

Riktlinje för användning av generativ AI inom Svenska Röda Korset

1 Bakgrund

1.1 Syfte och omfattning

Denna riktlinje reglerar användningen av artificiell intelligens (AI) av Svenska Röda Korsets medarbetare. Den omfattar användning av generativ AI och är ett komplement till gällande informationssäkerhets- och dataskyddsriktlinjer.

Syftet är att säkerställa att användningen av generativ AI är i enlighet med gällande lagstiftning, att den är säker, etisk och förenlig med organisationens humanitära uppdrag.

Utvecklingen inom AI går fort och detta dokument kommer behöva revideras regelbundet. Ökad intern användning av AI kommer också innebära ett lärande och behov av framtida revideringar.

Vad är generativ AI?

Generativ AI är en typ av artificiell intelligens som har kapacitet att generera text, bild, video, ljud, tabeller, diagram eller annan media som svar på förfrågningar. Systemet gör det genom en (språk)modell som lärt sig mönster och strukturer från det data som modellen tränats på och som sedan baserat på användarens förfrågningar genererar sannolika svar. Generativ AI kan inte likställas med en sökmotor eftersom de färdiga modellerna inte har tillgång till materialet de tränats på. Olika typer av AI har existerat sedan en tid tillbaka men generativ AI som till exempel ChatGPT representerar en stor förändring i hur vi nyttjar den typen av teknik.

1.2 Användning av generativ AI inom Svenska Röda Korset

Svenska Röda Korset ser många positiva möjligheter med användningen av generativ AI. Medarbetare och ledare uppmanas att utforska och använda sådana lösningar där det kan optimera arbetsflöden, förbättra kommunikationen, öka intäkterna och stärka det humanitära hjälparbetet, inklusive skyddet för våra målgrupper. Riskerna med att använda generativ AI som ett verktyg i arbetet är dock inte tillräckligt utrett och därför bör vi använda generativ AI med försiktighet.

Om du använder generativ AI i ditt arbete är det viktigt att du delar dina erfarenheter med din chef och dina kollegor så att vi på Svenska Röda Korset säkerställer öppenhet, kunskapsdelning och erfarenhetsutbyte.

2 Principer för användning av generativ AI

2.1 Ansvar och källkritik

När du som anställd använder generativ AI i ditt arbete ska du vara medveten om att det fortfarande är du och Svenska Röda Korset som är ansvariga för användningen av det AI-genererade innehållet på samma sätt som för material du tar fram utan AI-stöd. **Använd alltid ditt sunda förnuft och var källkritisk när du använder generativ AI i ditt arbete.**

Använd aldrig AI-genererat material om du inte själv har kompetensen och erfarenheten att kvalitetsgranska det (i teorin själv kan generera motsvarande slutresultat).

Generativ AI baserar sina svar på mönster och strukturer från träningsdata. Om den data som använts innehåller felaktig eller partisk information finns det en risk för att modellen producerar felaktig information eller partiska svar vilket potentiellt leder till spridning av falskt eller missvisande innehåll. Du är alltid ansvarig för att kritisk granska det resultat som du får fram. Till exempel:

- Vid användning av AI-drivna översättningstjänster, så måste du alltid, beroende på hur översättningen ska användas och spridas, säkerställa att resultatet förmedlar motsvarande budskap som originaltexten. Såvida du själv inte behärskar översättningsspråket tillräckligt väl för att bedöma om översättningen är fullgod för den planerade användningen behöver du säkerställa detta på annat sätt.
- AI kan generera utkast på rapporter, ansökningar, protokoll, tabeller, diagram och redaktionella texter antingen baserat på specifika källor som du anger eller baserat på det material tjänsten är tränad på, men du behöver själv kunna verifiera sanningshalten och rimligheten i slutresultatet och kunna ta ansvar för dokumentet i sin helhet. Du även behöver t.ex. säkerställa att en text speglar Svenska Röda Korsets varumärke och personlighet.
- Om det behövs kod för en digitalfunktion, men du inte är mjukvaruutvecklare, så bör du inte skapa ett script genom generativ AI, och använda det, eftersom du inte kan bedöma hur det verkligen kommer fungera.

För att använda AI-genererad text som en färdig produkt direkt mot målgrupp (dvs inte som ett utkast som sedan vidareutvecklas) så bör det alltid på något sätt vara tydligt, t.ex. att “du kommunicerar med en chatbot, alla svar är automatgenererade”, “detta är en transkribering genererad av tjänst XX”, “detta är ett AI genererat protokoll baserat på YY”, “detta är en AI-genererad översättning genererad av tjänst ZZ”, “detta är en AI-genererad artikel baserad på följande frågor”.

2.2 GDPR och dataskydd

I användandet av den publika versionen av OpenAI's ChatGPT (och andra motsvarande tjänster) så finns inga legala skydd för Svenska Röda Korsets intressen. All data som du laddar upp i verktyget har OpenAI rätt att förfoga över och återanvända på det sätt de vill. All information och även dokument som laddas upp i den publika versionen av OpenAI's ChatGPT blir del av deras insamling av data för att träna verktyget och kan även bli tillgänglig för externa aktörer. Du godkänner detta i och med att du godkänner användarvillkoren. Motsvarande gäller även för de olika tillägg som kan laddas ner gratis via Microsoft AppSource till Office-programmen.

Det är inte tillåtet att ladda upp/dela känslig data och information som tillhör Svenska Röda Korset till publika generativa AI tjänster. (detta inkluderar affärskritiska uppgifter – som kontrakt, anteckningar om affärsutveckling och annan konfidentiell Röda Korset-data.) Genom våra policys och riktlinjer så förbinder vi oss att inte sprida känslig information om verksamheten. Rådgör med din chef om du är osäker.

Det är aldrig tillåtet att ladda upp data eller information som innehåller personuppgifter i någon form i generativa AI-tjänster. Organisationen är skyldig att skydda integriteten och säkerheten för all individdata i enlighet med dataskyddslagar och Svenska Röda Korsets riktlinje för skydd av personuppgifter.

Var medveten om att du även riskerar att kränka andras upphovsrätt om du inkluderar ett annat företags eller konsults arbete (som bilder, texter, illustrationer etc.) i någon av de många öppna AI-språkmodellerna eller använder det AI-genererade materialet därifrån utan samtycke/kontrakt.

2.3 AI genererade bilder

Generellt ska alla bilder och videos som Svenska Röda Korset producerar och publicerar följa våra etiska riktlinjer oavsett om de är genererade med AI eller inte. Följande gäller för användning av AI-generat bildmaterial.

Redaktionella bilder/video: Alla bilder och videos vi producerar och publicerar ska vara autentiska och äkta. **Vi använder alltså inte AI genererade bilder och videos när vi utger oss för att kommunicera en verklig händelse (oavsett syfte).**

Reklam: När vi inte utger oss för att skildra en autentisk händelse kan vi tillåta att bilder eller videos är AI genererade eller modifierade.

Om du använder bilder, illustrationer, video och liknande genererade av AI ska dessa alltid märkas tydligt i samma utsträckning som vi t.ex. anger vem som tagit ett foto, så att det är klart för målgruppen att materialet genererats av AI. Till exempel med tillhörande texter som 'Denna bild är skapad av artificiell intelligens (AI)'.

Tillgång till generativa AI-tjänster

Säkra generativa AI-tjänster

Svenska Röda Korset har idag inga företags/betalkonton till några av de publika generativa AI-tjänsterna som finns tillgängliga.

Det vi erbjuder till alla är:

Microsoft Azure OpenAI - en intern AI-chat som ligger i vår egen Microsoftmiljö och som inte sparar eller delar någon information utanför organisationen. Funktionen är baserad på Microsoftkomponenter och under ständig utveckling, prestandan förbättras löpande, nya funktioner kan kopplas på och den underliggande språkmodellen kan uppdateras när bättre versioner görs tillgängliga. Du når den via ai.redcross.se

Microsoft Bing Copilot - en AI-chat som är ihopkopplad med en webbsökmotor och därför genererar svar med referenser till aktuella webb-länkar. Tjänsten är tillgänglig på alla SRK-datorer om du använder Microsofts webbläsare Edge samt är inloggad i webbläsaren med ditt Röda Korskonto. Denna tjänst är Microsoft-standard men ligger i Röda Korssets egna Microsoft-miljö så så länge du inte slår på några plugin-program så delas ingen information utanför organisationen.

För att ta ställning till eventuella framtida företagskonton och/eller egna installationer av motsvarande AI-modeller, t.ex. inom vår egen Microsoft miljö, i syfte att skydda vår data, behöver vi göra en avvägning av nyttan ställt mot kostnaderna för detta. Samma sak gäller när generativa AI-modeller erbjuds som en integrerad del av olika tjänster och verktyg, t.ex. Microsoft Co-pilot kopplat till olika Microsoft-tjänster, Googles Gemini, AI-stödd

översättning, AI-drivna chatbott-lösningar eller AI-tjänster som genererar redaktionellt material.

Upphandlade säkra tjänster inkluderar idag:

DeepL - översättningstjänst, några användare per avdelning har tillgång till betaltjänsten, den plug-in som finns i Word är en gratisversion.

Köp av AI-baserade lösningar

Den som ser behov av ett betalkonto i någon specifik generativ AI-tjänst ska precis som vid inköp av andra digitala tjänster alltid lyfta detta till IT som gör en bedömning av riskerna och tar beslut om köp av nya tjänster (eftersom det alltid är en kostnad som belastar IT-budgeten.)

Gratistjänster

Många generativa AI-tjänster är idag öppet tillgängliga för vem som helst, ofta är allt som krävs ett gratiskonto. **Det är inte tillåtet att använda sin redcross-mejl för att skapa gratiskonton** varken direkt eller indirekt genom t.ex. ett Googlekonto kopplat till en redcross-mejl. Svenska Röda Korset vill inte att ägarna av tjänsterna ska kunna spåra och analysera de förfrågningar som Svenska Röda Korsets medarbetare gör.

