Riktlinje för AI i Söderköpings kommun

Diarienummer: KS/2025-00501

Antagen: KS 2025-10-13, § 166

Reviderad:

Dokumentansvarig förvaltning: Kommunstyrelsens förvaltning

Dokumentet gäller för: Söderköpings kommun

Dokumentet gäller till och med: Tillsvidare

Postadress
Söderköpings kommun
614 80 Söderköping

Besöksadress
Kommunhuset
Storängsallén 20

Kontakt
0121-181 00
kommun@soderkoping.se

Org.nr och webbplats
212000-0464
www.soderkoping.se

Informationsklass: Intern Ansvarig: Elin Hofsten Dokumentet är endast giltigt vid utskriftstillfället. Senast giltiga version finns på Kanalen och på www.soderkoping.se

Versionshantering:

Version 01.00	Antagen 2025-10-13

1. Inledning

Detta dokument gäller för alla medarbetare i Söderköpings kommun och syftar till att vara ett stöd vid användningen av AI.

På samma sätt som med alla digitala system är du ansvarig för ditt eget handlande när du använder AI. Det innebär att du har ansvar för den information som du matar in i ett AI-system och att du har ansvar för hur du använder resultatet från ett AI-system.

Vid all hantering av information ska kommunens informationssäkerhetsriktlinje följas. Varje nämnd ansvarar för att genomföra erforderliga analyser innan ett AI-system används.

I ett försämrat säkerhetspolitiskt läge är det också extra viktigt att vara vaksam på desinformation. Ett ökat användande av AI kan innebära att mängden desinformation ökar och blir mer komplex i sitt innehåll och kontext.

2. Definition av AI-tjänster

Artificiell intelligens (AI) - Ett AI-system är ett maskinbaserat system som, för uttryckliga eller underförstådda ändamål, utifrån de indata det tar emot drar slutsatser om hur man genererar utdata, t.ex. förutsägelser, innehåll, rekommendationer eller beslut som kan påverka fysiska eller virtuella miljöer.

AI-system – med ordet "system" avses appar, programvaror samt andra digitala tjänster och verktyg. Med "AI-system" avses system innehållande större eller mindre inslag av artificiell intelligens (AI). Om det specifikt handlar om generativ AI är detta utskrivet.

Generativ AI – AI-system som kan skapa nytt material baserat på enkla instruktioner eller exempel kallas för generativ AI, eller skapande AI. Det finns en stor mängd AI-baserade datorprogram, system och IT-tjänster som genererar nytt material i form av till exempel text, bild, video och programmeringskod. Många av dem är tillgängliga via internet för allmänheten.

Allmänt tillgängliga AI-system – AI-system som är tillgängliga för vem som helst att använda t.ex. via webbsida eller via en app. Allmänt tillgängliga AI-system är inte anskaffade, skapade eller införda specifikt för kommunens behov och är därför inte heller granskade av kommunen.

Öppen information – information som kan spridas fritt inom och utanför Söderköpings kommun.

Intern information – information som inte är avsedd att spridas utan är ägnad att hantera inom kommunen.

Konfidentiell information . information som om den sprids till obehöriga kan innebära allvarliga eller mycket allvarliga konsekvenser för samhälle, ekonomi, verksamhet eller individ. Till exempel sekretessbelagd information eller känsliga personuppgifter.

3. Informationssäkerhet

I de flesta fall kräver användningen av ett AI-system att du matar in någon form av information i AI-systemet. Kommunens information är en av våra viktigaste tillgångar. Vår information omfattas av lagstiftning som rör bland annat personuppgiftsbehandling och sekretesskydd. Det finns även annan information som, om den sprids till obehöriga, kan orsaka skada för verksamhet, samhälle, ekonomi, individ och förtroendet för kommunen. Kommunens information ska hanteras i system som är godkända utifrån genomförd systemklassning. Det gäller även AI-system. Se rutin för systemklassning på Kanalen, Systemklassning - Intranet.

När kommunen anskaffar eller utvecklar system ska vi ställa krav på systemet utifrån den information som ska hanteras i det och de risker som kan uppkomma. Det kan inte kommunen göra när det gäller allmänt tillgängliga system som var och en kan använda. Där gäller i stället de användarvillkor du accepterar när du använder systemet. Av denna anledning är det endast tillåtet att mata in öppen information i allmänt tillgängliga AI-system.

För att intern och konfidentiell information ska kunna hanteras i ett AI-system så krävs att erforderliga analyser genomförts och att systemet uppfyller identifierade krav.
Nedan kan du läsa om vad som gäller för behandling av personuppgifter, sekretessbelagd information, annan känslig information och upphovsrättsskyddat material i AI-system.

A. Allmänt tillgängliga AI-system får enbart användas för öppen information (K0 enligt Söderköpings kommuns klassningsmodell).

B. Innan användning, utveckling och inköp av AI-system så måste erforderliga analyser genomföras i enlighet med rutinen för systemklassning.

3.1 Personuppgifter

Att mata in personuppgifter i ett AI-system är att behandla personuppgifter. All behandling av personuppgifter ska följa de grundläggande principerna i GDPR. Om och hur principerna uppfylls måste analyseras innan AI-systemet används för den tänkta behandlingen. Eftersom AI är en ny teknik ska det i de flesta fall också göras en konsekvensbedömning (DPIA) innan personuppgifter används i ett AI-system. Om det inte har gjorts en analys av behandlingen utifrån GDPR är det olagligt att mata in personuppgifter i ett AI-system. I rutinen för systemklassning ingår de olika aktiviteter som behöver genomföras ur ett dataskyddsperspektiv.

När det gäller allmänt tillgängliga AI-system saknas tillräckliga möjligheter för kommunen att kontrollera leverantörens behandling eller säkerhet. Därför får inga personuppgifter matas in i allmänt tillgängliga AI-system. Om detta ändå sker är det fråga om en personuppgiftsincident som omedelbart ska utredas enligt rutin. Här finns med information om hur du ska gå tillväga om du upptäcker eller misstänker att en personuppgiftsincident inträffat Incidentrapportering - personuppgifter - Intranet.

C. Innan personuppgifter matas in i ett AI-system måste det göras en analys av behandlingen enligt GDPR i enlighet med rutinen för systemklassning.

D. Personuppgifter får inte matas in i allmänt tillgängliga AI-system.

E. Om personuppgifter matas in i ett allmänt tillgängligt AI-system så är det fråga omen personuppgiftsincident som ska hanteras enligt rutin.

3.2 Sekretess

Enligt offentlighets- och sekretesslagen får information som omfattas av sekretess inte röjas för utomstående. När information görs tillgänglig för en IT-leverantör, t ex genom att laddas upp i en molntjänst, är informationen röjd. Om informationen är krypterad på ett sådant sätt att leverantören inte kan göra den läsbar är den dock inte röjd. Att röja sekretessbelagd information är bara tillåtet om det finns en sekretessbrytande bestämmelse.

Om du röjer sekretessbelagda uppgifter genom att mata in dem i ett AI-system kan det vara straffbart enligt brottsbalken som brott mot tystnadsplikten.

F. Mata aldrig in sekretessbelagd information i ett AI-verktyg som inte har analyserats och konstaterats vara säkert för sådan information.

G. Sekretessbelagd information får inte matas in i allmänt tillgängliga AI-system.

3.3 Annan känslig information

Annan känslig information är till exempel information som, om den kommer i fel händer, riskerar skada individer, kommunens verksamhet, ekonomi eller samhället i stort. Allmänt tillgängliga AI-system som använder inmatad information för att träna systemet kan ta med din information i svar till andra användare. Det vill säga den information som du matar in riskerar att göras tillgänglig för obehöriga. Därför är det inte tillåtet att mata in annan känslig information i allmänt tillgängliga AI-system

H. Mata aldrig in annan känslig information i ett AI-verktyg som inte har analyserats och konstaterats vara säkert för sådan information.

I. Annan känslig information får inte matas in i allmänt tillgängliga AI-system.

4. Förtroende och transparens

För att bevara eller stärka kommunens förtroende när AI används måste all användning ske på ett ansvarfullt sätt, med stöd av välutvecklade processer, ett systematiskt arbetssätt och med god dokumentation. Kommunens verksamheter måste alltid följa den lagstiftning som gäller. I verksamhet som är lagstyrd behöver man analysera om det man tänkt göra är lagligt innan man börjar använda AI-systemet.

4.1 Offentlighetsprincipen

Sverige har med offentlighetsprincipen ett väl utvecklat system för transparens inom offentlig sektor. En förutsättning är att allmänna handlingar ska diarieföras eller hållas ordnade så att de kan begäras ut. Därför måste det övervägas om det skapas nya allmänna handlingar genom användning av AI-systemet och hur de i så fall ska lagras och diarieföras.

J. En verksamhet som använder ett AI-system behöver ta ställning till om AI-systemet skapar nya allmänna handlingar och hur de i så fall ska lagras och diarieföras.

4.2 Förvaltningslagen

Enligt förvaltningslagen är det ett krav att en myndighet ska motivera sina beslut och förklara vad som gjort att myndigheten nått sin slutsats. Därför behöver varje verksamhet som använder ett AI-system på ett sätt som påverkar kommunens invånare eller brukare kunna förstå och förklara sitt AI-systems funktioner på en lämplig nivå. Utmaningen ligger i att ett AI-system kan bestå av tusentals, eller till och med miljontals, numeriska värden som AI-systemet lär sig under dess träningsfas. Det är därför ofta inte möjligt, ens för den som utvecklat systemet, att förklara exakt hur systemet når en viss slutsats. Därför kan det i vissa sammanhang vara olämpligt att använda AI av det skälet.

K. En verksamhet som använder ett AI-system på ett sätt som påverkar kommunens invånare eller brukare behöver kunna förstå och förklara sitt AI-systems inre funktioner.

L. Det är i dagsläget inte lämpligt att använda AI-system för att helt automatisera beslut i ärenden.

4.3 Använda resultatet från generativ (skapande) AI

M. Du är alltid ansvarig för att texten är faktamässigt korrekt, har rätt tonalitet och är fri från fördomar och annan bias.

N. Du kan inte överlåta din yrkesmässiga bedömning till ett generativt AI-system.

O. Du ansvarar för texten som om du själv hade skrivit den

4.4 Transparens

P. Har AI använts vid handläggningen av ett ärende ska det anges i tjänsteutlåtande och underlagshandlingar. Det ska även framgå vilket AI-verktyg som använts.

Q. I anslutning till bilder som genererats med hjälp av AI ska texten *Denna bild har genererats med hjälp av AI* infogas.

5. Upphovsrättsskyddat material

Innan du matar in upphovsrättsskyddat material i en AI-tjänst eller använder AI-genererat material är det viktigt att tänka på upphovsrätten. Läs igenom användarvillkoren för den tjänst du använder då de kan skilja sig åt. I de användarvillkor som man godkänner när man till exempel använder ChatGPT och Copilot anges att man själv ansvarar för att inte bryta mot någon annans upphovsrätt. Det innebär att det är du själv, och inte leverantören av systemet, som gör dig skyldig till upphovsrättsbrott om det skulle visa sig att det du gör strider mot upphovsrätten. Eftersom AI-tjänsterna är tränade på material från internet kan resultatet bli väldigt likt en bild eller text som är vanlig på internet. Om du sprider en AI-genererad bild eller text som är för lik någons upphovsrättsskyddade verk kan du begå ett upphovsrättsbrott och bli skadeståndsskyldig.

R. Det är tillåtet att använda AI-genererade bilder förutsatt att du tagit hänsyn till upphovsrätten och förutsatt att du, i direkt anslutning till bilden, anger en bildtext där det framgår att bilden är AI-genererad.

6. Anskaffa och införa AI-system

Att anskaffa AI-system eller att införa befintliga AI-system i nya verksamheter är en del i Söderköpings kommuns digitaliseringsarbete och integreras med övrig styrning och utveckling i kommunen. Genom att vi anskaffar, skapar och inför AI-system enligt befintliga processer ger vi förutsättningar för att alla relevanta aspekter utreds.

När kommunen anskaffar eller börjar använda befintliga AI-system i nya processer utvärderas systemets och leverantörens förmåga att skydda informationen utifrån de krav på skydd av informationen som identifierats genom analyserna. Det ställs också krav på funktioner och avtalsvillkor som säkrar kommunens kontroll över informationen.

S. Befintliga processer och metodiker för digitaliserings- och innovationsarbete ska användas oavsett om den tänkta lösningen innehåller AI eller inte. Se rutin för systemklassning.
T. Risker ska analyseras innan ett AI-system införs i verksamheten och lämpliga åtgärder ska vidtas för att eliminera eller minska riskerna. Om återstående risker inte kan accepteras är det inte tillåtet att använda AI-systemet.

7. Allmänt tillgängliga AI-system

I takt med teknikutvecklingen lanseras nya AI-system som görs allmänt tillgängliga för användning. Allmänt tillgängliga AI-system kan tillföra värde för kommunen genom att underlätta vissa arbetsuppgifter och effektivisera arbetet.

Eftersom kommunen inte har något avtal med leverantören av allmänt tillgängliga AI-system finns det ingen möjlighet för kommunen att påverka eller få insyn i leverantörens användning av informationen i AI-systemet. Kommunen är därför skyldig att begränsa vilken information som får matas in i allmänt tillgängliga AI-system.

U. Allmänt tillgängliga AI-system får enbart användas för öppen information.
V. Använd inte ett AI-system som är helt nytt eller kommer från en okänd avsändare i ditt arbete.
W. Ge aldrig ett allmänt tillgängligt AI-system åtkomst till information på dina enheter som arbetstelefon och dator
X. Du får under inga omständigheter mata in personuppgifter, sekretessbelagd information eller annan känslig information i AI-system tillgängliga för allmänheten.

Y. Du får inte använda allmänt tillgängliga AI-system för ändamål som är förbjuda eller högrisk enligt AI-förordningen.

7.1 Microsoft Copilot

Microsoft Copilot är Microsofts digitala assistent med AI-teknik. Den är en del av kommunens Microsofts 365-miljö (M365) vilket innebär att den information du matar in i denna digitala assistent hanteras i Microsofts molntjänst på samma sätt som informationen i OneDrive eller Teams. Till skillnad från OneDrive och Teams så har inte erforderliga analyser genomförts av kommunen rörande Copilot. Innan analyser är genomförda så gäller därför samma regler för användning av Copilot som för allmänt tillgängliga AI-system, såsom Chat GPT. Det vill säga, endast öppen information och inga personuppgifter får matas in i Copilot.

8. Risker med användning av generativ AI

Användning av generativ AI kan medföra flera olika risker. Nedan ges några exempel. Observera att listan nedan, som kommer från Myndigheten för digital förvaltning (DIGG) inte är uttömmande.

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.