# Exempel på annoteringsuppgifter

Detta papper beskriver de olika sätten vi annoterar och genererar data i projektet. Syftet med dessa beskrivningar är att ge er en uppfattning om vad det innebär rent praktiskt att annotera och vad syftet är med olika typer av annotering.

Beskrivningarna är generiska och kan tillämpas inom vilken del av verksamheten som helst. Det är bara innehållet som ändras, och experter från olika verksamheter kan skapa data för just sitt expertområde.

I följande avsnitt kommer vi att beskriva olika typer av annoteringar, inklusive utvärderingsdata, instruktionsdata och annotering av konversationer. Vi hoppas att dessa beskrivningar kommer att vara till hjälp för er i ert arbete med att planera vilka som ska delta i annoteringen.

Var och en av dessa annoteringsuppgifter har fler detaljer och delsteg än vad som står beskrivet i detta dokument. Dessa går vi igenom allteftersom vi arbetar tillsammans med er. Alla annoteringsuppgifter kräver också en mindre utbildning för att den som annoterar ska lära sig utföra uppgiften på bästa sätt.

Om ni har frågor är ni välkomna att höra av er i Teams!

# Utvärderingsdata för RAG-systemet

## Bakgrund

För att säkerställa att vår digitala assistent fungerar effektivt och ger korrekta svar, behöver vi utvärdera dess prestanda. Men hur gör vi det? Till skillnad från andra datorsystem, där outputen kan vara entydig och lätt att bedöma, är det svårare att utvärdera en textbaserad AI som vår digitala assistent. Därför är det avgörande att hitta metoder för att utvärdera systemet, annars riskerar vi att jobba i blindo när vi försöker förbättra det.

Utvärderingsdata är nyckeln till att förstå och förbättra Svea. Det är den typen av data som gör det möjligt för oss att bedöma om systemet ger korrekta och relevanta svar. Men för att kunna utvärdera systemet, behöver vi först ha goda exempel på hur vi vill att systemet ska skriva sina svar. Dessa exempel måste skrivas av människor med expertis inom området, och det är en komplex och tidskrävande process.

Men när vi väl har en stor och högkvalitativ uppsättning utvärderingsdata, kan vi snabbt och säkert förbättra systemet. Vi kan testa nya funktioner och ändringar i systemet och se om de leder till bättre resultat. Utvärderingsdata är därför en av de viktigaste delarna av utvecklingsprocessen, och medarbetarnas bidrag till detta är ett fantastiskt arbete och helt avgörande för att kunna skapa en bättre digital assistent.

## Mål

Att skapa exempel på frågor och svar som kan användas för att utvärdera och förbättra RAG-systemets prestanda.

* Frågor som är relevanta för den offentliga sektorn och som RAG-systemet bör kunna svara på
* Svar som är korrekta och relevanta för varje fråga
* Referenser till dokument som stödjer svaren

## Annoteringsprocessen

### Steg 1: Samla dokument för vektordatabasen

#### Syfte

Samla dokument som rör avdelningen och som ska finnas i vektordatabasen.

#### Ansvarig

Experter från verksamheten.

#### Process

Experterna samlar 300–3000 dokument som rör deras avdelning och som ska finnas i vektordatabasen. Dessa dokument kommer att användas i annoteringsarbetet.

### Steg 2: Definiera annoteringsriktlinjer

#### Syfte

Definiera vad ett bra svar från RAG-funktionen ska innehålla och hur det ska skrivas.

#### Ansvarig

Experter från verksamheten och AI Sweden.

#### Process

Experterna och AI Sweden träffas i en workshop för att definiera annoteringsriktlinjer. De kommer att diskutera och besluta om vad som är ett bra svar från RAG-funktionen och hur det ska skrivas.

## Steg 3: Annotera frågor, svar och referenser

#### Syfte

Skapa exempel på frågor och svar som RAG-funktionen ska kunna hantera, och lägga till referenser till dokument som stödjer svaren.

#### Ansvarig

Experter från verksamheten.

#### Process

Experterna arbetar självständigt och skriver först frågor som de vill att RAG-funktionen ska kunna svara på. De skriver sedan ett "guldstandard-svar" för varje fråga, det vill säga det svaret som de vill att RAG-funktionen ska ge. Samtidigt kommer de att lägga till referenser till de dokument som stödjer svaren.

## Steg 4: Granska och utvärdera annoterade data

#### Syfte

Granska och utvärdera annoterade data för att säkerställa att de är korrekta och av hög kvalitet.

#### Ansvarig

AI Sweden och experter från experter från verksamheten.

#### Process

AI Sweden och verksamhetsexperterna går igenom de data som skrivits och utvärderar dem enligt ett antal kriterier.

## Steg 5: Använda annoterade data för att testa RAG-funktionen

#### Syfte

Använda annoterade data för att testa RAG-funktionen och se till att den ger korrekta och relevanta svar.

#### Ansvarig

AI Sweden

#### Process

AI Sweden använder de annoterade data för att testa RAG-funktionen och se till att den ger korrekta och relevanta svar. AI-ingenjörerna kan testa nya metoder och ändra inställningar i systemet för att se om det ger bättre resultat. De utvärderingsdata som tagits fram kommer också att användas för att utvärdera systemet efter att AI-modellerna har finjusterats.

#### Exempel

Här nedan är ett verkligt exempel på ett stycke utvärderingsdata som skrivits av metodutvecklare från Hemtjänsten i en kommun. I etapp 1 skrevs ca 200 sådana här utvärderingsdata.

|  |  |
| --- | --- |
| Fråga | Jag är omvårdnadspersonal och har upptäckt att en vårdtagare har trycksår, ska jag skriva en avvikelse? |
| Svar | Om du upptäcker att en vårdtagare har trycksår ska du omedelbart vidta åtgärder för att undanröja större skada. Du ska kontakta legitimerad personal och du ska rapportera in avvikelsen i DF Respons Avvikelse brukare. En händelse som inte stämmer med normal rutin och som har medfört eller som kunde ha medfört obehag, risk eller skada för den enskilde skall rapporteras. Exempel på händelser som kan leda till en rapport enligt Lex Maria är trycksår. Det är enhetschef och MAS/MAR som avgör om det ska utredas enligt Lex Maria. Läs mer om vad som ska rapporteras in som avvikelser i rutinen för avvikelsehantering. Du kan också läsa mer om trycksår i vårdhandboken. |
| Riktlinje | Rutin för avvikelsehantering |
| Referenser | 2.4 Vad ska rapporteras  En händelse som inte stämmer med normal rutin och som har medfört eller som kunde ha medfört obehag, risk eller skada för den enskilde skall rapporteras.  Exempel på händelser som kan leda till en rapport enligt Lex Maria   * Fall, trycksår och brister i det förebyggande arbetet.   2.5 Tillvägagångssätt när en avvikelse inträffat  All avvikelsehantering hanteras i DF Respons Avvikelse Brukare. 1. När en avvikelse inträffat vidtas omgående åtgärder av rapportören för att undanröja ytterligare risk för missförhållande/vårdskada. 2. Legitimerad personal ska alltid informeras direkt vid avvikelser som omfattar hälso- och sjukvårdsinsatser. 3. Enhetschef informeras om avvikelsen dagtid, sker avvikelsen under jourtid meddelas enhetschef nästkommande vardag. 4. Avvikelsen ska dokumenteras i socialjournal samma dag. Senast dagen efter händelsen rapporteras den i avvikelsehanteringssystemet.  MAS/MAR Har en rådgivande och samordnande funktion i utredning av avvikelser som riskskattats till 8 eller mer eller klassificering 3-4. Ansvarar för att bedöma om utredningen visar att händelsen är så allvarlig att den skall anmälas till IVO enligt Lex Maria. Ansvarar för samordning av avvikelser i vårdsamverkan. Ansvarar för att redovisa antalet Lex Maria utredningar till nämnd. |

# Träningsdata för RAG-systemet

## Bakgrund

Vi kan skapa träningsdata som kan förbättra systemets förmåga att hitta rätt svar i databasen. Det handlar om att träna embeddingmodellen, som är en AI-modell som ansvarar för att hitta information i RAG-databasen som matchar användarens fråga. Även om embeddingmodellen redan har en grundläggande förmåga att hitta svar, kan den förbättras ytterligare genom träning med svenska data som är specifika för den offentliga sektorn. Genom att skapa träningsdata som är anpassade till vår specifika kontext, kan vi förbättra systemet så att det blir bättre på att hitta rätt svar i era dokument.

## RAG-systemet

Nedan ser du hur RAG-systemet fungerar. Det är embeddingmodellen som hämtar ”kontext” (svar på användarens fråga) från databasen. Sedan skickas prompten och kontexten till den stora språkmodellen som sedan generar ett svar.

En bild som visar text, skärmbild, diagram

Automatiskt genererad beskrivning

## Mål

Att skapa en uppsättning frågor som besvaras av ett dokument i RAG-databasen, samt samla den text från dokumentet som besvarar frågan.

|  |  |  |
| --- | --- | --- |
| Dokument | **Frågor som besvaras av dokumentet** | **Textpartier som besvarar frågan** |
| Fråga 1 | Text från dokument 1 |
| Fråga 2 | Text från dokument 2 |
| Fråga 3 | Text från dokument 3 |
| Osv | Osv |

## Annoteringsprocessen

### Steg 1: Välj ett dokument

Välj ett dokument från RAG-databasen. Det kan vara vilket dokument som helst som inte annoterats förut.

### Steg 2: Skriv frågor

Skriv frågor som besvaras av dokumentet. Det ska gärna vara sådana frågor som ofta ställs i verksamheten. Men det kan vara andra frågor också, ju fler desto bättre.

### Steg 3: Identifiera text

För varje fråga, ta ut de delar av dokumentet som besvarar frågan.

# Donera konversationer

## Bakgrund

När användare interagerar med Svea, skapas värdefulla data i form av frågor och svar. Dessa data kan ge oss viktig information om användarnas behov och kan användas för att förbättra språkmodellen. Men eftersom användarna kan skriva vad som helst som fråga, finns det en risk att frågorna innehåller personuppgifter eller annan känslig information.

För att skydda denna känsliga information, måste vi se till att den inte delas med någon som inte har behörighet. Detta innebär att vi inte kan använda dessa data direkt i utvecklingen av Svea. Men vi kan fortfarande använda dem på ett säkert sätt.

Genom att låta medarbetare granska sina egna konversationer med Svea, kan vi identifiera de data som är säkra att delas med utvecklingsteamet. Dessa data kan sedan användas för att förbättra språkmodellen och göra Svea mer effektiv. På det här sättet kan vi få tillgång till en viktig källa till data, samtidigt som vi skyddar användarnas integritet.

*Det är viktigt att notera att vi för närvarande inte tillåter användning av personuppgifter eller känslig information i Svea. Om du har frågor om detta, vänligen kontakta din administratör eller läs i PUB-avtalet för mer information.*

## Mål

Målet med denna annoteringsuppgift är att identifiera och markera konversationer mellan användare och Svea som är säkra att delas med utvecklingsteamet. Detta innebär att medarbetare ska granska sina egna konversationer och markera de som inte innehåller personuppgifter eller annan känslig information.

Genom att göra detta, kan vi:

* Skydda användarnas integritet och säkerställa att känslig information inte delas med obehöriga
* Få tillgång till en viktig källa till data som kan användas för att förbättra språkmodellen och göra Svea mer effektiv
* Säkerställa att utvecklingsteamet har tillgång till relevanta och säkra data för att förbättra Svea

Målet är att skapa en säker och effektiv process för att hantera och använda data från konversationer mellan användare och Svea.

## Annoteringsprocessen

**Steg 1: Välj en konversation**

Välj en konversation bland dina sparade konversationer som visas i den vänstra panelen i Svea.

**Steg 2: Granska**

Läs igenom konversationen i sin helhet och granska den för att identifiera eventuella personuppgifter eller sekretessbelagd information.

**Steg 3: Märk konversationen**

Konversationer som inte innehåller personuppgifter eller sekretessbelagd information kan markeras för delning. När en konversation markeras skickas den automatiskt till en särskild databas.

# Instruktionsdata

## Bakgrund

Instruktionsdata är en viktig typ av träningsdata för stora språkmodeller. Dessa data består av frågor och svar och syftar till att lära språkmodellen hur den ska bete sig och svara på ett användbart sätt.

När en språkmodell är klar med sin grundträning, behöver den lära sig hur den ska ge relevanta svar och klara av långa konversationer. Men den kan inte lära sig detta själv utan behöver goda exempel skrivna av människor. I det här projektet vill vi skapa instruktionsdata på svenska som ger modellen exempel på frågor och svar som är relevanta för den svenska offentliga sektorn.

Vårt mål är att finjustera stora språkmodeller som Llama 3.1 så att de blir ännu mer hjälpsamma och flexibla för sektorn. Det är ett banbrytande arbete eftersom svenska instruktionsdata i princip inte existerar idag, särskilt inte med fokus på den svenska offentliga sektorn.

#### Hur skapas instruktionsdata?

Instruktionsdata består av frågor och svar, men hur dessa frågor och svar är utformade är avgörande för hur språkmodellen kommer att bete sig och hur användbar den kommer vara. Vad är ett bra svar? Vilka frågor är relevanta? Hur vill vi att språkmodellen ska bete sig? Dessa frågor är inte enkla att svara på.

Processen för att skapa instruktionsdata är komplex och avancerad. Det kräver många människor, ibland hundratals, som hjälper till att skriva data som är konsekventa, korrekta, välskrivna, lätta att förstå och på en form och längd som är lämplig för den fråga som ställs. Ofta krävs också experter från många olika domänområden för att kunna ge svar av hög kvalitet på vissa frågor.

#### Hur väljer vi ut frågor?

Domänexperter kan komma på exempelfrågor själva som de tror att assistenten ska kunna svara på, men ett ännu bättre sätt är att titta på de frågor som användarna faktiskt skriver och skriva svar till de vanligaste och viktigaste av dessa. Det är vår ambition att vi ska göra just detta i det här projektet.

### Mål

Målet med denna annoteringsuppgift är att skapa frågor-och-svar par baserat på relevanta frågor från användningen av Svea. Vi vill identifiera de vanligaste och viktigaste frågorna som användarna ställer och skriva svar till dessa frågor. Dessa svar ska vara av hög kvalitet och utformade på ett sätt som är lämpligt för den fråga som ställs. Genom att skapa dessa instruktionsdata vill vi hjälpa språkmodellen att lära sig hur den ska bete sig och svara på ett användbart sätt, och på så sätt förbättra användarupplevelsen av Svea.

## Annoteringsprocessen

### Steg 1: Workshop

I detta steg träffas experter från er verksamhet med AI Sweden i en workshop för att diskutera och besluta om hur instruktionsdata ska skrivas. Frågor som "Vad är ett bra svar?", "Vilka frågor är relevanta?" och "Hur vill vi att språkmodellen ska bete sig?" kommer att behandlas.

### Steg 2: Annoteringsriktlinjer

Baserat på resultaten från workshopen skapas detaljerade och väl genomarbetade riktlinjer för hur instruktionsdata ska skrivas. Dessa riktlinjer är avgörande för att annoteringen ska ge ett bra resultat.

### Steg 3: Annotering

Verksamhetsexperter skriver antingen frågor själva eller ges frågor som kommer från användarna av Svea och svarar på dessa frågor utifrån annoteringsriktlinjerna. De använder sin expertis och erfarenhet för att skapa ett så bra svar som möjligt på olika frågor.

### Steg 4: Utvärdering

Efter att ett antal instruktionsdata har skrivits utvärderas dessa på olika sätt. Utvärderingen kan ske tekniskt eller genom att låta andra verksamhetsexperter granska och kommentera de svar som givits.

**Steg 5: Iterera**

Skrivandet av instruktionsdata är en iterativ process som kräver att man går tillbaka och ändrar riktlinjerna efter den återkoppling som ges i utvärderingssteget. Kvalitet är avgörande för dessa data, och det är viktigare att se till att man skapar bra data än att försöka skapa mycket.