GDPR vid användning av generativ AI

Integritetsskyddsmyndighetens del av regeringsuppdraget att ta fram vägledande riktlinjer för användningen av generativ AI inom offentlig förvaltning

Diarienummer
IMY-2024-9162

Datum
2025-02-05

Diarienummer:
IMY-2024-9162

Datum:
2025-02-05

GDPR vid användning av generativ AI

Innehållsförteckning

Postadress:
Box 8114
104 20 Stockholm

Webbplats:
www.imy.se

E-post:
imy@imy.se

Telefon:
08-657 61 00

Sammanfattning

Offentlig förvaltning kan förena användningen av generativ AI med dataskyddsregelverket (GDPR). I huvudsak det följande behöver i så fall göras. Vad som anges här är detsamma som rutorna nedan inledningsvis i respektive avsnitt.

Beakta dataskyddsregelverket som utgångspunkt. Verksamheter som avser att använda generativ AI behöver ofta se till att användningen är förenlig med dataskyddsregelverket, det vill säga i huvudsak dataskyddsförordningen (GDPR) och kompletterande lagstiftning.

Använd generativ AI enligt de dataskyddsrättsliga principerna. Användning av generativ AI kan ske på ett sätt som är förenligt med de grundläggande principerna för dataskydd. Detta förutsätter att åtgärder vidtas bland annat för att minimera användningen av personuppgifter och att behandlingen av personuppgifter sker på ett öppet sätt i förhållande till den enskilde. En verksamhet måste säkerställa att principerna efterlevs innan personuppgiftsbehandlingen påbörjas och också tillämpa dem löpande så länge behandlingen pågår.

Klargör rollfördelningen mellan personuppgiftsansvarig och personuppgiftsbiträde. Den verksamhet som använder generativ AI för att utföra sitt uppdrag är personuppgiftsansvarig för personuppgiftsbehandlingen. En leverantör av ett AI-system som behandlar personuppgifter för verksamhetens räkning kan vara personuppgiftsbiträde. I sådana fall krävs ett personuppgiftsbiträdesavtal mellan den ansvarige och biträdet.

Se till att det finns rättslig grund och annat rättsligt stöd för användningen. När generativ AI används för att uppfylla verksamhetens uppdrag kan det finnas rättslig grund för att behandla personuppgifter. Kravet på nödvändighet kan vara uppfyllt om användningen sker för att effektivisera verksamheten. Ju större riskerna är med den behandling av personuppgifter som sker vid användning av generativ AI, desto högre krav ställs på den rättsliga grunden.

Överväg om automatiserat beslutsfattande och överföring till tredje land förekommer. Vid användning av generativ AI kan bestämmelser i dataskyddsförordningen om automatiserat beslutsfattande och överföring av personuppgifter till tredjeland aktualiseras. För att myndigheter ska kunna använda generativ AI för automatiserat beslutsfattande behöver förvaltningslagens krav på proportionalitet, objektivitet och legalitet följas. Om användningen av AI innebär att personuppgifter förs över till ett land utanför EU/EES måste särskilda villkor enligt dataskyddsförordningen vara uppfyllda.

Säkerställ enskildas rättigheter vid användningen av generativ AI. Verksamheter behöver vidta åtgärder för att säkerställa enskildas rättigheter enligt dataskyddsförordningen vid användning av generativ AI. Verksamheten behöver till exempel överväga om det krävs uppdateringar av verksamhetens integritetspolicy eller liknande dokument som informerar enskilda om behandlingen av deras personuppgifter. Vidare bör verksamheten sträva efter att använda generativ AI med inbyggt skydd för enskildas rättigheter.

Bedöm riskerna med användningen och säkerställ lämplig säkerhet. Det är möjligt att förena användning av generativ AI med dataskyddsförordningens bestämmelser om säkerhet. Vilka säkerhetsåtgärder som behöver vidtas är beroende

av risken med behandlingen. Verksamheter som avser att använda generativ AI behöver alltså utföra riskbedömningar innan en behandling med generativ AI påbörjas. Användningen behöver dessutom kontinuerligt följas upp ur ett säkerhetsperspektiv. En konsekvensbedömning enligt dataskyddsförordningen behöver genomföras om personuppgiftsbehandlingen sannolikt medför höga risker för enskilda. Exempel på risker med användning av generativ AI innefattar övertro på AI-systemets förmågor, bristande förståelse för dess begränsningar och risk för dataläckage. Dessa risker kan kräva åtgärder såsom en tydlig styrningsstruktur, mänsklig kontroll och tekniska säkerhetsåtgärder.

Särskilt om användning av allmänt tillgänglig och integrerad generativ AI. Användningen av allmänt tillgängliga generativa AI-tjänster eller integrerade generativa AI-tjänster bör ske under verksamhetens ledning och översyn. Det bör vara tydligt för medarbetarna vad tjänsten får användas till och vilken information som får behandlas i tjänsten. Vid användning av integrerade generativa AI-tjänster, som innebär att tjänsten får tillgång till verksamhetsdata, krävs god informationshantering och behörighetsstyrning. Det är viktigt att ha kunskap om verktyg som används och dess möjligheter och begränsningar.

Inledning

Bakgrund

Under sommaren 2024 fick Integritetsskyddsmyndigheten (IMY) tillsammans med Myndigheten för digital förvaltning (Digg) i uppdrag av regeringen att ta fram vägledande riktlinjer för att främja effektiv och ändamålsenlig användning av generativ artificiell intelligens (AI) inom den offentliga förvaltningen. Enligt uppdraget skulle IMY i egenskap av oberoende dataskyddsmyndighet ta fram och besluta den del av vägledningen som avsåg dataskydd. Digg skulle samordna myndigheternas arbete även i den del som avsåg dataskydd.[1] I januari 2025 publicerades resultatet av uppdraget i form av vägledande riktlinjer från både Digg och IMY på Diggs webbplats: *digg.se/ai*.[2]

Rapporten

Denna rapport innehåller den del av de vägledande riktlinjerna som IMY har tagit fram och beslutat, det vill säga den del som rör dataskydd. Skälen för att innehållet publiceras även i denna rapport är att IMY vill tillmötesgå en efterfrågan på rapport-formatet, bland annat för att det underlättar hänvisning till innehållet, och för att kunna publicera innehållet i IMY:s egna kanaler och därigenom öka tillgängligheten.

Frågor som berörs

Verksamheter som avser att använda generativ AI behöver överväga dataskydd och ofta se till att användningen är förenlig med dataskyddsregelverket. Frågor som då behöver besvaras ges vägledning kring i denna rapport. De frågorna är översiktligt följande.

- Behöver dataskyddsregelverket beaktas vid användning av generativ AI?
- Är användningen förenlig med de grundläggande dataskyddsrättsliga principerna?
- Vem är personuppgiftsansvarig? Finns det personuppgiftsbiträden?
- Finns det rättslig grund och annat rättsligt stöd för behandlingen av personuppgifter?
- Sker det automatiserat beslutsfattande eller överföring av personuppgifter till tredje land vid användningen och finns i så fall förutsättningar för det?
- Hur ska enskildas rättigheter tillgodoses vid användningen?
- Hur ska lämplig säkerhet uppnås vid användningen?
- Vilka risker från ett dataskyddsperspektiv medför användningen av generativ AI? Vilka åtgärder kan vidtas för att begränsa riskerna?

Särskilt om allmänt tillgänglig och integrerad generativ AI
I ett särskilt avsnitt ges närmare vägledning för användningen av generativ AI som är allmänt tillgänglig via internet och för generativ AI som kan finnas integrerad i befintliga kontors- och företagsprogramvaror.

[1] Regeringsbeslut 2024-07-04, Fi2024/01535, tillgängligt här.
[2] Se https://www.digg.se/ai-for-offentlig-forvaltning/riktlinjer-for-generativ-ai.

Vänder sig till offentlig förvaltning

Riktlinjerna är avsedda för verksamheter inom offentlig förvaltning. Det hindrar inte att riktlinjerna kan vara användbara för andra som använder generativ AI i sin verksamhet.

Utveckling och finjustering omfattas inte

Att utveckla generativa AI-system, det vill säga att ta fram en grundmodell som kan användas för att lösa olika typer av uppgifter, är en resurskrävande process som kräver betydande investeringar, stora mängder data och en omfattande teknisk infrastruktur. Verksamheter inom offentlig förvaltning har generellt sett begränsade möjligheter att själva utveckla sådana system från grunden.

Ett alternativ till att utveckla en ny grundmodell är att finjustera en befintlig modell för att anpassa den till verksamhetens specifika behov. Även finjustering av befintliga grundmodeller är en avancerad process som kräver specialistkompetens, tillgång till större datamängder och betydande tekniska resurser. Detta är något som de flesta verksamheter inom offentlig förvaltning saknar idag.

Mot denna bakgrund behandlar riktlinjerna varken utvecklingen av nya grundmodeller eller finjustering av befintliga grundmodeller. Istället fokuserar riktlinjerna på hur offentliga verksamheter kan använda existerande generativa AI-system på ett ansvarsfullt och rättssäkert sätt.

Beakta dataskyddsregelverket som utgångspunkt

Verksamheter som avser att använda generativ AI behöver ofta se till att användningen är förenlig med dataskyddsregelverket, det vill säga i huvudsak dataskyddsförordningen (GDPR) och kompletterande lagstiftning.

Dataskyddsregelverket ska beaktas om personuppgifter behandlas

En stor del av den offentliga förvaltningens användning av generativ AI kan antas innebära att personuppgifter behandlas. Det sker till exempel en behandling av personuppgifter när generativ AI förses med data som innehåller personuppgifter. Om personuppgifter behandlas vid användning av generativ AI innebär det att dataskyddsregelverket ska beaktas.

Dataskyddsregelverket bör som utgångspunkt beaktas även vid annan användning

Även i situationer där verksamheter inte avser att behandla personuppgifter med generativ AI kan dataskyddsregelverket behöva beaktas vid användningen. Detta eftersom en grundmodell (även kallad AI-modell för allmänna ändamål eller AI-modell) som finns i generativa AI-system ofta har personuppgifter inlärda sedan utvecklingen av modellen. Sådana personuppgifter kan behandlas till exempel genom att en medarbetare avsiktligt eller oavsiktligt instruera AI-systemet på ett sätt som får det att ange personuppgifter som finns inlärda. Det kan också ske genom att en extern part attackerar systemet och då får fram personuppgifter som finns inlärda.

Som utgångspunkt kan därför verksamheter inom offentlig förvaltning utgå från att dataskyddsregelverket bör beaktas vid användning av generativ AI. I vissa fall kan dock åtgärder ha vidtagits för att säkerställa anonymisering av AI-system, vilket innebär att dataskyddsregelverket inte är tillämpligt. För att åstadkomma anonymisering krävs att risken är försumbar för att såväl användare av systemet som externa parter kan få ut personuppgifter ur AI-systemet.[3]

Vad avses med dataskyddsregelverket?

Med dataskyddsregelverket avses här regler som rör behandling av personuppgifter. Dessa regler är framför allt EU:s dataskyddsförordning, vanligen kallad GDPR, och kompletterande nationell lagstiftning. Exempel på sådan kompletterande nationell lagstiftning är den så kallade dataskyddslagen och sektorspecifik reglering av personuppgiftsbehandling, så kallade registerförfattningar. Mer information om dataskyddsregelverket finns på IMY:s webbplats, till exempel om vad som allmänt gäller enligt dataskyddsförordningen och hur de dataskyddsrättsliga lagarna hänger ihop.

[3] Se Europeiska dataskyddsstyrelsen (EDPB) *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*, antaget den 17 december 2024, https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf, avsnitt 3.1 – 3.2, särskilt p. 31 och 43.

Använd generativ AI enligt de dataskyddsrättsliga principerna

Användning av generativ AI kan ske på ett sätt som är förenligt med de grundläggande principerna för dataskydd. Detta förutsätter att åtgärder vidtas bland annat för att minimera användningen av personuppgifter och att behandlingen av personuppgifter sker på ett öppet sätt i förhållande till den enskilde. En verksamhet måste säkerställa att principerna efterlevs innan personuppgiftsbehandlingen påbörjas och också tillämpa dem löpande så länge behandlingen pågår.

Principen om ansvarsskyldighet

Verksamheten har ansvaret för användningen
Principen om ansvarsskyldighet innebär att verksamheten har ansvaret för att säkerställa och kunna visa att dataskyddsförordningen följs. Användning av generativ AI bör godkännas av verksamheten i förväg för att säkerställa detta. Det måste också vara tydligt för medarbetarna vad generativ AI får användas till och vilken information som får behandlas i generativa AI-system.

Ta fram styrdokument för användning av generativ AI
En verksamhet som avser att använda generativ AI bör ta fram och implementera en policy eller liknande styrdokument som reglerar användningen. Dokumenten bör tydligt beskriva hur generativ AI får användas inom verksamheten och vilka krav som ställs för att säkerställa en ansvarsfull hantering. Dokumenten bör bland annat klargöra om och under vilka förutsättningar personuppgifter får behandlas vid verksamhetens användning av generativ AI. Verksamheten bör beskriva de överväganden och åtgärder som vidtagits för att säkerställa att användningen sker i enlighet med dataskyddsregelverket. För att säkerställa efterlevnad bör alla berörda inom organisationen ges information om och få tillgång till dessa dokument.

Verksamheten bör också ta fram rutiner eller liknande styrdokument för att hantera risker systematiskt och säkerställa regelbundna revisioner av att AI-systemet fungerar på ett ändamålsenligt sätt.

Dokumentera användningen
En viktig del i ansvarsefterlevnaden är att dokumentera personuppgiftsbehandlingen. Om behandlingen sannolikt leder till hög risk för enskildas rättigheter och friheter behöver verksamheten göra en konsekvensbedömning. Läs mer om detta i avsnittet *Bedöm riskerna med användningen och säkerställ lämplig säkerhet,* särskilt under rubriken *Konsekvensbedömning för generativ AI*.

Verksamheter som använder generativ AI bör också införa lämpliga mekanismer för spårbarhet vid behandling av personuppgifter. Detta gör det möjligt att verifiera och härleda hur AI-systemet har använts.

Principen om laglighet, korrekthet och öppenhet

Personuppgifter ska behandlas på ett lagligt sätt
Principen om laglighet innebär bland annat att behandlingen av personuppgifter ska ha rättslig grund. Mer information om det finns i avsnittet om *Se till att det finns rättslig grund och annat rättsligt stöd för användningen.*

Personuppgifter ska behandlas på ett korrekt sätt
Principen om korrekthet innebär att behandlingen ska vara rättvis, skälig och rimlig och att uppgifterna behandlas på ett sätt som enskilda rimligen kan förvänta sig. Principen ställer också krav på proportionalitet. Det innebär att integritetsintrånget som behandlingen innebär behöver vara rimligt i förhållande till nyttan med behandlingen.

Bias
För att behandla personuppgifter på ett korrekt sätt krävs till exempel åtgärder för att minimera partiskhet, så kallad bias, som kan förekomma i ett AI-systems utdata. Bias kan uppstå av flera olika orsaker och kan exempelvis härstamma från kvaliteten på de data som grundmodellen har tränats på. Om grundmodellens träningsdata till exempel inte har utgjort ett representativt urval av uppgifter för ändamålet, kommer modellens utdata att återspegla dessa skevheter. Detta kan i sin tur leda till diskriminering eller en orättvis behandling av enskilda eller grupper av individer. Verksamheter bör därför utvärdera riskerna med en behandling och om den innebär att modellen kan generera snedvriden eller partisk information.

I avsnittet *Bedöm riskerna med användningen och säkerställ lämplig säkerhet* finns mer information om bias och åtgärder som kan vidtas för att begränsa riskerna i det avseendet.

Personuppgifter ska behandlas på ett öppet sätt
Principen om öppenhet ställer krav på transparens om hur AI-systemet används för att behandla personuppgifter och vilka konsekvenser det kan få för enskilda. Att enskilda får information när deras personuppgifter behandlas är nödvändigt för att de ska kunna vidta åtgärder för att skydda sina rättigheter. Samtidigt kan det exempelvis vara utmanade att på ett lättbegripligt sätt förklara hur ett AI-system fungerar eller har kommit fram till ett visst svar. Eftersom tekniken utvecklas snabbt är det viktigt att enskilda får uppdaterad information om användningen och hur det påverkar dem.

I avsnittet *Säkerställ enskildas rättigheter vid användningen av generativ AI* ges mer vägledning om hur personuppgifter kan behandlas på ett öppet sätt.

Principen om ändamålsbegränsning

Specificera ändamålet
Principen om ändamålsbegränsning innebär att personuppgifter ska samlas in för särskilda, uttryckligt angivna och berättigade ändamål. Uppgifterna får inte senare behandlas på ett sätt som är oförenligt med dessa ändamål.

Generativ AI är en typ av teknik som kan användas för flera olika ändamål. Att använda generativ AI är dock inte ett ändamål i sig, utan ett medel för att behandla information. Det är viktigt att verksamheten specificerar för vilka ändamål som AI-systemet ska användas och att detta tydliggörs för verksamhetens medarbetare.

Överväg tekniska begränsningar
Verksamheten kan också överväga att införa åtgärder för att avgränsa AI-systemets användningsområde, exempelvis genom att tekniskt begränsa vilka instruktioner och vilken information användarna kan ge till systemet genom så kallade promptar. I vissa fall kan AI-systemets funktionalitet behöva begränsas så att det endast är tekniskt möjligt att utföra de uppgifter som är relevanta för syftet med behandlingen.

Principen om uppgiftsminimering

Begränsa behandlingen av personuppgifter till det nödvändiga för användningen
Verksamheten måste kunna visa att omfattningen av den personuppgiftsbehandling som sker med hjälp av generativ AI är nödvändig i förhållande till ändamålet och att samma resultat inte kan uppnås med en mindre mängd personuppgifter. Genom att endast ange de mest relevanta personuppgifterna i en prompt kan personuppgiftsbehandlingen begränsas till det som är nödvändigt för att kunna utföra uppgiften. Om verksamheten använder tekniker som innebär att ett AI system ges annan åtkomst till personuppgifter bör verksamheten utvärdera och om lämpligt begränsa den mängd personuppgifter som systemet kan få åtkomst till när det används.

Säkerställ behörighetsstyrningen
Behörighetsstyrning som säkerställer vilka uppgifter som specifika användare har tillgång till ska återspeglas vid användningen av AI-system. Motsvarande gäller när verksamheten använder generativa AI-system som är integrerade i verksamhetens befintliga programvara, till exempel tilläggstjänster som kan aktiveras i programvaran. På så sätt säkerställs att personuppgifter endast hämtas från relevanta källor och att åtkomsten begränsas till källor som användaren har behörighet till.

Principen om riktighet

Risk för hallucinationer
Verksamheter som behandlar personuppgifter måste se till att uppgifterna är korrekta och om nödvändigt uppdaterade. Generativ AI skapar sina resultat baserat på statistiska sannolikheter, vilket innebär att den kan skapa innehåll som är faktamässigt felaktigt eller påhittat. Detta brukar kallas för hallucinationer. Det finns flera orsaker till varför ett AI-system kan hallucinera. Det kan exempelvis bero på hur grundmodellen har tränats, vilka parametrar som använts samt kvaliteten och mängden data som använts under träningen. Hallucinationer kan även uppstå på grund av hur frågor ställs till AI-systemet, då frågor avsiktligt eller oavsiktligt kan ställas på ett sätt som leder AI-systemet till att generera svar som innehåller hallucinationer. Den som använder ett generativt AI-system bör därför granska och kontrollera dess svar och inte utgå från att systemets utdata är korrekta och pålitliga.

För att generera mer korrekta och träffsäkra svar kan tekniker som *retrieval-augmented generation* (RAG) vara till hjälp. RAG kombinerar grundmodellens generativa förmåga med ytterligare information från utvalda källor, vilket ger modellen tillgång till mer uppdaterad information än vad som finns i dess ursprungliga träningsdata. RAG kan därmed förbättra de svar som AI-systemet genererar. Det bör dock betonas att det fortfarande är nödvändigt att noggrant granska och kontrollera AI-systemets svar eftersom hallucinationer förekommer även när RAG används.

Läs mer om hallucinationer och åtgärder för att säkerställa riktighet i avsnittet *Bedöm riskerna med användningen och säkerställ lämplig säkerhet.*

Principen om lagringsminimering

Undvik att spara irrelevanta personuppgifter
Principen om lagringsminimering innebär att verksamheten behöver se till att personuppgifter inte behandlas under en längre tid än nödvändigt. För att effektivt kunna använda generativ AI på verksamhetens data kan det i vissa fall krävas att all

data behandlas (exempelvis vektoriseras) för att göras sökbar och tillgänglig för AI-systemet. Det kan därför vara utmanande att göra en avvägning mellan å ena sidan behovet av att lagra data som kan vara relevanta för syftet med användningen, och å andra sidan kravet på lagringsminimering. Verksamheten behöver säkerställa att AI-systemet inte behandlar personuppgifter som inte längre är relevanta på ett sätt som innebär risker för enskilda. När det inte längre är nödvändigt att behandla personuppgifterna utifrån ändamålet ska de raderas eller anonymiseras. Det gäller även för sådana data som exempelvis ett generativt AI-system använder samt eventuella kopior av datauppsättningar som lagras någon annanstans än internt i verksamheten.

Principen om integritet och konfidentialitet

Riskbaserat förhållningssätt
Det finns ingen allmängiltig standard för hur lämplig säkerhet ska uppnås när personuppgifter behandlas i samband med användningen av generativa AI-system. Vilken nivå av säkerhet och vilka säkerhetsåtgärder som behöver vidtas beror på risken i det enskilda fallet. Verksamheten måste identifiera de risker som den planerade användningen av generativ AI medför och, i förekommande fall, vidta lämpliga säkerhetsåtgärder för att minska dessa risker.

Eftersom AI-system ofta integreras i komplexa kedjor av andra system, dataflöden och processer, bör verksamheten ta ett helhetsgrepp för att säkerställa säkerheten för personuppgifterna. Tekniken för generativ AI utvecklas snabbt, liksom metoderna för att behandla personuppgifter på ett säkert sätt. Som ett led i att efterleva principen om integritet och konfidentialitet bör verksamheten aktivt övervaka och förhålla sig till den senaste teknikutvecklingen, inklusive de metoder som ökar generativa AI-modellers motståndskraft mot olika typer av attacker.

Läs mer om säkerheten för personuppgifter i avsnittet *Bedöm riskerna med användningen och säkerställ lämplig säkerhet.*

Vidare läsning

På IMY:s webbplats finns allmän vägledande information om de grundläggande principerna.

Klargör rollfördelningen mellan personuppgiftsansvarig och personuppgiftsbiträde

Den verksamhet som använder generativ AI för att utföra sitt uppdrag är personuppgiftsansvarig för personuppgiftsbehandlingen. En leverantör av ett AI-system som behandlar personuppgifter för verksamhetens räkning kan vara personuppgiftsbiträde. I sådana fall krävs ett personuppgiftsbiträdesavtal mellan den ansvarige och biträdet.

Personuppgiftsansvar

För att säkerställa ett effektivt skydd av personuppgifter finns det alltid en eller flera personuppgiftsansvariga, det vill säga den som ansvarar för efterlevnaden av dataskyddsregelverket. Det är den som bestämmer ändamål och medel för behandlingen av personuppgifter, det vill säga hur och varför personuppgifterna behandlas, som är personuppgiftsansvarig. I offentlig förvaltning är det normalt en myndighet som är personuppgiftsansvarig.

Om en verksamhet behandlar personuppgifter för en personuppgiftsansvarigs räkning kallas den verksamheten för personuppgiftsbiträde. Dataskyddsförordningen ställer krav på att det finns ett personuppgiftsbiträdesavtal mellan den ansvarige och biträdet.

Verksamheten är personuppgiftsansvarig för sin användning av generativ AI

Den verksamhet inom offentlig förvaltning som använder generativ AI för att utföra sitt uppdrag är som utgångspunkt personuppgiftsansvarig för den personuppgiftsbehandling som sker i samband med användningen. Med detta ansvar följer en skyldighet att välja generativa AI-system som gör det möjligt för verksamheten att säkerställa att behandlingen sker i enlighet med dataskyddsförordningen.

Leverantörer av AI-systemet är som utgångspunkt personuppgiftsbiträde

Om en verksamhet anlitar en leverantör för att tillhandahålla ett AI-system för en viss uppgift, är leverantören personuppgiftsbiträde för den personuppgiftsbehandling som leverantören utför för verksamhetens räkning. En förutsättning för att leverantören inte också ska anses vara personuppgiftsansvarig är att den inte har något eget ändamål med behandlingen av personuppgifter, till exempel att använda uppgifterna för att vidareutveckla AI-systemets grundmodell. Enbart leverantörens intresse av att sälja ett AI-system är inte ett ändamål som i sig medför ett personuppgiftansvar.

Ansvar för oförutsedda behandlingar

Det kan vara svårt att förutse hur ett generativt AI-system kan komma att fungera i praktiken. En verksamhet är dock som utgångspunkt ansvarig för den behandling av personuppgifter som sker med generativ AI även när personuppgifter behandlas på ett oförutsett sätt. Undantag kan gälla för funktioner som verksamheten uttryckligen har motsatt sig och där tydliga instruktioner har getts till leverantören om att sådana

funktioner inte ska ingå i AI-systemet. Det är därför viktigt att överväga vilka begränsningar som bör byggas in i systemet.

Vidare läsning

På IMY:s webbplats finns allmän vägledande information om personuppgiftsansvar och personuppgiftsbiträden.

Se till att det finns rättslig grund och annat rättsligt stöd för användningen

När generativ AI används för att uppfylla verksamhetens uppdrag kan det finnas rättslig grund för att behandla personuppgifter. Kravet på nödvändighet kan vara uppfyllt om användningen sker för att effektivisera verksamheten. Ju större riskerna är med den behandling av personuppgifter som sker vid användning av generativ AI, desto högre krav ställs på den rättsliga grunden.

AI är ett medel, inte ett ändamål

Varje behandling av personuppgifter ska ske för ett eller flera berättigade ändamål. Ändamålen behöver vara särskilda och uttryckligt angivna. Att använda AI är inte ett sådant ändamål i sig, utan ett medel för att behandla information. Den behandling av personuppgifter som sker vid användning av generativ AI kan däremot utföras för flera olika ändamål samtidigt.

Använd generativ AI för att tillgodose verksamhetens uppgifter

Utgångspunkten är uppdraget
För varje behandling av personuppgifter behöver det finnas en rättslig grund. Den rättsliga grunden för behandling av personuppgifter i offentlig förvaltning är normalt att verksamheten ska utföra en uppgift av allmänt intresse. Uppdrag som ges till myndigheter och andra offentliga organ är uppgifter av allmänt intresse. Sådana uppdrag kan till exempel ges genom lag eller förordning eller ett regeringsbeslut. I den mån generativ AI bidrar till att verksamheten utför sitt uppdrag, kan det finnas stöd för att behandla personuppgifter med sådan teknik.

Effektivitet är centralt
För att personuppgiftsbehandling ska få ske för en uppgift av allmänt intresse krävs att behandlingen är *nödvändig* för att utföra uppgiften. Kravet på nödvändighet kan vara uppfyllt om användning av generativ AI bidrar till att effektivisera verksamheten. Effektivitet är ett krav som gäller för stora delar av den offentliga förvaltningen, till exempel till följd av bestämmelser i myndighetsförordningen, förvaltningslagen och bestämmelser om ekonomisk förvaltning i kommunallagen.

Överväg hur riskfylld behandlingen är

För att den rättsliga grunden *uppgift av allmänt intresse* ska kunna användas krävs det att uppgiften av allmänt intresse har fastställts till exempel i en lag eller annan författning eller ett regeringsbeslut. Ju större riskerna är med den tilltänka behandlingen, desto högre krav ställs på detta rättsliga stöds tydlighet, precision och förutsebarhet.

Mindre riskfyllda behandlingar
Behandlingar med mindre risker, till exempel behandling av ett mindre antal icke-känsliga personuppgifter, kan ske med stöd av ett mer allmänt hållet rättsligt stöd. Det kan exempelvis vara fallet om generativ AI används för att sammanfatta innehållet i en publik rapport där personuppgifter endast förekommer i form av namn på rapportens författare och ansvariga beslutsfattare. I dessa fall kan det utgöra en effektivitetsvinst

att sammanfatta rapporten med generativ AI och det krävs då inga närmare överväganden om den rättsliga grunden.

Mer riskfyllda behandlingar
För behandlingar som innebär större risker, till exempel behandling av stora mängder känsliga personuppgifter, ställs det högre krav på det rättsliga stödet ifråga om tydlighet, precision och förutsebarhet. I sådana fall kan det krävas en särskild reglering som gör det förutsebart att behandlingen kommer att ske.

Känsliga personuppgifter

Känsliga personuppgifter, till exempel uppgifter om hälsa, är endast tillåtna att behandla om det finns ett särskilt stöd för det. Det gäller även vid användning av generativa AI-system.

EU-domstolen har gjort en bred tolkning av vad som ska anses vara känsliga personuppgifter i relation till teknik som har möjlighet att behandla stora mängder information.[4] Vid användning av generativ AI kan det innebära att verksamheten behöver bedöma om det finns stöd för att behandla känsliga personuppgifter, även om avsikten inte är att behandla sådana uppgifter.[5]

Exempel

I rapporten Utlämnande av allmänna handlingar med hjälp av AI analyserade IMY användningen av generativ AI hos en kommun för att underlätta hanteringen av begäranden av allmänna handlingar enligt offentlighetsprincipen, bland annat i form av en så kallad RAG-lösning (eng. *retrieval-augmented generation*). Det rättsliga stödet för att behandla vanliga personuppgifter och känsliga personuppgifter analyserades i rapporten. Två olika fall berördes: ett mer omfattande fall kallat helhetstjänsten och ett snävare fall kallat maskeringstjänsten.

Sammantaget bedömde IMY att mycket talade för att det fanns stöd för att behandla personuppgifter, inklusive känsliga uppgifter, i den snävare maskeringstjänsten. Däremot ansågs övervägande skäl tala emot att det fanns rättsligt stöd för behandlingen av personuppgifter i den mer omfattande helhetstjänsten. Det rättsliga stödet i de två fallen var i huvudsak detsamma. Avgörande för skillnaden i bedömningen var de olika risker som de två fallen medförde, vilket är ett exempel på att ju större riskerna är med den tilltänka behandlingen, desto högre krav ställs på det rättsliga stödet.

Vidare läsning

På IMY:s webbplats finns allmän vägledande information om rättslig grund och känsliga personuppgifter.

[4] Se EU-domstolens dom 2023-07-04, Bundeskartellamt, C-252/21, ECLI:EU:C:2023:537, fråga 2 a särskilt p. 73. Se även EU-domstolens dom 2024-10-04, Lindenapotheke, C-21/23, ECLI:EU:C:2024:846, fråga 2 punkterna 86-88.
[5] Se t.ex. Claudio Novelli et al., Computer Law & Security Review, Volume 55, November 2024, https://doi.org/10.1016/j.clsr.2024.106066, avsnitt 3.1.2 på s. 5 f.

Överväg om automatiserat beslutsfattande och överföring till tredje land förekommer

Vid användning av generativ AI kan bestämmelser i dataskyddsförordningen om automatiserat beslutsfattande och överföring av personuppgifter till tredjeland aktualiseras. För att myndigheter ska kunna använda generativ AI för automatiserat beslutsfattande behöver förvaltningslagens krav på proportionalitet, objektivitet och legalitet följas. Om användningen av AI innebär att personuppgifter förs över till ett land utanför EU/EES måste särskilda villkor enligt dataskyddsförordningen vara uppfyllda.

Automatiserat beslutsfattande

Generativ AI kan användas för att fatta beslut om enskilda. Det finns bestämmelser i dataskyddsförordningen som reglerar i vilka situationer automatiserat beslutsfattande är tillåtet. Bestämmelserna är dock endast tillämpliga på beslut som har rättsliga följder för eller i betydande grad påverkar den enskilde, till exempel vid myndighetsutövning.

Vad är automatiserat beslutsfattande?
Automatiserat beslutsfattande innebär att beslut fattas utan mänsklig inblandning. För att något ska anses vara ett automatiserat beslutsfattande enligt dataskyddsförordningen måste själva beslutsfattandet ske genom automatiserad behandling av personuppgifter. Om generativ AI används som stöd i beslutsfattandet men det är en människa som fattar själva beslutet, är det inte fråga om ett automatiserat beslutsfattande enligt dataskyddsförordningen. Om en människa i praktiken förlitar sig på det underlag som ett AI-system genererar på ett avgörande sätt, kan det dock betraktas som ett automatiserat beslutsfattande enligt dataskyddsförordningen även om det formella beslutet fattas av personen ifråga.[6]

Är automatiserat beslutsfattande tillåtet?
Huvudregeln i dataskyddsförordningen är att automatiserat beslutsfattande är förbjudet. Det finns dock ett antal undantag från förbudet. Ett undantag är att det finns reglering som tillåter automatiserat beslutsfattande och att denna reglering innehåller lämpliga skyddsåtgärder. Regeringen har bedömt att den svenska regleringen i förvaltningslagen, kommunallagen och viss speciallagstiftning ger förutsättningar för automatiserat beslutsfattande enligt dataskyddsförordningen.[7]

För att generativ AI ska kunna ligga till grund för automatiserat beslutsfattande inom myndigheter krävs att förvaltningslagens krav på proportionalitet, objektivitet och legalitet följs. Det behöver därför säkerställas att beslutsfattandet följer gällande regler, är förutsebart och att ovidkommande hänsyn inte tas. Detta innebär att det bland annat måste säkerställas att partiskhet, så kallad bias, och hallucinationer inte påverkar enskilda på ett negativt sätt. Om detta inte kan säkerställas i tillräcklig utsträckning behöver andra lösningar väljas, till exempel ett regelbaserat system.

Vidare läsning
På IMY:s innovationsportal finns mer information om automatiserat beslutsfattande.

[6] Se EU-domstolens dom 2023-12-07, SCHUFA Holding (Scoring), C-634/21, ECLI:EU:C:2023:957, p. 73.
[7] Se t.ex. prop. 2017/18:95 s. 100 och prop. 2021/22:125 s. 58.

Överföring av personuppgifter till tredjeland

Användning av generativ AI kan många gånger innebära att personuppgifter överförs till en extern tjänsteleverantör. Överförs personuppgifterna då till ett land utanför EU/EES, det vill säga ett tredjeland, är överföringen endast tillåten om särskilda förutsättningar i dataskyddsförordningen är uppfyllda. För överföringar av personuppgifter till USA kan det finnas sådana förutsättningar om leverantören omfattas av det så kallade EU-U.S. Data Privacy Framework. En verksamhet som överväger att använda generativ AI som inbegriper överföring av personuppgifter till ett tredjeland behöver kontrollera om det finns förutsättningar för det.

På IMY:s webbplats finns mer information om överföring av personuppgifter till tredjeland.

Säkerställ enskildas rättigheter vid användningen av generativ AI

Verksamheter behöver vidta åtgärder för att säkerställa enskildas rättigheter enligt dataskyddsförordningen vid användning av generativ AI. Verksamheten behöver till exempel överväga om det krävs uppdateringar av verksamhetens integritetspolicy eller liknande dokument som informerar enskilda om behandlingen av deras personuppgifter. Vidare bör verksamheten sträva efter att använda generativ AI med inbyggt skydd för enskildas rättigheter.

Inledning

Enligt dataskyddsförordningen ska en verksamhet på eget initiativ lämna viss information till enskilda om verksamhetens behandling av personuppgifter. Detta brukar kallas för rätten till information. Enskilda har också vissa rättigheter som de kan utöva på eget initiativ, till exempel rätten till tillgång, rättelse och radering.

På IMY:s webbplats finns allmän vägledande information om dessa så kallade registrerades rättigheter, till exempel om de tidsfrister som gäller för att besvara en begäran från en enskild. På IMY:s webbplats finns även mer specifik vägledande information om till exempel svarta lådan och rätten till information, där bland annat information om automatiserat beslutsfattande berörs.

Personuppgifter i indata och utdata

Nedan finns riktlinjer för personuppgifter i indata till och utdata från generativa AI-system. Indata avser främst instruktioner och kontext som tillförs AI-systemet. Med utdata avses främst det som AI-systemet genererar, det vill säga resultatet.

Exempel angående rätten till information
Enligt rätten till information ska verksamheten informera enskilda om bland annat vilka mottagare eller kategorier av mottagare som får tillgång till deras personuppgifter. Om en verksamhet exempelvis använder generativ AI med en RAG-lösning (eng. *retrieval-augmented generation)* som innefattar behandling av personuppgifter, kan det innebära att uppgifterna överförs till en extern leverantör. Denna leverantör behandlar då personuppgifterna för verksamhetens räkning och agerar som personuppgiftsbiträde i detta avseende. I sådana fall är leverantören en sådan typ av mottagare som verksamheten är skyldig att informera enskilda om.

Rätten till information innebär också att enskilda ska informeras om bland annat överföringar av personuppgifter till ett tredjeland, det vill säga ett land utanför EU/EES, och det rättsliga stödet för överföringen. Det kan bli aktuellt vid användning av generativ AI.

Dessa krav på tydlig information kan innebära att en verksamhet som använder generativ AI behöver se över och uppdatera sina integritetspolicyer och liknande dokument med information till enskilda om hur deras personuppgifter behandlas vid användningen av generativ AI.

Exempel för rätten till tillgång
Rätten till tillgång innebär att enskilda som huvudregel har rätt att på begäran få veta om en verksamhet behandlar personuppgifter som rör dem. Om så är fallet har de

också rätt att få en kopia av dessa uppgifter samt information om hur de används. Denna rätt omfattar i regel också de personuppgifter som används som indata och som genereras som utdata av ett generativt AI-system.

Det finns dock undantag från rätten till tillgång som kan vara tillämpliga på både indata och utdata. Rätten till tillgång gäller inte personuppgifter som inte får lämnas ut till den enskilde på grund av sekretess eller tystnadsplikt. Undantag gäller också för personuppgifter i löpande text som ännu inte fått sin slutliga utformning vid tidpunkten för begäran, samt för personuppgifter som finns i minnesanteckningar eller liknande handlingar. Undantagen för löpande text och minnesanteckningar gäller dock endast om handlingarna inte har lämnats ut till en tredje part. I detta sammanhang räknas personuppgiftsbiträden inte som tredje part.

Det är möjligt att prompter som innehåller personuppgifter kan betraktas som löpande text som ännu inte fått sin slutliga utformning eller som en minnesanteckning. Detta skulle innebära att prompten kan undantas från rätten till tillgång, förutsatt att den inte har lämnats ut till en tredje part. Utdata som genereras baserat på prompten kan på samma sätt betraktas som löpande text eller en minnesanteckning. Om utdata däremot diarieförs eller expedieras omfattas de i regel av rätten till tillgång, men kan ändå vara undantagen om informationen är föremål för sekretess eller tystnadsplikt.

Personuppgifter som kan finnas i AI-modeller

Det kan vara utmanande att tillgodose enskildas rättigheter
Generativa AI-system kan ha personuppgifter inlärda från utvecklingen av grundmodellen. Det kan röra sig om uppgifter om väldigt många personer. Det är i dagsläget närmast omöjligt för en verksamhet som använder ett generativt AI-system att kunna avgöra vilkas personuppgifter som har behandlats i samband med utvecklingen av grundmodellen och som eventuellt finns inlärda i modellen. Detta gör det utmanande för verksamheter som planerar att använda generativ AI att kunna tillgodose enskildas rättigheter enligt dataskyddsförordningen. Mot denna bakgrund ges följande vägledande riktlinjer.

Rätten till information

Ge allmän information om hur generativ AI används i verksamheten
Verksamheter behöver inte på eget initiativ lämna information om sin användning av ett generativt AI-system till alla enskilda vars personuppgifter kan ha använts som träningsdata för att utveckla systemets grundmodell. Detta skulle nämligen medföra en oproportionerlig ansträngning enligt dataskyddsförordningens bestämmelser. Istället bör verksamheten tillhandahålla information för allmänheten om vilka generativa AI-system eller grundmodeller som används, exempelvis genom att inkludera sådan information i en integritetspolicy eller AI-policy som är allmänt tillgänglig.

Rätten till tillgång, rättelse och radering

Välj AI-system med inbyggt dataskydd som standard
Verksamheter bör prioritera att använda generativa AI-system som gör det möjligt att tillgodose en begäran om utövande av enskildas rättigheter enligt dataskyddsförordningen. Vid valet av AI-system bör verksamheten beakta den senaste tekniska utvecklingen på området, genomförandekostnaderna och de risker som kan uppstå för personuppgifter i samband med användning av AI-systemet.

Underlätta utövandet av enskildas rättigheter och förklara ett eventuellt avslag
Om en verksamhet bedömer att den inte kan tillgodose en begäran från en enskild om utövande av sina rättigheter enligt dataskyddsförordningen, bör verksamheten ändå sträva efter att underlätta för den enskilde. Det kan till exempel göras genom att hänvisa till lämpliga kanaler för en motsvarande begäran hos leverantören av det AI-system som verksamheten använder. Verksamheten bör även förklara skälen till varför en begäran inte kan tillgodoses, men verksamheten är inte skyldig att besvara begäranden som är uppenbart ogrundade eller orimliga.

Bedöm riskerna med användningen och säkerställ lämplig säkerhet

Det är möjligt att förena användning av generativ AI med dataskyddsförordningens bestämmelser om säkerhet. Vilka säkerhetsåtgärder som behöver vidtas är beroende av risken med behandlingen. Verksamheter som avser att använda generativ AI behöver alltså utföra riskbedömningar innan en behandling med generativ AI påbörjas. Användningen behöver dessutom kontinuerligt följas upp ur ett säkerhetsperspektiv. En konsekvensbedömning enligt dataskyddsförordningen behöver genomföras om personuppgiftsbehandlingen sannolikt medför höga risker för enskilda. Exempel på risker med användning av generativ AI innefattar övertro på AI-systemets förmågor, bristande förståelse för dess begränsningar och risk för dataläckage. Dessa risker kan kräva åtgärder såsom en tydlig styrningsstruktur, mänsklig kontroll och tekniska säkerhetsåtgärder.

Metod för lämplig säkerhet

Dataskyddsförordningen ställer krav på att det vidtas säkerhetsåtgärder som är lämpliga i förhållande till risken med personuppgiftsbehandlingen. Vid högre risker krävs mer omfattande och långtgående åtgärder och vid mindre riskfylld behandling är kraven lägre. Viktiga led i arbetet med att säkerställa en lämplig säkerhet är:

- *Analysera risker på förhand:* Analysera och bedöm risker med användningen innan AI-systemet införs i verksamheten. Syftet med detta är att i förväg vidta lämpliga åtgärder för att minska riskerna samt att inte genomföra behandlingar som innebär för hög risk.
- *Följ upp och hantera risker under användning:* Följ kontinuerligt upp riskerna samt vidta lämpliga åtgärder för att minska riskerna. I förekommande fall, upphör med behandlingar som visar sig vara alltför riskfyllda.
- *Hantera incidenter:* Identifiera, analysera och åtgärda incidenter. Informera berörda personer och IMY vid behov.

Risker med användning av generativ AI

Användning av generativ AI kan medföra flera olika risker. Nedan ges exempel.

Övertro och olämplig användning
En risk med generativ AI är att användare i verksamheten har bristande förståelse för tekniken, särskilt när det gäller dess begränsningar. Det kan leda till att medarbetare som använder tekniken i sitt arbete förlitar sig i alltför stor utsträckning på de resultat som AI-systemet genererar. Det kan också leda till att tekniken används på sätt som är olämpliga utifrån ett informationssäkerhetsperspektiv.

Dataläckage
Ytterligare en risk är att personuppgifter som överförs till en extern generativ AI-tjänst kan hamna i orätta händer eller används på ett obehörigt sätt utanför verksamhetens kontroll.

Svarta lådan-problematiken
En utmaning med generativ AI är den så kallade svarta lådan-problematiken (eng. *the black box problem*), vilket innebär att det inte alltid går att fullt ut förstå hur tekniken fungerar eller på vilka grunder ett svar genereras. Detta kan leda till säkerhetsrisker

som vi ännu inte känner till eller förstår. Offentliga verksamheter bör därför överväga om användningen av generativ AI är nödvändig eller om en mer förutsebar, regelbaserad IT-lösning kan fylla samma funktion. Enklare AI-lösningar som kan ge bättre kontroll över och förståelse för behandlingen bör också övervägas, till exempel en liten språkmodell (eng. *small language model, SLM*) som kan hanteras närmare verksamheten.

Hallucinationer
Hallucinationer innebär att generativ AI skapar utdata som inte stämmer överens med verkligheten. Exempel på hallucinationer kan vara svar eller handlingar som innehåller vilseledande eller felaktiga uppgifter. Vissa hallucinationer kan vara mycket svåra att skilja från fakta. Ju svårare det är att identifiera det felaktiga, desto större kan risken och konsekvenserna av hallucinationen bli. Hallucinationer kan medföra svårigheter att följa dataskyddsförordningens princip om riktighet.

Bias
Bias innebär att ett generativt AI-systemet producerar resultat som är diskriminerande eller snedvridna. Detta beror ofta på att skevheter i träningsdatan har påverkat grundmodellens inlärning. Exempelvis om grundmodellens träningsdata visar att kvinnor generellt sett tjänar mindre än män, kan modellen felaktigt lära sig att detta är en norm och återskapa eller förstärka sådana mönster i sina resultat. Bias kan medföra svårigheter att följa dataskyddsförordningens princip om korrekthet.

Lämpliga åtgärder för att begränsa risker

Flera olika åtgärder kan vara lämpliga för att begränsa risker med användningen av generativ AI. Nedan ges exempel.

Styrningsstruktur
För att minska riskerna med generativ AI är det vanligtvis lämpligt att etablera en tydlig styrningsstruktur, både internt och i förhållande till leverantörer av generativa AI-system. En sådan styrningsstruktur kan omfatta riktlinjer och rutiner gällande användning, säkerhet och hantering av personuppgifter. Det bör exempelvis finnas en tydlig styrning kring vilka personuppgifter som får användas som indata och vilka personuppgifter som en grundmodell får ges tillgång till på andra sätt.

Utbildning
Verksamheten bör utbilda de medarbetare som ska använda generativ AI i sitt arbete. Medarbetarna bör lära sig att identifiera risker och det bör betonas att tekniken är ett stödverktyg som inte ersätter det mänskliga omdömet.

Mänsklig kontroll
En annan lämplig åtgärd är att upprätthålla mänsklig kontroll (eng. *human-in-the-loop*) genom att säkerställa mänsklig inblandning i viktiga steg vid användning av generativ AI. Mänsklig kontroll innebär att resultaten ska kunna granskas och bedömas av en människa, vilket kan vara en utmaning i vissa fall när AI-systemet genererar komplexa eller svårtolkade svar. Exempelvis kan risken för hallucinationer begränsas genom att medarbetare granskar det innehåll som generativ AI producerar. Granskningen kan underlättas om AI-systemets grundmodell använder metoden *chain-of-thought* (CoT) som innebär att modellen redovisar sitt resonemang steg för steg, vilket gör dess processer mer transparenta och lättbegripliga.

Tekniska säkerhetsåtgärder
För att skydda personuppgifter, både i vila och under överföring, är det viktigt att använda stark kryptering. Vidare bör lösningar för åtkomstkontroll och behörighetsstyrning implementeras för att säkerställa att endast behöriga har tillgång till de personuppgifter som överförs till AI-systemet, särskilt vid användning av en extern tjänst. Utöver detta är det viktigt att införa säkerhetsåtgärder som kan upptäcka och förhindra obehörig användning av personuppgifter, såsom övervakning av systemaktivitet och larm vid misstänkta avvikelser.

För att ytterligare stärka säkerheten bör verksamheten regelbundet utvärdera prestanda och funktionalitet hos generativa AI-system för att identifiera och åtgärda eventuella brister.

Sannolikhetströsklar och temperatur
Ett generativt AI-system är ofta konstruerat för att generera svar oavsett om svaret är faktamässigt korrekt eller inte. Till skillnad från en människa, som kan uttrycka osäkerhet eller medge att man inte vet eller förstår, tenderar generativ AI att alltid generera ett svar, även vid osäkerhet. Ett sätt att motverka problemet är att använda så kallade sannolikhetströsklar (eng. *confidence thresholds*). AI-systemet kan då konfigureras till att avstå från att generera ett svar om sannolikheten för att svaret är korrekt ligger under ett visst tröskelvärde. Gränsen för när ett svar ska anses ligga under tröskelvärdet kan inte fastställas generellt och beror bland annat på hur hög den så kallade temperaturen tillåts vara.

Temperaturen är en parameter som styr hur ”kreativt” ett AI-system tillåts vara när det genererar utdata. En låg temperatur gör systemet mer förutsägbart genom att prioritera de mest sannolika svaren, medan en hög temperatur skapar mer varierade och kreativa resultat. Konfigurering av temperaturen kan sålunda också vara en lämplig åtgärd i sammanhanget.

RAG
För att generera mer korrekta och träffsäkra svar kan tekniken *retrieval-augmented generation* (RAG) användas. RAG kombinerar grundmodellens generativa förmåga med extern eller verksamhetsspecifik information från utvalda källor. Detta gör att systemet inte bara genererar mer kvalitativa svar, utan att det också kan besvara frågor som bygger på verksamhetens egna data. Det är viktigt att notera att RAG inte är någon garanti mot bias eller hallucinationer, men denna teknik kan ändå bidra till att minska riskerna för dessa problem.

Det bör samtidigt framhållas att RAG medför egna risker. Till exempel krävs avancerade processer som vektorisering och indexering av data för att systemet ska kunna använda informationen. Det kan i sig innebära en komplex bearbetning av personuppgifter, vilket ställer krav på att analyser genomförs för att säkerställa att behandlingen är nödvändig och proportionerlig i enlighet med dataskyddsförordningen. Dessutom krävs åtgärder såsom behörighetsstyrning där det noggrant bör övervägas vilken information RAG-lösningen ska ha tillgång till.

PET
Olika integritetshöjande teknologier, ofta kallade *privacy-enhancing technologies* (PET), kan vara en viktig del av arbetet med att minska riskerna med generativ AI. PET kan säkerställa efterlevnaden av de grundläggande dataskyddsprinciperna, till exempel uppgiftsminimering. PET kan också öka säkerheten, särskilt vid avancerade personuppgiftsbehandlingar i samband med användning av generativ AI.

Konsekvensbedömning för generativ AI

För behandling av personuppgifter som sannolikt innebär en hög risk för enskilda krävs att verksamheten gör en konsekvensbedömning avseende dataskydd. Det kan beskrivas som en process för att identifiera risker med behandlingen av personuppgifter och för att ta fram åtgärder för att hantera dessa risker.

Analysen av om en konsekvensbedömning behöver genomföras innan en verksamhet börjar använda ett generativt AI-system, skiljer sig inte från motsvarande analys för annan personuppgiftsbehandling. Att genomföra en riskanalys inför användning av teknik som är ny för verksamheten kan dock kräva särskild kompetens. Exempel på faktorer som kan aktualiseras vid användning av generativ AI och som talar för att en konsekvensbedömning behöver göras är att det kan röra sig om användning av ny teknik, behandling av personuppgifter i stor omfattning, bearbetning av känsliga personuppgifter och automatiserat beslutsfattande.

Användning av generativ AI kan alltså innebära att en konsekvensbedömning måste genomföras, men det är inte alltid fallet. En bedömning måste göras från fall till fall.

Vidare läsning

På IMY:s webbplats finns vägledande information om bland annat inbyggt dataskydd och dataskydd som standard, informationssäkerhet, konsekvensbedömningar och förhandssamråd samt personuppgiftsincidenter. Fördjupande information finns därutöver i till exempel IMY:s rapport Utlämnande av allmänna handlingar med hjälp av AI, se särskilt avsnitt 6.

Särskilt om användning av allmänt tillgänglig och integrerad generativ AI

Användningen av allmänt tillgängliga generativa AI-tjänster eller integrerade generativa AI-tjänster bör ske under verksamhetens ledning och översyn. Det bör vara tydligt för medarbetarna vad tjänsten får användas till och vilken information som får behandlas i tjänsten. Vid användning av integrerade generativa AI-tjänster som innebär att tjänsten får tillgång till verksamhetsdata krävs god informationshantering och behörighetsstyrning. Det är viktigt att ha kunskap om verktyg som används och dess möjligheter och begränsningar.

Allmänt tillgänglig generativ AI

Utvecklingen av maskininlärning och datorkraft har lett till lanseringen av en rad nya tjänster där användare kan få tillgång till generativ AI via användarvänliga gränssnitt i molntjänster och öppna API:er som kan användas för en mängd olika ändamål. Genom denna typ av tjänster, som idag huvudsakligen utvecklas och tillhandahålls av amerikanska leverantörer, har generativ AI blivit tillgänglig för allmänheten. För att börja använda dessa typer av öppna tjänster behöver användaren oftast bara skapa ett konto eller teckna en prenumeration hos leverantören.

Att använda ett AI-system som bygger på generativ AI kan innebära att personuppgifter behandlas som gör att kraven i dataskyddsförordningen och annan dataskyddslagstiftning måste iakttas.

Användning av generativ AI behöver ske kontrollerat
För allmänt tillgängliga AI-tjänster är det sällan möjligt att ställa krav på systemets funktioner, säkerhet och informationshantering. Om medarbetare ska få använda sådana typer av tjänster i arbetet måste verksamheten först ta ställning till hur detta ska få ske. Användningen kan innebära att känslig information överförs till AI-tjänsten och en risk för att den blir en del av en grundmodells träningsdata. Verksamheten bör ta fram interna riktlinjer som reglerar användningen i syfte att skydda verksamhetens information och för att säkerställa att användningen sker på ett säkert sätt som också är förenligt med dataskyddslagstiftningen.

Behandling av personuppgifter i allmänt tillgängliga generativa AI-tjänster
Att använda allmänt tillgängliga generativa AI-tjänster på ett sätt som innebär att personuppgifter skrivs in i promptar eller att ge tjänsten instruktioner som leder till att personuppgifter skapas utgör personuppgiftsbehandling. Sådana behandlingar måste ske i enlighet med dataskyddsförordningen. En analys av om och hur användningen lever upp till dataskyddsförordningens krav behöver göras innan tjänsterna tas i bruk. Det behöver finnas tydliga interna riktlinjer om och under vilka förutsättningar personuppgifter får användas som indata eller genereras som utdata.

Personuppgiftsansvar för användning av allmänt tillgängliga generativa AI-tjänster
Generativa AI-tjänster som är allmänt tillgängliga erbjuds ofta i olika versioner, från kostnadsfria alternativ till prenumerationsbaserade tjänster eller sådana som regleras av särskilda licensavtal. Verksamheter som använder dessa tjänster bör alltid säkerställa att personuppgifter behandlas lagligt och är tillräckligt skyddade, oavsett vilket alternativ som används.

Allmänt tillgängliga generativa AI-tjänster utan personuppgiftsbiträdesavtal
Vid skapandet av ett konto hos en allmänt tillgänglig generativ AI-tjänst kan det förekomma att personuppgiftsbiträdesavtal saknas för användningen av tjänsten. Detta beror ofta på att villkoren i vissa versioner av dessa tjänster förutsätter att kontot registreras av en användare i egenskap av privatperson, i syfte att använda tjänsten för eget bruk. Om en medarbetare använder en sådan tjänst i sitt arbete och i samband med det behandlar personuppgifter, kan arbetsgivaren bli personuppgiftsansvarig för behandlingen. Eftersom arbetsgivaren i dessa fall inte har ingått ett personuppgiftsbiträdesavtal med leverantören, kan användningen strida mot dataskyddsförordningen. Arbetsgivare bör därför tydliggöra att dessa tjänster inte får användas av medarbetare i arbetet utan föregående godkännande och inte utan ett personuppgiftsbiträdesavtal som lever upp till kraven i dataskyddsförordningen.

Allmänt tillgängliga generativa AI-tjänster med personuppgiftsbiträdesavtal
Vid skapande av ett konto hos en allmänt tillgänglig generativ AI-tjänst där avtalsvillkoren riktar sig mot användning som inte sker för privat bruk fungerar leverantören i många fall som ett personuppgiftsbiträde i förhållande till den användande verksamheten. Leverantören av sådana tjänster erbjuder ofta personuppgiftsbiträdesavtal där det ska framgå hur leverantören behandlar personuppgifter för den användande verksamhetens räkning. Det kan till exempel avtalas om att leverantören inte får behandla personuppgifter som leverantören får tillgång till för egna ändamål. Ett vanligt problem är dock att leverantören många gånger använder standardiserade personuppgiftsbiträdesavtal och användarvillkor där det finns begränsade eller inga möjligheter att förhandla och anpassa villkoren till verksamhetens behov. I dessa situationer behöver verksamheten bedöma om det är möjligt att följa dataskyddsförordningen med leverantörens villkor och annars avstå från att använda tjänsten.

Generella råd vid användning av allmänt tillgängliga generativa AI-tjänster
Att medarbetare på eget initiativ använder generativa AI-tjänster för arbetsrelaterade syften kan innebära att arbetsgivaren blir personuppgiftsansvarig för behandlingen. Användningen av dessa tjänster bör därför i stället ske under arbetsgivarens kontroll. Tjänsten bör upphandlas av verksamheten utifrån den tilltänkta användningen och avtalsvillkoren behöver garantera en tillräcklig skyddsnivå för personuppgifterna.

Många leverantörer av allmänt tillgängliga generativa AI-tjänster lägger ansvaret på användarna att säkerställa att tjänsten används på ett lagligt sätt. Det innebär att verksamheten måste säkerställa att tjänsten är korrekt inställd och konfigurerad för att användningen ska uppfylla kraven i dataskyddsförordningen.

Överväg lämpligheten av användningen
En verksamhet bör överväga om generativa AI-tjänster är lämpliga att införa på arbetsplatsen baserat på syftet med användningen. Det bör också övervägas om tjänsten ska användas inom hela verksamheten eller endast i vissa delar. Dessa överväganden bör dokumenteras i interna styrdokument som reglerar hur tjänsten får användas och vilken typ av information som får behandlas i den.

Integrerad generativa AI

Förutom att generativ AI utvecklas i nya tjänster har det även blivit vanligare att generativ AI integreras i befintliga kontors- och företagsprogramvaror. Det kan handla om enklare AI-tjänster som automatiskt genererar förslag på texter och innehåll, men också mer avancerade lösningar där AI-tjänsten fungerar som en personlig, digital

assistent som kan användas för allt från att analysera data till att föreslå mer relevanta och personliga sökresultat baserat på användarens arbetskontext och preferenser.

Dessa tjänster fungerar som ett tillägg till befintlig kontorsprogramvara som kan aktiveras av användaren. Många verksamheter inom offentlig förvaltning använder idag kontorsprogramvara med möjlighet till att aktivera denna typ av tilläggstjänster. Sådana verktyg innebär då vanligtvis att en AI-tjänst ansluts till den användande verksamhetens data och ges åtkomst till information som verksamheten behandlar. I enklare fall av integrerade generativa AI-tjänster kan det handla om att integrera en chatbot som svarar på generella frågor utan att ha direkt åtkomst till verksamhetens interna information.

Många av dessa tilläggsfunktioner av generativa AI-tjänster befinner sig i ett tidigt utvecklingsstadium och erbjuder ofta begränsade möjligheter till lokala och flexibla anpassningar. Det kan exempelvis handla om begränsade möjligheter att bestämma vilka data som AI-tjänsten ska ha tillgång till. Om verksamhetens informationshantering och behörighetsstyrning är bristfällig kan det leda till att en användare får åtkomst till information som denne inte borde ha åtkomst till. Detta ökar i sin tur risken för att också AI-tjänsten kan komma åt och behandla data som tjänsten inte borde hantera.

Behandling av personuppgifter i integrerade generativa AI-lösningar
Att aktivera en generativ AI-tjänst som redan finns integrerad i befintlig programvara kan innebära att AI-tjänsten får tillgång till information som behandlas i programvaran. Om denna information innehåller personuppgifter innebär aktiveringen att personuppgifter behandlas av AI-tjänsten. Denna omständighet innebär inte nödvändigtvis att en ny behandling av personuppgifter uppstår eftersom AI-tjänsten kan vara avsedd att uppfylla samma funktion som den programvara som redan används i verksamheten. Huruvida användningen av AI-tjänsten utgör en ny behandling av personuppgifter måste bedömas från fall till fall.

En generativ AI-tjänst, exempelvis en digital assistent med tillgång till verksamhetens data, kan användas för att behandla personuppgifter på många olika sätt och för varierande ändamål. En verksamhet som avser att aktivera en integrerad generativ AI-tjänst bör tydligt definiera ändamålet för användningen. Om AI-tjänsten är byggd som en extrafunktion till den befintliga programvaran kan ändamålet med användningen av AI-tjänsten ofta kopplas till samma ändamål som för användningen av den aktuella programvaran.

Nya behandlingar av personuppgifter kan uppstå
Aktiveringen av en integrerad generativ AI-tjänst kan leda till nya personuppgiftsbehandlingar. Det kan vara fallet om AI-tjänsten som aktiveras kan samla in användardata genom att registrera interaktioner med tjänsten, inklusive de instruktioner som användaren skapar och de svar som tjänsten genererar.

En verksamhet som planerar att aktivera en integrerad generativ AI-tjänst bör först analysera om användningen medför nya behandlingar av personuppgifter och dokumentera detta. Om fler personuppgifter behandlas eller kombineras på nya sätt genom användningen av AI-tjänsten, är det viktigt att kartlägga dessa nya behandlingar. Verksamheten kan också behöva uppdatera sin integritetspolicy eller liknande dokument.

Personuppgiftsansvar för integrerade generativa AI-tjänster
Användning av en integrerad generativ AI-tjänst i befintlig programvara innebär i många fall att verksamheten blir personuppgiftsansvarig för de behandlingar som sker genom AI-tjänsten. Ansvaret för att medarbetarna använder integrerade AI-tjänster på ett lagligt och integritetsvänligt sätt ligger vanligtvis på verksamheten. Därför krävs det att verksamheten har tillräckliga kunskaper om tjänstens funktionalitet för att kunna säkerställa en korrekt användning.

Generella råd vid användning av en integrerad generativ AI-tjänst
Generativa AI-tjänster integrerade i befintliga programvaror skiljer sig från traditionella digitala assistenter genom att de kombinerar funktionaliteten hos grundmodeller med tillgång till verksamhetens egna data. Användningen av dessa AI-tjänster kräver att verksamheten har god översikt och kontroll över sin information. AI-tjänsten bör inte få mer åtkomst än den data som den individuella användaren har behörighet till. Verksamheten behöver ha en adekvat informationshantering på plats innan AI-tjänsten ges tillgång till verksamhetens data. Detta är en grundläggande förutsättning för att kunna leva upp till dataskyddsförordningen.

Informationskartläggning kan vara nödvändig
För att minska risker, och för att kunna efterleva principen om ansvarsskyldighet, bör verksamheten genomföra en informationskartläggning. Syftet med en sådan kartläggning är bland annat att identifiera vilken information som finns, var den lagras och vilka inom verksamheten som har åtkomst till den. En noggrant genomförd informationskartläggning är också avgörande för att säkerställa korrekta behörigheter vid aktiveringen av en AI-tjänst som får tillgång till verksamhetens data.

Verksamheten bör etablera rutiner för att kontinuerligt klassificera information och efterleva eventuella lagkrav på informationshantering. Detta kräver ett samlat och noggrant arbete från hela verksamheten.

Överväg nya strukturer för att anpassa verksamheten
Eftersom användningen av AI-tjänsten kan innebära att verksamhetens information hanteras på nya sätt, kan befintliga processer behöva anpassas. Det är viktigt att verksamheten skaffar sig kunskap om tjänstens funktionalitet för att kunna integrera den på ett integritetsvänligt sätt.

Aktiveringen av en generativ AI-tjänst i befintliga programvaror innebär ofta att verksamheten behöver införa nya rutiner, riktlinjer och utbildningsinsatser. Syftet med det är att skapa en djupare förståelse för AI-tjänstens kapacitet och samtidigt öka medvetenheten om generativ AI:s potential, begränsningar och risker.