, dom av den 24 februari 2022, mål C-175/20, Valsts ierņēmumu dienests, EU:C:2022:124, p. 63–66.

aggregerade personuppgifter, och att resultatet eller uppgifterna inte används till stöd för åtgärder eller beslut som avser en särskild fysisk person. Enligt 4 kap. 2 § lagen (2018:218) med kompletterande bestämmelser till EU:s dataskyddsförordning (dataskyddslagen) får personuppgifter som behandlas för enbart statistiska ändamål användas för att vidta åtgärder gällande den registrerade om det finns synnerliga skäl med hänsyn till den registrerades vitala intressen.

Begreppen "statistisk undersökning" och "statistiska resultat" definieras inte i dataskyddsförordningen. Regeringen har uttalat att statistik är metoder för att samla in, bearbeta, utvärdera och analysera data eller information. Även resultatet av ett sådant arbete, ofta redovisat i numerisk form, benämns ofta som statistik.¹⁰ I en rekommendation från Europarådet definieras statistiska resultat som "information som har framställts genom behandling av personuppgifter i syfte att karaktärisera ett kollektivt fenomen i en given population".¹¹

Aggregerad information kan beskrivas som information om enskilda individer som har slagits samman för att skapa en samlad, statistisk, bild på gruppnivå. Det innebär att huvudfokus ligger på att identifiera mönster, trender och sammanfattningar över en kollektiv grupp eller befolkning, snarare än på de specifika egenskaperna hos en enskild person.

Behandling av personuppgifter för statistiska ändamål förekommer inom såväl offentlig som privat sektor, både som en självständig verksamhet och som en uppföljande åtgärd till annan verksamhet.¹² Vissa myndigheter har dock tilldelats uppgiften att ta fram officiell statistik och annan reglerad statistik som en uppgift av allmänt intresse med stöd av artikel 6.1 e i dataskyddsförordningen.

4.2.2. IMY:s kommentarer

Familjens Jurist har uppgett att ändamålet med den aktuella personuppgiftsbehandlingen är att utveckla en AI-modell som ska kunna analysera ärendedata för att effektivisera och förbättra framtida klientarbete och Familjens Jurists verksamhet. Detta kan ske antingen genom att personuppgifterna används för att träna AI-modellen eller för att skapa syntetiska data som i sin tur används för att träna modellen.

Vidarebehandling av personuppgifter för statistiska ändamål är särskilt gynnad i dataskyddsförordningen, genom att vidarebehandlingen ska anses vara förenlig med insamlingsändamålen.

Med statistiska ändamål avses i dataskyddsförordningen sådan personuppgiftsbehandling som sker för statistiska undersökningar eller för framställning av statistiska resultat. Slutresultatet ska vara aggregerade uppgifter och inte personuppgifter.

Syftet med Familjens Jurists tänkta personuppgiftsbehandling är inte att genomföra en statistisk undersökning eller att framställa statistiska resultat i den mening som normalt avses med dessa uttryck, utan att träna AI-modellen respektive framställa syntetiska data. Behandlingen av personuppgifter innebär inte heller att det ska ske en

¹⁰ Prop. 2017/18:107, Ändringar i vissa författningar inom Finansdepartementets ansvarsområde med anledning av EU:s dataskyddsreform, s. 19.

¹¹ Recommendation No. R (97) s. 18.

¹² SOU 2017:39 s. 234.

aggregering av uppgifter som skapar en bild på gruppnivå.¹³ Mot den bakgrunden bedömer IMY att mycket talar för att vidarebehandlingen av personuppgifter inte sker för statistiska ändamål.

För att den ytterligare behandlingen av personuppgifter ska kunna genomföras måste det istället göras en bedömning av om ändamålet med behandlingen är förenlig med de ursprungliga ändamålen enligt artikel 6.4 i dataskyddsförordningen.

4.3. Är behandlingen förenlig med de ursprungliga ändamålen enligt artikel 6.4 i dataskyddsförordningen?

4.3.1. Bedömningen av förenliga ändamål enligt dataskyddsförordningen

Enligt artikel 5.1 b i dataskyddsförordningen får personuppgifter efter insamlingen inte vidarebehandlas på ett sätt som är oförenligt med de ursprungliga ändamålen. All behandling som sker efter den ursprungliga behandlingen utgör en vidarebehandling av uppgifterna. Frågan om en vidarebehandling är förenlig med de ursprungliga ändamålen uppkommer dock endast om de nya ändamålen inte är identiska med de ursprungliga.¹⁴ I sådana fall måste en förenlighetsprövning göras.

I artikel 6.4 i dataskyddsförordningen anges ett antal kriterier som den personuppgiftsansvarige ska beakta vid bedömningen av om vidarebehandlingen är förenlig med det ändamål för vilket personuppgifterna ursprungligen samlades in:

- a) Kopplingar mellan de ändamål för vilka personuppgifterna har samlats in och ändamålen med den avsedda vidarebehandlingen.
- b) Det sammanhang inom vilket personuppgifterna har samlats in, särskilt förhållandet mellan de registrerade och den personuppgiftsansvarige.
- c) Personuppgifternas art, särskilt huruvida särskilda kategorier av personuppgifter behandlas i enlighet med artikel 9 eller huruvida personuppgifter om fällande domar i brottmål och lagöverträdelser som innefattar brott behandlas i enlighet med artikel 10.
- d) Eventuella konsekvenser för registrerade av den planerade fortsatta behandlingen.
- e) Förekomsten av lämpliga skyddsåtgärder, vilket kan inbegripa kryptering eller pseudonymisering.

Av skäl 50 i dataskyddsförordningen framgår att för att fastställa om ett ändamål med vidarebehandlingen är förenligt med det ändamål för vilket personuppgifterna ursprungligen insamlades bör den personuppgiftsansvarige särskilt beakta de registrerades rimliga förväntningar till följd av förhållandet till den personuppgiftsansvarige.

EU-domstolen har ansett att kriterierna i artikel 6.4 i förordningen ger uttryck för behovet av ett konkret, logiskt och tillräckligt nära samband mellan ändamålet med den ursprungliga insamlingen av personuppgifterna och den senare behandlingen av dem, och att kriterierna även gör det möjligt att säkerställa att sådan senare behandling inte avviker från registrerades berättigade förväntningar på hur deras

¹³ Jfr uttalanden från den franska dataskyddsmyndigheten (CNIL) i artikeln Ensuring the lawfulness of the data processing - In case of re-use of data, carrying out the necessary additional tests and verifications den 7 juni 2024 (<https://www.cnil.fr/en/ensuring-lawfulness-data-processing-re-use>).

¹⁴ EU-domstolen, dom av den 20 oktober 2022, mål C-77/21, Digi, EU:C:2022:805, p. 34.

uppgifter kan komma att användas. Vidare menade EU-domstolen att kriterierna i artikel 6.4 gör det möjligt att säkerställa en balans mellan, å ena sidan, behovet av förutsägbarhet och rättssäkerhet vad gäller de ändamål dessa uppgifter ursprungligen samlats in för och, å andra sidan, medgivandet av en viss flexibilitet till förmån för den personuppgiftsansvarige vid hanteringen av uppgifterna.¹⁵

Av artiklarna 13.3 och 14.4 framgår att om den personuppgiftsansvarige avser att ytterligare behandla personuppgifterna för ett annat syfte än det för vilket de insamlades, ska den personuppgiftsansvarige före denna ytterligare behandling ge den registrerade information om detta andra syfte samt ytterligare relevant information enligt punkten 2 i artiklarna. För att informationsskyldigheten ska inträda förutsätts att vidarebehandlingen är tillåten. I artiklarna 13.4 och 14.5 finns vissa undantag från informationsskyldigheten.

4.3.2. IMY:s kommentarer

Träning av AI-modellen med personuppgifter

Familjens Jurist har ursprungligen samlat in uppgifterna, som kan användas för träning av AI-modellen, för att handlägga byråns vårdnadsärenden. Extraheringen av personuppgifter ur Familjens Jurists ärendesystem för att skapa träningsdata innebär att det sker en vidarebehandling av personuppgifterna. Ändamålet med vidarebehandlingen är att träna en AI-modell som ska kunna sammanställa och analysera vårdnadsärenden i syfte att effektivisera och förbättra klientarbetet.

Det nya ändamålet är enligt IMY inte identiskt med det ursprungliga ändamålet. Det behöver därför göras en bedömning enligt artikel 6.4 för att avgöra om behandlingen är förenlig med det ursprungliga ändamålet.

IMY bedömer att det finns en naturlig koppling mellan ändamålen, eftersom vidarebehandlingen syftar till att förbättra den verksamhet som uppgifterna har samlats in i. En sådan koppling talar i många fall för att en vidarebehandling av personuppgifter för verksamhetsutvecklingsändamål kan anses vara förenlig med de ursprungliga ändamålen.

I detta fall finns dock särskilda omständigheter som talar för att de registrerade inte kan förvänta sig att uppgifterna vidarebehandlas för verksamhetsutvecklingsändamål. Av betydelse är att personuppgifterna har samlats in inom ramen för ett juridiskt ombuds klientverksamhet, vilket innebär att de registrerade har berättigade förväntningar att uppgifterna behandlas med stor aktsamhet. Detta gäller i synnerhet som personuppgifterna avser vårdnadsärenden, som får anses vara av särskilt privat och känslig natur. Aktuella ärendedata innehåller dessutom uppgifter om andra än byråns klienter, till exempel motparter, som inte själva har lämnat sina uppgifter till Familjens Jurist.

Vidare är Familjens Jurists ärendedata av mycket känsligt slag. Den innehåller känsliga personuppgifter enligt artikel 9 i dataskyddsförordningen, uppgifter om lagöverträdelse enligt artikel 10, uppgifter om enskildas ekonomi samt uppgifter om barn.

¹⁵ EU-domstolen, dom av den 20 oktober 2022, C-77/21, Digi, EU:C:2022:805, p. 36–37.

Sammantaget anser IMY att det finns flera omständigheter som pekar i riktningen att behandlingen av personuppgifter för att träna AI-modellen inte kan anses vara en tillåten vidarebehandling.

För att en vidarebehandling ska anses vara tillåten kan därför särskilda skyddsåtgärder behöva övervägas. Exempelvis kan informationen som ska lämnas enligt artiklarna 13 och 14 förenas med en så kallad opt-out lösning, det vill säga en möjlighet för de registrerade att välja att inte låta personuppgifterna vidarebehandlas. Med en opt-out lösning finns det mycket som talar för att behandlingen av personuppgifter skulle kunna vara förenlig med de ursprungliga ändamålen.

Behandling av personuppgifter för syntetisering

Att skapa syntetiska uppgifter på det sätt som är aktuellt i detta projekt innebär en behandling av personuppgifter. Behandlingen behöver därmed uppfylla kraven i dataskyddsförordningen. Detta innebär bland annat att behandlingen behöver ske för ett specifikt ändamål, ha stöd i en rättslig grund och i övrigt leva upp till kraven i dataskyddsförordningen. Om de syntetiska uppgifterna som skapas genom behandlingen inte går att härleda till enskilda individer är dataskyddsförordningen däremot inte tillämplig på behandlingen av de uppgifterna. IMY utgår i sin bedömning från att syntetiseringen innebär att personuppgifterna blir anonyma.

Behandlingen av personuppgifter från Familjens Jurists ärendesystem för att skapa syntetiska uppgifter innebär att det sker en vidarebehandling av personuppgifter. Familjens Jurist har uppgett att syftet med syntetiseringen är att skapa anonymiserade uppgifter som ska användas för träning av AI-modellen. Ändamålet är, enligt IMY:s bedömning, inte identiskt med det ursprungliga ändamålet. Det behöver därför göras en förenlighetsbedömning enligt artikel 6.4 för att avgöra om behandlingen är tillåten.

IMY bedömer att det finns en naturlig koppling mellan ändamålen, eftersom vidarebehandlingen syftar till att förbättra den verksamhet som uppgifterna har samlats in i. Jämfört med att använda personuppgifter för träning av AI-modellen innebär alternativet att syntetisera data att åtgärder vidtas för att tydligt minska riskerna för enskildas rättigheter. Enligt IMY:s bedömning är detta något som typiskt sett kan vara tillräckligt för att det ska vara fråga om en tillåten vidarebehandling.

När det gäller det aktuella fallet finns dock, som anförts ovan om träningen av AI-modellen med personuppgifter, särskilda omständigheter som avser sammanhanget för uppgiftsinsamlingen, förhållandet mellan Familjens Jurist och de registrerade, de registrerades förväntningar och personuppgifternas känslighet. Detta innebär att ytterligare skyddsåtgärder kan behöva övervägas för att säkerställa att det blir fråga om en förenlig vidarebehandling. Som beskrivs ovan skulle en sådan åtgärd kunna vara att förena den information som ska lämnas enligt artiklarna 13 och 14 med en så kallad opt-out-möjlighet. Till skillnad från att behandla faktiska personuppgifter är lösningen att syntetisera data en mindre riskfylld behandling och det bör således finnas större förutsättningar att komma fram till slutsatsen att vidarebehandlingen är tillåten.

5. Avslutande kommentar

Möjligheten att vidarebehandla personuppgifter är relevant för många verksamheter som vill utveckla och använda AI. För att en vidarebehandling av personuppgifter ska få ske för att träna en AI-modellen med personuppgifter, eller för att skapa syntetiska uppgifter, ställer principen om ändamålsbegränsning krav på att behandlingen inte är oförenlig med de ursprungliga ändamålen. Det är ofta möjligt att vidarebehandla personuppgifter för verksamhetsutvecklingsändamål. Om vidarebehandlingen sker genom att personuppgifter i verksamheten används för att framställa syntetiska data, som inte längre är personuppgifter, är det enligt IMY:s bedömning typiskt sett fråga om en förenlig vidarebehandling.

I detta projekt finns dock särskilda omständigheter, bland annat sammanhanget som uppgifterna har samlats in i och uppgifternas känslighet, som gör att ytterligare skyddsåtgärder kan behöva övervägas. IMY ser dock att en vidarebehandling av personuppgifter skulle kunna anses vara tillåten om skyddsåtgärder i form av opt-out möjligheter infördes. I detta sammanhang vill IMY understryka att träning av en AI-modell med syntetiska uppgifter minskar riskerna för de registrerades rättigheter och att syntetiseringen får anses utgöra ett bättre alternativ än att träna AI-modellen med faktiska personuppgifter.